جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم ظهر المهراز فاس
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

N° d'ordre : 41/2024

Centre des Etudes : *Sciences et Techniques et Sciences Médicales*

Formation doctorale : *Sciences et Technologies de l'Information et de Communication*

Discipline : *Informatique*

Spécialité : *Informatique*

Laboratoire : *Laboratoire d'Informatique, Signaux, Automatique et Cognitivisme (LISAC)*

THÈSE DE DOCTORAT

Sécurité routière au Maroc : Analyse, modélisation et prédiction basée sur des approches multivariées, par apprentissage statistique et automatique

Présentée par :

MBAREK Abdelilah

Soutenue le 22/07/2024 devant le jury composé de :

Nom et prénom	Etablissement	Grade	Qualité
Pr. TAIRI Hamid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Président
Pr. EL BANNAY Omar	ENSA, Khouribga	MCH	Rapporteur & examinateur
Pr. EL FALLAHI Abdellah	ENSA, Tétouan	PES	Rapporteur & examinateur
Pr. RIFFI Jamal	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Rapporteur & examinateur
Pr. BOUMHIDI Jaouad	Faculté des Sciences Dhar El Mahraz, Fès	PES	Examinateur
Pr. MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinateur
Pr. SATORI Khalid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Examinateur
Pr. SABRI My Abdelouahed	Faculté des Sciences Dhar El Mahraz, Fès	PES	Co-Directeur de thèse
Pr. YAHYAOUY Ali	Faculté des Sciences Dhar El Mahraz, Fès	PES	Directeur de thèse

Année universitaire : 2023-2024

REMERCIEMENTS

En préambule, je tiens à exprimer ma profonde gratitude envers mon éminent directeur de thèse Pr. YAHYAOUY Ali, et mon co-directeur Pr. SABRI My Abdelouahed, dont la guidance constante a été l'épine dorsale de l'ensemble de ce travail de recherche. Leur engagement indéfectible, soutien éclairé, et leurs conseils érudits ont constitué des piliers fondamentaux ayant grandement contribué à l'aboutissement fructueux de cette thèse. Leur encadrement et leur expertise ont été des ressources inestimables, fournissant une direction éclairée qui a enrichi la substance même de cette thèse.

Je tiens à exprimer ma sincère gratitude envers l'ensemble des membres du jury, en particulier à l'honorable président, Pr. TAIRI Hamid, ainsi qu'aux distingués rapporteurs, Pr. EL BANNAY Omar, Pr. EL FALLAHI Abdellah et Pr. RIFFI Jamal, ainsi qu'aux éminents examinateurs Pr. BOUMHIDI Jaouad, Pr. MAHRAZ Mohamed Adnane et Pr. SATORI Khalid, pour leurs engagements inestimables dans l'évaluation de ma contribution. Leurs observations perspicaces et leurs commentaires éclairés ont grandement enrichi mon travail, reflétant leurs profondes compréhensions des enjeux académiques sous-jacents.

Je tiens également à exprimer cette même gratitude à Madame JIBER Mouna pour sa contribution inestimable à cette recherche, tant par la mise à disposition des données que par son assistance tout au long des différentes étapes de l'élaboration de cette thèse.

Je remercie par ailleurs toutes les personnes et institutions qui ont généreusement apporté leur soutien financier, matériel et intellectuel tout au long de la réalisation de ce travail de recherche. L'appui inestimable que j'ai reçu a grandement facilité la concrétisation de cette entreprise académique, témoignant de la bienveillance de ceux qui ont contribué à la pérennité de mon engagement. Les ressources fournies ont joué un rôle crucial dans la conduite de mes investigations, démontrant ainsi la solidarité et le soutien de la communauté académique et institutionnelle.

Il me tient à cœur d'exprimer ma sincère reconnaissance envers l'ensemble de ma famille, mes amis et toute personne qui a gracieusement offert un soutien émotionnel tout au long de mon périple académique. Leur présence bienveillante et leurs encouragements constants ont constitué un pilier essentiel, contribuant de manière significative à ma résilience et à ma persévérance dans les moments de défis intellectuels.

RÉSUMÉ

Cette thèse présente deux aspects interdépendants de la sécurité routière. D'une part, elle examine la gestion des données manquantes du trafic routier et la prédiction à court terme du trafic routier. En effet, la gestion des données manquantes pose un défi majeur en raison des interruptions fréquentes des compteurs installés le long des routes. Pour remédier à cette problématique, trois modèles ont été employés afin d'estimer les données manquantes d'un dispositif défectueux, à savoir SARIMA, la régression linéaire et SVR. Ces modèles utilisent le trafic routier recueilli par des dispositifs adjacents au compteur en question. Les résultats indiquent que SARIMA a démontré une grande précision dans la régénération des données manquantes. Par ailleurs, anticiper le trafic routier futur est crucial pour les décideurs, en raison de son rôle prépondérant dans la résolution proactive des problèmes de circulation. Dans cette optique, nous proposons un modèle de prédiction hybride combinant les techniques ELM et d'ensemble pour anticiper à court terme le trafic horaire sur un tronçon routier de Tanger. En s'appuyant sur des données historiques des cinq années précédentes, ce modèle a pu surpasser les algorithmes comparés en termes de précision et de généralisation.

D'autre part, cette thèse se concentre sur l'analyse des accidents routiers au Maroc, en exposant deux méthodologies distinctes, notamment l'identification des points noirs des accidents et leur classification en fonction des facteurs infrastructurels. Tout d'abord, un cadre analytique s'appuyant sur l'extraction de règles d'association est utilisé pour discerner les caractéristiques des points noirs. L'indice de gravité pondéré a été calculé pour chaque section afin de déterminer leurs niveaux de gravité. L'algorithme Apriori est appliqué pour découvrir les associations entre les caractéristiques infrastructurelles et les niveaux de gravité. Parallèlement, un modèle, intégrant l'algorithme ELM et la technique du Bagging, est développé pour classer avec précision les points noirs selon leurs niveaux de gravité, sur la base de facteurs infrastructurels. Ce modèle de prédiction a atteint une précision remarquable de 98,6%, alors que le modèle d'identification des facteurs de risque a mis en évidence l'importance des éléments de l'infrastructure, tels que la largeur de la chaussée et des accotements, le tracé rectiligne, le bon revêtement et la présence des ponts, dans l'occurrence des accidents mortels.

Mots-clés : accident de la route, apprentissage automatique, circulation routière, dommage humain, facteurs de risque, flux du trafic, indice de gravité pondéré, mortalité, prédiction, points noirs, risque d'accident, sécurité routière, taux de létalité.

ABSTRACT

This thesis presents two interrelated aspects of road safety. Firstly, it examines the management of road traffic missing data and the short-term traffic prediction. Indeed, handling missing data is a major challenge due to the frequent interruptions of roadside counters. To address this issue, three models were employed to estimate the missing data from a faulty device, namely SARIMA, linear regression and SVR. These models are based on traffic data collected by devices adjacent to the faulty one. The findings indicate that the SARIMA model demonstrated high accuracy in generating missing data. On the other hand, anticipating future road traffic is of crucial importance to decision-makers of this sector, due to its prominent role in proactively solving traffic problems. In this context, we propose a hybrid prediction model combining ELM and ensemble techniques to anticipate short-term hourly traffic on a road section in Tangier. The study uses historical data from the previous five years. Our proposed model was compared with well-known algorithms, and the results showed that it improved performance in terms of accuracy and generalization.

Secondly, this thesis focuses on the analysis of road accidents in Morocco, by outlining two different methodologies, namely the identification of accident black spots and their classification according to infrastructural factors. Initially, an analytical framework based on the extraction of association rules is used to discern the characteristics of accident black spots, focusing on infrastructural factors. The weighted severity index was calculated for each section to determine their severity levels. The Apriori algorithm is applied to identify associations between infrastructural characteristics and severity levels. Simultaneously, a model integrating the ELM algorithm and Bagging technique is developed to accurately classify black spots according to their severity levels, based on infrastructural factors. This predictive model achieved a significant accuracy of 98.6%. Concurrently, the risk factor identification model highlighted the significance of specific infrastructure elements, such as road and shoulders width, straight alignment, and the presence of bridges, in the occurrence of fatal accidents.

Keywords: road accident, machine learning, road traffic, human damage, risk factors, traffic flow, weighted severity index, fatality, prediction, black spots, accident risk, road safety, lethality rate.

ملخص

تعرض هذه الأطروحة جانبين مترابطين من جوانب السلامة الطرقية. أولاً، تتطرق إلى إدارة بيانات حركة المرور المفقودة والتنبؤ بحركة المرور على المدى القصير. في الواقع، يمثل التعامل مع البيانات المفقودة تحدياً كبيراً بسبب الانقطاع المتكرر للعدادات على جانب الطريق. لتدارك هذه المشكلة، تم استخدام ثلاثة نماذج لتقدير البيانات المفقودة في الجهاز المعطل وهي نموذج SARIMA، الانحدار الخطي، و SVR. تستخدم هذه النماذج بيانات حركة المرور على الطرق التي يتم جمعها من الأجهزة المجاورة للعدّاد المعني. تشير النتائج إلى أن نموذج SARIMA أظهر دقة عالية في تحديد البيانات المفقودة. وبالإضافة إلى ذلك، يكتسي توقع حركة المرور المستقبلية على الطرق أهمية قصوى لاتخاذ القرار لدى الفاعلين في قطاع النقل، نظرًا لدوره الرائد في حل مشاكل المرور بشكل استباقي. في هذا السياق، اقترحنا نموذجًا تنبّؤيًا هجينًا يجمع بين ELM والتقنيات التجميعية لتوقع حركة المرور على المدى القصير على مقطع طرقي في طنجة. باستخدام البيانات التاريخية من خمس سنوات سابقة، تمكن هذا النموذج من التفوق على الخوارزميات المقارنة من حيث الدقة والتعميم.

ثانيًا، تركز هذه الأطروحة على تحليل حوادث الطرق في المغرب، من خلال تقديم منهجيتين مختلفتين، وهما تحديد النقاط السوداء للحوادث وتصنيفها وفقًا لعوامل البنية التحتية. في البداية، يتم استخدام إطار تحليلي يعتمد على استخراج قواعد الارتباط لتمييز خصائص نقاط الحوادث السوداء، مع التركيز بشكل خاص على عوامل البنية التحتية. تم حساب مؤشر الخطورة المرجّح لكل مقطع طرقي لتحديد مستويات الفتك فيها. كما تم تطبيق خوارزمية Apriori لتحديد الارتباطات بين خصائص البنية التحتية ومستويات الفتك. في الوقت نفسه، يتم تطوير نموذج تنبؤي يدمج خوارزمية ELM وتقنية Bagging. حيث يصنف هذا النموذج بدقة نقاط الحوادث السوداء بناءً على مستويات فتكها، باستخدام عوامل البنية التحتية. أظهر هذا النموذج التنبؤي معدل دقة ملحوظ بلغ نسبة 98.6%. وفي الوقت نفسه، أكد نموذج تحديد عوامل الخطر على أهمية عناصر البنية التحتية، مثل عرض الطريق والجوانب، وشكل مسار الطرق ووجود الجسور، في وقوع الحوادث المميتة.

كلمات مفتاحية: حوادث الطرق، التعلم الآلي، حركة السير، الأضرار البشرية، عوامل الخطر، تدفّق حركة المرور، مؤشر الخطورة المرجّح، الوفيات، التنبّؤ، النقاط السوداء، مخاطر الحوادث، السلامة الطرقية، معدل الفتك.

Tables des matières

Liste des figures ..... **xi**

Liste des tableaux ..... **xii**

Liste des abréviations ..... **xiii**

Introduction générale **1**

1. Contexte général **1**

2. Problématiques de recherche et objectifs de la thèse **3**

3. Méthodologie générale **6**

3.1. Gestion des données manquantes du trafic routier 6

3.2. Prédiction à court terme du futur trafic routier 7

3.3. Analyse des données d'accidentologie 7

4. Structure de la thèse **9**

CHAPITRE 1. SYSTEMES DE TRANSPORT INTELLIGENTS : CONCEPTS ET APERÇUS AU MAROC **11**

1. Introduction **11**

2. Systèmes de transport intelligents **12**

2.1. Avantages des systèmes de transport intelligents 12

2.2. Applications des systèmes de transport intelligents 12

2.3. Utilisation des systèmes de transport intelligents dans la sécurité routière 13

3. Apprentissage automatique et sécurité routière **17**

3.1. Concepts clés 17

3.1.1. Apprentissage automatique supervisé 17

3.1.2. Apprentissage automatique non supervisé 18

3.1.3. Apprentissage automatique semi-supervisé 18

3.1.4. Apprentissage par renforcement 19

3.2. Études antérieures pour l'apprentissage automatique dans la sécurité routière 19

4. Concepts clés du trafic routier et de la sécurité routière **22**

4.1. Mesures fondamentales 22

4.1.1. Densité 22

4.1.2. Vitesse moyenne de libre circulation 23

4.1.3. Vitesse moyenne temporelle 23

4.1.4. Vitesse moyenne spatiale 24

4.1.5. Flux 24
4.1.6. Temps du trajet 24
4.1.7. Capacité de la route 25
4.1.8. Facteurs temporels 25
4.2. Diagramme fondamental du trafic routier 26
4.2.1. Principes du diagramme fondamental 26
4.2.2. Dynamiques du trafic routier 27
4.3. Infrastructure routière 29
4.3.1. Fondements essentiels des infrastructures 29
4.3.2. Types de routes au Maroc 30
4.3.3. Description des éléments de l'infrastructure routière 32
4.3.4. Description des paramètres de conception géométrique des routes 33
4.3.4.1. Visibilité 33
4.3.4.2. Profil en travers 34
4.3.4.3. Tracé en plan 35
4.3.4.4. Profil en long 35
4.4. Aspects clés de la sécurité routière 36
4.4.1. Points noirs des accidents 36
4.4.2. Taux d'accidents 36
4.4.3. Facteurs accidentogènes 37
5. Conclusion 38

CHAPITRE 2. COLLECTE DES DONNEES DE TRAFIC ROUTIER ET GESTION DES DONNEES MANQUANTES 39

1. Introduction 39
2. Exploration des méthodes de collecte de données 41
2.1. Méthodes de collecte de données du trafic routier 41
2.1.1. Méthodes intrusives 41
2.1.2. Méthodes non intrusives 42
2.2. Avantages, limites et problèmes courants liés à la collecte des données 43
3. Description de l'étude 45
3.1. Contexte spécifique de la collecte des données au Maroc 45
3.2. Présentation de l'étude d'estimation des données manquantes 47
4. Méthodes d'estimation des données 47
4.1. Prétraitement des données 47

4.2. Régression linéaire - Moindres carrés ordinaires 50
4.3. Machine à vecteurs de support pour la régression 52
4.4. Moyenne mobile intégrée autorégressive saisonnière 53
5. Évaluation des résultats et discussion 54
5.1. Analyse de corrélation du trafic routier des capteurs adjacents 54
5.2. Discussion et interprétabilité 56
5.2.1. Interprétabilité des modèles 56
5.2.2. Interprétation du modèle de régression linéaire 57
5.2.3. Interprétation du modèle SARIMA 58
5.2.4. Interprétation du modèle SVR 60
5.2.5. Comparaison entre les différents modèles 62
6. Conclusion 63
CHAPITRE 3. MODELE DE PREDICTION DU TRAFIC ROUTIER A COURT TERME A L'AIDE DE L'EXTREME LEARNING MACHINE 65
1. Introduction 65
2. Méthodes existantes de prédiction 66
3. Modèle de prédiction du trafic routier basé sur l'ELM 69
3.1. Description du modèle 69
3.2. Méthodologie de recherche 69
3.2.1. Prétraitement des données 69
3.2.2. Variables d'entrée 72
4. Méthodes 73
4.1. Extrême learning machine 74
4.2. Systèmes d'ensemble dans la prise de décision 76
4.3. Moyenne mobile intégrée autorégressive 77
4.4. Régression à vecteur de support 77
4.5. Perceptron multicouche 78
5. Entraînement du modèle ELM 79
5.1. Critères d'évaluation 79
5.2. Ajustement du modèle proposé et évaluation de la prédiction 81
5.3. Prédiction du trafic routier 83
6. Comparaison des performances des modèles prédictifs 84
7. Analyse des résultats et discussion 86

8. Conclusion **87**

CHAPITRE 4. ANALYSE STATISTIQUE SUR L'IMPACT DE L'INFRASTRUCTURE ROUTIERE SUR LA MORTALITE ROUTIERE AU MAROC **89**

1. Introduction **89**

2. Contexte théorique et historique **91**

3. Préparation des données **93**

4. Méthodologie de recherche **94**

5. Facteurs de risque des accidents **96**

5.1. Répartition des accidents sur les différentes conditions routières 97

5.2. Répartition des taux de létalité 101

5.3. Analyse des facteurs infrastructurels 104

5.4. Évaluation de l'impact de chaque facteur sur la mortalité routière 105

6. Discussion des résultats et recommandations **108**

7. Conclusion **114**

CHAPITRE 5. IDENTIFICATION DES POINTS NOIRS D'ACCIDENTS PAR LA METHODE D'INDICE DE GRAVITE PONDERE **115**

1. Introduction **115**

2. Méthodes d'identification des points noirs **117**

3. Modèle d'identification des points noirs d'accidents **119**

3.1. Méthodologie utilisée 119

3.2. Préparation de données 119

3.3. Détermination des poids de l'indice de gravité pondéré 123

3.4. Indice de gravité pondéré 125

4. Analyse de sections à haut risque d'accident **125**

4.1. Classement des sections en fonction du nombre de victimes et d'accidents 125

4.2. Résultats de l'importance des caractéristiques 129

4.3. Classement des tronçons routiers selon l'indice de gravité pondéré 129

4.4. Résultats du regroupement des points noirs 132

5. Discussion **133**

6. Conclusion **136**

CHAPITRE 6. ANALYSES DES FACTEURS DE RISQUES ET PREDICTION DES POINTS NOIRS DES ACCIDENTS SUR LES ROUTES RURALES 137

1. Introduction 137

2. Approches existantes d'identification et d'analyse des points noirs 140

3. Méthodologie de recherche 144

4. Méthodes 145

4.1. Prétraitement de données 145

4.2. Indice de gravité pondéré 148

4.3. Extraction de règles d'association 149

4.4. Méthode de rééchantillonnage Bootstrap 150

4.5. Sélection des règles générales 151

5. Résultats d'identification des facteurs de risque 152

5.1. Résultats des poids de la méthode de l'indice de gravité pondéré 152

5.2. Résultats de la méthode de l'indice de gravité pondéré 152

5.3. Résultats d'extraction des règles d'association 153

5.4. Analyse des règles d'association par diagramme de dispersion 158

6. Résultats de prédiction des points noirs 158

7. Discussions et recommandations 161

8. Conclusion 162

Conclusion générale 164

Bibliographies 167

Liste des publications 178

Annexes 179

Liste des figures

Figure 1. Méthodologie générale utilisée dans l'analyse des points noirs des accidents 8
Figure 2. Diagramme fondamental du trafic routier 27
Figure 3. Diagramme fondamental : débit en fonction de concentration 28
Figure 4. Diagramme fondamental : vitesse en fonction du débit 29
Figure 5. Diagramme fondamental : vitesse en fonction de la concentration 30
Figure 6. Répartition des stations permanentes et périodiques autour de la ville de Marrakech 48
Figure 7. Distribution des données avec et sans valeurs aberrantes. 50
Figure 8. Flux de trafic de chaque station dans l'ordre chronologique 55
Figure 9. Matrice de corrélation entre les différentes stations 56
Figure 10. Résultats de la régression linéaire 57
Figure 11. Histogramme des poids estimés par la régression linéaire 58
Figure 12. Résultats du modèle SARIMA 59
Figure 13. Résultats de la décomposition saisonnière et tendancielle à l'aide du Loess 60
Figure 14. Graphique "Beeswarm" Résultats du SHAP pour le modèle SVR 61
Figure 15. Comparaison des performances de trois modèles de prédiction du trafic routier 63
Figure 16. Structure du modèle ELM 70
Figure 17. Cadre proposé pour la prédiction du trafic à l'aide d'une méthode d'ensemble 70
Figure 18. Structure du réseau SLFN 74
Figure 19. Exemple d'un réseau de neurones à trois couches 79
Figure 20. Évolution des erreurs moyennes dans le modèle ELM en fonction du nombre d'essais 82
Figure 21. Erreur Résiduelle Absolue de la phase de test 84
Figure 22. Résultats des prévisions de trafic des différents modèles 86
Figure 23. Taux de mortalité dû aux accidents de la route pour 100 000 personnes 90
Figure 24. Résultats du test post-hoc avec le niveau de signification 106
Figure 25. Taux de mortalité accidentelle pour 100 000 personnes 116
Figure 26. Méthodologie de recherche 120
Figure 27. Route nationale 1 121
Figure 28. Modèle de génération de l'importance des caractéristiques 124
Figure 29. Répartition des sections de la route par IGP 130
Figure 30. Carte générée affichant les points noirs 134
Figure 31. Méthodologie de recherche 145
Figure 32. Visualisation par graphique à bulle des règles générées 159

Liste des tableaux

Tableau 1. Évolution du parc automobile, du réseau routier, des accidents et des tués 2
Tableau 2. Synthèse des avantages et des inconvénients des méthodes intrusives et non intrusives 44
Tableau 3. Structure d'un fichier généré par un compteur pour une route à deux voies 48
Tableau 4. Structure de données après traitement 49
Tableau 5. Évaluation des modèles MCO, SARIMA et SVR 62
Tableau 6. Échantillon de données provenant d'un compteur permanent 72
Tableau 7. Classification des véhicules en fonction de leur longueur 72
Tableau 8. Comparaison des performances de prédiction par l'AIC, RMSE, MAE et MAPE 85
Tableau 9. Répartition des accidents sur les conditions routières 100
Tableau 10. Taux de létalité en fonction des facteurs infrastructurels 102
Tableau 11. Statistique descriptive des facteurs infrastructurels 104
Tableau 12. Résultats du test de Krukal-Wallis et du test de post-hoc 107
Tableau 13. Structure des données utilisées 123
Tableau 14. Classement des sections en fonction du nombre de tués, de blessés et d'accidents 126
Tableau 15. Classement des sections selon DH, RA et TL 128
Tableau 16. Poids générés par les modèles XGBoost 130
Tableau 17. Classement des sections selon l'IGP 131
Tableau 18. Classification des points noirs à partir de la moyenne et de l'écart-type de IGP 132
Tableau 19. Statistiques des points noirs de niveau de sévérité élevé 135
Tableau 20. Statistiques des points noirs de niveau de sévérité moyen 135
Tableau 21. Statistiques des points noirs de niveau de sévérité faible 135
Tableau 22. Description des facteurs étudiés 147
Tableau 23. Classification des points noirs sur la base de la moyenne et de l'écart-type du IGP 149
Tableau 24. Poids des caractéristiques obtenus par le modèle XGBoost 152
Tableau 25. Règles générales générées sur les sections de niveau de gravité élevé 155
Tableau 26. Règles générales générées sur les sections de niveau de gravité moyen 155
Tableau 27. Règles générales générées sur les sections de niveau de gravité faible 156
Tableau 28. Règles générales générées sur les sections sûres 157
Tableau 29. Propriétés de l'ensemble de données final 159
Tableau 30. Comparaison des performances entre B-ELM, ELM, SVM et MLP 160

Liste des abréviations

ACF : Autocorrelation Function
ACP : Analyse en Composantes Principales
AIC : Akaike's Information Criterion
ANN : Artificial Neural Network
ANOVA : Analysis of Variance
ARIMA : Autoregressive Integrated Moving Average
BIC : Bayes Information Criterion
BP : Backpropagation
CamVid : Cambridge-Driving Labeled Video Database
CH : Cluster-Head
CMAFS : Carrefour Multiple Avec Feux ou Stop
CMSFS : Carrefour Multiple Sans Feux ou Stop
CNER : Centre National d'Études Routières
CNN : Convolutional Neural Network
CSAFS : Carrefour Simple Avec Feux ou Stop
CSSFS : Carrefour Simple Sans Feux ou Stop
DGRTT : Direction Générale des Routes et des Transports Terrestres
DM : Dommages Humains
DNNC : Deep Neural Network Multi-Class Classifier
DPETL : Direction Provinciale de l'Équipement, du Transport et de la Logistique
DR : Direction des Routes
DRETL : Direction Régionale de l'Équipement, du Transport et de la Logistique
ELM : Extreme Learning Machine
ERA : Extraction de Règles d'Association
FNN : Feedforward Neural Network
HOG : Histogram of Oriented Gradients
HQIC : Hannan-Quinn Information Criterion
IGP : Indice de Gravité Pondéré
IoT : Internet of Things

IPB : Itinéraires Prioritaires pour les Autobus

IQR : Interquartile Range

KNN : K-Nearest Neighbors

LiDAR : Light Detection and Ranging

MCO : Moindres Carrés Ordinaires

MEE : Ministère de l'Équipement et de l'Eau

MLP : Multilayer Perceptron

NARSA : Agence Nationale de la Sécurité Routière

OMS : Organisation Mondiale de la Santé

PACF : Partial Autocorrelation Function

PCI : Pavement Condition Index

PIB : Produit Intérieur Brut

PSAS : Programme Spécial d'Amélioration de la Sécurité Routière

QoS : Quality of Service

RA : Risque d'Accident

RBANet : Residual Balanced Attention Network

SAPO : Safety Potential

SARIMA : Seasonal Autoregressive Integrated Moving Average

SEPD : Systèmes d'Ensemble dans la Prise de Décision

SHAP : Shapley Additive Explanations

SIG : Systèmes d'Information Géographique

Smart DLSP : Smart Digital Logistic Services Provider

SPEE : Service de Planification des Études Économiques

STI : Système de Transport Intelligent

SVM : Support Vector Machines

SVR : Support Vector Regression

TL : Taux de Létalité

TMJA : Trafic Moyen Journalier Annuel

VANET : Vehicle Ad Hoc Network

XDLL : Explained Deep Learning LiDAR-Based Localization

XGBoost : Extreme Gradient Boosting

Introduction générale

1. Contexte général

Dans l'ère numérique où chaque instant sur la route compte, la sécurité routière émerge comme le pilier essentiel de la santé publique, impliquant un ensemble d'actions et de mesures visant à protéger les usagers de la route et à réduire au minimum les conséquences en termes de décès et de blessures. Cela englobe la conception de routes sécurisées, la mise en place de lois de circulation efficaces et la gestion du trafic routier. L'objectif étant de créer un environnement de conduite sécurisé et ordonné qui minimise les perturbations de la circulation normale, telles que la congestion et les accidents. Ces phénomènes sont considérés comme des défis majeurs en matière de sécurité routière et font l'objet de nombreux projets, études et conférences de portée internationales.

Dans ce contexte, le trafic routier représente un paramètre crucial dans la planification et la gestion des projets routiers, nécessitant une surveillance attentive de son évolution et de sa composition par les entités responsables des transports. La croissance rapide de la population et l'urbanisation croissante contribuent à l'augmentation du nombre de véhicules sur les routes, impactant la mobilité, l'économie et la sécurité. Cette expansion du trafic est malheureusement liée à une hausse des accidents de la route, entraînant des pertes humaines, des souffrances pour les familles et une pression sur les services médicaux et les infrastructures routières. En 2023, l'organisation mondiale de la santé a signalé une légère diminution du nombre annuel de décès liés aux accidents de la route, atteignant 1,19 million, soulignant la nécessité continue d'investir dans des efforts soutenus pour améliorer la sécurité routière.

Pour remédier à cette problématique, il devient impératif que les autorités spécialisées entreprennent des initiatives adéquates visant à rehausser la sécurité routière, à investir dans les infrastructures et à promouvoir des alternatives de transport durables. La collaboration entre les gouvernements, les citoyens et les entités commerciales est cruciale pour établir un environnement routier plus sécurisé, tout en préservant la vie des usagers de la route et en favorisant un développement économique durable. Des investissements ciblés dans les infrastructures routières, accompagnés de mesures incitatives pour favoriser l'utilisation de moyens de transport durables, peuvent significativement contribuer à atténuer les risques associés à la circulation routière.

Au Maroc, l'évolution constante du paysage routier est attribuable à la croissance démographique, à l'essor économique et à l'augmentation significative du parc automobile. D'évidence, les grandes villes telles que Casablanca, Rabat et Tanger font face à une densité de circulation élevée, surtout aux heures de pointe. Les axes routiers majeurs reliant ces centres urbains subissent également une augmentation significative du trafic, entraînant une congestion croissante. Cette réalité souligne la nécessité de solutions stratégiques pour optimiser la gestion du trafic et promouvoir la mobilité durable. En ce sens, le gouvernement marocain a pris des mesures pour améliorer la sécurité routière, incluant le renforcement des contrôles, des campagnes de sensibilisation, des améliorations des infrastructures et des mesures répressives contre les comportements dangereux. Dans ce cadre, le Programme Spécial d'Amélioration de la Sécurité routière (PSAS), lancé dans le cadre de la stratégie nationale 2017-2026, vise à traiter les zones à risque d'accidents. Malgré ces efforts, le nombre élevé de décès liés aux accidents souligne la nécessité de recherches approfondies pour comprendre les facteurs contributifs et élaborer des solutions plus efficaces.

Ci-après, le tableau 1 examine les tendances de 2010 à 2019 concernant le parc automobile, le réseau routier, les accidents mortels, et les tués. On observe une augmentation constante du nombre de véhicules en circulation, passant de 2 759 651 en 2010 à 4 315 941 en 2019. La longueur du réseau routier revêtu a légèrement augmenté de 41 737 km à 46 015 km sur la même période. Cependant, les chiffres liés aux accidents mortels varient de 3 181 à 3 247, et il y a une légère diminution du nombre de tués (3 622 en 2019). Aussi, ces fluctuations suggèrent-elles l'influence possible de facteurs tels que les mesures de sécurité routière, la réglementation, et la sensibilisation. Bien que le réseau routier ait augmenté, des investissements supplémentaires peuvent être nécessaires pour assurer des conditions de circulation sûres face à la croissance continue du parc automobile.

Tableau 1. Évolution du parc automobile, du réseau routier, des accidents et des tués[1]

Année	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Parc-auto	2 759 651	2 920 308	3 088 261	3 247 629	3 396 847	3 546 998	3 735 952	3 926 341	4 120 233	4 315 941
Réseau routier revêtu	41 737	42 057	42 531	42 531	43 678	43 755	45 096	45 118	46 015	46 015
Accident mortel	3 181	3 636	3 531	3 265	3 021	3 365	3 317	3 274	3 292	3 247
Tués	3 778	3 778	4 167	3 832	3 489	3 776	3 785	3 726	3 736	3 622

[1] Recueil des statistiques des accidents corporels de la circulation routière 2019, https://www.narsa.ma/fr/etudes-et-statistiques ; Recueil du trafic routier 2019, https://www.equipement.gov.ma/Infrastructures-Routieres/Reseau-Routier-du-Royaume/Pages/Statistiques-des-accidents.aspx.

2. Problématiques de recherche et objectifs de la thèse

La prédiction constitue l'essence même des Systèmes de Transport Intelligents (STI) et est extrêmement importante dans le processus de prise de décision. Les chercheurs ont accordé une grande attention au domaine de l'analyse du trafic routier et de la sécurité routière en raison de leur importance. Ils ont envisagé de nombreuses méthodes appropriées, des modèles de trafic aux approches statistiques ou d'apprentissage automatique. Les analyses prédictives, basées sur des techniques telles que la modélisation prédictive et l'exploration de données [1], utilisent des statistiques historiques et actuelles pour estimer des résultats futurs [2]. Dans le contexte des systèmes de transport intelligents, la gestion et la prédiction du trafic routier jouent un rôle crucial, influençant le développement des infrastructures et des projets, ainsi que la régulation du trafic et la sécurité routière.

Pour ce faire, de nombreux organismes et agences existent pour planifier, concevoir, construire et entretenir le système de transport du pays. Au Maroc, la gestion du trafic routier est supervisée par le Centre National des Études Routières (CNER) qui est rattaché au Ministère de l'Équipement et de l'Eau (MEE). Le centre est responsable de la collecte des données sur les flux de trafic et les accidents de la route. Dans ce contexte, le défi consiste aujourd'hui à tirer parti des dernières technologies dans ce domaine, afin de développer des analyses intelligentes du grand volume de données collectées. L'une des questions fondamentales de la sécurité routière consiste à prévoir avec une grande précision les futurs états possibles du trafic routier. Dans ces conditions, les informations prédites permettent d'éviter des incidents tels que les embouteillages ou d'autres anomalies routières [3]. De plus, la capacité à anticiper le flux de trafic peut potentiellement contribuer à planifier des interventions visant à améliorer les infrastructures routières.

Les informations relatives à la circulation routière sont systématiquement collectées au moyen de dispositifs de comptage permanents installés le long des voies de circulation. Toutefois, ces instruments sont susceptibles de présenter des défaillances ou des altérations, induisant ainsi des lacunes dans les données enregistrées. Ces dispositifs de comptage encourent des risques inhérents tels que des problèmes d'alimentation électrique, des défaillances techniques, ou des altérations physiques, engendrant ainsi des lacunes dans l'enregistrement des données et une carence en informations cruciales. Cette problématique a été identifiée lors de l'analyse des données effectuée au sein de notre étude. Plusieurs cas ont été observés, parmi lesquels se distingue l'absence d'enregistrement sur une seule voie d'une route, tandis que la voie opposée présente des observations normales. Un autre scénario notable concerne des données

manquantes simultanément sur les deux voies de circulation. Ces constatations soulignent la complexité et la variabilité des problèmes rencontrés par les dispositifs de comptage, mettant en évidence la nécessité d'une approche approfondie pour comprendre et résoudre ces irrégularités dans la collecte des données de circulation routière.

La finalité de notre étude dans le domaine du trafic routier était initialement de proposer une solution au défi des lacunes de données. Pour remédier à cette problématique, l'approche adoptée repose sur l'utilisation des données issues des dispositifs de comptage adjacents afin d'estimer les lacunes de données résultant d'un dysfonctionnement particulier d'un dispositif. L'objectif central est de fournir une solution fiable et précise pour pallier les données manquantes, tout en facilitant une utilisation optimale de la masse d'informations sur le trafic routier par les décideurs et les entités responsables, sans perte d'information.

Par ailleurs, la seconde motivation de notre projet réside dans l'exploitation des données historiques du trafic routier en vue de prédire de manière fiable et précise l'évolution future de la circulation. Ce modèle prédictif offre aux parties prenantes la possibilité de prendre des décisions proactives, basées sur des prévisions robustes émanant de données historiques, afin d'optimiser la gestion du trafic et d'accroître l'efficacité des interventions en temps réel. Cette approche anticipative contribuera à améliorer la planification opérationnelle et la réactivité face aux éventuelles perturbations, renforçant ainsi la pertinence et l'efficacité globale du système de gestion du trafic routier.

La seconde partie de notre étude se consacre à l'analyse approfondie des accidents mortels, visant à examiner la sécurité routière au Maroc. Bien que de nombreuses études aient été menées étudier les facteurs humains tels que l'âge, le sexe du conducteur et le taux d'alcool dans les accidents mortels de la route, la littérature actuelle présente certaines lacunes en ce qui concerne l'exploration de la relation entre l'infrastructure routière et la sévérité des accidents. Par conséquent, la motivation de cette partie de recherche visera à examiner la relation entre l'état physique des routes et le taux de létalité des accidents, qui est actuellement inconnue et sous-étudiée. L'objectif est d'identifier les principaux facteurs infrastructurels qui contribuent le plus à la « guerre sur les routes ». Pour atteindre cet objectif, les données recueillies sur les accidents de la route au cours des dernières années ont été soumises à une analyse approfondie, à l'aide des techniques statistiques et intelligentes bien connues. Les résultats des analyses montrent que les facteurs étudiés ont un effet significatif sur la mortalité due aux accidents. L'analyse présentée apporte une contribution significative à l'amélioration de la sécurité routière,

représentant ainsi une ressource indispensable pour les acteurs opérant dans le domaine de la sécurité routière.

Précisons que de nombreux accidents routiers se produisent de manière récurrente dans des zones spécifiques, parmi lesquelles certains peuvent être associés à des incidents graves, voire mortels. Ces accidents engendrent des coûts économiques substantiels, notamment en termes de soins médicaux, de réparations d'infrastructures et de véhicules. Pour remédier à ces problématiques, notre attention s'est focalisée sur l'identification des segments routiers les plus dangereux du pays, communément désignés dans le cadre de la gestion de la sécurité routière sous l'appellation de « points noirs ». Ce terme fait référence aux zones où l'historique des accidents a été significativement concentré. Ces points noirs ne résultent pas de simples coïncidences fortuites, mais plutôt de divers facteurs explicatifs qui contribuent à la récurrence des accidents. L'objectif étant de développer un modèle préventif précis et fiable permettant l'identification des points noirs existants. À ce titre, cela implique la conception d'un modèle robuste d'identification, intégrant l'analyse de caractéristiques spatiales telles que l'infrastructure routière et l'environnement. Le processus envisagé vise à classer les points noirs en quatre catégories distinctes, à savoir les sections à forte, moyenne et faible létalité, ainsi que les sections ne présentant pas de danger significatif. L'objectif sous-jacent est de réduire de manière significative la létalité liée aux accidents de la route, ainsi que d'orienter judicieusement les investissements dans l'infrastructure routière en concentrant les ressources sur les sections à forte létalité.

L'infrastructure routière, du fait de sa relative stabilité et de sa résistance aux changements fréquents, offre une opportunité précieuse pour la prédiction des points noirs d'accidents. En effet, cette stabilité intrinsèque permet de considérer les caractéristiques durables, telles que la configuration des routes et les conditions environnementales, comme des indicateurs clés dans le développement de modèles de prédiction. Ces modèles, en intégrant ces données d'infrastructure, peuvent de ce fait anticiper de manière proactive les zones présentant un risque accru d'accidents. Dans ce contexte, un modèle intelligent de prédiction des points noirs sur la base des infrastructures et conditions de la route a été proposé. Un tel modèle apporte par ailleurs de nombreux avantages dans le domaine de la sécurité routière, notamment la prévention des accidents en identifiant les zones susceptibles de devenir des points noirs, la planification des infrastructures à long terme pour améliorer l'infrastructure routière, l'allocation des ressources en dirigeant les efforts vers les zones prévues comme étant à haut risque, ainsi que l'évaluation de l'efficacité des interventions effectuées sur des zones jugées à haut risque.

Aussi, dans notre thèse, nous nous attacherons à fournir des solutions éclairées aux interrogations suivantes :

- Comment construire des données basées sur des données de détection fixes ?
- Comment traiter notre ensemble de données sur le trafic routier à l'aide des algorithmes d'apprentissage automatique et statistiques ?
- Comment améliorer la prédiction/précision de nos modèles de prédiction ?
- Quels sont les facteurs, liés à l'état de la route, qui contribuent le plus à la létalité des accidents ?
- Quels sont les points noirs de la route la plus meurtrière ?
- Quelles sont les caractéristiques qui caractérisent les points noirs des accidents ?
- Comment prédire les points noirs d'accidents sur la base des informations d'infrastructure ?

3. Méthodologie générale

Cette section sera dédiée à la méthodologie globale adoptée pour la gestion des données incomplètes, détaillant les techniques employées dans le traitement des données relatives au trafic routier. Afin d'éviter toute confusion entre les deux études de prédiction du trafic routier, bien qu'elles reposent sur une source de données commune, elles sont présentées de manière distincte. La première étude se fonde sur les informations de trafic de quatre stations adjacentes en vue de prédire les données manquantes d'un compteur voisin endommagé. En revanche, la seconde étude anticipe le trafic futur d'une unique station en se basant sur l'historique passé du trafic routier enregistré par ladite station. En derniers lieux, la méthodologie adoptée pour l'identification et l'analyse des données des accidents de la route a été brièvement anticipée.

3.1. Gestion des données manquantes du trafic routier

Notre recherche s'appuie sur une méthodologie qui exploite les bases de données fournies par la Direction Générale des Routes et des Transports Terrestres (DGRTT). Ces bases de données comprennent des tables d'enregistrements du trafic provenant de diverses stations de comptage routier. Chaque station est associée à plusieurs tables qui détaillent le nombre de véhicules de chaque catégorie circulant sur chaque voie de la route. Pour mener cette étude, une sélection minutieuse des stations est opérée, et leurs données sont divisées en deux séries distinctes : la première série correspond à la première voie, tandis que la seconde série concerne la seconde voie. Chaque série représente le nombre total de véhicules de

toutes les catégories enregistrées par heure. Trois méthodes bien connues dans la littérature ont été utilisées pour la prédiction des données incomplètes du trafic routier. À savoir : la moyenne mobile intégrée autorégressive saisonnière (SARIMA), la régression à vecteurs de support (SVR) et la régression linéaire. L'objectif de notre étude est de proposer une solution au problème des données manquantes d'un dispositif défectueux ou hors service pendant une période de maintenance en s'appuyant sur les données des dispositifs de comptage voisins.

3.2. Prédiction à court terme du futur trafic routier

Cette section présentera de manière générale la méthodologie que nous avons élaborée afin de procéder à des prévisions à court terme du trafic routier ; cette approche reposant sur l'exploitation des données historiques de circulation provenant d'un compteur permanent installé à la ville de Tanger sur la route nationale 1. Dans le cadre de la méthodologie employée, la variable à prédire est le trafic routier de l'ensemble des catégories de véhicules circulant sur toutes les voies de la route.

Pour enrichir notre modèle prévisionnel, nous incorporerons diverses variables temporelles, telles que le jour de la semaine, l'heure du jour et les jours de vacances, ainsi que des variables retardées par jour, semaine et mois. Après traitement de données, le modèle Extreme Learning Machine (ELM) a été entraîné, testé et comparé avec des modèles bien connus dans la littérature, notamment la moyenne mobile intégrée autorégressive, la régression à vecteurs de support et le perceptron multicouche. Notre démarche prévisionnelle s'ancre fermement dans l'utilisation judicieuse des données historiques du trafic, permettant d'anticiper avec précision les évolutions futures.

3.3. Analyse des données d'accidentologie

Dans notre thèse, nous avons mené une série d'études visant à analyser les points noirs d'accidents de la route à partir des données d'accidents du Maroc. Chaque étude a été réalisée de manière systématique en suivant un processus de traitement des données standardisé, depuis la collecte des données par des services de collecte de données spécialisés jusqu'à la classification des sections routières en fonction de leur gravité (figure 1).

Dans un premier temps, les variables spatiales décrivant les conditions routières et environnementales ont été soigneusement collectées. Dans cette phase, les routes ont été divisées en sections d'un kilomètre, en se concentrant uniquement sur les accidents mortels. Ensuite, les décès et les blessures survenus dans chaque section ont été méticuleusement enregistrés en conjonction avec les

variables spatiales correspondantes. Le Risque d'Accident (RA) a ensuite été évalué pour chaque section à l'aide d'une formule simple qui traduit le nombre de décès en nombre total d'accidents. Cet indice a été utilisé pour entraîner un modèle Extreme Gradient Boosting (XGBoost) avec les types de victimes comme données d'entrée. Le choix du nombre de types de décès dépendait du délai entre l'accident et le décès. Les blessures ont été divisées en deux catégories dans toutes les études, tandis que les tués ont été divisés en 2 ou 4 catégories, selon le cas. Lorsque deux catégories ont été définies, la variable tuée a été divisée en « tués immédiat » et « tués à la suite des blessures ». Dans le cas de quatre catégories, les tués ont été divisés en « tués sur le champ », « tués pendant le transfert à l'hôpital », « tués dans les 7 jours » et « tués dans 30 jours de l'accident ».

Le modèle XGBoost a ensuite déterminé les poids associés à chaque entrée. L'Indice de Gravité Pondéré (IGP) a ensuite été déterminé pour chaque section en multipliant chaque type de victime par le poids qui lui est associé. En conséquence, chaque section de route s'est vu attribuer une valeur IGP unique, ce qui a permis de classer les sections en fonction de leur niveau de danger. Enfin, un système de catégorisation en quatre classes basées sur la moyenne et l'écart-type des valeurs IGP a été construit, permettant de définir les catégories suivantes : gravité élevée, moyenne, faible et sûre. Cette méthodologie générale expose les éléments partagés entre les diverses études menées au sein de cette thèse. La démarche spécifique de chaque étude est ensuite exposée en détail dans le chapitre qui lui est dédié.

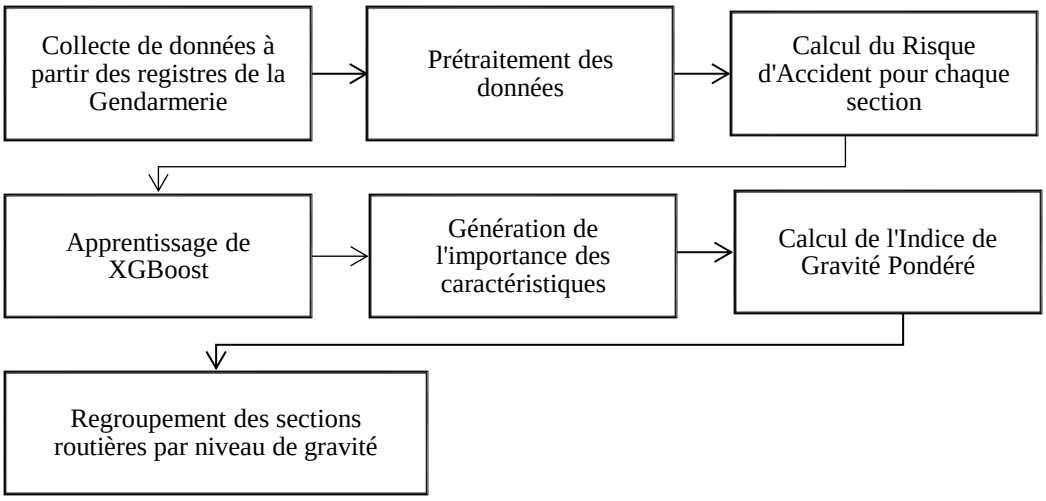


Figure 1. Méthodologie générale utilisée dans l'analyse des points noirs des accidents

4. Structure de la thèse

Subséquemment, la composition de notre thèse englobe une exploration détaillée de diverses problématiques.

Le premier chapitre de notre mémoire se consacre à l'introduction des concepts fondamentaux liés aux STI, à l'apprentissage automatique, et à la sécurité routière. Une analyse approfondie sera entreprise afin de comprendre de quelle manière les STI sont mise à contribution pour optimiser la gestion du trafic et renforcer la sécurité routière. De plus, une révision exhaustive des études antérieures portant sur l'application de ces technologies au Maroc sera effectuée. Ce chapitre constitue une base conceptuelle essentielle, établissant le cadre nécessaire pour l'exploration des aspects liés à la sécurité routière et à l'utilisation des technologies émergentes.

Le deuxième chapitre sera consacré à la gestion des données du trafic routier, une dimension cruciale pour analyser le comportement du trafic. Une attention particulière sera accordée à une étude spécifique visant à estimer les données manquantes, avec une mise en évidence des méthodes utilisées à cet effet. Par ailleurs, une discussion approfondie sur la gestion des données manquantes viendra compléter cette analyse. Le chapitre accorde une attention particulière aux méthodes d'estimation, exposant les étapes impliquées et évaluant les résultats. Cette analyse constitue une étape essentielle pour développer une méthodologie solide dans la gestion des données de trafic, préparant ainsi le terrain pour les futures analyses de la thèse.

Quant au troisième chapitre, ce dernier se concentrera sur l'application du modèle ELM pour la prédiction du trafic routier à court terme. Une étude de cas sera présentée, détaillant un modèle basé sur l'ELM pour anticiper le trafic routier sur la route nationale N1. La méthodologie de recherche, incluant le prétraitement des données et les méthodes de prédiction, sera exposée. Une analyse comparative sera réalisée pour évaluer les performances du modèle par rapport à d'autres approches. Cette évaluation critique vise à déterminer l'efficacité de l'ELM en tant qu'outil de prédiction du trafic routier. En fin de compte, ce chapitre aspire à contribuer à l'avancement des connaissances dans le domaine de la prédiction du trafic routier en exploitant les capacités de l'ELM.

Le quatrième chapitre de la thèse porte sur l'évaluation statistique de l'impact de l'infrastructure routière sur la létalité routière. En mettant en lumière les facteurs d'infrastructure présumés contribuant aux accidents, une analyse approfondie des statistiques des accidents mortels sera entreprise. Les

méthodes d'analyse statistique seront présentées en détail, soulignant la rigueur méthodologique de l'évaluation. Les résultats seront soigneusement examinés, fournissant des conclusions enrichissantes pour la compréhension des relations complexes entre l'infrastructure et la sécurité routière. Ancré dans une méthodologie robuste, ce chapitre vise à éclairer les stratégies futures en matière de sécurité routière.

Le cinquième chapitre sera dédié à l'identification des points noirs des accidents routiers à travers l'IGP. En s'appuyant sur une analyse approfondie de la littérature, un modèle novateur sera proposé pour détecter ces points noirs. La méthodologie, comprenant la préparation des données et le calcul du poids de l'IGP, sera détaillée, mettant l'accent sur l'identification des tronçons routiers à haut risque. Le processus de calcul de l'IGP et son application pour repérer les zones propices aux accidents seront expliqués en détail. Les résultats seront discutés, offrant des implications approfondies pour la sécurité routière. Ce chapitre vise à apporter une contribution significative à l'identification et à la gestion des points noirs des accidents routiers.

Enfin, notre sixième chapitre examinera les facteurs de risque liés aux points noirs des accidents, en utilisant les règles d'association et en prédisant ces points sur les routes rurales à travers le modèle ELM. La méthodologie comprend une analyse approfondie des composants d'infrastructure via les règles d'association pour évaluer l'influence de chaque élément sur la formation des points noirs. Les étapes, y compris la préparation des données, le calcul de l'IGP, l'utilisation des règles d'association et l'entraînement du modèle ELM, seront détaillées. Les résultats seront discutés en détail, offrant une compréhension approfondie des facteurs de risque. Une comparaison approfondie des performances du modèle ELM sera réalisée pour évaluer sa pertinence et son efficacité par rapport à d'autres stratégies de prédiction.

En guise de conclusion, cette thèse présente une synthèse des résultats les plus importants et met en évidence leurs contributions substantielles. Une analyse critique des lacunes de l'étude permettra d'identifier des axes potentiels d'amélioration dans la gestion et la sécurité du trafic routier. Cette section offrira une réflexion globale sur les enseignements tirés de la thèse, mettant en avant les découvertes clés et les avancées méthodologiques. Des recommandations stratégiques seront formulées pour orienter des pistes de recherche futures afin de combler les lacunes identifiées et d'approfondir la compréhension des dynamiques du trafic routier.

Chapitre 1. Systèmes de transport intelligents : concepts et aperçus au Maroc

1. Introduction

Le domaine de la sécurité routière, par sa nature vaste et complexe, implique la collaboration de divers secteurs, notamment le secteur décisionnel, de construction et de gestion. Ces domaines d'intervention convergent vers un objectif commun ; à savoir la préservation de la vie humaine et la prévention d'irrégularités telles que la congestion routière et les accidents de la circulation. La coordination entre ces secteurs revêt une importance cruciale pour garantir un environnement routier sécurisé et efficace.

Dans ce chapitre, nous exposerons les concepts fondamentaux dans le domaine de la sécurité routière, notamment les systèmes de transport intelligents qui proposent des solutions avancées, ainsi que l'apprentissage automatique qui facilite le processus décisionnel. En outre, nous examinerons en détail les concepts clés relatifs au trafic routier et à l'infrastructure routière. Ces éléments revêtent une importance cruciale en offrant des solutions novatrices pour résoudre les défis complexes liés à la mobilité et à la sécurité routière. Dans un premier temps, nous nous intéresserons aux principes essentiels des systèmes de transport intelligents, en mettant en lumière leurs diverses applications dans les problématiques liées à la sécurité routière. Ensuite, nous explorerons les bases de l'apprentissage automatique et ses applications spécifiques dans le contexte marocain en matière de sécurité routière. Par la suite, nous explorons les concepts clés liés au trafic routier, aux infrastructures routières et aux accidents de la circulation. In fine, nous présentons un ensemble de mesures simples, issues de la théorie du trafic, qui peuvent être utilisées pour évaluer le trafic routier, ainsi que les dynamiques et les mécanismes de congestion. De plus, nous exposons les éléments essentiels de l'infrastructure routière, qui sont indispensables à tout projet routier, ainsi que les facteurs de risque associés aux accidents de la route.

Ainsi, ce chapitre met en lumière les éléments cruciaux liés à la sécurité routière et à l'apprentissage automatique. De nombreux éléments discutés dans ce chapitre auront des implications directes ou indirectes dans les chapitres subséquents. L'exposition des concepts fondamentaux dans ce chapitre est destinée à faciliter la compréhension des travaux de recherche à venir, que ce soit dans le

domaine du prétraitement des données, de la modélisation, de l'interprétation des résultats ou la formulation des recommandations.

2. Systèmes de transport intelligents

2.1. Avantages des systèmes de transport intelligents

Vers les années 1960, la communauté scientifique a commencé à se mobiliser pour lutter contre les effets négatifs de l'urbanisation, qui entraînait souvent une hausse des demandes d'infrastructures de transport ; ceci se traduisant par la surcharge des routes et la longueur des trajets. Par conséquent, une augmentation de la consommation de carburant et du niveau de pollution de l'air est constatée, contribuant ainsi à la hausse des concentrations de gaz à effet de serre. Pour atténuer ces conséquences, les autorités ont opté pour des systèmes intelligents pour réduire les effets suscités. Par exemple la mise en place des systèmes de gestion de la congestion qui permet d'améliorer la qualité de vie des usagers de la route.

Pour faire face aux défis actuels des transports, les parties prenantes du secteur comme les gestionnaires, opérateurs et ministères de tutelle, se tournent désormais vers les STI. Il s'agit d'une combinaison de technologies de l'information et de communication avancées qui offrent la possibilité de transformer la manière dont les réseaux routiers et autoroutiers sont habituellement exploités et gérés. Ils sont utilisés pour assurer la sécurité, l'efficacité et la durabilité des réseaux de transport routier, et pour améliorer l'expérience de conduite à travers la réduction des embouteillages. Ces systèmes sont dits « intelligents » parce qu'ils utilisent des technologies avancées pour l'analyse de données. Ils sont capables de collecter et de traiter des informations en temps réel afin de prendre des décisions éclairées et d'optimiser le flux de véhicules et de personnes. Ces données concernent, entre autres, l'état de la circulation, les conditions météorologiques ainsi que toute information qui a un impact sur la circulation routière. Aussi, ces systèmes intègrent des caractéristiques telles que des feux de circulation intelligents, des véhicules connectés et des systèmes de péage automatisés, qui contribuent tous à un réseau de transport plus intelligent et interconnecté.

2.2. Applications des systèmes de transport intelligents

Notons que les STI se retrouvent dans divers domaines d'activité, notamment l'optimisation de l'utilisation des infrastructures de transport, l'amélioration de la sécurité routière et le développement de divers services. Ces nouveaux systèmes contribuent à la maîtrise de la mobilité en favorisant, entre autres,

l'abandon de la voiture au profit de modes de transport plus respectueux de l'environnement. Grâce aux STI, les systèmes d'information avancés aux passagers et les systèmes conçus pour le confort et le divertissement ont connu un développement important.

Les STI sont en ce sens une composante d'un concept global plus vaste qui comprend à la fois la congestion du trafic routier et le développement de nouvelles technologies de l'information, en particulier dans le domaine des communications et de la connectivité. Effectivement, ils permettent de gérer et d'optimiser le trafic routier en utilisant les données en temps réel sur les conditions de circulation, les capteurs de véhicules et les systèmes de communication avancés. Parallèlement, les nouvelles technologies de l'information telles que l'internet des objets (IoT) et la 5G permettent une connectivité améliorée entre les véhicules, les infrastructures et les utilisateurs. Les principales applications des STI qui contribuent à l'exploitation des réseaux routiers sont, entre autres, comme suit :

- La gestion du trafic et du réseau routier.
- Les systèmes d'information des voyageurs.
- Les systèmes de transport public.
- Les applications pour les véhicules commerciaux.
- Les applications relatives à la sécurité des véhicules.
- Les applications de gestion de la maintenance et de la construction.
- La gestion des urgences.
- La gestion des données archivées.

2.3. Utilisation des systèmes de transport intelligents dans la sécurité routière

Dans cet environnement, les autorités de transport, en qualité d'acteurs centraux dans le domaine des systèmes de transport intelligents, portent la responsabilité primordiale de garantir la sécurité de ces systèmes. À cet égard, il est impératif qu'elles intensifient leurs efforts de recherche, entreprennent des études approfondies sur les comportements des usagers et consacrent des investissements substantiels à l'amélioration de la sécurité et des STI. Cette démarche s'inscrit dans une perspective globale visant à renforcer l'intégrité, la fiabilité et la sécurité de ces derniers.

De nombreuses études sur la sécurité routière ont été menées à l'échelle mondiale. Cependant, en raison du manque d'études dans ce domaine au Maroc, les chercheurs et les universitaires ont investi davantage de temps et d'efforts pour améliorer l'expérience des usagers de la route dans le pays. Les STI

et les réseaux ad hoc véhiculaires (VANET) figurent parmi les technologies les plus utilisées dans ces initiatives. Dans une contribution récente [4], certains auteurs ont présenté un système d'aide à la conduite basé sur les systèmes multiagents. Ce système vise à améliorer l'adaptation de la vitesse lors des virages en fournissant aux conducteurs les informations nécessaires pour calculer la limite de vitesse à ne pas dépasser, assurant ainsi un équilibre optimal dans les virages. Les auteurs de [5], ont élaboré un nouveau cadre intitulé Smart Digital Logistic Services Provider (Smart DLSP) visant à améliorer la communication entre les véhicules. Ce cadre permet d'identifier les risques ayant un impact sur la qualité de service (QoS) du système de communication sans fil VANET en utilisant des approches telles que le réseau bayésien et la logique floue. L'article présente également l'architecture d'Ecodid, un écosystème logistique intelligent, collaboratif et hautement connecté qui exploite des techniques d'intelligence artificielle et des réseaux de véhicules pour optimiser l'efficacité des livraisons et influencer les facteurs économiques pour les industries. L'étude citée dans [6] aborde le développement d'un système de transport intelligent basé sur des capteurs souples intelligents destinés à analyser le trafic routier et à assister le comportement du conducteur, en particulier dans le contexte des villes intelligentes. Les auteurs ont conceptualisé un système d'assistance qui surveille la position du véhicule, dirigé par un capteur de vision (caméra) permettant une perception de l'environnement. Le système repose sur l'utilisation de capteurs intelligents intégrant des techniques de traitement d'image et d'intelligence artificielle pour détecter les marquages routiers et estimer l'angle de braquage, fournissant ainsi un appui au conducteur.

La catégorisation précise des véhicules demeure néanmoins cruciale pour les systèmes de transport intelligents, car elle contribue substantiellement à l'optimisation de la gestion et de la sécurité du trafic routier. De ce fait, diverses stratégies novatrices ont été élaborées et rendues publiques dans le but d'améliorer l'efficacité de l'identification et de la classification des véhicules en temps réel. Les auteurs de [7] ont élaboré un réseau de neurones convolutif (CNN) dédié à la classification en temps réel des véhicules d'urgence. L'objectif ultime sous-tendant cette démarche étant d'améliorer la préservation de vies grâce à une classification efficace. La méthodologie adoptée consiste dans ce cas en l'évaluation de huit architectures de réseaux CNN, en vue de classifier les véhicules d'urgence dans un contexte de surveillance du trafic routier en temps réel. Les résultats démontrent que, parmi les modèles examinés, DenseNet121 se distingue comme le classifieur optimal en termes de précision. Par ailleurs, un autre article propose une nouvelle approche inspirée de la mécanique quantique pour la classification des types de véhicules en contexte de streaming vidéo [8]. La méthodologie proposée se structure autour de trois

phases principales : l'acquisition et le prétraitement du flux vidéo, la détection et le suivi des véhicules, ainsi que la classification des véhicules suivis. Cette dernière étape englobe l'extraction de caractéristiques à l'aide de l'histogramme de gradient orienté (HOG) et de l'Analyse en Composantes Principales (ACP). Elle est suivie par l'entraînement d'un classificateur à perceptron multicouche (MLP) visant à estimer les probabilités associées à chaque classe. La désignation des classes s'effectue par le biais d'un intégrateur de probabilités d'inspiration quantique, permettant la fusion des informations provenant de chaque image. L'efficacité de cette approche a été démontrée par des performances supérieures aux méthodes conventionnelles, atteignant une précision allant jusqu'à 96 %.

L'utilisation répandue de techniques d'apprentissage automatique dans le domaine de la sécurité routière a été remarquable. Des applications significatives incluent la prédiction du trafic routier, la détection d'obstacles, de panneaux de signalisation, de nids-de-poule, ainsi que l'identification du comportement du conducteur, entre autres. En la matière, les auteurs référencés dans [9] ont formulé un modèle de prédiction du trafic routier basé sur l'apprentissage fédéré. Ce modèle innovant présente l'avantage de maintenir la confidentialité des données utilisées pendant la phase d'apprentissage, ce qui se traduit par un accès restreint des utilisateurs aux seules informations du modèle entraîné. Le système proposé a démontré une grande précision dans ses prédictions tout en préservant la vie privée et les informations personnelles des utilisateurs. Dans le cadre de la référence [10], une méthodologie améliorée a été avancée par les auteurs, s'appuyant sur l'intelligence artificielle pour la détection des panneaux de signalisation routière. Cette approche intègre diverses techniques, notamment la transformation de forme, la sélection de couleur, la détection des contours, ainsi que des méthodes fondées sur les CNN. Dans le cadre de l'étude présentée dans [11], les chercheurs ont préconisé l'adoption du modèle nommé Residual Balanced Attention Network (RBANet), un réseau de neurones convolutif robuste et léger, pour effectuer la segmentation sémantique des scènes de trafic en temps réel. La segmentation sémantique des scènes de trafic se révèle essentielle pour des applications tangibles telles que le suivi du flux de circulation, l'estimation de la vitesse des véhicules, la gestion des panneaux de signalisation et la détection des dangers sur la voie publique. L'évaluation du RBANet a été conduite sur le jeu de données Cambridge-Driving Labeled Video Database (CamVid), et les résultats obtenus ont démontré une performance satisfaisante en termes de requis en paramètres et de précision comparativement aux travaux connexes. Les rédacteurs de l'article [12] ont formulé un cadre conceptuel intitulé Explained Deep Learning LiDAR-Based Localization (XDLL), représentant une méthode de localisation et de cartographie. XDLL repose sur les principes de l'apprentissage profond et utilise la

technologie Light Detection and Ranging (LiDAR) pour anticiper la position des véhicules, notamment dans le contexte des véhicules autonomes. Le modèle élaboré démontre une capacité à atteindre une erreur de positionnement absolue moyenne inférieure à un mètre, ce qui se traduit par une amélioration significative de la localisation des véhicules, notamment dans des environnements présentant des restrictions d'accès.

Par ailleurs, l'IoT a été examiné dans le contexte des Systèmes de Transport Intelligents. Les auteurs de la référence [13] ont élaboré une méthode consistant à déployer un système intelligent de surveillance du conducteur fondé sur les technologies de l'IoT afin de détecter et d'identifier les comportements de conduite agressifs. Ce système collecte diverses données, notamment la vitesse, le couple du moteur et les coordonnées géographiques, qu'il transmet ensuite à un serveur web. Ces données sont ensuite interprétées et affichées au moyen d'une interface graphique, permettant ainsi la surveillance à distance et l'archivage de l'historique des déplacements et de la conduite du véhicule. Les résultats expérimentaux mettent en évidence l'efficacité de l'approche proposée pour la surveillance du comportement de conduite.

Parallèlement, diverses stratégies ont été développées afin de faciliter la prise de décision pour les acteurs impliqués dans le choix de la trajectoire optimale des véhicules connectés. Dans l'étude référencée sous [14], une application mobile a été suggérée pour la prédiction du chemin le plus court pour les taxis, reposant sur la détermination des itinéraires les moins encombrés. L'algorithme A* a été appliqué pour identifier les parcours les plus efficaces sur les cartes, permettant ainsi aux conducteurs d'économiser du temps en évitant les congestions. Les auteurs mentionnés dans [15] ont présenté un nouveau système visant à optimiser la collecte du lait en suivant le trajet le plus court et en réduisant les coûts associés. Cette approche exploite les technologies de l'IoT et du Big Data pour améliorer la prise de décision des entreprises laitières lors de la collecte auprès des fournisseurs de lait. En outre, l'article [16] préconise une stratégie efficace pour la planification des voies de communication entre les véhicules en temps réel. Cette méthode repose sur une approche coopérative de partage d'informations entre les véhicules et les unités de bord de route, avec pour objectif principal la réduction des coûts et de la redondance des communications.

3. Apprentissage automatique et sécurité routière

3.1. Concepts clés

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle qui se concentre sur la création de modèles informatiques et d'algorithmes capables d'apprendre et de s'améliorer à partir de données. Ces algorithmes sont conçus pour analyser de grandes quantités de données, identifier des modèles et faire des prédictions ou prendre des décisions. En fait, l'apprentissage automatique « apprend aux ordinateurs à apprendre » et, à leur tour, à agir et à réagir de la même manière que les humains.

In fine, les techniques d'apprentissage automatique utilisent des programmes qui s'adaptent chaque fois qu'ils sont exposés à différents types de données d'entrée. Ce processus est connu sous le nom d'« apprentissage », car le programme affine continuellement sa compréhension et ses performances au fil du temps. En analysant les modèles et les relations dans les données d'entrée, les algorithmes d'apprentissage automatique peuvent identifier des tendances et faire des prédictions de plus en plus précises. Au fur et à mesure que les données sont fournies, l'algorithme acquiert une meilleure connaissance et peut prendre des décisions plus éclairées, ce qui permet d'obtenir de meilleurs résultats et d'améliorer l'efficacité. Ce processus itératif d'apprentissage permet à l'algorithme de s'adapter et d'évoluer en fonction des nouvelles informations qu'il reçoit. En outre, les algorithmes d'apprentissage automatique peuvent également détecter les anomalies et les valeurs aberrantes dans les données, ce qui peut s'avérer utile pour identifier les erreurs ou les modèles inhabituels qui peuvent nécessiter une enquête plus approfondie. Les algorithmes d'apprentissage automatique constituent ainsi une approche incontournable visant à automatiser le processus de prise de décision. Ces algorithmes se répartissent principalement en quatre catégories distinctes, à savoir l'apprentissage supervisé, semi-supervisé, non supervisé et par renforcement.

3.1.1. Apprentissage automatique supervisé

Dans l'apprentissage automatique, le modèle le plus fréquemment utilisé est l'apprentissage supervisé. Celui-ci implique des variables d'entrée et une variable de sortie, ainsi que la fonction de correspondance entre l'entrée et la sortie. Afin de prédire les variables de sortie pour de nouvelles données d'entrée, cet algorithme tente d'approximer la fonction de correspondance. Il s'agit du processus d'acquisition des connaissances par un algorithme à partir d'un ensemble de données d'apprentissage qui sert de fondement à cette procédure. Dans ce cadre, le résultat associé à chaque entrée est préalablement connu. L'algorithme effectue des prédictions itératives sur l'ensemble des données d'apprentissage et est

soumis à des corrections par un « enseignant ». Le processus d'apprentissage persiste jusqu'à ce que l'algorithme atteigne un niveau de performance jugé acceptable, à partir duquel le processus d'apprentissage est conclu. Selon le type de données de sortie, les algorithmes d'apprentissage supervisé sont généralement utilisés pour les problèmes de classification et de régression. Par la suite, la classification est utilisée pour les données qui peuvent être classées, séparées en groupes ou classes spécifiques, tandis que les techniques de régression sont efficaces pour prédire des données continues. Dans l'apprentissage supervisé, les algorithmes courants comprennent la régression logistique, les bayes naïfs, les machines à vecteurs de support, les réseaux de neurones artificiels et les forêts aléatoires.

3.1.2. Apprentissage automatique non supervisé

La distinction fondamentale entre l'apprentissage supervisé et non supervisé réside dans le fait que ce dernier ne comporte pas de variable de sortie préalablement définie ; ce qui implique une exploration des données sans connaissance a priori des résultats. L'apprentissage non supervisé, dans ce contexte, s'articule autour de l'objectif d'acquérir une compréhension approfondie des données en modélisant leur structure ou distribution sous-jacente. Dès lors, cette démarche confie aux algorithmes la tâche d'explorer et de mettre en évidence la structure significative des données. Toutefois, il est important de noter que l'apprentissage non supervisé, en comparaison avec son homologue supervisé, se caractérise par une complexité informatique élevée et une moindre précision dans les prédictions.

Les problématiques de regroupement (clustering), de réduction de dimensionnalité et d'association constituent des domaines d'application privilégiés pour les algorithmes d'apprentissage non supervisé. Parmi ces algorithmes, on retrouve fréquemment le K-means, K-medians, la maximisation de l'espérance, et le regroupement hiérarchique, qui sont largement employés pour les tâches de regroupement. En ce qui concerne la réduction de la dimensionnalité, l'ACP émerge comme une méthode prioritaire visant à projeter les données dans une dimension réduite tout en maximisant la variance. Par ailleurs, dans le domaine des algorithmes d'association, des méthodes telles que Apriori et Eclat sont souvent employées ; ils utilisent une approche de recherche en profondeur pour déterminer les éléments fréquents au sein des données.

3.1.3. Apprentissage automatique semi-supervisé

L'apprentissage semi-supervisé repose sur une combinaison de l'apprentissage supervisé et non supervisé, utilisée lorsque la quantité de données étiquetées disponibles est limitée. Il vise à utiliser toutes les données étiquetées pour classer certaines données non étiquetées. L'ensemble de données non

étiquetées doit être beaucoup plus important que l'ensemble de données étiquetées. Dans le cas contraire, des techniques d'apprentissage supervisé peuvent aussi être utilisées pour résoudre le problème. Plusieurs algorithmes sont utilisés dans le cadre de l'apprentissage semi-supervisé, parmi lesquels on retrouve l'algorithme Label Propagation, qui consiste à trouver des similarités entre les données étiquetées et non étiquetées, l'algorithme Trivial Co-Training consistant à entraîner deux algorithmes sur les données étiquetées afin d'échanger les informations sur les données non étiquetées.

3.1.4. Apprentissage par renforcement

L'apprentissage par renforcement constitue une technique d'apprentissage automatique caractérisée par la présence de quatre composants fondamentaux, à savoir l'environnement, l'agent, l'action, et la récompense. L'environnement, qui peut prendre la forme d'une simulation ou d'une situation réelle, interagit avec l'agent en acceptant les actions de ce dernier et en fournissant une réponse correspondante. L'agent, quant à lui, représente l'entité responsable des actions au sein de l'environnement. La récompense constitue le signal essentiel utilisé pour évaluer la pertinence de l'action précédemment entreprise par l'agent.

Dans ce contexte, l'agent interagit avec l'environnement par le biais d'une approche itérative impliquant des essais, des erreurs, et la réception de rétroactions. Il est important de souligner que ce dernier ne reçoit pas d'instructions explicites sur les actions à entreprendre ; au contraire, il apprend progressivement quelles actions génèrent les récompenses les plus favorables au travers de l'expérience et des itérations. L'objectif ultime de l'agent résidant dans la maximisation de sa récompense, conduisant à une préférence pour certaines actions plutôt que d'autres, basée sur les résultats obtenus au cours de ses interactions passées avec l'environnement.

3.2. Études antérieures pour l'apprentissage automatique dans la sécurité routière

L'analyse et la prédiction des flux de trafic occupent désormais une position centrale au sein des systèmes de transport intelligents. Diverses méthodes d'apprentissage et de prédiction du trafic ont été développées, parmi lesquelles figurent des heuristiques fondées sur une analyse statistique des données historiques. Ces approches, telles que l'algorithme k plus proches voisins (KNN), la régression linéaire et non linéaire, ainsi que les modèles linéaires autorégressifs comme la moyenne mobile intégrée autorégressive (ARIMA), sont reconnues pour leur simplicité et leur praticité dans le contexte de la prédiction du trafic [17][18].

Au sein des méthodes statistiques, les modèles ARIMA sont théoriquement considérés comme la catégorie la plus générale pour la prédiction des données dans le contexte des séries temporelles. Leur potentiel et leur efficacité résident dans leur aptitude à anticiper de manière substantielle les données futures, tirant avantage de leur praticité, de la précision inhérente à leurs prédictions, d'une exigence modérée en termes de données, et de la simplicité de leur calcul [19][20]. Concernant les défis liés aux problèmes de régression, le modèle de régression à vecteur de support a démontré son efficacité en tant qu'outil puissant pour l'estimation des données réelles [21]. En qualité d'algorithme d'apprentissage supervisé au sein du domaine de l'apprentissage automatique, le modèle SVR se distingue par l'utilisation d'une fonction de perte symétrique, laquelle inflige des pénalités aux décisions se situant aux limites.

Récemment, les réseaux neuronaux artificiels (ANN) ont suscité un vif intérêt dans le domaine de la prédiction du trafic. Certains éléments historiques tels que les horaires scolaires, les jours de congé, et d'autres événements périodiques sont exploités pour anticiper le volume de circulation au niveau de chaque intersection significative de la ville, en utilisant des données historiques et recourant aux capacités des réseaux neuronaux artificiels [22]. Les chercheurs ont par ailleurs expérimenté diverses configurations de réseaux neuronaux profonds dans le but de modéliser et prédire les flux de trafic [23][24]. Dans l'étude référencée sous [24], un modèle de réseau neuronal profond a été développé en vue de prédire l'évolution future du flux de trafic. Ainsi, la structure proposée comprend un classifieur multiclasses de réseau neuronal profond, désigné comme le « Deep Neural Network Multi-Class Classifier » (DNNC). Les données sous-jacentes à cette étude englobent les flux de trafic enregistrés au cours des années 2016 et 2017. Ce modèle, s'appuyant sur la somme des volumes de trafic horaires, procède à la classification du trafic routier en quatre catégories (très faible, faible, moyen, élevé, très élevé) en fonction du niveau de trafic horaire. Dans ces conditions, les auteurs de l'article [25] exposent un système de classification des types de véhicules exploitant la technologie de l'apprentissage profond, mettant en œuvre le traitement d'images et la reconnaissance des formes. Ce système se structure autour de deux phases distinctes : l'augmentation des données, visant à traiter les ensembles de données déséquilibrés, et la construction d'un modèle CNN avec différentes architectures. L'application résultante offre une gestion automatisée des feux de circulation, reposant sur la détection automatique du type de véhicule.

À n'en pas douter, l'augmentation constante du nombre d'accidents sur les voies de circulation a induit une hausse proportionnelle du nombre de victimes humaines, stimulant ainsi l'adoption généralisée

de caméras dans le cadre des activités quotidiennes liées à la gestion du trafic routier. L'étude référencée sous [26] se focalise sur la prédiction de la gravité des accidents de la circulation et évalue l'impact potentiel de ces incidents sur la logistique du transport d'urgence en milieu urbain. En ce sens, l'analyse procède à une comparaison entre plusieurs algorithmes d'apprentissage automatique, dans le but de classer et prédire la gravité des accidents routiers. Les résultats obtenus indiquent que l'algorithme de Forêt Aléatoire atteint une précision de 93 % dans la classification des accidents. Ainsi, les auteurs de l'étude référencée sous [27] introduisent un cadre de fusion d'ensemble destiné à la prédiction en temps réel des accidents de la route, en mettant l'accent sur l'intégration de quatre catégories de caractéristiques en temps réel, à savoir les signaux physiologiques, les manœuvres du conducteur, la cinématique du véhicule et les variables météorologiques. Précisons que ce cadre adopte une approche orientée des données, générant une diversité dans les combinaisons d'apprenants telles que les apprenants bayésiens, le k plus proche voisin (k-NN), la machine à vecteur de support (SVM) et le perceptron multicouche. L'ensemble de fusion atteint un score F1 maximal de 93,66 % avec le classifieur Naïf Bayésien, les réseaux bayésiens, k-NN, SVM et MLP en tant que métaclassificateur. Les chercheurs de l'étude référencée sous [28] ont investigué les accidents de la route impliquant des conducteurs novices, notamment dans des scénarios de visibilité réduite. La méthodologie adoptée visant à évaluer l'impact des précipitations légères et fortes sur la sécurité routière, en utilisant des ANN et des arbres de décision. Les conclusions démontrent la supériorité des modèles ANN par rapport aux arbres de décision en termes de précision et d'AUC pour l'ensemble des variables météorologiques examinées.

De son côté, le modèle VANET est un réseau sans fil prometteur pour la communication entre les véhicules en mouvement, dans le but d'améliorer la sécurité sur les routes. Les protocoles de routage de clustering jouent un rôle crucial dans la distribution efficace de la bande passante et l'évolutivité. Toutefois, ces algorithmes souffrent souvent de l'instabilité du réseau et d'une connectivité discontinue. Dans l'étude citée sous [29], un algorithme récent de regroupement a été conçu pour les réseaux VANET ; lequel repose sur la mobilité et la fiabilité des véhicules. Ce modèle utilise un système de réseau de neurones artificiel distribué afin de former des regroupements. L'algorithme prend en considération plusieurs paramètres tels que le taux de fiabilité, la vitesse, la différence de distance entre les nœuds, et le degré d'apprentissage pour la sélection d'un chef de cluster, communément appelé Cluster-Head (CH). Les résultats expérimentaux démontrent toutefois que le protocole proposé surpasse les solutions existantes en termes de durée de vie du CH, de durée de vie des clusters, de pourcentage de sélection des véhicules anormaux, de bande passante, et de délai de bout en bout. Les auteurs mentionnés dans la

référence [30] ont développé un nouveau cadre conceptuel en s'appuyant sur l'apprentissage par renforcement au moyen de réseaux Deep-Q afin de réguler les déplacements des véhicules autonomes dans un contexte complexe impliquant la présence simultanée de véhicules et de piétons. Le véhicule autonome assimile une politique d'actions diverses visant à atteindre sa destination de manière sécurisée. La méthodologie proposée est évaluée au moyen de simulations et est entraînée sur le simulateur CARLA. Les résultats démontrent la capacité du modèle à acquérir des connaissances au fil des épisodes, et l'approche suggérée se traduit par des performances améliorées en termes de récompense moyenne, de taux de réussite, et de taux de collision. Dans leur étude, les chercheurs référencés sous [31] ont formulé un système destiné à surveiller les manœuvres de conduite aux intersections routières, faisant appel à une logique basée sur des règles et à des techniques de vision par ordinateur intégrant l'apprentissage profond. Ce système procède à la détection et à la classification des mouvements de virage en ligne, tout en identifiant des infractions telles que l'omission de respecter les panneaux-STOPS et la négligence du droit de priorité vis-à-vis des autres usagers de la route. À cet égard, une méthode de détection d'objets est proposée, reposant sur un réseau CNN à coût modéré. Cette approche intègre une logique de raisonnement, la soustraction de l'arrière-plan, et un détecteur d'objets basé sur le CNN.

4. Concepts clés du trafic routier et de la sécurité routière

Le trafic routier reste néanmoins un phénomène complexe qui a demandé de nombreuses années pour le comprendre et le quantifier. La congestion est l'événement le plus marqué du trafic routier, ce qui a suscité de nombreuses recherches de la part des experts. Nous présentons dans cette section un ensemble de mesures simples, issues de la théorie du trafic, qui peuvent être utilisées pour mesurer le trafic routier et mieux comprendre les principales dynamiques sous-jacentes à l'apparition de la congestion.

4.1. Mesures fondamentales

4.1.1. Densité

La densité du flux de trafic (concentration), notée K, constitue une caractéristique spatiale déterminant le niveau de contrainte du trafic ainsi que la charge de la voie de circulation. Celle-ci s'évalue en fonction du nombre de véhicules circulant sur une section de route donnée. La mesure de la densité permet d'évaluer l'état de la circulation. Dans ce cadre, une valeur élevée de densité peut signaler une situation congestionnée, tandis qu'une valeur basse peut indiquer une circulation fluide. Le calcul de la densité s'effectue en considérant le ratio de véhicules par unité de longueur, exprimé en véhicules par

kilomètre (Veh/km) ou en véhicules par mètre (Veh/m). Si M est le nombre de véhicules existants entre x et $x + \Delta x$ à un instant t, alors la densité est exprimée par :

$$K_{\Delta x}(t) = K(x \to x + \Delta x, t) = \frac{M}{\Delta x} \quad (1)$$

4.1.2. Vitesse moyenne de libre circulation

La vitesse moyenne de libre circulation est la vitesse à laquelle les véhicules peuvent circuler en douceur et sans restriction sur une route sans être perturbés par des embouteillages. C'est la vitesse à laquelle un véhicule peut garder une distance de sécurité avec le véhicule qui le précède sans avoir à freiner ou à ralentir brusquement. Comme la circulation est souvent affectée par des ralentissements, des arrêts fréquents, la vitesse moyenne de circulation est généralement plus élevée que la vitesse moyenne réelle enregistrée dans des conditions de circulation réelles. Cette vitesse est essentielle pour la planification des infrastructures routières et l'évaluation de l'efficacité du trafic. La vitesse moyenne de libre circulation est déterminée par la formule :

$$\bar{V} = \frac{\sum_1^n V_i}{n} \quad (2)$$

Où V_i est la vitesse instantanée de la $i^{ème}$ voiture dans cette section, mesurée en km/h ; et n est le nombre de voitures dont les vitesses sont mesurées.

4.1.3. Vitesse moyenne temporelle

Pour sa part, la vitesse moyenne temporelle (v_t) est une mesure de la vitesse des véhicules qui « passent » en un capteur spécifique de la route pendant une période spécifique. Cette mesure est obtenue en prenant la moyenne arithmétique de toutes les vitesses observées pendant cette période. Il est important de noter que la vitesse moyenne temporelle peut varier considérablement en fonction de l'heure de la journée, de la densité de la circulation et d'autres facteurs. En outre, soulignons que la vitesse moyenne temporelle est souvent utilisée comme indicateur de la qualité de la circulation routière et peut être utile pour évaluer l'efficacité des politiques de transport et d'autres initiatives visant à améliorer la mobilité urbaine. La vitesse moyenne temporelle est calculée comme suit :

$$v_t = \frac{1}{n}\sum_1^n v_i \quad (3)$$

Où v_i est la vitesse du i^{ème} véhicule, et n est le nombre de véhicules.

4.1.4. Vitesse moyenne spatiale

Enfin, la vitesse moyenne spatiale (v_s) est une valeur qui mesure la moyenne harmonique des vitesses des véhicules passant en un point précis de la route au cours d'une période définie. Cette mesure est utilisée pour évaluer la fluidité du trafic sur une route ou une autoroute, ce qui peut être utile pour la planification et la gestion des flux de circulation. En général, une vitesse moyenne spatiale élevée indique une fluidité de trafic élevée, tandis qu'une vitesse moyenne spatiale faible peut indiquer des congestions ou des ralentissements sur la route. Il est donc important de surveiller régulièrement la vitesse moyenne spatiale pour assurer une circulation efficace et sécurisée sur les voies de circulation. La vitesse moyenne spatiale est calculée par la formule suivante :

$$v_s = \frac{n}{\sum_1^n \left(\frac{1}{v_i}\right)} = \frac{nL}{\sum_1^n t_i} \tag{4}$$

Où n est le nombre de véhicules, v_i est la vitesse du i^{ème} véhicule (m/s), L est la longueur de la section (m), $t_i$est le temps mis par le i^{ème} véhicule pour traverser un tronçon de route (s).

4.1.5. Flux

Le flux (débit) du trafic Q est le nombre de véhicules qui passent par un point donné de la route pendant une période spécifique. Généralement, il exprime le trafic horaire et est mesuré par le nombre de véhicules par heure pour exprimer le taux de passage à un moment précis. Le flux du trafic permet de détecter les heures de congestion, il est utilisé pour la conception des routes et pour la planification urbaine ainsi pour le contrôle de la circulation.

Il est exprimé en nombre de véhicules par unité de temps (généralement par Veh/h ou Veh/s). Soit N le nombre de véhicules qui passent par un point donné x pendant une période définie Δt, le flux est exprimé comme suit :

$$Q_{\Delta t}(x) = Q(x, t \to t + \Delta t) = \frac{N}{\Delta t} \tag{5}$$

4.1.6. Temps du trajet

Le temps de trajet, souvent appelé « temps de déplacement » ou « durée du voyage », est le temps total nécessaire pour parcourir un itinéraire ou une distance spécifique entre deux points sur une route ou

dans un réseau de transport. Ce temps comprend le temps de déplacement ainsi que tout temps d'arrêt causé par le trafic, les feux de circulation, les arrêts de bus, les embouteillages, les conditions météorologiques, etc. Il peut varier considérablement en fonction des variables telles que la distance parcourue, la vitesse moyenne du véhicule, la densité du trafic et les conditions routières et météorologiques. Ce temps peut être exprimé en minutes, en heures ou en toute autre unité de temps.

Concrètement, la mesure du temps de trajet est essentielle pour analyser les performances des réseaux de transport, développer des itinéraires appropriés, prévoir les horaires de déplacement et prendre des décisions éclairées en matière de gestion du trafic. Ce paramètre est largement utilisé par les systèmes de navigation GPS et les applications cartographiques pour offrir aux utilisateurs des prévisions en temps réel des temps de trajet en fonction des conditions de circulation actuelles.

4.1.7. Capacité de la route

La capacité de la route représente le nombre maximal de véhicules pouvant franchir une section donnée de la route dans une unité de temps donnée. Il est toutefois essentiel de reconnaître que les caractéristiques intrinsèques des véhicules et des conducteurs peuvent influer sur le débit de circulation autant que les paramètres routiers eux-mêmes. Les conditions météorologiques exercent manifestement une influence significative sur la capacité réelle de la route, laquelle tend à diminuer, notamment en cas de fortes pluies, de brouillard, de fortes chutes de neige et de verglas.

Dès lors, la méthode la plus simple pour calculer la capacité est basée sur un modèle dynamique simplifié qui considère le flux comme une série de voitures similaires uniformément réparties sur la voie de la route. Cette méthode prend en considération la distance de sécurité, le temps de réaction du conducteur (environ 1 s) et la différence des décélérations maximales sur une chaussée sèche en asphalte-béton lors d'un freinage d'urgence de voitures particulières similaires (environ 2 m/s^2).

4.1.8. Facteurs temporels

Les facteurs temporels du trafic routier font référence aux éléments qui exercent une influence sur la congestion du trafic en fonction de l'heure de la journée. Ils déterminent la manière dont les caractéristiques du trafic, telles que la densité, le débit et la vitesse, varient au fil du temps. Ces facteurs temporels induisent des fluctuations temporaires dans le contexte du trafic routier, parmi ces facteurs on retrouve :

- Les heures de pointe : ce sont les moments de la journée où les embouteillages sont les plus importants. Cette période dépend de la ville ou de la région, mais se situe généralement en début et en fin de journée, lorsque les gens se rendent au travail et en reviennent.
- Les jours de la semaine : ils exercent également une influence significative sur l'intensité du trafic routier. En semaine, on observe généralement une augmentation du volume de circulation par rapport aux jours fériés ou aux fins de semaine.
- Les événements : les événements spéciaux tels que les concerts, des rencontres sportives ou les festivals peuvent également provoquer des embouteillages. Ces événements ont la capacité d'engendrer des ralentissements significatifs sur les voies de circulation en raison de l'afflux massif des participants et des activités connexes qui peuvent perturber le flux habituel du trafic.
- Les conditions météorologiques : celles-ci peuvent également avoir une incidence sur les embouteillages. Par exemple, des phénomènes tels que la neige ou le verglas peuvent occasionner des retards et des accidents, contribuant ainsi à la formation d'embouteillages.

Les facteurs temporels sont essentiels à la planification des infrastructures routières, à la gestion du trafic et à l'élaboration de stratégies de contrôle du trafic. Comprendre l'évolution de l'intensité du trafic au fil du temps permet aux autorités compétentes d'anticiper efficacement les périodes de congestion, de moduler les heures de travail, de réduire le flux de trafic et de développer des solutions visant à optimiser le flux de trafic et à améliorer la mobilité.

4.2. Diagramme fondamental du trafic routier

4.2.1. Principes du diagramme fondamental

Le diagramme fondamental du trafic constitue une illustration graphique de la relation entre la densité du trafic, la vitesse des véhicules et le flux du trafic sur une route ou une autoroute. Cette relation est fondamentale pour comprendre le comportement du transport dans différentes conditions et son évolution en fonction de la densité du trafic. Ainsi, il aide les planificateurs, les ingénieurs et les chercheurs à développer des infrastructures routières plus efficaces et à améliorer la gestion du trafic. En général, le diagramme montre la densité du trafic (habituellement en nombre de voitures par unité de distance) sur l'axe vertical et la vitesse sur l'axe horizontal. Le flux (débit) de trafic, qui est le nombre

de voitures traversant une section de route par unité de temps, est directement lié à la densité (concentration) du trafic et à la vitesse (voir figure 2).

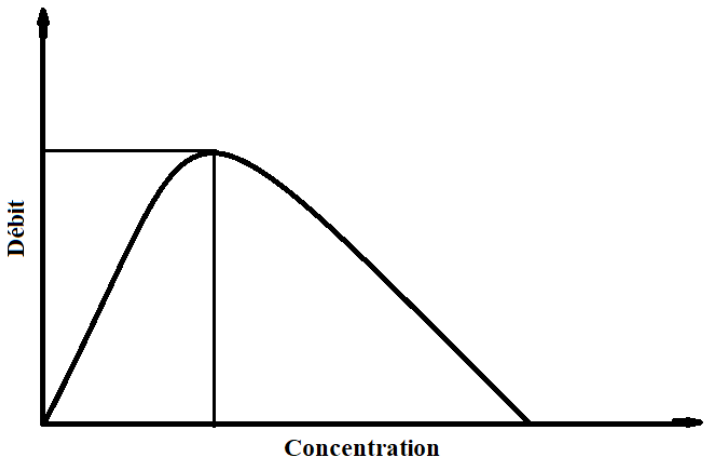


Figure 2. Diagramme fondamental du trafic routier

Les caractéristiques du diagramme fondamental varient en fonction de la densité du trafic. Lorsque la densité du trafic est faible, la vitesse des véhicules est généralement élevée et le trafic est relativement faible. À mesure que la densité du trafic augmente, la vitesse diminue en raison de l'interaction entre les voitures. Il en résulte une densité critique à laquelle la vitesse est minimale et le débit est maximal. Au-delà de cette densité critique, si la densité du trafic augmente encore, la vitesse diminue encore et la capacité commence à baisser en raison de la congestion et de l'interaction entre les véhicules. Cette zone de trafic dense est de nature à provoquer des embouteillages et un trafic lent et instable.

4.2.2. Dynamiques du trafic routier

Au niveau macroscopique, le trafic routier est décrit par les variables du flux (débit) routier, de la densité (concentration) routière et de la vitesse moyenne. La définition de ces trois variables du trafic routier est comme suit :

- Débit Q : Définit le nombre de véhicules N passant par un point donné dans un intervalle de temps donné T. Le débit est exprimé en Véhicules/h et explique l'intensité du trafic routier. Le débit d'une section routière est calculé par la formule suivante :

$$Q = \frac{N}{T} \tag{6}$$

- Concentration K : Aussi appelée densité, elle décrit la quantité de véhicules occupant une zone spécifique à un moment donné. Elle définit la répartition des véhicules dans une section et s'exprime en termes de nombre de véhicules par kilomètre (Veh/km).
- Vitesse U : Elle caractérise la moyenne des vitesses de véhicules occupants à un moment donné une zone spécifique de la route et s'exprime en kilomètres par heure (km/h).

Ces trois variables sont interdépendantes selon la relation suivante :

$$Q = K \times U \tag{7}$$

La figure 3 illustre le débit en fonction de la concentration. Elle établit que lorsque la concentration augmente, le débit s'accroît également, cette relation est presque linéaire dans la zone de trafic fluide, tan que la capacité maximale de la route n'a pas encore été atteinte. Au-delà du point critique, la congestion apparaît. Le trafic routier connait une limitation du débit entraînant une diminution du débit accompagné par une augmentation de la concentration.

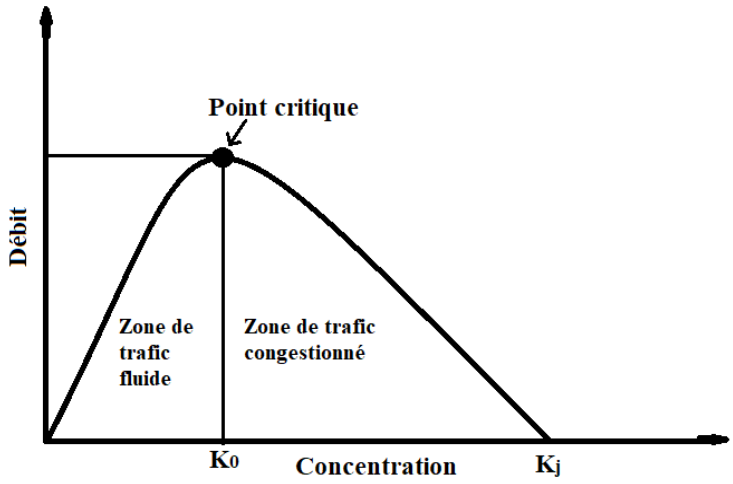


Figure 3. Diagramme fondamental : débit en fonction de concentration

Dans la zone de trafic fluide, le débit part de zéro jusqu'à la densité critique. Un débit nul signifiant l'absence de véhicules sur la route. La valeur maximale du débit est la valeur critique qui correspond au nombre maximal de véhicules qui peuvent se déplacer à la vitesse libre. De ce fait, lorsque le nombre de véhicules augmente, la vitesse commence à diminuer et la circulation à vitesse libre n'est plus assurée. Dans la zone de trafic congestionnée, le débit diminue de la valeur critique à zéro. Dans ce

cas, la concentration atteint la valeur maximale, soit le cas où la route est remplie au maximum de sa capacité et où aucun véhicule ne circule (leur vitesse est nulle).

La figure 4 montre le diagramme fondamental montrant le graphique de la vitesse en fonction du débit. Nous y constatons que la vitesse égale la vitesse libre Uf dans la zone de trafic fluide. Lorsque la densité atteint la valeur critique K0, la route se rapproche de sa capacité maximale Qm et la vitesse du flux de trafic diminue jusqu'à atteindre U0. Lorsque la concentration atteint le seuil de congestion Kj, tant le débit que la vitesse se réduisent à zéro. La circulation est considérée comme saturée lorsque les concentrations dépassent la valeur critique.

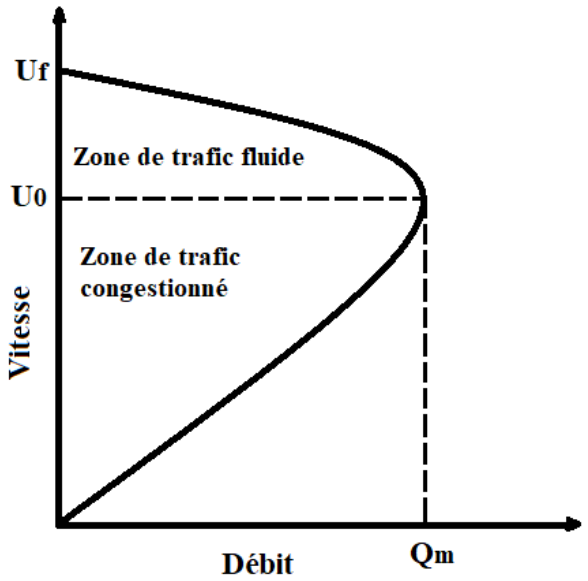


Figure 4. Diagramme fondamental : vitesse en fonction du débit

La figure 5 présente un diagramme essentiel qui représente la relation entre la vitesse et la concentration à travers un graphique. Dans la zone de trafic fluide, la concentration est faible et les véhicules circulent librement avec une vitesse élevée. Au point critique, la concentration augmente et la vitesse commence à diminuer à cause de la saturation de la route. Dans la zone de trafic congestionné, la concentration atteint sa valeur maximale et la vitesse s'annule, les véhicules sont bloqués et ne peuvent plus circuler, à ce point la circulation est congestionnée.

4.3. Infrastructure routière

4.3.1. Fondements essentiels des infrastructures

Par nature, les infrastructures de transport comprennent tous les équipements fixes qui permettent aux voitures de se déplacer et, plus généralement, aux systèmes de transport de fonctionner efficacement.

Ces infrastructures vont des routes, autoroutes et ponts aux chemins de fer, aéroports et ports maritimes. En outre, elles englobent également les installations de base nécessaires, telles que les stations-service, les dépôts d'entretien et les systèmes de contrôle du trafic. Ces éléments vitaux travaillent ensemble pour faciliter la circulation des biens et des personnes, favorisant la croissance économique et la connectivité au sein des régions et entre les nations.

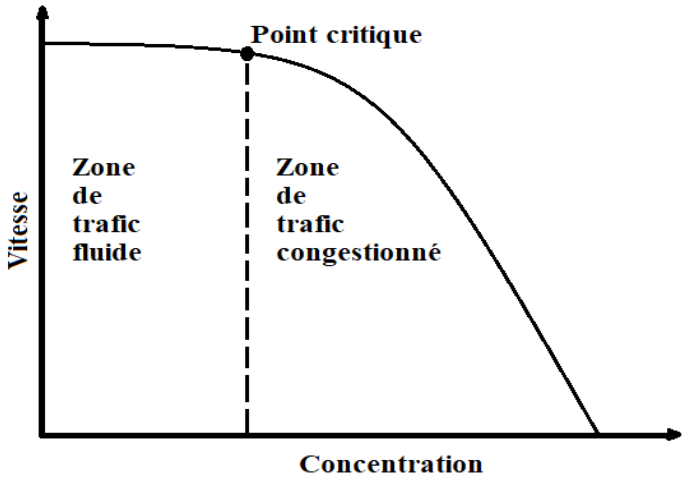


Figure 5. Diagramme fondamental : vitesse en fonction de la concentration

Par définition, une route demeure un moyen de transport qui relie différents endroits et permet la circulation des véhicules, des piétons et des marchandises. Elle est généralement constituée d'une surface dure en asphalte ou en béton et est équipée de divers éléments tels que des voies, des panneaux de signalisation et des signaux afin de garantir un transport sûr et efficace. Les routes reliant souvent diverses destinations, telles que les villes et les zones rurales constituent un lien essentiel pour les déplacements domicile-travail et le transport de marchandises. Ce composant joue un rôle essentiel dans la facilitation du commerce, du tourisme et des besoins quotidiens en matière de transport, ce qui en fait un élément essentiel d'un système d'infrastructure de transport bien développé.

4.3.2. Types de routes au Maroc

Concernant le système routier du Royaume du Maroc, ce dernier est structuré en quatre classifications, sous l'administration du ministère de l'Équipement et de l'Eau : les autoroutes, les routes nationales, les routes régionales et les routes provinciales. Ce réseau joue un rôle prépondérant en facilitant la mobilité, supportant approximativement 90 % du flux de passagers et 75 % du transport de

marchandises[2]. Cette segmentation hiérarchique permet une gestion efficace des différents niveaux de trafic, contribuant ainsi à la fluidité et à la pérennité des déplacements sur le territoire marocain.

Les routes nationales sont reconnaissables à leurs bornes kilométriques rouges et leur nom commence par N. Elles appartiennent à l'état via le ministère chargé des transports. Le réseau routier national est constitué de routes reliant les principaux pôles du pays, permettant des échanges d'importance nationale ou internationale jusqu'aux frontières avec les pays voisins. Il comprend également des routes ayant des implications importantes en termes de développement régional. En 2021, ce réseau comprend 15 619 km, soit environ 27 % du réseau routier total, avec un taux de revêtement d'environ 87,6 %[3].

Les routes régionales se définissent comme les voies de circulation localisées à l'intérieur des municipalités, caractérisées par leur capacité étendue. Ces infrastructures ont pour vocation première de canaliser le flux de trafic en direction des autoroutes, constituant ainsi un maillage essentiel au sein du réseau routier global. En 2021, la longueur totale de ce réseau s'élevait à 11 199 km, dont 9 813 km étaient revêtus[4], démontrant l'engagement envers le développement et l'entretien de ces voies cruciales pour la connectivité interurbaine. L'importance stratégique des routes régionales est manifeste dans leur rôle de facilitateur de la circulation, participant ainsi à la fluidité des déplacements sur le territoire et au désenclavement de certaines régions.

Les voies provinciales constituent un réseau routier affilié aux différentes provinces, relevant de la responsabilité des collectivités locales pour ce qui est de la planification, de la construction et de l'entretien. En vertu de cette attribution, les autorités provinciales assument le rôle crucial de gestion de ces infrastructures routières. En 2018, ce réseau s'étendait sur une distance totale de 30 513 km, dont 20 684 km pourvus d'un revêtement[5], soulignant ainsi l'importance accordée à l'amélioration et à la durabilité de ces voies provinciales. L'engagement des collectivités locales dans la préservation et l'optimisation de ces liaisons routières est manifeste à travers ces chiffres, témoignant de leur contribution significative à la connectivité et à la mobilité à l'échelle provinciale.

[2] Ministère de l'équipement et de l'eau, https://www.equipement.gov.ma/Infrastructures-Routieres/Reseau-Routier-du-Royaume/Pages/Importance-du-reseau.aspx.
[3] Idem.
[4] Idem.
[5] Idem.

4.3.3. Description des éléments de l'infrastructure routière

Les infrastructures de transport constituent un élément nécessaire et indispensable de la société moderne. Elles permettent non seulement la circulation des personnes et des marchandises, mais facilitent également la croissance et le développement économiques. D'évidence, sans une infrastructure de transport bien développée, les entreprises auraient du mal à communiquer avec leurs clients, les chaînes d'approvisionnement seraient perturbées et les individus auraient du mal à accéder à des services et des opportunités vitaux.

Ces infrastructures sont généralement organisées en réseaux composés de nœuds, de plateformes et de corridors. Les nœuds sont des points de connexion au sein du réseau, tels que les gares ou les aéroports. Les plateformes sont de grands centres de transport où les différents modes de transport se croisent, ce qui permet de passer d'un itinéraire à l'autre en toute fluidité. Les corridors sont les itinéraires qui relient ces nœuds et ces plateformes et qui permettent la circulation des personnes et des marchandises sur l'ensemble du réseau. La gestion efficace des nœuds, des plateformes et des corridors est essentielle pour garantir des services de transport fluides et fiables, au bénéfice des particuliers et des entreprises. Parmi les éléments de l'infrastructure routière, nous citerons :

- Les accotements qui sont les parties d'une route située entre les limites de la chaussée.
- Les trottoirs qui constituent un élément de liaison essentiel des réseaux piétonniers, car ils permettent d'isoler le piéton des dangers de la circulation routière.
- La signalisation verticale : dont les panneaux de police (danger) et les panneaux directionnels. Elle est destinée à informer les usagers de la route de manière suffisante pour leur permettre de se déplacer en toute sécurité d'un point à un autre.
- La signalisation routière horizontale : composée de marquages au sol. Elle est de couleur blanche. On distingue les types suivants :
    - La ligne continue : dépassement strictement interdit.
    - La ligne discontinue : possibilité de franchir la ligne.
    - Les lignes mixtes : peuvent être traversées du côté de la ligne discontinue.
    - Les passages piétons : un passage piéton est une zone désignée sur une route ou à une intersection où les piétons peuvent traverser en toute sécurité. Il est généralement signalé par des lignes blanches ou des bandes zébrées.

- L'éclairage public couvre tous les systèmes d'éclairage utilisés dans les espaces publics tels que les rues, les parcs, les places et les bâtiments publics. Ces systèmes d'éclairage sont conçus pour assurer la visibilité et la sécurité des piétons et des automobilistes, ainsi que pour améliorer l'attrait esthétique des zones environnantes.

4.3.4. Description des paramètres de conception géométrique des routes

Les caractéristiques de la route, leurs interactions avec le conducteur ou le véhicule, la visibilité de la route et l'homogénéité de l'itinéraire sont également des éléments importants pour la sécurité routière. Les conducteurs sont souvent impactés par les éléments qui constituent l'environnement de la circulation. Afin de réduire ou d'éliminer l'incertitude et l'inattendu pour les conducteurs, et donc de ne pas briser leurs attentes, il est nécessaire de considérer la cohérence des éléments de conception géométrique de la route.

Dans ces conditions, les conducteurs engagent des interactions en fonction des conditions environnementales, de sorte que tant que les paramètres techniques sont favorables, les interactions conducteur-environnement demeurent propices. À ce titre, les autorités responsables de la sécurité routière soulignent l'importance de la sûreté des éléments d'infrastructure afin de minimiser, voire d'éliminer les accidents. La conception constitue le fondement de la construction et de la reconstruction des voies de circulation. La qualité du projet de construction influe directement sur la fiabilité, la durabilité et la sécurité routière des voies à venir. Parmi les aspects de conception de l'infrastructure routière figurent des éléments tels que la visibilité, le profil en long, et le tracé en plan, etc.

4.3.4.1. Visibilité

Afin d'assurer la sécurité et le confort des usagers de la route, la conception géométrique des routes doit permettre une visibilité optimale, aussi bien pour les points singuliers que pour les sections droites. Les concepteurs routiers doivent ainsi trouver un juste équilibre entre les besoins en visibilité et les contraintes liées au projet. La distance de visibilité requise constitue un élément essentiel en vue d'assurer la sécurité routière. Précisons que cette distance est tributaire de divers paramètres, dont la vitesse de déplacement et la distance requise pour accomplir une manœuvre. Par exemple, en cas de freinage brusque, un espace d'arrêt plus étendu est nécessaire à des vitesses élevées. De plus, la distance de visibilité nécessaire peut varier en fonction des conditions météorologiques, exigeant un temps de réaction et un espace augmenté par temps pluvieux ou neigeux. Ainsi, la conception géométrique des routes doit prendre en compte ces multiples facteurs pour assurer une sécurité routière optimale.

4.3.4.2. Profil en travers

Le profil en travers d'une route revêt une importance significative en tant qu'élément de conception géométrique crucial visant à garantir la sécurité routière et la fluidité de la circulation. Il se présente comme une représentation en réduction du profil transversal de la route, établie par un plan vertical perpendiculaire à l'axe de la voie. En ce sens, la représentation graphique du profil en travers doit mettre en lumière divers composants routiers tels que le nombre de voies, la largeur des voies, les accotements, les trottoirs, et le terre-plein central. Aussi, ces éléments revêtent-ils une importance fondamentale dans la conception routière en contribuant à assurer la sécurité des usagers de la route.

- Le nombre de voies : leur nombre nécessaire pour une route est déterminé après une analyse minutieuse des données de trafic, des objectifs de niveau de service et des éléments économiques. Cette analyse implique plusieurs étapes, telles que la collecte et l'analyse de données de trafic en temps réel, la prise en compte des tendances de trafic futures, ainsi que l'identification des objectifs de niveau de service pour la route en question.
- La largeur des voies : c'est un élément crucial dans la conception des routes, car elle dépend du type de route et de son usage. En effet, les routes urbaines sont généralement plus étroites que les routes rurales, afin de s'adapter à l'environnement urbain et à la circulation locale. De plus, la largeur des voies doit également prendre en compte les exigences de sécurité, telles que la visibilité, la distance de freinage et le nombre de voies.
- Les accotements ce sont des parties importantes de toute route, car ils sont cruciaux pour la sécurité routière. En effet, ils sont utilisés pour plusieurs raisons, notamment pour l'arrêt temporaire des véhicules en cas d'urgence ou de panne, pour le placement de matériaux destinés à la réparation des routes, et pour permettre aux véhicules de ralentir ou de s'arrêter afin de laisser passer des piétons ou des cyclistes. La largeur des accotements varie en fonction de la largeur de la route et peut donc être différente d'une route à l'autre.
- Les trottoirs : ils constituent un élément clé des réseaux piétonniers, car ils offrent une séparation essentielle entre le piéton et les dangers de la circulation routière. Effectivement, il est important de prendre en compte plusieurs facteurs lorsqu'il s'agit de déterminer la largeur d'un trottoir. Parmi ces facteurs, on peut notamment citer la fréquentation piétonne du trottoir et la vitesse maximale autorisée sur la route qui longe le trottoir.
- Le terre-plein central : cette partie de la route sépare les chaussées indépendantes les unes des autres. Ce séparateur empêche les mouvements de franchissement hasardeux des véhicules et

d'opérer un demi-tour de l'autre côté, ce qui pourrait perturber la circulation sur les voies en sens inverse. Les caractéristiques du terre-plein central dépendent principalement de l'environnement, des usages de la route et de la vitesse autorisée.

Nonobstant ces différents composants, le profil en travers doit mettre en évidence les différents éléments de la route pour les usagers. En outre, il est également important d'inclure des caractéristiques telles que les marquages routiers, les feux de signalisation, les panneaux de signalisation et les dispositifs de sécurité pour assurer une conduite sûre et fluide sur la route. Par conséquent, en concevant un profil en travers efficace, les ingénieurs en génie civil peuvent garantir une infrastructure routière de qualité supérieure pour tous les usagers de la route.

4.3.4.3. Tracé en plan

Le tracé en plan est un composant crucial dans la conception des routes. Il s'agit là de la projection verticale de la route sur un plan horizontal. Il est composé d'une série de courbes, de lignes droites et d'accordements progressifs qui peuvent être séparés ou non. Le but de ce composant est de garantir de bonnes conditions de sécurité et de confort pour les usagers, tout en s'intégrant harmonieusement dans la topographie du site. Ainsi, il doit prendre en compte les caractéristiques du terrain, comme les pentes, les courbes de niveau, les obstacles naturels et artificiels, pour offrir une route fluide et adaptée au site. En outre, il doit répondre à des normes et à des réglementations en matière de sécurité routière, de signalisation, d'éclairage et d'accessibilité pour les personnes à mobilité réduite.

4.3.4.4. Profil en long

Le profil en long est une représentation graphique hypothétique d'une section d'une route par un plan vertical qui passe par son axe. Il est composé avec des composantes rectilignes qui se caractérisent par leur déclinaison (c'est-à-dire la pente ou la rampe), ainsi que des raccordements circulaires ou paraboliques qui sont caractérisés par leur rayon. Ces éléments peuvent être utilisés pour déterminer la configuration de la route, comme les zones à forte pente ou les virages serrés, et pour concevoir des routes plus sûres et plus efficaces. La création d'un profil en long précis représente une étape importante dans la planification et la conception de routes, car cela permet de s'assurer que la route est adaptée aux besoins des usagers et qu'elle est construite pour durer.

Aussi, est-il primordial de s'assurer de la coordination adéquate entre le tracé en plan et le profil en long pour garantir des conditions de visibilité optimales. Pour les routes neuves, il est également important de prêter attention à l'esthétique du tracé, en évitant de créer des angles trop brusques ou des

discontinuités qui pourraient perturber les conducteurs. À cette fin, il est généralement nécessaire de faire coïncider les courbes du tracé en plan et du profil en long, en prévoyant des rayons de courbure plus importants pour le profil en long que pour le tracé en plan.

4.4. Aspects clés de la sécurité routière

4.4.1. Points noirs des accidents

Un point noir est une section de route qui a enregistré un nombre anormalement élevé d'accidents et de victimes au cours d'une période donnée. L'identification d'un point noir est conditionnée par le dépassement de certains seuils quantitatifs en ce qui concerne le nombre d'accidents enregistrés et le nombre de victimes impliquées. Cette caractérisation met en relief la nécessité de définir des critères objectifs et quantifiables afin d'isoler les tronçons de route considérés comme des points noirs.

Au Maroc, les autorités considèrent qu'une portion de route d'un (1) kilomètre ou plus ayant enregistré au moins dix (10) accidents ayant fait plus de 10 morts ou blessés graves au cours des 5 dernières années est un point noir. Cette définition tient compte de la gravité des accidents, mais elle peut sembler trop restrictive dans certains cas. Pour cette raison, les autorités ont opté pour un lissage de 500 mètres entre chaque 2 km suivant, avec un chevauchement de 500 m entre deux zones de calcul. Concrètement, cela signifie que plutôt que de diviser les routes en sections de seulement 1 kilomètre, les autorités ont choisi d'élargir la distance entre les points d'analyse à 2 km. Ensuite, ils étendent cette distance en ajoutant 500 mètres à chaque extrémité des sections de 2 km. En outre, le chevauchement de 500 mètres entre les sections signifie que les sections de 2 km se superposent sur une distance de 500 mètres. Cette approche permet de prendre en compte les accidents qui se produisent à proximité des points noirs, et d'identifier les sections de route qui présentent un risque élevé pour les usagers. Cette définition concise et cohérente autorise ainsi les instances compétentes à concentrer leurs initiatives de sécurité routière sur les zones identifiées comme les plus vulnérables, dans le but de diminuer l'incidence d'accidents et le nombre de victimes à l'échelle de l'ensemble du réseau routier.

4.4.2. Taux d'accidents

Le taux d'accidents est un indicateur crucial pour évaluer la gravité des points noirs. Effectivement, il permet de mesurer le nombre d'accidents qui se produisent par an pour 100 millions de kilomètres parcourus. Cela quantifie la probabilité qu'un accident se produise sur un site donné, ce qui constitue une mesure du risque individuel. Par ailleurs, ce taux peut également être utilisé pour évaluer l'efficacité des mesures de sécurité routière mises en place sur un site. Si le taux d'accidents diminue

après la mise en place de ces mesures, cela peut indiquer que les mesures sont efficaces. En outre, le taux d'accidents peut être utilisé pour comparer différents sites et identifier ceux qui présentent le plus de risques pour les conducteurs. Pour ce faire, il est important d'analyser le taux d'accidents sur une période donnée pour identifier les tendances et les évolutions à long terme. Cela permet de mettre en place des mesures préventives pour réduire les risques d'accident à l'avenir. Le taux d'accidents s'évalue au moyen de la formule suivante :

$$t = \frac{N \times 10^8}{L \times T \times 365 \times n} \quad (8)$$

Où N est le nombre d'accidents pendant n années, L est la longueur de la section étudiée en kilomètres ; T est le trafic moyen journalier annuel en véhicules/jour, et n est le nombre d'années étudiées.

4.4.3. Facteurs accidentogènes

Les facteurs accidentogènes des accidents sont les facteurs qui contribuent à la survenue d'accidents de la route graves et récurrents dans une zone géographique donnée. Ils incluent la densité du trafic, la conception de la route, la signalisation inadéquate, l'état de la route, la présence de zones de construction ou d'autres obstacles, et le comportement des conducteurs tels que la vitesse excessive ou la conduite sous l'influence de drogues ou d'alcool. L'identification et la compréhension de ces facteurs sont cruciales pour mettre en place des mesures de sécurité routière efficaces visant à réduire le nombre d'accidents de la route et à améliorer la sécurité de tous les usagers de la route.

Force est de conster que la sécurité routière est un sujet complexe et multidimensionnel qui implique des éléments de l'environnement physique tels que l'infrastructure routière, l'environnement naturel et le trafic ainsi que des éléments mécaniques tels que les véhicules et des facteurs humains comme les usagers de la route. À cet égard, des études sur les accidents ont montré le rôle important de l'environnement routier sur les accidents de la circulation et leur gravité. Ainsi, il est crucial d'adopter des aménagements qui permettent de guider les actions, de contraindre les comportements et d'offrir un environnement de conduite sûr. Par conséquent, il est essentiel de tenir compte de tous les éléments qui peuvent influencer la sécurité routière et de mettre en place des mesures pour prévenir les accidents de la route. Un système routier sûr est un système prenant en considération les besoins des usagers de la route qui est continuellement évalué et amélioré pour garantir la sécurité de tous.

5. Conclusion

Dans ce chapitre, nous avons présenté une vue d'ensemble approfondie des STI, de l'apprentissage automatique, du trafic routier et de la sécurité routière, en mettant l'accent sur le contexte marocain. Nous avons examiné les bénéfices des STI ainsi que leurs applications en termes de sécurité routière, soulignant leur importance cruciale dans l'assurance d'une circulation sécurisée et d'une mobilité durable. Plusieurs applications visant à améliorer les STI ont été identifiées dans la littérature, avec des implications variées pour le contexte marocain, incluant notamment l'assistance à la conduite, la communication inter-véhiculaire et des applications dans divers secteurs comme le secteur commercial et le domaine médical. L'intégration de ces solutions innovantes joue un rôle fondamental dans l'optimisation de la sécurité routière en réduisant les incidents tels que la congestion et les accidents, facilitant ainsi une gestion efficace du flux des usagers de la route, des véhicules et des marchandises, tout en minimisant les risques d'accidents ou d'autres anomalies.

Nous avons également abordé les concepts fondamentaux de l'apprentissage automatique en relation avec la sécurité routière. Cette technologie devient de plus en plus cruciale à intégrer dans les solutions relatives à la sécurité routière, du fait qu'elle permet d'explorer de vastes ensembles de données en temps réel de manière robuste et fiable. Cette approche améliore les interventions, l'utilisation des ressources, la fluidité du trafic et la préservation des vies. Plusieurs applications de l'apprentissage automatique ont été présentées, notamment la détection des panneaux de signalisation, des comportements agressifs, des divers utilisateurs de la route, la prédiction du trafic routier, des accidents, ainsi que l'assistance à la conduite, entre autres.

Nous avons établi une base solide pour la compréhension des chapitres à venir en ce qui concerne le trafic routier et les accidents de la circulation. Les principes fondamentaux de la circulation et de la sécurité routières ont été explorés, mettant en avant les mesures essentielles du trafic routier et le diagramme fondamental de la circulation, qui illustre la corrélation entre ces différentes mesures. De plus, l'infrastructure routière du Maroc a été examinée en détail, étant un élément crucial de la sécurité routière. Les composantes de cette infrastructure ont été clairement définies, et les aspects essentiels de la sécurité routière, tels que les points noirs des accidents, les taux d'accidents et les facteurs contribuant aux accidents, ont été présentés de manière approfondie.

Chapitre 2. Collecte des données de trafic routier et gestion des données manquantes

1. Introduction

Nul doute que la croissance démographique urbaine et l'accroissement de la mobilité de la population ont engendré une expansion rapide du parc automobile, induisant ainsi une augmentation substantielle de la circulation routière. La nécessité de réguler le flux ordinaire de personnes et de marchandises a conduit à l'impératif de gestion du trafic routier. Toutefois, cette gestion est actuellement confrontée à des défis tels que les embouteillages et les accidents. En réponse à ces problèmes, les professionnels du domaine ont élaboré des stratégies pour les réseaux routiers, axées sur la planification des itinéraires et le contrôle du trafic. L'élément central de l'élaboration de telles stratégies réside dans l'analyse des données en constante augmentation relatives au trafic, ainsi que dans la mise en place d'un système de prévision du trafic d'une grande précision. Aussi, la prédiction des flux de trafic implique -t-elle l'anticipation de l'état subséquent (volume, vitesse, densité, ou comportement) du trafic, en se basant sur des données historiques et/ou en temps réel. Elle représente une composante essentielle des STI et s'avère indispensable dans le processus de prise de décision.

Au cours des récentes années, la gestion mondiale du trafic routier a connu une transformation significative, attribuable à l'émergence de technologies innovantes telles que l'IoT, les systèmes de transport intelligents et les systèmes informatiques décisionnels. De manière continue, de nouveaux composants sont intégrés dans les systèmes de transport à l'échelle mondiale, et des approches novatrices sont déployées pour exploiter et maintenir ces systèmes de manière sécurisée et efficiente. L'objectif prédominant de ces systèmes visant à faciliter la mise en œuvre des principales stratégies de transport. Les STI englobent la collecte, le traitement et l'analyse de données en vue de fournir un outil de prise de décision efficace [24]. L'évolution de ces systèmes repose considérablement sur la qualité et la quantité des données liées au trafic routier. Généralement, les informations relatives à la circulation, telles que la vitesse des véhicules ou le flux de circulation, sont recueillies au moyen de capteurs stationnaires positionnés le long du réseau routier [32].

Dans ce cadre, la collecte de données du trafic routier constitue une étape cruciale dans l'analyse et la gestion des flux routiers. Elle se révèle être une source fondamentale pour la compréhension des

schémas de trafic, la prévision des conditions de circulation, et la mise en œuvre de stratégies visant à améliorer l'efficacité et la sécurité du réseau routier. L'acquisition systématique de ces données, provenant de points spécifiques du réseau routier, permettant d'alimenter des modèles analytiques et des outils de simulation, contribuant ainsi à une prise de décision éclairée. La précision et la régularité des données collectées par ces capteurs sont essentielles pour garantir la validité des analyses et des prédictions effectuées, renforçant ainsi l'efficacité des efforts déployés dans l'optimisation des flux de circulation et la planification du réseau routier.

Les avancées rapides dans le domaine de l'Intelligence Artificielle (IA) ont induit une transformation significative dans la gestion du trafic. Désormais, l'IA permet de prédire et de contrôler avec une grande précision le flux de personnes, d'objets, de véhicules et de marchandises à divers points du réseau de transport. En plus d'améliorer les services rendus aux citoyens, elle contribue à la réduction des accidents en optimisant les flux aux intersections et en renforçant la sécurité lors des périodes de fermeture des routes en raison de travaux ou d'autres événements. De surcroît, sa capacité à traiter et à analyser d'importantes quantités de données a facilité une analyse efficiente, incluant la prédiction et éventuellement la reconstruction des données manquantes [33].

Parallèlement, les chercheurs ont consacré une attention soutenue au domaine de l'analyse du trafic routier et de la sécurité routière, en raison de leur importance cruciale. Ils ont exploré diverses méthodes pertinentes, allant des modèles de trafic aux approches statistiques et aux techniques d'apprentissage automatique. Le défi fondamental lié à la sécurité routière étant d'anticiper avec une précision exceptionnelle les états futurs potentiels du trafic routier. En ce sens, la prédiction des flux de trafic joue un rôle essentiel dans la prévention d'incidents tels que les embouteillages et d'autres anomalies sur les routes [32]. Dans cette perspective, diverses méthodes d'apprentissage et de prédiction du trafic routier ont été élaborées, parmi lesquelles figurent des heuristiques fondées sur l'analyse statistique des données historiques. Ces approches, telles que l'algorithme KNN, la régression linéaire et non linéaire, ainsi que les modèles autorégressifs linéaires comme ARIMA, se révèlent adéquates et pratiques pour la prédiction des conditions de circulation [17][18]. Au sein des méthodes statistiques, les modèles ARIMA représentent, en principe, la catégorie la plus étendue pour la prédiction des données de séries temporelles. En effet, leur efficacité réside dans leur capacité théorique à anticiper de manière potentiellement précise les données futures, étant donné leur commodité, leur précision, leur faible exigence en termes de données d'entrée, et leur simplicité de calcul [19][20].

Dans notre étude, les données horaires de trafic routier recueillies auprès de quatre stations adjacentes de la ville de Marrakech ont été confrontées à la problématique des données manquantes. Ces lacunes résultant de défaillances dans les systèmes d'acquisition, telles que la panne d'un capteur survenue pendant une période donnée, entraînant ainsi une perte de données susceptible d'affecter la fiabilité des systèmes ultérieurs de prédiction et de prise de décision. Dans cette optique et dans le but de remédier à ces limitations, un modèle a été élaboré pour reconstruire les données de trafic manquantes d'une section à partir des enregistrements des capteurs des sections voisines. À cet effet, trois méthodes d'apprentissage automatique ont été mises en œuvre pour anticiper le flux de trafic horaire du capteur défaillant, en nous basant sur les capteurs adjacents. Ainsi, l'étude repose sur des données historiques recueillies entre 2014 et 2017 sur quatre segments de routes nationales à l'entrée de la ville de Marrakech, où les capteurs étudiés sont positionnés.

2. Exploration des méthodes de collecte de données

2.1. Méthodes de collecte de données du trafic routier

L'évolution des STI repose considérablement sur la qualité et la quantité des données relatives au trafic routier. En général, ces données, englobant des paramètres tels que la vitesse des véhicules ou le volume de circulation, sont recueillies au moyen de détecteurs fixes judicieusement positionnés le long du réseau routier à des emplacements stratégiques. À l'échelle mondiale, plusieurs prestataires de services exploitent de manière étendue tant les sources mobiles que fixes afin de fournir aux utilisateurs des informations de grandes qualités concernant l'état actuel du trafic routier [34]. Cette utilisation combinée de sources de données contribue significativement à une offre d'informations robustes et fiables dans le domaine de la gestion du trafic.

Les détecteurs sur site, également désignés sous le terme de « technologies in situ, » englobent les informations relatives au trafic obtenues grâce à des dispositifs de détection implantés le long des axes routiers. Ces technologies de comptage du trafic sont couramment classifiées en deux catégories distinctes, à savoir les méthodes intrusives et les méthodes non intrusives. Chacune présentant des caractéristiques spécifiques.

2.1.1. Méthodes intrusives

Les méthodes intrusives : consistent essentiellement en un enregistreur de données doté d'un capteur placé sur ou dans la route. Ces dispositifs peuvent mesurer divers paramètres tels que la pression

ou le poids des véhicules qui les traversent, fournissant ainsi des données précises sur le volume de trafic, la classification des véhicules et d'autres informations pertinentes. Ces entités sont ancrées dans l'histoire et les plus significatives d'entre elles sont succinctement exposées ci-après :

- Les tubes routiers pneumatiques : des tubes en caoutchouc sont installés sur les routes pour détecter les véhicules sur la base des changements de pression produits lorsqu'un type de véhicule passe sur le tube. L'impulsion d'air créée est enregistrée et traitée par un compteur situé sur le bord de la route.
- Les capteurs piézoélectriques : les capteurs sont installés dans une rainure le long de la surface de la chaussée des voies surveillées. Ce système peut mesurer à la fois le poids et la vitesse des véhicules.
- Les boucles magnétiques : il s'agit de la méthode la plus utilisée pour recueillir des données sur le trafic. Les boucles sont disposées en carré sur les routes afin de créer un champ magnétique. Les données sont ensuite envoyées à un dispositif de comptage installé sur le bord de la route.

2.1.2. Méthodes non intrusives

Les méthodes non intrusives : se fondent sur des observations réalisées à distance, ce qui signifie qu'elles ne nécessitent pas d'interférence directe avec le flux de circulation ou les utilisateurs. Ces approches s'appuient sur des capteurs, des caméras, ou d'autres dispositifs de détection à distance afin de recueillir des données de manière passive, sans perturber l'environnement étudié. Bien que le comptage manuel soit la méthode la plus couramment utilisée, de nouvelles technologies sont apparues récemment et semblent très prometteuses :

- Les comptages manuels : cette méthode implique que des observateurs recueillent des données sur le trafic, notamment le taux d'occupation des véhicules, les déplacements des piétons, et la classification des véhicules.
- L'infrarouge passif et actif : la détection de la présence, de la vitesse, et du type de véhicules repose sur l'utilisation de l'énergie infrarouge émise dans la zone de détection.
- Les capteurs magnétiques passifs : ces capteurs, installés sous ou sur la chaussée, enregistrent le nombre de véhicules, leur modèle, et leur vitesse.
- Le radar à micro-ondes : cette technologie novatrice détecte les véhicules en mouvement, enregistrant des données telles que le nombre, la vitesse, et la classification des véhicules.

- Les capteurs ultrasons et acoustiques passifs : ces dispositifs utilisent des ondes sonores pour détecter les véhicules. Ces capteurs offrent des informations sur le nombre, la vitesse, et la classification des véhicules.
- La détection d'images vidéo : les caméras vidéo enregistrent et analysent le nombre, le type, et la vitesse des véhicules en utilisant diverses techniques, telles que la ligne de déplacement et le suivi.

2.2. Avantages, limites et problèmes courants liés à la collecte des données

Les capteurs routiers, qu'ils soient intrusifs ou non intrusifs, comportent divers avantages et inconvénients. Les méthodes intrusives reposent sur des observations à distance, impliquant de petites unités électroniques logées dans des boîtiers résistants aux conditions météorologiques et positionnées à divers emplacements [35]. Elles se distinguent par leur coût réduit, leur autonomie, leur facilité de déploiement, leur utilisation répandue, leur précision acceptable, et leur capacité à effectuer une classification basée sur l'essieu. Cependant, ces méthodes sont sujettes à des inconvénients significatifs, tels que la perturbation du trafic, pouvant engendrer des problèmes de sécurité et de performance, ainsi que la potentialité de réduire la durée de vie des chaussées en cas d'installation défaillante du capteur intrusif. De plus, l'installation de ces capteurs comporte des risques pour les équipes chargées de la collecte des données, et en présence de conditions routières défavorables, les capteurs pourraient ne pas fonctionner correctement, entraînant une réduction de leur durée de vie [36]. Ces dispositifs, espacés à des intervalles prédéfinis, exigent des calibrations fréquentes afin de prévenir des mesures imprécises [37].

D'un autre côté, les méthodes non intrusives présentent des avantages significatifs en termes d'installation, d'accessibilité et d'entretien. Elles se caractérisent par leur facilité de déploiement, d'accès et de maintenance, offrant la possibilité de capturer de manière permanente et visuelle toutes les informations souhaitées concernant le trafic. Elles font preuve d'une précision élevée, permettant la détection de zones multiples, fonctionnant de manière efficace tant de jour que de nuit, et offrant la capacité de mesurer les vitesses directes. Cependant, leur installation peut entraîner des coûts élevés, potentiellement perturbateurs pour le flux de circulation. De plus, leur capacité limitée à détecter le trafic dans les voies les plus éloignées se manifeste particulièrement lors de l'étude de grandes routes à plusieurs voies. Le tableau 2 [34][38][39] synthétise de manière exhaustive les avantages et les inconvénients inhérents aux méthodes intrusives et non intrusives. Il présente cinq méthodes de comptage

de trafic intrusives et cinq méthodes de comptage de trafic non intrusives, offrant ainsi une comparaison détaillée des différentes méthodes.

Tableau 2. Synthèse des avantages et des inconvénients des méthodes intrusives et non intrusives

Méthode	**Avantages**	**Désavantages**
Méthodes intrusives		
Tubes pneumatiques	• Faible coût, facilité d'entretien • Installation rapide pour l'enregistrement permanent/temporaire des données • Faible consommation d'énergie	• Couverture limitée des voies • Inefficace pour les flux à faible vitesse • L'efficacité dépend des conditions météorologiques, de la température et du trafic. • Comptage d'essieux imprécis pour les volumes élevés de camions et d'autobus • Commutateur d'air sensible à la température • Tubes coupés à la suite d'actes de vandalisme et de l'usure produite par les pneus des camions
Capteurs piézoélectriques	• Détecte le passage du pneu et non du véhicule sur le capteur • Convient pour mesurer le poids et la vitesse des véhicules • Différencie les véhicules individuels avec une extrême précision	• Interruption de la circulation pour l'installation, la réparation et les pannes • La surface de la route et les réparations des services publics peuvent nécessiter sa réinstallation. • Sensible à la température de la chaussée et à la vitesse du véhicule
Boucles magnétiques	• Insensible aux mauvaises conditions météorologiques • Insensible aux contraintes de la circulation	• Fermeture de la voie pendant l'installation • Durée de vie courte • Endommagement par les véhicules lourds • Coûts de mise en œuvre et d'entretien élevés
Mouvement de pesage	• Peut servir à la fois à la collecte de données sur le trafic et à la mesure du poids • Plus grande précision que les systèmes piézoélectriques • Coût inférieur à celui des systèmes à cellule de charge • Ne nécessite pas le remplacement complet du capteur, à l'exception d'une remise à neuf tous les 5 ans.	• Moins précis que les systèmes à cellule de charge • Plus chers que les systèmes piézoélectriques • Mesures
Magnétique passif	• Moins sensible que la perte aux contraintes du trafic • Peut être utilisé là où les boucles ne sont pas réalisables, par exemple sur les tabliers de pont.	• Les capteurs rencontrent des difficultés à différencier les véhicules proches les uns des autres.
Méthodes non intrusives		
Comptage manuel	• Convient aux petits échantillons de données	• La sécurité des recenseurs n'est pas prévisible la nuit, ce qui rend difficile l'utilisation du système la nuit.

	• Préférable pour les périodes de comptage courtes • Insensible aux caractéristiques géométriques	• Les données sont moins précises que celles des comptages automatisés. • Susceptible d'erreur humaine
Infrarouge passif et actif	• L'installation ne nécessite pas de procédure de revêtement invasive • Transmettre plusieurs faisceaux pour des mesures précises de la position, de la vitesse et de la classe du véhicule • Détection de présence sur plusieurs voies disponible	• Couverture limitée des voies de circulation • Mauvaises performances par mauvais temps et en cas d'ensoleillement
Ultrasons et acoustique passive	• Installation dans des conditions météorologiques défavorables • Insensible aux contraintes de la circulation	• Affecté par la température et le mauvais temps • Les mesures d'occupation peuvent être affectées par de grandes périodes de répétition des impulsions.
Micro-ondes radar	• N'est pas affecté par les conditions météorologiques • Mesure directe de la vitesse • Modèles à voies multiples disponibles	• Ne détecte pas les véhicules arrêtés • Performances médiocres aux intersections en tant que compteurs de volume
Détection par image vidéo	• Sensible au contraste de couleur entre le véhicule et la route • Haute précision	• Sensible aux conditions météorologiques. • Vulnérable aux obstacles, aux ombres, etc. • Peut être sensible aux mouvements de caméra. • Nécessite une hauteur de montage de 15 à 18 mètres.

3. Description de l'étude

3.1. Contexte spécifique de la collecte des données au Maroc

Historiquement, le système de comptage routier a été introduit au Maroc au cours des années 1970, évoluant progressivement pour s'adapter aux exigences changeantes. Les données recueillies par les comptages routiers sont essentielles pour déterminer le Trafic Moyen Journalier Annuel (TMJA), ainsi que le trafic de pointe. Actuellement, deux types de compteurs sont déployés pour mesurer le trafic au Maroc. Les stations de comptage « mono boucle » enregistrent exclusivement le flux de véhicules dans toutes les directions, tandis que les stations « multiboucles » permettent d'enregistrer le flux de véhicules, la vitesse des véhicules, et la longueur des véhicules par direction et par voie.

Dans ce cadre, le ministère d'Équipement et de l'Eau au Maroc utilise deux approches distinctes pour surveiller le trafic routier. D'une part, le comptage permanent est effectué à l'aide de compteurs multiboucles installés dans des stations fixes sur l'autoroute ou le réseau routier. Ces stations fonctionnent de manière automatisée et continue 24 heures par jour. En 2017, 145 stations permanentes, hors autoroutes, ont été déployées sur l'ensemble du réseau routier. D'autre part, le comptage périodique

est réalisé à l'aide de compteurs monoboucle ou multiboucles, installés dans des stations périodiques. Ces stations fonctionnent généralement pendant 8 jours par semestre, conformément à un calendrier établi par chaque Direction régionale de l'Équipement, du Transport et de la Logistique (DRETL) ou Direction provinciale de l'Équipement, du Transport et de la Logistique (DPETL). En 2017, 636 stations périodiques ont été déployées sur l'ensemble du réseau routier.

Il convient de noter que chaque station périodique de comptage routier est interconnectée à une station permanente, permettant ainsi la réalisation des ajustements requis en fonction des facteurs saisonniers. Soulignons par ailleurs que le choix des emplacements pour les stations permanentes et/ou périodiques est déterminé par le volume de trafic traversant la station, indépendamment de la classification actuelle du réseau routier. Enfin, le CNER fournit annuellement à ses directions régionales et provinciales un recueil exhaustif de données de trafic. Ce document permet de synthétiser les résultats des comptages routiers, mettant en relief divers éléments tels que les tendances du trafic, les classes de trafic, les variations hebdomadaires et mensuelles du trafic, les variations horaires du trafic, ainsi que les tendances du trafic, du réseau routier, du parc de véhicules et de la consommation de carburant, entre autres.

Dans le cadre de la régionalisation du système de comptage routier, les services du Service de Planification des Études Économiques (SPEE) de chaque direction régionale sont responsables de la collecte des données de trafic routier auprès de chaque direction provinciale. Ils assurent également le transfert et le traitement des disquettes provenant des compteurs, transmettant ensuite les fichiers résultants au CNER pour un traitement ultérieur. En vue de faciliter la collecte des données de trafic routier, le CNER, en collaboration avec la Division des Opérations de la Direction des Routes, a initié un marché pour l'acquisition de compteurs permanents dotés de fonctionnalités de transfert automatique des données. Dans ce contexte, le CNER a procédé à l'acquisition de 14 stations de comptage routier et 3 stations de pesage dynamique, toutes équipées de dispositifs de transfert automatique de données.

Les postes de comptage routier avec transfert automatique permettent la synthèse de données telles que la date de passage du véhicule (incluant la date, l'heure, la minute et la seconde), la vitesse en kilomètres par heure, la voie (direction), la longueur en mètres, le taux d'occupation, ainsi que la classification en fonction des distances inter-essieux et/ou de la technique de reconnaissance de forme de la chaussée, de la température de roulement et de la température de l'air, entre autres paramètres.

De plus, le poste de comptage routier dédié à la classification des véhicules et au pesage automatique offre la possibilité d'obtenir des données supplémentaires, telles que la date de pesée du véhicule (comprenant la date, l'heure, la minute et la seconde), la vitesse en kilomètres par heure, la longueur en mètres, le taux d'occupation, la classification selon les distances inter-essieux et/ou les techniques de reconnaissance de formes, le nombre d'essieux, le temps intervéhicule, la distance entre les véhicules, les températures de la surface de la route et de l'air, le poids total, ainsi que le poids de chaque essieu.

3.2. Présentation de l'étude d'estimation des données manquantes

Notre étude a impliqué l'utilisation de quatre stations de comptage permanent. Ces dispositifs de mesure ont enregistré le flux de circulation routière horaire dans la ville de Marrakech tout au long de l'année. Chaque compteur a enregistré le nombre de véhicules pour chaque catégorie de véhicule sur chaque voie. La figure 6 présente la localisation de chaque station examinée dans le réseau routier, avec les stations permanentes représentées en rouge et les stations périodiques en bleu. L'analyse de la figure révèle que le réseau routier de Marrakech est dense et bien développé. Ce dernier se composant de plusieurs axes routiers majeurs qui convergent au cœur du centre-ville. Plus précisément, le poste permanent 137 se trouve à l'intersection de la route nationale 7 (N7) et de la route provinciale 2008 (P2008), le poste 39 est situé sur la route nationale 8 (N8) à l'est de la ville, le poste 138 est positionné sur la route nationale 8 (N8) à l'ouest de la ville, et enfin le poste 74 est localisé sur la route nationale 9 (N9). Ces stations de comptage permanent sont stratégiquement positionnées le long des axes principaux, favorisant ainsi la collecte de données exhaustives relatives au trafic routier sur l'ensemble de la ville. L'analyse approfondie de ces données contribuera à une meilleure compréhension des flux de circulation urbaine et à l'identification précise des zones présentant des risques.

4. Méthodes d'estimation des données

4.1. Prétraitement des données

Les données mobilisées dans notre étude proviennent du service relevant de la Direction des Routes du Ministère d'Équipement et de l'Eau. Effectivement, cette entité gouvernementale met en œuvre des dispositifs de mesure, sous la forme de capteurs, destinés à recueillir des données sur la circulation routière à l'échelle nationale. L'ensemble de données préliminaires se composant de fichiers distincts, chacun correspondant à un capteur permanent. Les informations encapsulées dans ces fichiers

incluent la date de début et de fin de l'enregistrement, le numéro de la route, le nombre de voies, la classe du véhicule et les relevés horaires du trafic. Les colonnes du fichier contiennent les classes de véhicules, tandis que les lignes indiquent le nombre de véhicules comptés. La structure des fichiers d'enregistrement est succinctement présentée dans le tableau 3 (pour une expansion du contenu présenté, veuillez consulter l'Annexe B).

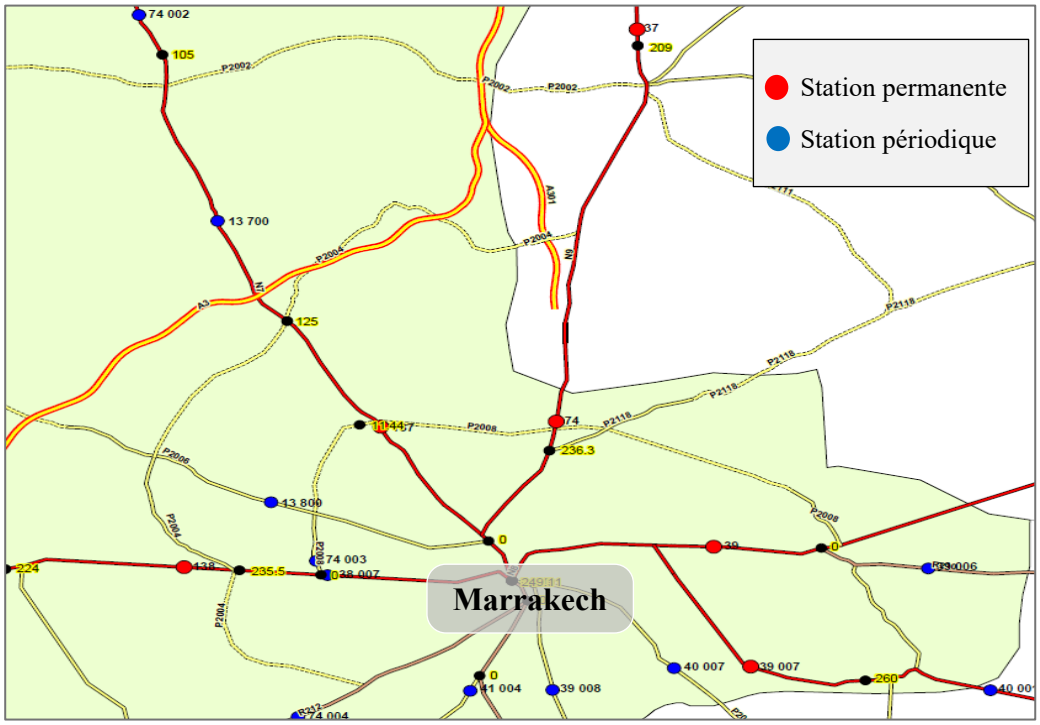


Figure 6. Répartition des stations permanentes et périodiques autour de la ville de Marrakech

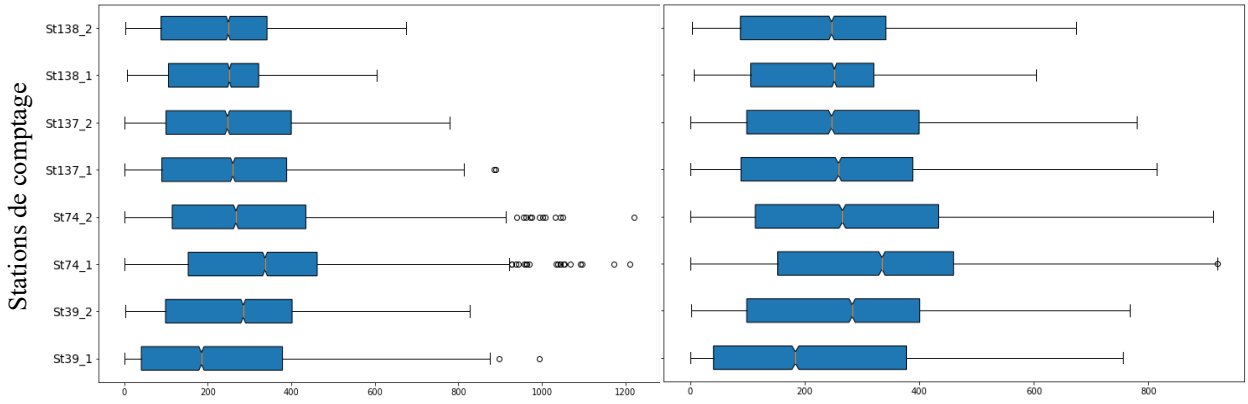
Tableau 3. Structure d'un fichier généré par un compteur pour une route à deux voies

Date	Heure	Voie 1	Classe 1	Classe 2	Classe 3	Classe 4	Voie 2	Classe 1	Classe 2	Classe 3	Classe 4
01/01/16	22 h	181	85	45	34	17	202	46	63	52	41
01/01/16	23 h	389	182	152	42	13	136	46	41	25	24
02/01/16	0 h	418	156	124	87	51	118	42	45	19	12
02/01/16	1 h	438	205	140	52	41	134	52	43	24	15
02/01/16	2 h	316	132	85	67	32	370	196	92	45	37

Le corpus de données exploité au sein de notre recherche englobe les flux de circulation enregistrés sur une période de quatre années, de 2014 à 2017, associés à quatre stations de comptage routières désignées par les identifiants 138, 137, 74 et 39 ; il se compose d'un ensemble distinct de fichiers. Les fichiers issus des capteurs ont été soumis à un traitement initial afin d'assurer leur fonctionnalité. En première instance, une consolidation des fichiers de chaque capteur a été effectuée, regroupant ainsi les enregistrements du trafic routier avec les informations relatives à la date et l'heure des observations au sein d'un unique fichier. Subséquemment, l'agrégation des données du trafic horaire par voie a été réalisée pour l'ensemble des catégories de véhicules. Par la suite, les fichiers correspondants aux quatre stations routières ont été fusionnés en un seul fichier consolidé. La structure de ce tableau résultant est explicitement détaillée dans le tableau 4, où les colonnes reflètent les volumes de trafic routier pour chaque voie au sein des différentes stations. Les données ont été préalablement soumises à un traitement consistant en l'élimination des données extrêmes ou aberrantes, cette démarche s'avérant nécessaire en raison du potentiel impact négatif de ces valeurs sur la qualité des résultats. La méthode employée pour évaluer la dispersion des données repose sur l'utilisation de l'« intervalle interquartile » (IQR), permettant l'exclusion des valeurs nettement éloignées des autres observations. Les figures 7.a et 7.b illustrent la distribution des données de trafic routier pour chaque station avant et après l'application de la méthode IQR. La figure 7.b met en évidence la suppression des valeurs extrêmes (indiquées par des cercles dans la figure a). Par ailleurs, seules la station 74 et la direction 1 des stations 137 et 39 présentent des valeurs aberrantes. Cette représentation graphique indique également que l'intervalle interquartile couvre 50 % du trafic enregistré, avec des valeurs comprises entre 40 et 460 véhicules par heure pour l'ensemble des stations.

Tableau 4. Structure de données après traitement

Date	**Voie 1 Station 39**	**Voie 2 Station 39**	**Voie 1 Station 74**	**Voie 2 Station 74**	**Voie 1 Station 137**	**Voie 2 Station 137**	**Voie 1 Station 138**	**Voie 2 Station 138**
2014-01-07 13 h	371	370	436	411	463	463	299	303
2014-01-07 14 h	370	364	471	471	405	405	274	360
2014-01-07 15 h	356	379	489	426	400	400	315	400
2014-01-07 16 h	476	378	563	433	412	412	302	494
2014-01-07 17 h	506	364	583	389	498	498	294	377



a. Distribution des données avec valeurs

b. Distribution des données sans valeurs

Figure 7. Distribution des données avec et sans valeurs aberrantes.

4.2. Régression linéaire - Moindres carrés ordinaires

En apprentissage automatique, un modèle de régression linéaire est un cadre analytique qui vise à établir une relation linéaire entre une variable, désignée comme la variable dépendante, et une ou plusieurs variables, appelées variables indépendantes. Dans ce contexte, la méthode des Moindres Carrés Ordinaires (MCO) représente une technique d'estimation des coefficients de la régression linéaire, décrivant les relations entre une ou plusieurs variables quantitatives et une autre variable cible.

Les moindres carrés se réfèrent à la minimisation de l'erreur quadratique. Cette méthode analytique permet d'estimer les coefficients de l'équation linéaire, impliquant une ou plusieurs variables indépendantes, afin de prédire au mieux les valeurs de la variable dépendante. La régression linéaire ajuste une ligne droite ou une surface qui minimise les écarts entre les valeurs prédites et réelles de la variable dépendante. Il existe des calculateurs de régression linéaire simples qui utilisent la méthode des moindres carrés pour déterminer la ligne la mieux adaptée à un ensemble de données appariées. Ainsi, pour un modèle comportant p variables explicatives (indépendantes), la formulation statistique du modèle de régression par la méthode MCO est exprimée comme suit :

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \qquad (9)$$

Où Y désigne la variable dépendante, β0, est la constante du modèle (intercept), X_j désigne les variables explicatives du modèle (pour j allant de 1 à p), β_j sont les coefficients associés aux variables

explicatives X_j, et ε est une erreur aléatoire de moyenne 0 et de variance σ^2. Étant donné n observations, l'estimation de la valeur de la variable Y pour l'observation i est donnée par l'équation suivante :

$$\widehat{Y_i} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_i X_{ij} \; (i = 1 \dots n) \tag{10}$$

Où $\widehat{Y_i}$ est l'estimation de la valeur de la variable dépendante Y pour l'observation i, $\hat{\beta}_0$ est l'estimation de l'intercept, $\hat{\beta}_i$ sont les estimations des coefficients associés aux variables explicatives X_{ij}. Cette équation représente la prédiction du modèle de régression linéaire pour une observation spécifique en utilisant les coefficients estimés à partir des données d'apprentissage.

La régression linéaire représente l'un des modèles interprétables les plus fréquemment utilisés pour aborder des problèmes de régression. Comme son appellation l'indique, la régression linéaire est un modèle de nature linéaire, ce qui signifie que l'association entre les caractéristiques et les résultats est modélisée de manière linéaire. Cette linéarité facilite l'interprétation du modèle, permettant une compréhension directe des poids attribués. Soulignons également que ce modèle est assujetti à des contraintes de monotonicité, garantissant ainsi que la relation entre une caractéristique et le résultat évolue toujours dans la même direction sur toute l'étendue de la caractéristique. Autrement dit, une augmentation de la valeur de la caractéristique entraînera systématiquement une augmentation ou une diminution du résultat. La propriété de monotonicité est cruciale pour l'interprétation des modèles, simplifiant ainsi la compréhension des relations.

Il est important de noter qu'il existe un intervalle de confiance (IC) associé à chaque poids dans l'équation linéaire. L'intervalle de confiance représentant une plage d'estimation du poids réel avec un certain niveau de confiance. Par exemple, un IC de 95 % pour un poids estimé indique que dans 95 % des cas, la valeur réelle du poids se situe dans cet intervalle si l'estimation est répétée. Cette notion d'intervalle de confiance apporte une mesure de la fiabilité des estimations des poids dans le modèle.

En plus de la monotonicité et des intervalles de confiance, le coefficient de détermination est un autre élément interprétable dans la régression linéaire. À ce titre, le coefficient de détermination, souvent noté R^2, constitue un élément fondamental dans l'évaluation d'un modèle de régression. Il revêt une importance significative en raison de sa capacité à quantifier la qualité d'ajustement du modèle aux données observées. Un R^2 élevé indique que le modèle parvient à expliquer une part substantielle de la variabilité inhérente aux données, suggérant ainsi une adéquation robuste entre les variables explicatives

et la variable dépendante. Cette mesure de performance offre dans ces conditions un indicateur quantitatif de la capacité prédictive du modèle, éclairant ainsi la qualité de la modélisation.

4.3. Machine à vecteurs de support pour la régression

La machine à vecteur de support pour la régression est un algorithme de régression qui applique une technique similaire à celle du SVM pour l'analyse de régression. Alors que les SVM sont traditionnellement employés pour des tâches de classification, le SVR s'applique à des données de régression impliquant des valeurs continues ou discrètes. Pour modéliser ces données, le SVR ajuste la ligne de régression de sorte qu'elle se rapproche des meilleures valeurs avec une marge donnée appelée ε-tube (epsilon-tube), où ε représente la largeur de ce tube.

La stratégie sous-jacente du SVR consiste à optimiser un hyperplan qui maximise le nombre de points tout en minimisant l'erreur associée. Cette erreur, mesurée par le ε-tube, tient compte de la complexité du modèle ainsi que du taux d'erreur ; ce dernier est défini comme la disparité entre les valeurs de sortie réelles et celles prédites. Les contraintes du problème SVR s'expriment formellement à travers l'inéquation suivante :

$$|y_i - (w_i x_i + b)| \leq \varepsilon \tag{11}$$

Où y_i représente la valeur réelle, w_i est le poids associé à la variable indépendante x_i, et b est le terme de biais. L'équation caractéristique de l'hyperplan dans le SVR est formulée par l'équation linéaire comme suit :

$$y = wx + b \tag{12}$$

L'expression mathématique ci-dessus illustre la recherche d'un compromis optimal entre la complexité du modèle et la précision des prédictions, mettant en avant la sophistication intrinsèque de l'approche SVR dans la modélisation des données de régression. L'objectif fondamental du SVR est de minimiser les coefficients conformément à la fonction objective définie comme suit :

$$min \frac{1}{2} \|w\|^2 \tag{13}$$

Dans cette fonction objectif, $\|w\|^2$ représente la norme euclidienne des poids w du modèle. La minimisation de cette quantité vise à favoriser un modèle plus simple tout en préservant la capacité de

bien s'ajuster aux données observées. Cette quête d'une solution optimale s'inscrit dans le cadre de l'approche SVR, où la sophistication algorithmique se conjugue avec une formulation mathématique élaborée pour obtenir une modélisation robuste des relations complexes entre les variables dans le contexte de la régression.

4.4. Moyenne mobile intégrée autorégressive saisonnière

L'une des méthodes de prédiction des séries temporelles les plus utilisées est la méthode ARIMA. Il s'agit là d'une moyenne mobile intégrée autorégressive. Ce modèle prédit les valeurs futures d'une série temporelle sur la base de certains aspects de la structure statistique de la série observée. D'après son nom, le modèle peut être divisé en plusieurs composantes, AR, MA et I. la composante AR désigne un modèle autorégressif qui représente un type de processus aléatoire. La sortie du modèle dépend linéairement de sa valeur précédente. C'est-à-dire d'un point de données décalé ou du nombre d'observations passées. La composante MA indique un modèle de moyenne mobile dont le résultat dépend linéairement des observations actuelles et passées d'un terme stochastique [40]. La composante I, signifie Intégré ; elle définit l'étape de différenciation pour générer des données de séries temporelles stationnaires. C'est-à-dire l'élimination des composantes saisonnières et tendancielles. Le modèle ARIMA est généralement désigné par ARIMA [p, d, q] tel que p est le nombre de décalages temporels du modèle autorégressif AR [p], d est le nombre de fois que les données ont été soustraites de la valeur passée, et q est l'ordre du modèle de moyenne mobile MA [q].

Le modèle SARIMA, ou ARIMA saisonnier, applique de manière analogue les valeurs passées tout en incorporant des modèles de saisonnalité. La distinction principale entre SARIMA et ARIMA réside dans la capacité du premier à prendre en considération les tendances saisonnières, conférant ainsi une puissance accrue pour la prédiction de données complexes. Ainsi, En intégrant la saisonnalité en tant que paramètre, le modèle SARIMA offre des performances supérieures à l'ARIMA lorsqu'il s'agit de prévoir des données caractérisées par des variations périodiques. En outre, le modèle SARIMA présente des paramètres additionnels, en l'occurrence P, D, et Q, qui dénotent des composantes saisonnières supplémentaires et contribuent à affiner la modélisation de séries temporelles. La prise en compte de ces paramètres renforce donc la flexibilité du modèle SARIMA, le rendant particulièrement adapté à la modélisation et à la prédiction de séries temporelles comportant des tendances saisonnières marquées. L'équation générale pour un modèle SARIMA est définie par :

$$\Phi_P(B)\Phi_p(B^\pi)(1-B)^D\ (1-B^\pi)^d\ X_t\ =\ \Theta_Q(B)\Theta_q(B^\pi)Z_t \qquad (14)$$

Où X_t est la série temporelle observée, B est l'opérateur de retard non saisonnier, B^π [Où π est la période de saisonnalité] est l'opérateur de retard saisonnier, $\Phi_P(B)$ est le polynôme autorégressif saisonnier d'ordre P, $\Phi_p(B^\pi)$ est le polynôme autoregressif non saisonnier d'ordre p, $(1-B)^D$ est le polynôme saisonnier de différenciation d'ordre D, $(1-B^\pi)^d$ est le polynôme non saisonnier de différenciation d'ordre d, $\Theta_Q(B)$ est le polynôme de moyenne mobile saisonnier d'ordre Q, $\Theta_q(B^\pi)$ est le polynôme non saisonnier d'ordre q, et Z_t est le bruit saisonnier, également appelé résidu ou erreur, et il représente la composante non expliquée par le modèle.

5. Évaluation des résultats et discussion

5.1. Analyse de corrélation du trafic routier des capteurs adjacents

Dans notre analyse, les simulations ont été exécutées en utilisant le langage de programmation Python 3.6 sur une plateforme informatique équipée d'un processeur Intel Core i5-7200U cadencé à 2,5 GHz.

Dans notre recherche, nous développerons un modèle apte à assimiler et à anticiper les données générées par le capteur numéro 138, en nous fondant sur les informations recueillies par trois capteurs adjacents, à savoir les capteurs numéro 29, 74 et 137. Dans un premier temps, nous avons appliqué trois méthodes largement établies dans la littérature, à savoir la méthode de la régression linéaire, le modèle SVR et le modèle SARIMA, afin d'examiner la corrélation entre les données enregistrées par les différents capteurs. Cette approche vise à reconstruire les données manquantes associées à des sections spécifiques de la route, en se fondant sur l'historique du trafic capté par les capteurs adjacents. Il est à noter que les capteurs sont fréquemment sujets à des dysfonctionnements entraînant une importante lacune de données, impactant ainsi la qualité et la fiabilité de l'analyse du trafic. À cet égard, la figure 8 montre, de manière chronologique, le flux de trafic de quatre stations de comptage permanentes situées dans la ville de Marrakech. La première observation, à savoir la similarité entre les courbes des différentes stations, suggère une répartition relativement uniforme du trafic routier dans la ville de Marrakech. Cette uniformité indique une distribution équitable du trafic entre les divers points de mesure. Cette constatation revêt une importance particulière, car elle laisse supposer que les données présentées dans la figure 8 sont représentatives du trafic global dans la ville. La seconde observation, concernant

l'augmentation du trafic routier en journée et sa diminution pendant la nuit, reflète une tendance fréquemment observée dans de nombreuses agglomérations. Cette variation pendant la journée s'explique par les déplacements plus fréquents des individus, que ce soit pour se rendre au travail, à l'école ou pour d'autres activités.

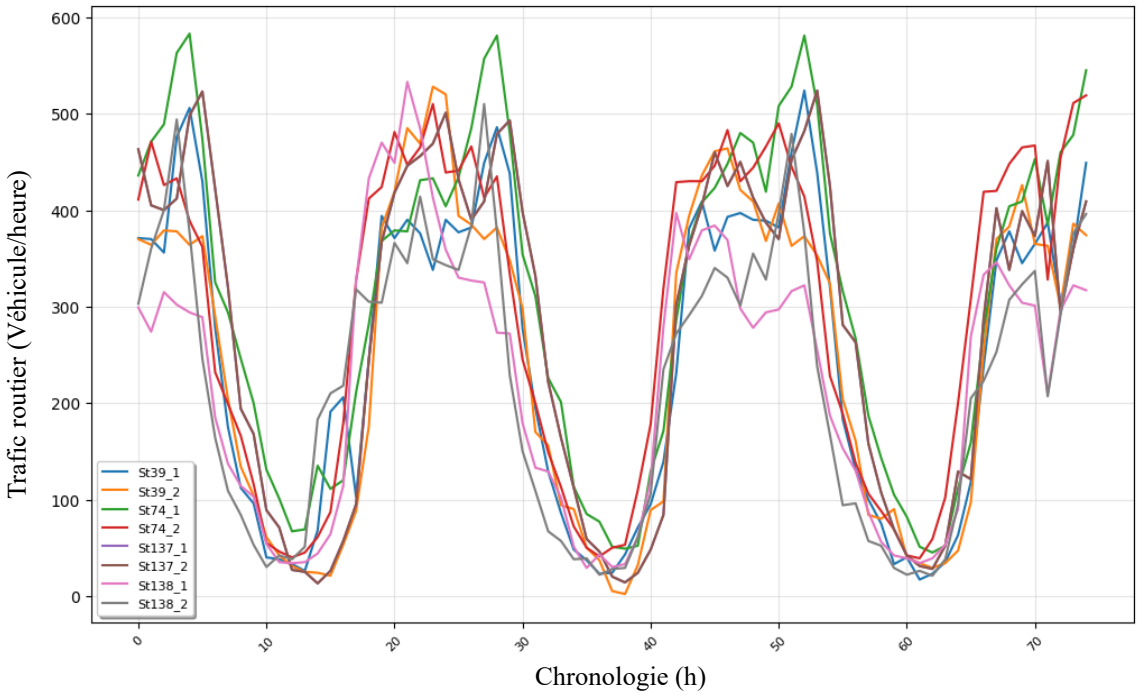


Figure 8. Flux de trafic de chaque station dans l'ordre chronologique

La figure 9 montre la corrélation entre les différentes stations à l'aide d'une carte thermique. Cette dernière est assistée par une barre de couleur qui rend les données facilement lisibles et compréhensibles. Celle-ci nous permet de visualiser les valeurs de corrélation à l'aide de couleurs ; ainsi, pour représenter les valeurs les plus courantes ou les activités les plus élevées, on utilise principalement des couleurs plus sombres, et pour représenter les valeurs ou les activités moins courantes, on utilise des couleurs plus vives. En examinant la carte thermique, nous constatons que la station St39_1 est hautement corrélée avec les stations St74_2 et St137_2, avec une corrélation de 0,84. Les stations St137_2 et St74_2 sont également fortement corrélées, avec une corrélation de 0,8. Cela signifie que ces deux stations ont tendance à avoir des valeurs similaires. Les stations St39_2 et St138_2 sont fortement corrélées, avec une corrélation de 0,81. Tandis que les stations St39_1 et St138_1 sont faiblement corrélées, avec une corrélation de 0,55. Cela signifie que ces deux stations tendent à avoir des valeurs différentes.

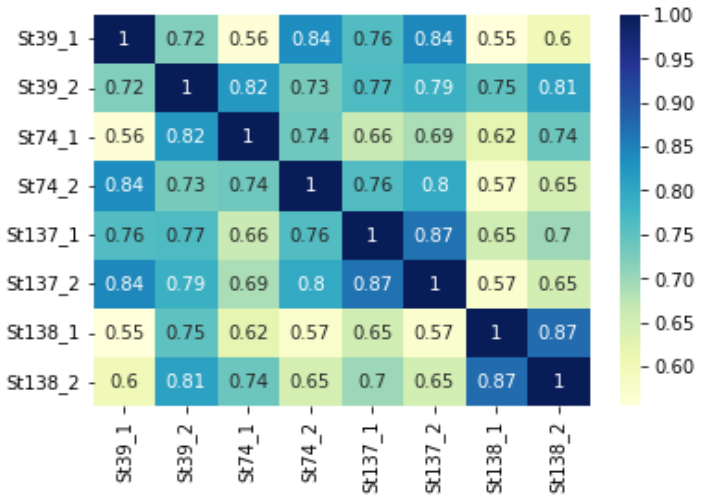


Figure 9. Matrice de corrélation entre les différentes stations

5.2. Discussion et interprétabilité

Cette étude examine la problématique des données manquantes dans le contexte du trafic routier et présente un modèle visant à réaliser des prédictions précises. Afin d'assurer l'efficacité de l'approche suggérée, notre recherche se focalisera initialement sur l'analyse de l'interprétabilité de trois modèles distincts. Par la suite, dans une seconde phase, nous procèderons à une comparaison systématique de ces modèles.

5.2.1. Interprétabilité des modèles

De nombreux défis et problèmes contemporains trouvent des solutions à travers l'emploi de modèles d'apprentissage automatique souvent qualifiés de « boîte noire ». Notons que ces modèles, en raison de leur complexité intrinsèque, agissent en tant que source d'information, reléguant parfois les données elles-mêmes au second plan. Néanmoins, l'impératif d'interprétabilité s'avère crucial dans cette démarche, puisqu'il nous autorise à extraire des informations complémentaires et à clarifier le fondement des décisions prises par le modèle. Ainsi, cette capacité à rendre clair le processus décisionnel de la « boîte noire » offre une perspective cruciale aux utilisateurs, aux praticiens et aux chercheurs, facilitant ainsi une confiance accrue dans les résultats générés par ces modèles complexes. La quête de l'interprétabilité s'inscrit donc comme un enjeu majeur dans l'exploitation des modèles d'apprentissage automatique, contribuant à renforcer la transparence, la compréhension et l'acceptabilité de ces solutions technologiques.

5.2.2. Interprétation du modèle de régression linéaire

Les résultats de la régression linéaire sont présentés dans la figure 10. Une corrélation positive est observée entre les données provenant des stations St39_1, St74_1, St74_2, St137_1, St138_1, et la station-cible St138_2. Celle-ci indique que l'accroissement du trafic au niveau de ces stations est associé à une augmentation du trafic à la station cible. À titre illustratif, pour la station St39_1, une augmentation du trafic de 100 véhicules sur la voie 1 de la route de la station 39 se traduit par une augmentation de 19 véhicules sur la voie 2 de la route de la station 138. Par ailleurs, une corrélation négative est observée entre les stations St137_2 et St39_2 et la station cible St138_2. Celle-ci suggère qu'une augmentation du trafic à ces stations est associée à une diminution du trafic à la station cible. À titre d'exemple, pour la station St39_2, une augmentation du trafic de 100 véhicules sur la voie 2 de la route de la station 39 se traduit par une diminution d'environ 7 véhicules sur la voie 2 de la route de la station 138.

F-statistic:		1.763e+04				
Prob (F-statistic):		0.00				
Log-Likelihood:		-28781.				
AIC:		5.758e+04				
BIC:		5.762e+04				
	coef	std err	t	P>|t|	[0.025	0.975]
St39_1	0.1910	0.022	8.865	0.000	0.149	0.233
St39_2	-0.0748	0.021	-3.522	0.000	-0.116	-0.033
St74_1	0.0766	0.013	5.734	0.000	0.050	0.103
St74_2	0.0998	0.012	8.002	0.000	0.075	0.124
St137_1	0.0449	0.011	4.077	0.000	0.023	0.066
St137_2	-0.0006	0.013	-0.048	0.961	-0.025	0.024
St138_1	0.5747	0.009	64.138	0.000	0.557	0.592
Omnibus:	666.615		Durbin-Watson:	0.626		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	1140.641		
Skew:	0.845		Prob(JB):	2.06e-248		
Kurtosis:	4.513		Cond. No.	35.3		

Figure 10. Résultats de la régression linéaire

La mesure du coefficient de détermination (R au carré) nous informe par ailleurs que 95,9 % de la variance totale du trafic routier cible est expliquée par le modèle de régression linéaire par la méthode des moindres carrés ordinaires. Cette valeur représente la proportion significative de la variabilité du résultat cible expliquée par le modèle, renforçant ainsi la validité et la pertinence du modèle dans la compréhension des relations entre les différentes stations de comptage du trafic routier.

Enfin, l'examen de l'histogramme des poids estimés, illustré dans la figure 11, révèle que St138_1 exerce une influence substantielle et positive sur la prédiction du nombre de véhicules dans St138_2. Cette tendance est suivie par St38_1, St74_2, St74_1, et St137_1, dans un ordre décroissant d'impact. Par ailleurs, une corrélation négative de faible amplitude est observée entre St39_2 et St138_2. Il est à noter que le poids attribué à la caractéristique St137_2 est très proche de zéro, indiquant ainsi que l'effet de St137_2 n'atteint pas de significativité statistique.

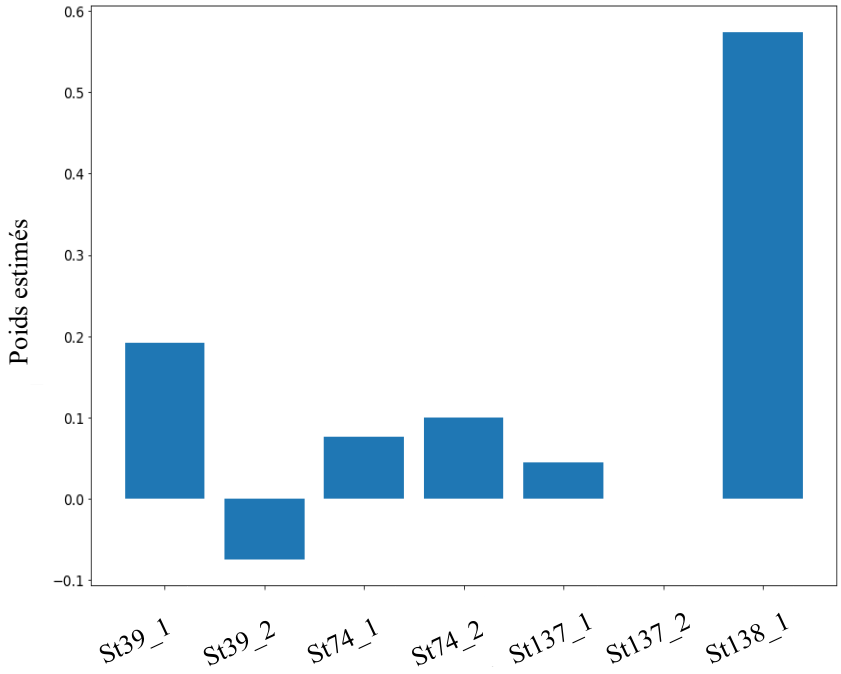


Figure 11. Histogramme des poids estimés par la régression linéaire

5.2.3. Interprétation du modèle SARIMA

La figure 12 expose les résultats obtenus à travers la modélisation SARIMA. La colonne des coefficients où les valeurs sous « Coef » sont les poids des termes dans l'équation du modèle. Nous remarquons que tous les coefficients des termes AR et MA ont une valeur P inférieure à 0,05 dans la colonne « P>|z| » ; ce qui signifie qu'ils sont tous hautement significatifs. Pour analyser les résultats du modèle SARIMA, il est nécessaire d'analyser les hypothèses du modèle à l'aide des statistiques de Ljung-Box, du chi-carré et de l'autocorrélation résiduelle ; d'évaluer si chaque terme est significatif à l'aide des valeurs p ; et d'identifier la précision du modèle à l'aide de l'erreur quadratique moyenne. Pour nous assurer que chaque terme du modèle est statistiquement significatif, chaque terme doit avoir une valeur p inférieure à 0,05, afin de rejeter l'hypothèse nulle avec des valeurs statistiquement significatives. L'hypothèse nulle dans cette section nécessitant que chaque coefficient ne soit pas statistiquement

significatif. Cependant, les résultats montrent que tous les termes sont statistiquement significatifs. Ensuite, il est nécessaire de s'assurer que notre modèle répond à l'hypothèse selon laquelle les résidus sont indépendants, connus sous le nom de bruit blanc, en utilisant la méthode de Ljung-Box. Ainsi, la statistique du test de Ljung-Box (L1) (Q) au lag 1 montre que la Prob(Q) est de 0,55 et la valeur p de 0,46. Étant donné que la probabilité est supérieure à 0,05, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle les résidus sont indépendants.

Le test d'hétéroscédasticité vise à vérifier si les résidus d'erreur présentent une variance constante. Les statistiques résumées indiquent une statistique de test de 1,14 avec une valeur p de 0,00. Cela indique que nous rejetons l'hypothèse nulle, ce qui confirme que nos résidus présentent une variance non constante. Ensuite, le test de Jarque-Bera est utilisé pour évaluer la normalité des erreurs, en testant l'hypothèse selon laquelle les données suivent une distribution normale. La statistique du test est de 249,24 avec une probabilité de 0. Cela indique que nous rejetons l'hypothèse nulle, ce qui signifie que les données ne suivent pas une distribution normale. De plus, selon les résultats du test de Jarque-Bera, la distribution présente une légère asymétrie et un aplatissement positif.

No. Observations:	7603
Log Likelihood	-40693.355
AIC	81406.711
BIC	81476.104
HQIC	81430.520

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	1.8194	0.007	276.063	0.000	1.807	1.832
ar.L2	-0.9009	0.007	-133.232	0.000	-0.914	-0.888
ma.L1	-0.7944	0.012	-68.642	0.000	-0.817	-0.772
ma.L2	-0.2270	0.014	-16.184	0.000	-0.254	-0.199
ma.L3	0.2406	0.014	17.083	0.000	0.213	0.268
ma.L4	0.2495	0.013	18.664	0.000	0.223	0.276
ma.L5	-0.0598	0.014	-4.346	0.000	-0.087	-0.033
ma.L6	-0.3884	0.011	-34.999	0.000	-0.410	-0.367
ar.S.L8	-0.4153	0.012	-34.222	0.000	-0.439	-0.391
sigma2	2648.91	38.849	68.185	0.000	2572.77	2725.06

Ljung-Box (L1) (Q):	0.55	Jarque-Bera (JB):	249.24
Prob(Q):	0.46	Prob(JB):	0.00
Heteroskedasticity (H):	1.14	Skew:	0.24
Prob(H) (two-sided):	0.00	Kurtosis:	3.75

Figure 12. Résultats du modèle SARIMA

La log-vraisemblance, l'AIC, le BIC et le HQIC sont des critères utilisés pour comparer les modèles entre eux, de sorte que les meilleurs modèles aient des valeurs plus faibles. La fonction de log-vraisemblance identifie une distribution qui correspond le mieux aux données échantillonnées. Le critère d'information d'Akaike (AIC) est utilisé pour déterminer la force du modèle de régression linéaire. Tandis que le critère d'information bayésien (BIC) est également utilisé pour juger de la qualité de l'ajustement des données. Quant au critère d'information de Hannan-Quinn (HQIC), il constitue un autre critère de sélection de modèles, utilisé comme alternative à l'AIC et au BIC.

La figure 13 illustre la décomposition des données analysées en tendance, composante saisonnière et résidu à l'aide de la méthode « Seasonal-Trend decomposition using LOESS » (STL). Cette approche permettant d'évaluer la présence éventuelle de saisonnalité dans les données. Les observations tirées de la figure indiquent une saisonnalité hebdomadaire dans le modèle, suggérant un cycle hebdomadaire dans le flux de trafic. Les données examinées révèlent une tendance à la hausse et à la baisse, avec une stabilité du flux de trafic dans la plage de 200 à 400 véhicules par heure. L'analyse des résidus met en évidence leur pertinence, avec des périodes de variabilité notable pendant les mois d'été.

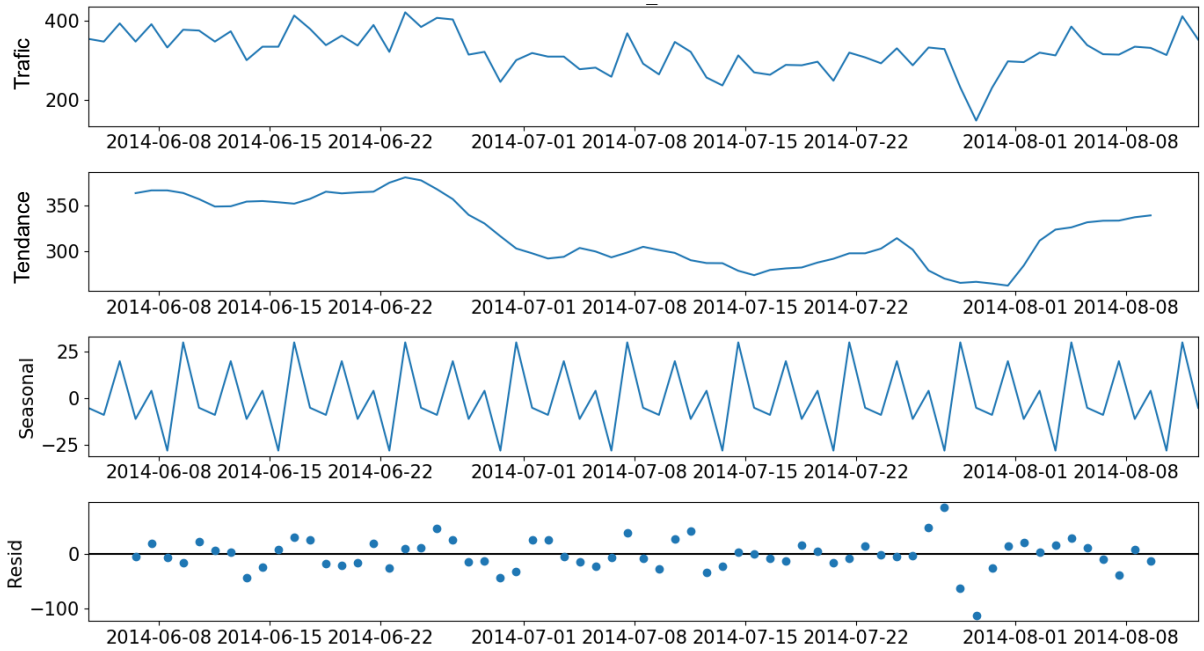


Figure 13. Résultats de la décomposition saisonnière et tendancielle à l'aide du Loess

5.2.4. Interprétation du modèle SVR

Les valeurs SHAP (Shapley Additive Explanations) [40] représentent des métriques qui quantifient l'impact individuel de chaque caractéristique sur les prédictions d'un modèle. Une valeur SHAP de 0

indique une absence d'effet, tandis qu'une valeur positive indique une corrélation positive et une valeur négative une corrélation négative. Pour visualiser les caractéristiques les plus importantes d'un modèle, les valeurs SHAP de chaque caractéristique sont tracées pour chaque échantillon. Dans le graphique « Beeswarm », représenté dans la figure 14, les caractéristiques sont classées selon la valeur absolue moyenne des valeurs SHAP pour tous les échantillons. Ce graphique est employé pour illustrer la répartition des impacts de chaque caractéristique sur les sorties du modèle. Chaque point du graphique correspondant à un enregistrement de trafic routier, sa position horizontale étant déterminée par sa valeur SHAP. Les points sont distribués le long de chaque ligne de caractéristique pour refléter leur densité. De plus, l'axe vertical classe les stations par ordre croissant d'effet, les couleurs indiquant quant à elles les niveaux de trafic (bleu pour faible trafic, rouge pour trafic élevé).

D'après la figure 14, toutes les stations exercent une influence significative sur les prédictions du modèle SVR. La station la plus influente est St138_1, tandis que la moins influente est St137_2. Il est observé que la station St138_1 affiche la corrélation la plus forte avec la station cible. Un faible niveau de trafic routier (bleu) à St138_1 étant fortement associé à un faible trafic à la station cible, et inversement. De plus, on observe que toutes les stations présentent une corrélation positive avec la variable cible. D'après la répartition des valeurs SHAP, on constate que celles-ci se situent dans la même plage sur l'axe horizontal pour toutes les caractéristiques, suggérant ainsi que le modèle SVR accorde une importance similaire à chacune d'elles pour les prédictions finales. Par conséquent, il découle du

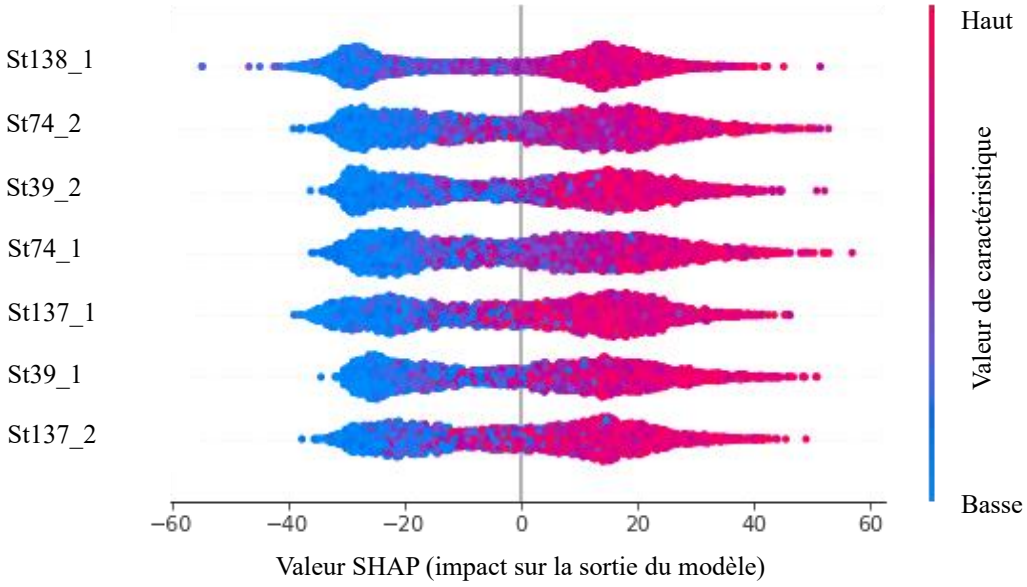


Figure 14. Graphique "Beeswarm" Résultats du SHAP pour le modèle SVR

modèle SVR que toutes les caractéristiques sont positivement corrélées à la sortie, signifiant qu'une augmentation du trafic routier aux stations d'entrée induit une augmentation correspondante du trafic à la station de sortie.

5.2.5. Comparaison entre les différents modèles

Pour comparer les performances des divers modèles, nous avons employé plusieurs métriques d'évaluation telles que l'erreur quadratique moyenne (MSE), la racine carrée de l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE), le pourcentage d'erreur absolue moyenne (MAPE), et le coefficient de détermination (R-carré). Dans ce cadre, les données présentées dans le tableau 5 révèlent que le modèle SARIMA surpasse la régression linéaire et le modèle SVR en affichant des erreurs nettement réduites. En outre, le modèle SARIMA affiche un R-carré ajusté élevé de 92,72, indiquant une meilleure concordance entre le modèle et les observations. Ce constat suggère donc que le modèle de régression linéaire offre une précision supérieure à celle du modèle SVR. Contrairement au modèle SVR qui accorde une importance uniforme à toutes les stations dans la prédiction, le modèle de régression linéaire attribue une importance différenciée à chaque station, soulignant ainsi l'impact individuel de chacune sur le modèle.

Tableau 5. Évaluation des modèles MCO, SARIMA et SVR

Modèle / Métriques	Régression linéaire	SARIMA	SVR
MSE	7973,51	1639,19	8052,05
RMSE	89,29	40,48	89,73
MAE	64,95	29,58	65,49
MAPE	0.4072	0.2515	0.4152
R-carre ajusté	62,96	92,72	26,60

Les résultats démontrent que le modèle SARIMA affiche un R-carré ajusté de 92,72, tandis que le modèle SVR présente un R-carré ajusté de 26,60 et la régression linéaire, de 62,96. Ceci suggère que le modèle SARIMA s'ajuste mieux à la distribution des données que le modèle SVR et la régression linéaire. Cependant, cela ne sous-entend pas nécessairement que le modèle SARIMA est plus précis que le SVR et la régression linéaire. Pour évaluer précisément la précision des modèles, il est indispensable de comparer d'autres mesures telles que le RMSE, le MSE ou le MAPE. Le modèle SARIMA affiche un RMSE de 40,48 unités, tandis que le modèle SVR présente une valeur de 89,73, et la régression linéaire,

de 89,29. Ceci implique que le modèle SARIMA est en moyenne 2,2 fois plus précis que le modèle SVR et la méthode de régression linéaire.

Par ailleurs, les résultats des prédictions du trafic routier pour les trois modèles sur une période de 48 heures sont présentés dans la figure 15. À ce propos, les observations révèlent que la régression linéaire et le SVR suivent de près les tendances réelles, mais avec des variations notables. Cependant, le modèle SARIMA a pu anticiper le trafic routier avec une précision supérieure à celle du modèle de régression linéaire et du modèle SVR, se traduisant par une moindre erreur de prédiction et une meilleure adéquation aux données observées. Cette performance élevée du modèle SARIMA était attendue, étant donné la saisonnalité manifeste du trafic routier. De plus, le modèle SARIMA parvient rapidement à identifier les changements de tendance lorsque ces derniers se manifestent. En comparaison avec le SVR et la régression linéaire, l'ajustement du modèle SARIMA est notablement plus satisfaisant, montrant ainsi sa capacité accrue à capturer les variations saisonnières. Par conséquent, le modèle SARIMA s'avère être le choix le plus approprié pour la modélisation des données présentant une saisonnalité marquée.

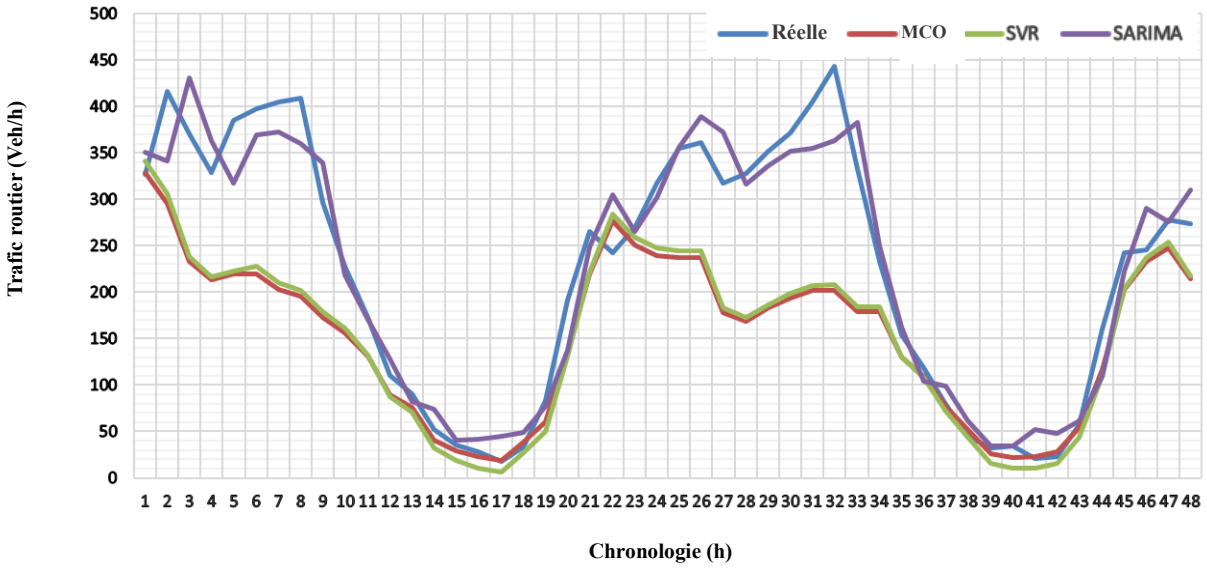


Figure 15. Comparaison des performances de trois modèles de prédiction du trafic routier

6. Conclusion

Ce chapitre expose l'application et l'examen de trois algorithmes d'apprentissage automatique pour la création d'un modèle de prédiction des données horaires de trafic routier. L'étude repose sur des données historiques couvrant la période de 2014 à 2017, provenant de quatre segments routiers à

Marrakech, au Maroc. La contribution principale de cette recherche réside dans l'exploitation exhaustive des données historiques sur le trafic routier, ainsi que de leur corrélation spatio-temporelle, en vue de la prédiction des données manquantes d'un capteur en fonction de ses stations environnantes. Les performances des modèles sont évaluées à l'aide de mesures d'erreur bien établies dans la littérature, et les résultats de la simulation démontrent que le modèle SARIMA surpasse les autres en termes de précision et de stabilité. Ce modèle de prédiction du trafic envisagé sera employé pour reconstruire les données manquantes ou endommagées grâce aux valeurs prédites. Cette démarche vise à pallier les lacunes inhérentes à la matrice de trafic utilisée pour calculer le trafic moyen journalier annuel. Enfin, en regardant vers l'avenir, il est opportun d'envisager l'exploitation des données en temps réel fournies par les capteurs déployés par le Ministère de l'Equipement et de l'Eau. Bien que leur déploiement complet ne soit pas encore achevé sur tout le territoire, ces données offrent un potentiel prometteur. Une analyse intelligente de ces données permettrait le développement de modèles de prédiction des flux de trafic à court terme, offrant ainsi aux usagers de la route des informations anticipées sur le trafic pour une gestion efficace de leur temps et de leurs déplacements, en évitant les périodes de congestion.

Chapitre 3. Modèle de prédiction du trafic routier à court terme à l'aide de l'Extreme Learning Machine

1. Introduction

Comme nous l'avons précédemment évoqué, la croissance des villes et la mobilité croissante de la population ont entraîné une augmentation rapide du nombre de véhicules sur les routes. La gestion du trafic routier est désormais confrontée à des défis tels que les embouteillages, les accidents et la pollution de l'air. Des statistiques récentes révèlent que la majorité des accidents de véhicules se produisent généralement aux abords des routes encombrées. Cela s'explique par le fait que les conducteurs ont tendance à rouler plus vite, avant ou après avoir rencontré des embouteillages, afin de compenser le retard subi. En conséquence, les chercheurs de l'industrie et les académiques se sont attachés à rendre les systèmes de gestion du trafic plus efficaces pour faire face aux problèmes susmentionnés. Dans ce contexte, les données sur le trafic augmentent rapidement et leur analyse est un élément clé dans l'élaboration d'une stratégie pour le réseau routier. La question la plus importante étant de savoir comment analyser et tirer profit de cette mine d'or d'informations afin de prédire les données futures. Dans cette perspective, un système de prévision du trafic précis est l'une des étapes critiques de l'exploitation d'un Système de Transport Intelligent et il est extrêmement indispensable aux praticiens pour planifier les itinéraires et réguler le trafic.

Dans ce contexte, notre étude cherche principalement à exploiter la masse d'informations historiques collectées sur une section spécifique de route afin de prédire le trafic horaire à venir. Outre cet objectif premier, d'autres motivations sous-tendent notre recherche. Tout d'abord, la gestion efficace du trafic routier revêt une importance cruciale pour la sécurité des usagers de la route ainsi que pour la fluidité des déplacements. En anticipant les variations de flux de circulation, il devient dès lors possible de mettre en place des mesures préventives pour réduire les risques d'accidents et de congestions, contribuant ainsi à améliorer la sécurité routière globale [3]. De plus, une meilleure compréhension des modèles de trafic permet d'optimiser l'utilisation des infrastructures existantes. En identifiant les périodes de pointe et les zones à risque de saturation, les autorités responsables peuvent prendre des décisions éclairées en matière de planification urbaine et d'investissement dans de nouvelles infrastructures. Cette anticipation des besoins en matière de transport peut également contribuer à réduire les coûts associés à la construction et à l'entretien des routes, tout en favorisant une utilisation plus

efficace et durable des ressources publiques. Par ailleurs, la prédiction précise du trafic routier revêt un intérêt économique non négligeable, notamment dans le domaine du transport de marchandises et de la logistique. En fournissant des estimations fiables du temps de parcours et des conditions de circulation, les entreprises optimisent leurs chaînes d'approvisionnement et réduisent les coûts liés au transport. De même, les voyageurs individuels bénéficient de ces prévisions pour planifier leurs déplacements de manière plus efficace, minimisant ainsi les pertes de temps et les frustrations associées aux embouteillages imprévus.

2. Méthodes existantes de prédiction

Dans les écrits académiques, le trafic est fréquemment assimilé à un flux en raison de ses similitudes avec les fluides. Par conséquent, lorsque nous abordons la prédiction du flux de trafic, notre objectif est d'anticiper l'état à venir, tel que le volume, la vitesse, la densité ou le comportement, du flux de circulation en se fondant sur des données historiques ou en temps réel. La capacité à anticiper ces paramètres est essentielle dans le domaine des systèmes intelligents, jouant un rôle crucial dans le processus décisionnel. Son importance découlant des nombreux avantages qu'elle peut offrir dans les secteurs de la gestion des transports, de la sécurité routière et de la planification urbaine.

En raison de son importance cruciale dans divers secteurs, le domaine de la prédiction du trafic routier a suscité un intérêt considérable de la part des chercheurs. Ces derniers ont exploré un large éventail de méthodes appropriées, allant des modèles de trafic aux approches statistiques et d'apprentissage automatique. L'analyse prédictive repose sur une gamme de techniques analytiques, notamment la modélisation prédictive et l'exploration de données [1], qui s'appuient sur des données historiques et actuelles pour estimer ou anticiper les résultats futurs [2]. L'apprentissage automatique, sous-domaine de l'intelligence artificielle, a émergé de l'étude de la reconnaissance des formes, explorant la capacité des algorithmes à apprendre à partir des données et à en faire des prédictions. En outre, ces algorithmes peuvent évoluer vers des décisions précises et fondées sur les données, dépassant les directives de programmation traditionnelles. On distingue dans cet environnement trois catégories d'algorithmes d'apprentissage automatique : l'apprentissage supervisé (incluant la régression, les arbres de décision, KNN, la régression logistique, etc.), l'apprentissage non supervisé (tel que l'algorithme Apriori, K-moyennes, etc.) et l'apprentissage par renforcement (englobant l'apprentissage par processus de décision de Markov, entre autres) [41]. Ces algorithmes sont largement employés pour résoudre une variété de problèmes dans divers domaines, y compris les applications médicales [42][43].

Les heuristiques fondées sur une analyse statistique des données historiques représentent une approche pratique et accessible pour la prédiction, impliquant des techniques telles que l'algorithme KNN, la régression linéaire et non linéaire, ainsi que les modèles linéaires autorégressifs comme ARIMA [17][18][44]. Les modèles ARIMA, en particulier, sont considérés comme la classe de modèles la plus générale pour la prévision des données de séries temporelles, car ils offrent une précision et une adaptabilité potentiellement élevées, ainsi qu'une relative simplicité de mise en œuvre [19][20]. Cependant, leur utilisation nécessite souvent la transformation de la série temporelle en une forme stationnaire, et leur nature linéaire limite leur capacité à capturer les modèles non linéaires inhérents aux séries temporelles [3].

Pour les problèmes de régression, le modèle de régression à vecteur de support se distingue comme un outil puissant, notamment pour l'estimation des données réelles [21]. En tant qu'algorithme d'apprentissage supervisé, le SVR utilise une fonction de perte symétrique qui pénalise les écarts par rapport aux bornes de décision. Cependant, le choix de la fonction de noyau appropriée demeure un défi, et le SVR peut être moins rapide que d'autres ANN pour des performances de généralisation comparables. Par ailleurs, les ANN ont récemment attiré une attention croissante dans le domaine de la prédiction du trafic routier. Effectivement, ces modèles exploitent des événements historiques tels que les horaires scolaires, les vacances et d'autres événements périodiques pour prédire le volume de trafic à partir des données historiques de chaque carrefour important de la ville [22].

L'Extreme Learning Machine, technique récente pour les réseaux neuronaux à propagation avant à une seule couche cachée (SLFN), a été largement utilisée comme classifieur puissant dans divers problèmes de référence [45]. Elle se distingue par sa structure simple et sa capacité à apprendre rapidement tout en offrant une meilleure performance de généralisation. Dans [46] un cadre combinant le lissage exponentiel et l'ELM a été développé pour prédire les flux de trafic routier. Cette étude cherche à améliorer la précision et l'efficacité de la prédiction en appliquant l'approche d'optimisation Taguchi. Les résultats démontrent l'efficacité de la méthode Taguchi pour déterminer la configuration optimale, et confirment que le modèle proposé atteint un haut niveau de précision. Par ailleurs, l'utilisation de l'ELM pour la prédiction des données de trafic en temps réel a été explorée dans [47], avec des résultats prometteurs. D'autres recherches ont également exploité l'ELM pour anticiper la congestion du trafic sur des sections spécifiques de route [48], puis pour généraliser ces prédictions à l'ensemble du réseau routier de la ville. Précisons qu'une approche alternative a été proposée dans [49], introduisant une Extreme Learning Machine séquentielle en ligne pour la prédiction à court terme du trafic routier, présentant des

performances élevées, bien que nécessitant un grand nombre de neurones et de données, et ne tenant pas compte de tous les facteurs influençant le trafic. Malgré ces avancées, certaines études ultérieures ont cherché à améliorer la prédiction à court terme en utilisant le modèle ELM, mais sans aborder de manière exhaustive les défis inhérents à ce modèle ou à minimiser les fluctuations qu'il peut générer [50].

Dans le domaine de la prédiction des flux de trafic, certains chercheurs ont exploré diverses configurations de réseaux neuronaux profonds, comme en témoignent les études référencées [23][24][51][52]. Les travaux de référence, tels que ceux détaillés dans l'article de Li & Lu [23], expose une méthodologie axée sur l'apprentissage profond pour la prédiction des flux de trafic lors d'événements spéciaux, soulignant ainsi la remarquable précision obtenue à court terme. Cependant, ces approches exigent un investissement temporel conséquent et une quantité importante de données sur le trafic routier afin d'assurer leur fiabilité. Certains scientifiques ont également envisagé des modèles hybrides, combinant plusieurs techniques de prédiction du trafic, mettant en évidence la nécessité d'une approche multifacette pour aborder toutes les situations de trafic [53]. Malgré leur variabilité en termes d'efficacité, ces modèles complexes nécessitent un investissement temporel significatif et présentent des défis non négligeables en matière de mise en œuvre.

Notre étude propose un modèle hybride associant les techniques de l'ELM et d'ensemble pour la prédiction du trafic routier. Comparativement aux réseaux neuronaux traditionnels et aux machines à vecteurs de support, l'ELM présente des avantages notables tels qu'une vitesse d'apprentissage élevée, une mise en œuvre simplifiée et une intervention humaine minimale [54][55]. Une comparaison avec des algorithmes d'apprentissage traditionnels tels que MLP, SVR et ARIMA a été réalisée, démontrant la supériorité de notre modèle en termes de précision de prédiction. Les résultats expérimentaux attestent de la robustesse de notre modèle face aux valeurs aberrantes et de sa capacité de généralisation supérieure.

Rappelons que dans notre thèse, nous aborderons les questions suivantes pour explorer en profondeur les principes clés de la prédiction du trafic routier. :

- Quelles sont les caractéristiques essentielles requises pour prédire le trafic routier ?
- Comment pouvons-nous élaborer un jeu de données en nous appuyant sur des données collectées par des capteurs fixes ?
- Quelles sont les méthodes appropriées pour traiter notre ensemble de données en utilisant des algorithmes d'apprentissage automatique ?

- Comment pouvons-nous améliorer la capacité prédictive de notre modèle ?

3. Modèle de prédiction du trafic routier basé sur l'ELM

3.1. Description du modèle

L'ELM représente une forme spécifique de réseau neuronal artificiel employée tant pour la classification que pour l'analyse de régression. Ce système repose sur une architecture à une seule couche de neurones, générés de manière aléatoire ; il se distingue par une méthode d'apprentissage analytique, différente des approches numériques conventionnelles. Cette approche novatrice de la formation confère à l'algorithme une capacité d'apprentissage et de traitement des données nettement supérieure à celle des modèles traditionnels, tout en maintenant un niveau de précision élevé.

Dans le cadre de notre recherche, le modèle ELM envisagé est constitué de 10 neurones dans la couche d'entrée (voir figure 16). Chaque neurone représentant une caractéristique essentielle qui sera exploitée dans la prédiction du trafic routier horaire subséquent. La couche cachée est formée de multiples neurones entièrement connectés, dont le nombre a été délibérément sélectionné après une série de tests approfondis. Par ailleurs, le modèle ELM est doté d'une couche de sortie contenant un unique neurone dédié à la valeur prédite du trafic horaire. Pour ce faire, divers modèles ELM ont été entraînés sur les données d'apprentissage et évalués sur les données de test (voir figure 17). Le choix du nombre de modèles ELM à inclure dans cet ensemble a été minutieusement établi. Les prédictions générées par l'ensemble des modèles ELM ont été agrégées afin d'obtenir la prédiction finale. Dans ce contexte, la méthode de la moyenne arithmétique a été adoptée pour combiner les prédictions des différents modèles ELM, en raison de sa simplicité et de son efficacité comparativement à l'utilisation d'un modèle ELM individuel.

3.2. Méthodologie de recherche

3.2.1. Prétraitement des données

L'analyse des données relatives au trafic est d'une importance capitale et stratégique dans l'élaboration des politiques nationales en matière d'infrastructures et de systèmes de transport, étant donné qu'elle fournit des informations cruciales pour la planification efficace des routes, l'optimisation de la circulation, la réduction des embouteillages et des accidents, ainsi que la conception de solutions

innovantes visant à améliorer la mobilité des citoyens et à favoriser le développement économique et social du pays.

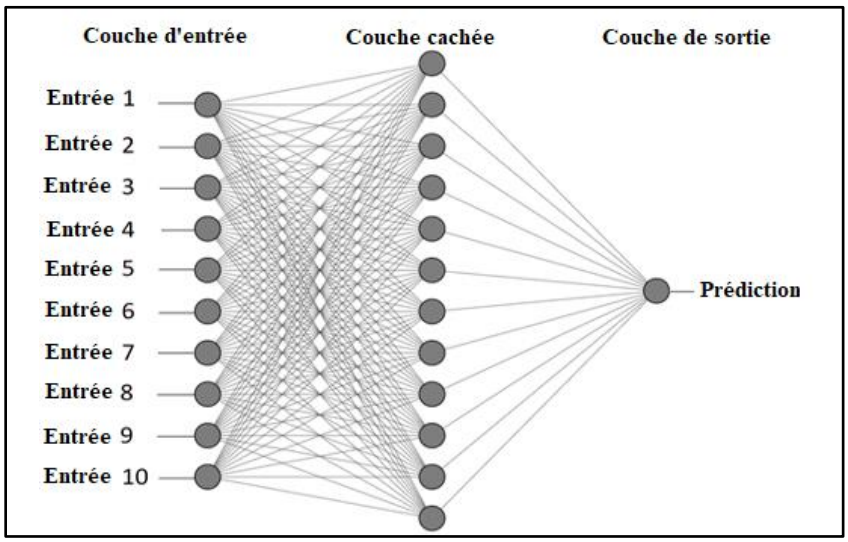


Figure 17. Structure du modèle ELM

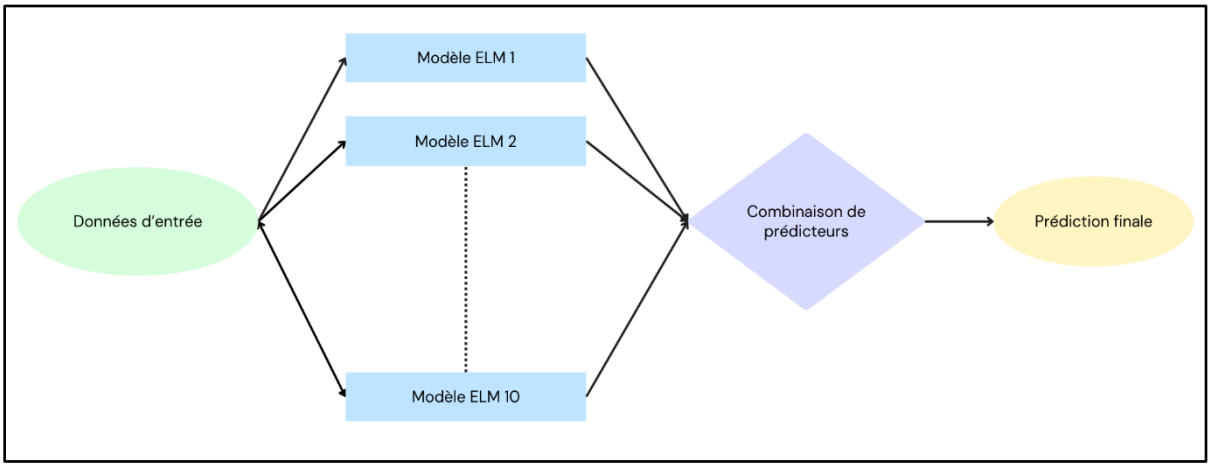


Figure 16. Cadre proposé pour la prédiction du trafic à l'aide d'une méthode d'ensemble

À ce titre, les données utilisées ont été collectées auprès du CNER, qui utilise des boucles inductives permanentes ou temporaires installées sous la surface des routes pour recueillir une quantité substantielle de données statistiques sur les flux de trafic observés sur certains des principaux axes routiers du Maroc. Ces sources de données comprennent des capteurs placés à des endroits fixes, capables de détecter la présence de véhicules à proximité. Initialement, les détecteurs à boucle inductive étaient prédominants, mais une grande variété de capteurs est désormais disponible [56]. Notons que l'ensemble de données utilisé englobe les flux de trafic enregistrés sur une période de cinq ans, de 2013 à 2017 ; il

est constitué d'une série de fichiers distincts. Chaque fichier renferme des informations sur le trafic pour une période donnée. Les métadonnées associées à ces fichiers incluent la date (jour et heure) du début et de la fin de l'enregistrement, le nombre de voies, ainsi qu'une répartition du trafic par classe de véhicule en fonction de leur longueur (voir les tableaux 6 et 7).

La première étape du processus de préparation de l'ensemble de données implique la consolidation des données de trafic routier dans un fichier unique. Dans notre étude, nous agrégerons la somme du trafic horaire pour toutes les classes de véhicules, au lieu de considérer le nombre de véhicules pour chaque classe distincte. Ensuite, la deuxième étape consistera à éliminer les enregistrements de données contenant des valeurs manquantes, car ces lacunes peuvent résulter d'une défaillance du capteur ou d'une absence de certaines données caractéristiques. Une fois cette étape franchie, le prétraitement des données se poursuit avec la normalisation des données. Cette technique permettant de placer toutes les valeurs dans une plage spécifique, ce qui permet d'évaluer l'impact des facteurs d'entrée sur le modèle en fonction des caractéristiques plutôt que des valeurs numériques. De ce fait, la normalisation présente deux avantages majeurs : elle standardise l'échelle du réseau et évite que les facteurs d'entrée dotés de valeurs numériques élevées ne prédominent sur ceux ayant des valeurs numériques plus faibles.

Toutefois, l'échelle des données peut influencer considérablement les performances des algorithmes d'apprentissage automatique, impactant ainsi la précision du modèle. L'utilisation de petits poids dans le modèle et la gestion des erreurs entre les valeurs prédites et les valeurs attendues soulignent l'importance cruciale de l'échelle des entrées et sorties utilisées pour entraîner le modèle. Ainsi, des variables d'entrée non normalisées peuvent entraîner un processus d'apprentissage lent ou instable, tandis que des variables cibles non normalisées dans le contexte de problèmes de régression peuvent provoquer une explosion du gradient, entraînant la défaillance du processus d'apprentissage. Subséquemment, il est fortement recommandé de normaliser les données dans une plage commune afin d'éliminer les disparités entre les caractéristiques présentant des plages différentes. Parmi les techniques de normalisation, la méthode Min-Max est couramment utilisée, définissant la plage de mise à l'échelle par défaut de 0 à 1. Les données sont normalisées par cette méthode comme suit : chaque valeur initiale x_i dans un ensemble de données sera transformée en une nouvelle valeur normalisée x_i' selon la formule (3.1) :

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{15}$$

Où x_{min}et x_{max} représentent respectivement la valeur minimale et maximale de l'ensemble de données d'origine. Au sein de notre étude, les caractéristiques ont été normalisées en utilisant la méthode de mise à l'échelle par défaut, et ce conformément à la formule précédemment mentionnée (3,1). En revanche, la variable cible a été normalisée dans l'intervalle [-1, 1] par l'équation suivante (3.2) :

$$x_i' = a + \frac{(b-a)(x_i - x_{\min})}{x_{max} - x_{min}} \quad (16)$$

Où a et b représentent respectivement les nouvelles valeurs minimale et maximale de la plage cible, qui sont, dans notre cas, fixées à -1 et 1.

Tableau 6. Échantillon de données provenant d'un compteur permanent

Date	Heure	Voie 1	Classe 1	Classe 2	Classe 3	Classe 4	Voie 2	Classe 1	Classe 2	Classe 3	Classe 4
01/01/16	22 h	181	85	45	34	17	202	46	63	52	41
01/01/16	23 h	389	182	152	42	13	136	46	41	25	24
02/01/16	00 h	418	156	124	87	51	118	42	45	19	12
02/01/16	01 h	438	205	140	52	41	134	52	43	24	15

Tableau 7. Classification des véhicules en fonction de leur longueur

Classe	1	2	3	4
Longueur (m)	L < 5.2	5.2 < L < 6.5	6.5 < L < 10,5	10.5< L

3.2.2. Variables d'entrée

Pour prétraiter les données, nous avons intégré plusieurs caractéristiques pertinentes susceptibles d'influencer le flux du trafic routier, formant ainsi la couche d'entrée de notre modèle (comme illustré dans la figure 16) :

- Caractéristique 1 : jour de la semaine (lundi, mardi, mercredi…)
- Caractéristique 2 : type de jour férié (nationale, religieuse ou autre)
- Caractéristique 3 : partie de la journée (matin ou soir)
- Caractéristique 4 : statut de l'école/vacances/jour férié
- Caractéristique 5 : trafic de l'heure précédente (heure-1)
- Caractéristique 6 : enregistrement de la même heure le jour précédent (jour-1)
- Caractéristique 7 : enregistrement de la semaine précédente (Semaine-1)

- Caractéristique 8 : valeur de la même heure le même jour du mois précédent (Mois-1)
- Caractéristique 9 : trafic actuel à l'heure donnée (tous types de véhicules)
- Caractéristique 10 : plage horaire (divisée en 7 périodes). Chaque journée est subdivisée en différentes périodes, chaque période présentant apparemment des classes de trafic distinctes :
    - Période 1 : 01 h 00 min - 5 h 59 min
    - Période 2 : 06 h 00 min - 8 h 59 min
    - Période 3 : 09 h 00 min - 10 h 59 min
    - Période 4 : 11 h 00 min - 14 h 59 min
    - Période 5 : 15 h 00 min - 17 h 59 min
    - Période 6 : 18 h 00 min - 21 h 59 min
    - Période 7 : 22 h 00 min - 00 h 59 min

Les caractéristiques choisies pour enrichir les données d'entrée de notre modèle ont été sélectionnées pour capturer divers aspects temporels et contextuels influençant les schémas de trafic routier.

La première caractéristique, le jour de la semaine, permet de tenir compte des variations hebdomadaires du trafic, notamment les différences entre les jours de semaine et les week-ends. Ensuite, le type de jour férié identifie les jours spéciaux, tels que les fêtes nationales ou religieuses, qui peuvent entraîner des modifications significatives des déplacements. La division en parties de la journée distingue les heures de pointe du matin et du soir, reflétant les variations du trafic tout au long de la journée. La quatrième caractéristique, le statut de l'école, des vacances ou des jours fériés, indique les périodes où les schémas de déplacement peuvent différer en raison de la fermeture des écoles ou des vacances. Enfin, les caractéristiques 5 à 8 intègrent des données historiques sur le trafic, fournissant des informations sur les tendances et les variations saisonnières du trafic à différentes échelles temporelles. La caractéristique 9 délivre un aperçu du niveau actuel de trafic, tandis que la caractéristique 10 divise la journée en périodes distinctes pour capturer les fluctuations du trafic tout au long de la journée. Toutes ces caractéristiques combinées permettent à notre modèle de prendre en compte les variations temporelles et contextuelles complexes du trafic routier lors de la prédiction.

4. Méthodes

Dans le cadre de cette partie consacrée aux méthodes, nous explorerons deux approches majeures dans le domaine de l'apprentissage automatique pour la prise de décision. Nous commencerons par

examiner l'Extreme Learning Machine, une méthode innovante et rapide pour les réseaux neuronaux SLFN. Ensuite, nous aborderons les Systèmes d'Ensemble dans la Prise de Décision (SEPD), une approche qui tire parti de la diversité des modèles individuels pour améliorer la précision et la robustesse des prédictions.

4.1. Extrême learning machine

L'Extreme Learning Machine constitue une méthode novatrice et rapide pour les réseaux neuronaux SLFN, où les poids d'entrée et les biais de la couche cachée sont sélectionnés de manière aléatoire (voir figure 18) [54]. Le problème d'optimisation associé à l'apprentissage des paramètres d'un modèle ELM peut être résolu de manière analytique, engendrant ainsi une solution fermée qui nécessite uniquement des opérations de multiplication et d'inversion de matrices [57]. Ce processus d'apprentissage est donc réalisable de manière efficace, évitant ainsi le recours à des algorithmes itératifs, tels que ceux utilisés dans le cadre des réseaux neuronaux de rétropropagation (BP-NN), ou à des méthodes de résolution de problèmes plus complexes, comme dans la formulation standard du SVM. Dans le cadre de l'ELM, les poids d'entrée, établissant la connexion entre la couche d'entrée et la couche cachée, ainsi que les biais cachés, sont sélectionnés aléatoirement, tandis que les poids de sortie, reliant la couche cachée à la couche de sortie, sont déterminés analytiquement en utilisant la méthode de Moore-Penrose.

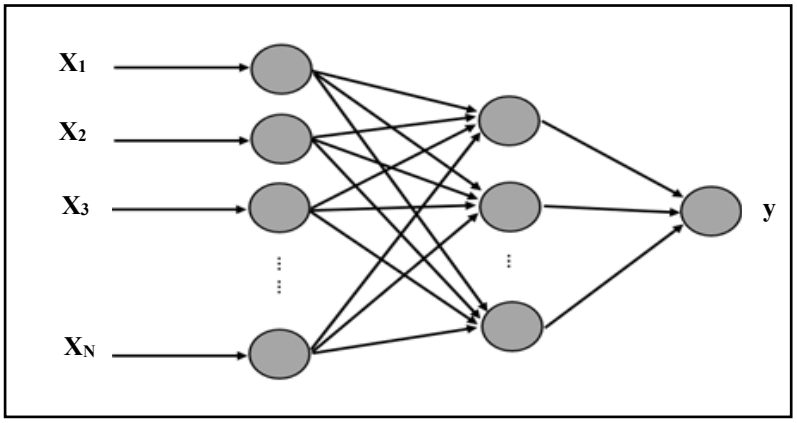


Figure 18. Structure du réseau SLFN

Étant donné un ensemble d'apprentissages fini $(x_j, y_j) \in \mathbb{R}^n \times \mathbb{R}^m$, où les x_j désignent les variables d'entrée et les y_j représentent les variables de sortie. Le réseau neuronal SLFN avec N neurones de la couche cachée s'écrit sous la forme générale suivante :

$$\sum_{i=1}^{N} \beta_i g_i(x_j) = \sum_{i=1}^{N} \beta_i g(w_i . x_j + b_i) \quad (17)$$

Où w_i sont les vecteurs de poids d'entrée associés à chaque neurone de la couche cachée. Les β_i sont les poids de sortie associés à chaque neurone de la couche cachée. Les b_i sont les biais de chaque neurone de la couche cachée, x_j représente l'entrée du réseau neuronal pour le j^{ème} exemple, et g est la fonction d'activation utilisée pour chaque neurone de la couche cachée.

L'objectif de l'ELM est de minimiser l'erreur d'apprentissage, c'est-à-dire de minimiser la différence entre les sorties prédites et les sorties réelles :

$$\sum_{j=1} \|o_j - y_j\| = 0 \quad (18)$$

Avec o_j est la sortie prédite pour la j^{ème} entrée et y_j est la sortie réelle de j^{ème} entrée, et $\|o_j - y_j\|$ représente une mesure de la distance entre la sortie prédite et la sortie réelle. L'objectif est d'ajuster les poids et les biais du réseau neuronal ELM pour réduire cette erreur d'apprentissage aussi près que possible de zéro. Ce qui équivaut à :

$$\sum_{i=1}^{N} \beta_i g(w_i . x_j + b_i) = y_j \quad (19)$$

L'équation ci-dessus peut être écrite comme suit :

$$H\beta = Y \quad (20)$$

Avec β est le vecteur des poids de la couche de sortie, Y est le vecteur de sortie, et H est la matrice de sortie de la couche cachée définie par :

$$H = \begin{bmatrix} g_1(x_1) \cdots \cdots g_N(x_1) \\ g_1(x_2) \cdots \cdots g_N(x_2) \\ \vdots \cdots \cdots \vdots \end{bmatrix} \quad (21)$$

Le vecteur β est calculé comme suit :

$$\beta = H^{\dagger} Y \quad (22)$$

Où $H^{\dagger}$ est l'inverse généralisé de Moore-Penrose de H.

4.2. Systèmes d'ensemble dans la prise de décision

Le SEPD constitue un processus fondé sur l'utilisation d'un ensemble de modèles ; chaque modèle étant construit à partir d'une technique spécifique d'apprentissage automatique appliquée à des données distinctes. Le système d'ensemble intègre ensuite ces modèles de manière différenciée en vue de parvenir à une prédiction finale. Notons que cette approche a connu une large adoption dans les domaines de l'apprentissage automatique et de la reconnaissance des formes en raison de ses performances prometteuses et de l'amélioration significative de la précision des prédictions par rapport aux modèles individuels. Il est vrai que les motivations à recourir au SEPD sont multiples, incluant des considérations statistiques, le traitement de grands volumes de données, la rareté des données et la fusion de données [58]. Plusieurs techniques sont disponibles, comprenant des méthodes qui combinent des règles telles que la combinaison algébrique, l'approche basée sur le vote, les modèles de décision et les méthodes basées sur les réseaux neuronaux [59]. La méthode adoptée dans notre travail est la moyenne arithmétique, privilégiée en raison de sa simplicité et sa capacité à générer des résultats robustes [1]. La formule générale de la moyenne arithmétique est donnée par :

$$\mu_i = \frac{1}{N}\sum_{k=1}^{N} y_i^k \tag{23}$$

Où N est le nombre total de modèles dans l'ensemble, y_i^k est la prédiction du modèle k pour la i^{ème} entrée, et μ_i est la prédiction moyenne à considérer pour la i^{ème} entrée.

Le modèle ELM produit des résultats variables pour chaque application sur l'ensemble de données de test en raison de la nature aléatoire de la génération de ses poids d'entrée et du biais de la couche cachée. Cependant, ces fluctuations dans les prédictions sont relativement insignifiantes par rapport à celles observées dans d'autres réseaux de neurones artificiels. Pour pallier ce problème, nous utiliserons des méthodes d'ensemble afin de réduire ces petites variations et d'améliorer la fiabilité des prédictions par rapport à un modèle simple, tout en augmentant la précision. À cette fin, nous générerons plusieurs modèles ELM, que nous appliquerons à nos données tout en maintenant les mêmes ensembles de données d'apprentissage et de test. Étant donné que les résultats de la prédiction diffèrent pour chaque modèle, nous nous servirons de la moyenne algébrique de toutes les prédictions obtenues comme valeur finale prédite à considérer.

4.3. Moyenne mobile intégrée autorégressive

Le modèle ARIMA est couramment employé dans la modélisation des séries temporelles en raison de sa robustesse et de sa facilité de mise en œuvre. Effectivement, la composante autorégressive (AR) du modèle ARIMA repose sur une régression linéaire des valeurs antérieures de la variable d'intérêt. Ce qui lui permet de prédire les valeurs futures. De manière similaire ; la composante moyenne mobile (MA) se fondant sur une régression linéaire des erreurs résiduelles des prévisions précédentes. Le terme « intégré » (I) dans ARIMA fait là référence à la différenciation, consistant à calculer les différences entre les observations consécutives afin de remédier à la non-stationnarité des données. Le modèle ARIMA (p, d, q) complet peut être écrit comme suit :

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (24)$$

Où y'_t est la valeur prédite à un moment *t* de la série temporelle, c est la constante de régression, les ϕ_i sont les coefficients des termes autorégressifs, les $y'_{t-1}, \ldots, y'_{t-p}$ sont les valeurs prédites aux instants précédents, θ_j sont les coefficients des termes de la moyenne mobile, et les $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}, \varepsilon_t$ sont les erreurs résiduelles aux instants précédents.

Pour déterminer les valeurs de p et q, on peut utiliser les tracés de la fonction d'autocorrélation (ACF) et de la fonction d'autocorrélation partielle (PACF). Le tracé du ACF montre l'autocorrélation entre les valeurs prédites y'_t et leurs valeurs retardées y'_{t-k} pour différents décalages *k*. Il aide à identifier la présence de structures de dépendance temporelle dans la série temporelle, telles que les tendances saisonnières ou les motifs cycliques. Le PACF représente la corrélation directe entre une observation à un moment donné y'_t et ses valeurs prédites à l'instant précédent y'_{t-1}, en éliminant les effets de toutes les autres valeurs prédites à des décalages plus éloignés dans le passé. Cette analyse permet ainsi de déterminer les retards significatifs à inclure dans le modèle ARIMA en vue de capturer les tendances et les motifs temporels de manière plus précise.

4.4. Régression à vecteur de support

La régression à vecteur de support est un algorithme de régression qui applique une technique similaire à celle du SVM pour l'analyse de régression. Comme nous le connaissons, les données de régression contiennent des valeurs numériques continues ou discrètes. Pour modéliser ces données, le modèle SVR ajuste la ligne de régression afin d'approcher les meilleures valeurs avec une marge donnée

appelée ε-tube (epsilon-tube, où ε identifie une largeur de tube). Cette marge ε prend en compte la complexité du modèle et le taux d'erreur, qui est défini comme la différence entre les valeurs réelles et les prédictions du modèle. L'objectif du SVR étant d'identifier l'hyperplan qui englobe le plus grand nombre possible de points tout en minimisant l'erreur globale. Les contraintes du problème SVR peuvent être formulées comme suit :

$$|y_i - (w_i x_i + b)| \leq \varepsilon \tag{25}$$

Où y_i représente la valeur cible, x_i correspond aux caractéristiques de l'observation, w_i et b sont les paramètres du modèle, et ε est la marge définie pour tolérer l'erreur. L'équation de l'hyperplan est définie comme suit :

$$y = w^T x + b \tag{26}$$

Où y représente la variable cible, x est le vecteur de caractéristiques, w est le vecteur de poids (coefficients) associés à chaque caractéristique, b est le biais (intercept) de l'hyperplan, et w^T désigne la transposée du vecteur de poids. La fonction objective du SVR est de minimiser les coefficients :

$$min \frac{1}{2} \|w\|^2 \tag{27}$$

4.5. Perceptron multicouche

Le perceptron multicouche, également connu sous le nom de réseau de neurones à propagation avant (FNN), est un modèle utilisant un algorithme d'apprentissage supervisé. La structure de base du MLP comprend trois couches ou plus : une couche d'entrée, une couche cachée et une couche de sortie (voir figure 19). Il est capable d'apprendre des approximations de fonctions non linéaires pour des tâches de classification ou de régression. Grâce à sa structure multicouche, le MLP est capable d'apprendre des représentations hiérarchiques des données. Ce qui peut améliorer sa capacité à extraire des caractéristiques pertinentes pour la tâche à accomplir. De plus, avec une mise en œuvre appropriée et une sélection judicieuse des hyperparamètres, le MLP peut être entraîné de manière efficace et produire des résultats précis et fiables. Cependant, l'utilisation du MLP présente plusieurs limitations. Tout d'abord, il nécessite la spécification de divers hyperparamètres tels que le nombre de neurones cachés, le nombre de couches et le nombre d'itérations. De plus, le MLP est sensible à la mise à l'échelle des caractéristiques. En ce qui concerne les problèmes de régression, le MLP utilise la rétropropagation (BP)

pour calculer les gradients et minimiser la fonction de perte associée à l'erreur quadratique. La rétropropagation est un processus complexe. Aussi, est-il recommandé de commencer avec un nombre restreint de neurones cachés et de couches cachées lors de l'apprentissage. Cependant, une faible configuration en termes de neurones peut conduire à un sous-apprentissage, entraînant ainsi une erreur d'apprentissage élevée. En conséquence, les performances de généralisation des réseaux de neurones basés sur la rétropropagation peuvent être sous-optimales [54][60].

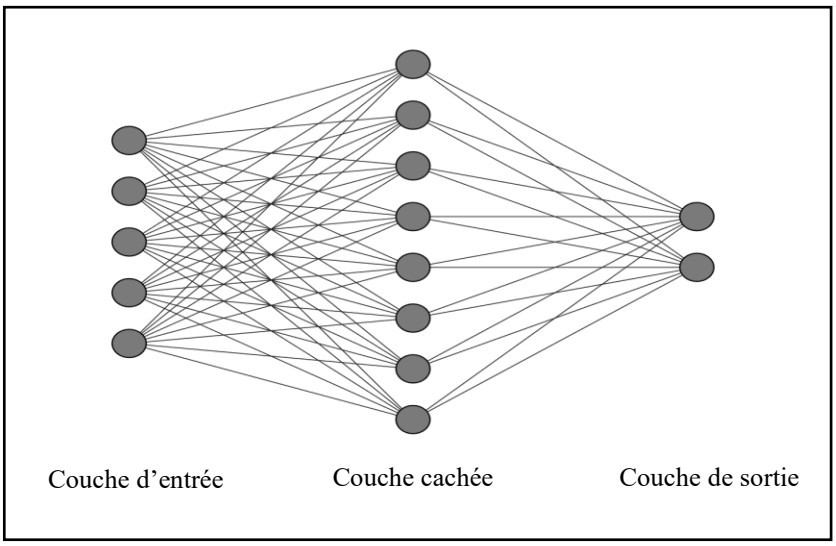


Figure 19. Exemple d'un réseau de neurones à trois couches

5. Entraînement du modèle ELM

Le processus d'entraînement du modèle ELM, revêt une importance particulière dans l'obtention de performances optimales pour la prédiction du trafic routier. Avant de plonger dans les détails de cette phase d'apprentissage, il convient d'examiner les critères d'évaluation utilisés pour évaluer la qualité du modèle construit, ainsi que les méthodes d'ajustement et d'évaluation de la prédiction proposée.

5.1. Critères d'évaluation

L'évaluation du modèle construit est une étape permettant d'estimer sa performance et son aptitude à prédire correctement les résultats. Dans le cadre des modèles de régression destinés à prédire des valeurs numériques, diverses mesures sont employées à cette fin. Parmi lesquelles nous citerons : le Root Mean Square Error (RMSE), le Mean Absolute Error (MAE) et le Mean Absolute Percentage Error (MAPE) [61]. Dans le cadre de notre étude, ces trois métriques ont été sélectionnées pour évaluer la précision du modèle. En outre, l'AIC a été incluse dans l'arsenal des critères d'évaluation afin de

permettre une comparaison plus approfondie des différents modèles. Cette approche exhaustive d'évaluation garantit donc une évaluation rigoureuse et complète de la performance du modèle construit, permettant ainsi de prendre des décisions éclairées quant à son utilisation et son adaptation aux besoins spécifiques de la tâche à accomplir. Les formules des métriques d'évaluation mentionnées sont les suivantes :

RMSE : la formule calcule l'erreur quadratique moyenne :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{28}$$

MAE : la formule de l'erreur absolue moyenne est la suivante :

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \tag{29}$$

MAPE : l'erreur absolue moyenne en pourcentage est donnée par la formule suivante :

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|x_i - \hat{x}_i|}{x_i} \tag{30}$$

Où N représente le nombre d'observations dans l'ensemble de données d'essai, x_i est la sortie réelle de la i^{ème} observation et $\hat{x}_i$ est la valeur prédite pour la i^{ème} observation.

AIC : ce critère [62] est utilisé pour vérifier la complexité des modèles. Il permet de comparer différents modèles en prenant en compte leur ajustement aux données ainsi que leur nombre de paramètres. Plus précisément, le modèle présentant la valeur AIC la plus basse est considéré comme le meilleur, car il parvient à expliquer les données de manière satisfaisante tout en conservant une complexité minimale. La formule de l'AIC est définie comme suit :

$$AIC = nlog\left(\frac{RSS}{n}\right) + 2K \tag{31}$$

Où RSS est la somme des carrés des résidus, K est le nombre de paramètres dans le modèle et n est le nombre d'observations.

5.2. Ajustement du modèle proposé et évaluation de la prédiction

Afin d'améliorer la précision du modèle ELM et de réduire ses fluctuations, nous avons envisagé l'utilisation de méthodes d'ensemble. À cet effet, notre approche combine les performances élevées du modèle ELM avec les avantages des méthodes d'ensemble. Dans un premier temps, nous ajustons le modèle ELM sur nos données d'apprentissage N fois, où N représente le nombre d'ensembles. Le choix du nombre de neurones dans la couche cachée de l'ELM ainsi que la valeur de N est effectué en fonction de la qualité des résultats obtenus. Ensuite, les sorties de chaque modèle de l'ensemble sont combinées algébriquement par calcul de la moyenne, ce qui réduit le risque associé à une mauvaise sélection d'un prédicteur défaillant.

Il est largement admis que plusieurs paramètres peuvent influencer la précision d'un modèle, notamment le nombre de neurones cachés, la taille des données et la fonction d'activation utilisée. Les chercheurs et spécialistes de ce domaine ont tendance à explorer plusieurs combinaisons de paramètres ou à adopter des approches déjà éprouvées. Aussi, avons-nous entrepris 20, 50 et 100 essais pour chaque taille fixe du modèle ELM, et la valeur retenue est la moyenne des résultats obtenus. Le réseau ELM est composé de 10 neurones dans la couche d'entrée (variables d'entrée), d'une couche cachée et d'un unique nœud dans la couche de sortie, fournissant ainsi la prédiction du trafic pour une durée d'une heure. Le choix de la taille de la couche cachée est basé sur les résultats de plusieurs tests ; le nombre de neurones conduisant à la meilleure performance est alors retenu.

En ce qui concerne la préparation de notre ensemble de données, nous avons ajouté des informations sur les caractéristiques pour mieux décrire les observations. Cela a entraîné des enregistrements incomplets, représentant 2,8 % de la taille de l'ensemble de données, qui ont été exclus de notre analyse. Notre modèle utilise ainsi 25 323 observations, dont 22 790 sont affectées à la phase d'apprentissage et 2 533 au test du modèle.

Nous avons par la suite cherché à déterminer les paramètres optimaux pour notre modèle, notamment le nombre de neurones cachés et le nombre d'ensembles N, en tenant compte des critères d'évaluation montrant le moins de fluctuation. L'algorithme ELM a été évalué plusieurs fois pour déterminer le meilleur nombre de neurones cachés, en testant différentes tailles de couche cachée entre 10 et 400. Pour chaque taille de couche cachée, le modèle a été entraîné, testé et évalué successivement 20, 50 et 100 fois. Les critères d'évaluation tels que RMSE, MAE et MAPE ont été calculés et leurs valeurs moyennes ont été retenues.

La figure 20.a présente l'erreur moyenne RMSE calculée lors de la phase de test. Comme illustré dans la figure, les sorties sont normalisées entre 0 et 1, représentant ainsi la RMSE de l'ensemble des données de trafic routier utilisées dans le test. À la suite de 10 tests, nous observons que le nombre optimal de nœuds cachés pour notre ensemble d'apprentissages varie entre 80 et 360, tel que documenté dans [54]. Nous remarquons aussi une diminution de la RMSE à mesure que le nombre de neurones augmente, atteignant une optimisation à 180 neurones dans la couche cachée. Étant donné la sélection aléatoire des neurones cachés, ces résultats sont générés par plusieurs essais, dont la valeur moyenne est retenue. De plus, la courbe des 20 essais présente des fluctuations pour les mesures RMSE et MAE, suggérant qu'elle n'est probablement pas optimale pour le nombre d'ensembles. En revanche, le temps

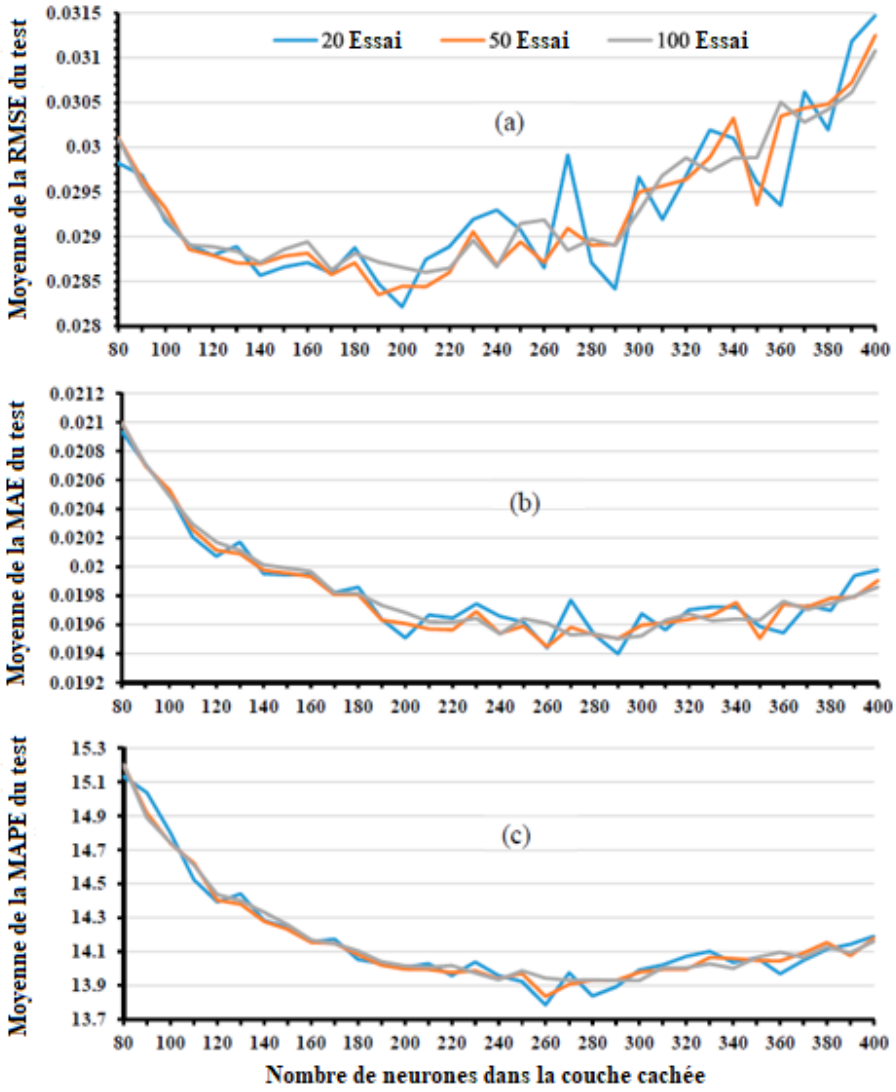


Figure 20. Évolution des erreurs moyennes dans le modèle ELM en fonction du nombre d'essais

d'exécution des 50 essais est considérablement plus rapide que celui des 100 essais. Ainsi, le nombre d'ensembles N retenu pour cette étude est de 50. L'analyse de l'évolution de la performance du modèle en fonction du nombre de neurones cachés pour 50 essais (voir les figures 20.b et 20.c) révèle que, pour notre ensemble d'apprentissages, le nombre optimal de neurones cachés générant de meilleurs résultats avec moins de fluctuations est de 250 neurones.

5.3. Prédiction du trafic routier

La procédure d'apprentissage du modèle d'ensemble SEPD est une méthode qui repose sur l'utilisation répétée d'un modèle ELM spécifique pour créer un ensemble de prédictions et les combiner afin d'améliorer la précision générale des résultats. Celle-ci est décrite étape par étape ci-après :

- Cycle d'apprentissage :

La boucle d'apprentissage est exécutée 50 fois (nombre d'ensembles). À chaque itération, un nouveau modèle ELM est créé avec 250 neurones cachés et un état aléatoire initialisé. Ce modèle est entraîné sur l'ensemble de données d'apprentissage. Le modèle ELM a été entraîné à partir de 90 % de données d'apprentissage.

- Prédiction et agrégation :

Une fois le modèle ELM entraîné, ce dernier est utilisé pour prédire les données de l'ensemble du test. Les prédictions sont ajoutées à un tableau pour chaque itération.

- Évaluation des prédictions :

À chaque itération, les mesures d'évaluation (MAE, RMSE, MAPE) sont mises à jour en accumulant les différences entre les prédictions et les valeurs réelles de l'ensemble de données de test. Les résultats de ces mesures sont utilisés pour évaluer les performances de chaque modèle ELM individuel.

- Évaluation de la moyenne des prédictions et des mesures d'évaluation :

Après avoir entraîné les 50 modèles ELM et agrégé leurs prédictions, celles-ci sont divisées par 50 afin de calculer une moyenne. De même, les moyennes des mesures d'évaluation (MAE, RMSE, MAPE) ont été calculées pour obtenir des mesures moyennes représentatives de tous les modèles ELM.

La procédure d'ensemble SEPD exploite les variations aléatoires dans l'initialisation du modèle ELM pour obtenir une prédiction plus robuste et plus précise. Cela peut contribuer à réduire le surapprentissage et à améliorer la généralisation du modèle sur de nouvelles données.

6. Comparaison des performances des modèles prédictifs

Le modèle ELM a été évalué en utilisant le nombre optimal de neurones déterminé pour la couche cachée. Les performances de ce modèle ont été comparées à celles des modèles SVR et ARIMA. Dans le cas du modèle SVR, différentes combinaisons des paramètres de coût C et de noyau ont été testées. Conformément aux recommandations de [59], la meilleure combinaison a été recherchée parmi les valeurs suivantes : $C = [2^{12}, 2^{11}, \dots, 2^{-1}, 2^{-2}]$ et $\gamma = [2^{4}, 2^{3}, \dots, 2^{-9}, 2^{-10}]$, en utilisant la méthode de validation croisée.

À cet égard, la figure 21 illustre la relation entre le trafic routier horaire prévu, disposé chronologiquement sur l'axe des abscisses, et les résidus des valeurs prédites par rapport aux valeurs réelles, représentées sur l'axe des ordonnées. L'observation révèle l'absence de schéma distinctif pour

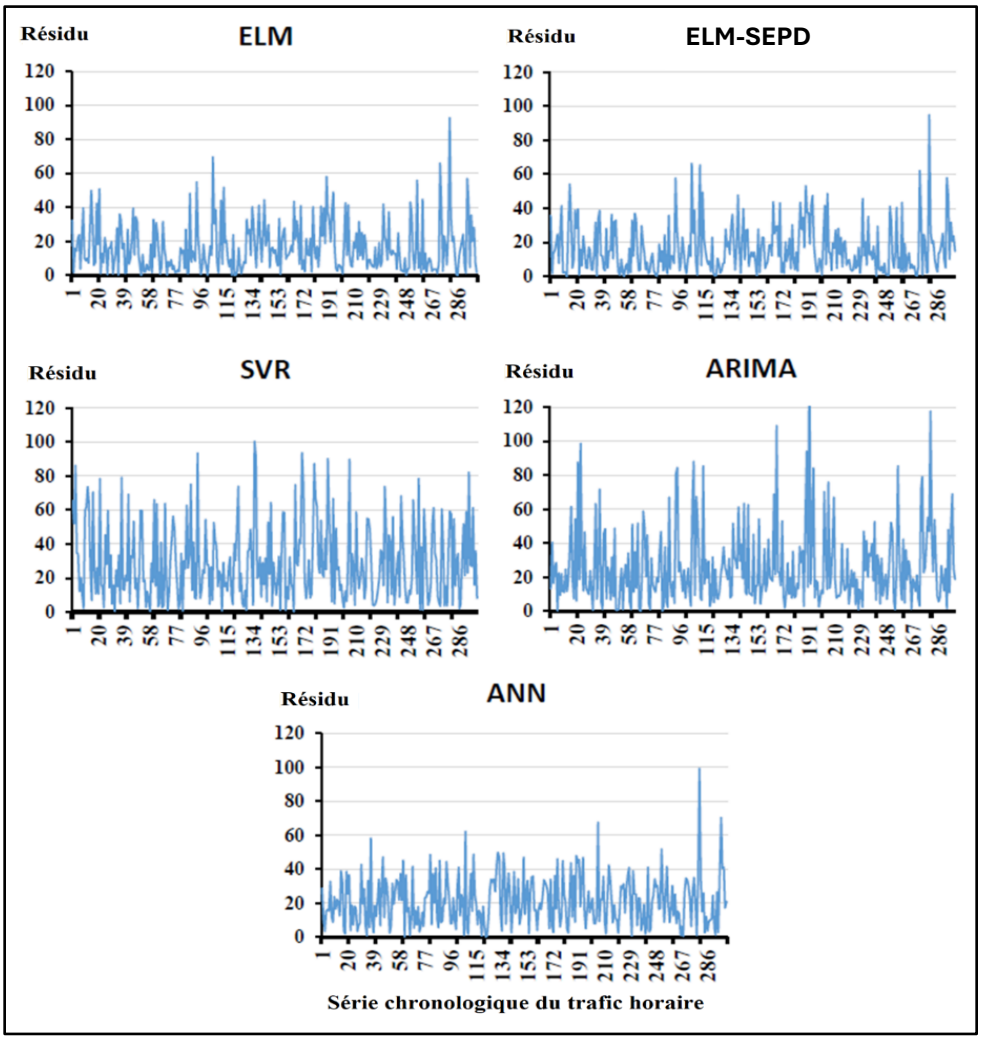


Figure 21. Erreur Résiduelle Absolue de la phase de test

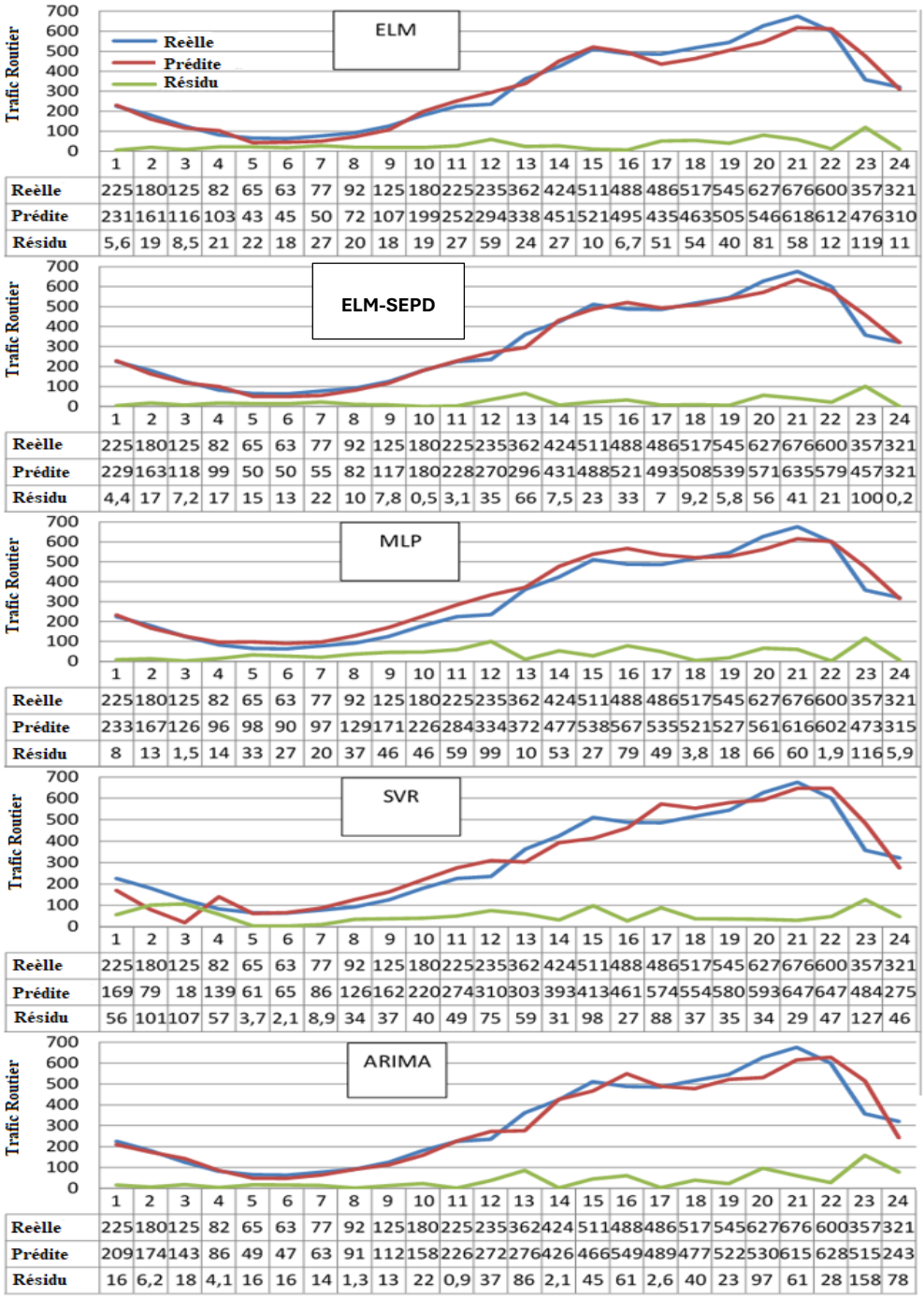
les cinq modèles, suggérant ainsi une capacité d'apprentissage satisfaisante. Les configurations des courbes mettent en évidence une nette amélioration apportée par le modèle ELM-SEPD hybride par rapport au modèle ELM simple. De plus, les résidus de ce dernier s'alignent étroitement autour de la ligne d'équation 0. Cette proximité indique une précision élevée du modèle. Ce résultat découle probablement d'une sélection judicieuse des paramètres pertinents, tels que les caractéristiques, la qualité élevée des données, et le choix du nombre de neurones de la couche cachée.

L'analyse de la figure 22 met en évidence que notre modèle, exploitant le système SEPD, démontre une capacité accrue à produire des prédictions plus précises par rapport au modèle ELM individuel. Sur l'axe des ordonnées, sont représentées à la fois les valeurs réelles et prédites, ainsi que les résidus résultant de la comparaison entre ces deux ensembles de données. Une observation significative est la capacité du SEPD à atténuer les fluctuations engendrées par la généralisation aléatoire des paramètres d'entrée. En effet, cette approche réduit efficacement l'erreur de prédiction associée à l'ELM, tout en maintenant un niveau de performance temporellement acceptable. Cette amélioration notable découle de l'aptitude inhérente du SEPD à ajuster ses prédictions pour tenir compte des légères fluctuations induites par l'ELM à chaque étape de prédiction.

Le tableau 8 montre le temps d'exécution, la complexité et la précision des différents modèles de prédiction. Selon les valeurs de l'indice AIC, qui tient compte de la complexité et de la performance du modèle, on observe que le modèle ELM-SEPD présente le plus faible AIC, suivi de près par le modèle ELM. Cela suggère que ces deux modèles offrent une meilleure balance entre complexité et performance par rapport aux modèles ARIMA, ANN et SVR. Le modèle ELM-SEPD est celui qui affiche la RMSE la plus basse et les erreurs MAE et MAPE les plus faibles, ce qui indique une meilleure précision de prédiction par rapport aux autres modèles.

Tableau 8. Comparaison des performances de prédiction par l'AIC, RMSE, MAE et MAPE

Modèle	Temps d'Exécution	AIC	RMSE	MAE	MAPE
ELM	**0 mn 1s. 82**	−171 407,88	0,028 8	0,019 7	0,143 8
ELM-SEPD	1 mn 27s. 90	**−172 066,57**	**0,027 5**	**0,018 9**	**0,137 3**
ARIMA	1mn 00s. 98	−151 855,69	0,034 0	0,024 9	0,205 2
SVR	0 mn 2s. 15	−154 447,35	0,189 2	0,152 8	1,886 4
ANN	0 mn 6s. 10	−163 396,78	0,184 9	0,149 1	1,842 0



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Reèlle	225	180	125	82	65	63	77	92	125	180	225	235	362	424	511	488	486	517	545	627	676	600	357	321
Prédite	231	161	116	103	43	45	50	72	107	199	252	294	338	451	521	495	435	463	505	546	618	612	476	310
Résidu	5,6	19	8,5	21	22	18	27	20	18	19	27	59	24	27	10	6,7	51	54	40	81	58	12	119	11

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Reèlle	225	180	125	82	65	63	77	92	125	180	225	235	362	424	511	488	486	517	545	627	676	600	357	321
Prédite	229	163	118	99	50	50	55	82	117	180	228	270	296	431	488	521	493	508	539	571	635	579	457	321
Résidu	4,4	17	7,2	17	15	13	22	10	7,8	0,5	3,1	35	66	7,5	23	33	7	9,2	5,8	56	41	21	100	0,2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Reèlle	225	180	125	82	65	63	77	92	125	180	225	235	362	424	511	488	486	517	545	627	676	600	357	321
Prédite	233	167	126	96	98	90	97	129	171	226	284	334	372	477	538	567	535	521	527	561	616	602	473	315
Résidu	8	13	1,5	14	33	27	20	37	46	46	59	99	10	53	27	79	49	3,8	18	66	60	1,9	116	5,9

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Reèlle	225	180	125	82	65	63	77	92	125	180	225	235	362	424	511	488	486	517	545	627	676	600	357	321
Prédite	169	79	18	139	61	65	86	126	162	220	274	310	303	393	413	461	574	554	580	593	647	647	484	275
Résidu	56	101	107	57	3,7	2,1	8,9	34	37	40	49	75	59	31	98	27	88	37	35	34	29	47	127	46

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Reèlle	225	180	125	82	65	63	77	92	125	180	225	235	362	424	511	488	486	517	545	627	676	600	357	321
Prédite	209	174	143	86	49	47	63	91	112	158	226	272	276	426	466	549	489	477	522	530	615	628	515	243
Résidu	16	6,2	18	4,1	16	16	14	1,3	13	22	0,9	37	86	2,1	45	61	2,6	40	23	97	61	28	158	78

Figure 22. Résultats des prévisions de trafic des différents modèles

7. Analyse des résultats et discussion

Le modèle ELM démontre une précision notable, caractérisée par des valeurs de RMSE, MAE et MAPE faibles, indiquant une proximité satisfaisante entre ses prédictions et les valeurs réelles, et un

temps d'exécution rapide. Cependant, son indice AIC est relativement élevé par rapport au modèle ELM-SEPD, suggérant une adaptation moins optimale aux données. En comparaison, le modèle ELM-SEPD, en intégrant une technique d'ensemble, améliore significativement la précision des prédictions, comme en témoignent des valeurs réduites de RMSE, MAE et MAPE. Néanmoins, son temps d'exécution plus long peut représenter un inconvénient pour les applications en temps réel ou les grands ensembles de données.

Concernant le modèle ARIMA, ce dernier affiche une précision adéquate avec des valeurs appropriées de RMSE, MAE et MAPE, bien que son temps d'exécution soit relativement long, comparable à celui du modèle ELM-SEPD. En termes de précision, le modèle ARIMA est performant, mais ses résultats sont légèrement inférieurs à ceux du modèle ELM-SEPD. En revanche, le modèle SVR présente des performances en termes de précision inférieures, malgré un temps d'exécution réduit. Quant au modèle ANN, ses performances en termes de précision sont similaires à celles du modèle SVR, mais son temps d'exécution est le plus long parmi les modèles testés. Toutefois, le modèle ELM-SEPD surpasse le modèle ELM simple en termes de précision, ce qui indique que la technique d'ensemble est avantageuse dans cette situation. Comparés aux autres modèles, les modèles SVR et ANN ont une précision relativement faible, ce qui peut les rendre moins intéressants pour des applications de prédiction spécifiques. Ainsi, le modèle ELM-SEPD, bien qu'il présente un temps d'exécution plus long, offre une augmentation de la précision qui peut justifier l'investissement supplémentaire en temps dans certains scénarios.

En guise de synthèse, le modèle ELM-SEPD semble être le choix optimal en termes de précision parmi les modèles examinés. Cependant, d'autres facteurs tels que les contraintes de temps d'exécution, la complexité du modèle, la taille des données et les objectifs spécifiques du projet auront une incidence sur la décision finale concernant le modèle à adopter. Des essais supplémentaires avec divers ensembles de données pourraient également être bénéfiques pour évaluer la généralisation des modèles dans différents contextes.

8. Conclusion

Dans ce chapitre, nous avons utilisé l'algorithme d'Extreme Learning Machine pour concevoir un modèle précis permettant de prédire les données horaires du trafic routier. Dans cette optique, l'étude utilise les données historiques des 5 dernières années sur une section de route à Tanger, au Maroc. La principale contribution de notre travail est le développement d'un modèle hybride appelé ELM-SEPD

qui montre une meilleure performance que l'ELM simple sans affecter la vitesse du processus d'apprentissage. Le modèle proposé a été évalué par comparaison avec les algorithmes bien connus comme MLP, SVR et ARIMA. Les résultats ont montré qu'il offrait de meilleures performances en termes de précision et de stabilité. Le modèle de prévision du trafic proposé représente un outil fiable pour les ingénieurs et les gestionnaires pour planifier les routes et réguler le trafic. Il permettra également aux usagers de la route de disposer d'informations préalables sur le trafic afin de mieux planifier leurs déplacements et de gérer leur temps efficacement en évitant les heures de forte congestion.

Des améliorations futures pourraient être envisagées dans le cadre de notre étude. Tout d'abord, il serait pertinent d'enrichir l'ensemble de données en y incluant d'autres variables pertinentes telles que les conditions météorologiques, afin d'améliorer la précision des prédictions. Ensuite, explorer d'autres techniques d'apprentissage automatique ou de modélisation statistique pourrait permettre de comparer leurs performances avec le modèle proposé et éventuellement découvrir des approches plus efficaces pour la prédiction du trafic routier. Pareillement, étendre l'application du modèle à d'autres zones géographiques permettrait également de vérifier sa généralisabilité et d'évaluer sa robustesse dans des contextes différents. En outre, une recherche plus approfondie des meilleurs hyperparamètres pour le modèle, en utilisant des techniques telles que l'optimisation bayésienne, serait de nature à améliorer davantage sa performance. Intégrer le modèle dans des systèmes de gestion de trafic en temps réel pour une évaluation pratique dans la planification et la régulation du trafic routier constituerait également une avancée significative.

Chapitre 4. Analyse statistique sur l'impact de l'infrastructure routière sur la mortalité routière au Maroc

1. Introduction

Les routes servent de voies de déplacement pour conduire, marcher ou voyager en tant que passager, mais également de moyens d'échange de biens et de services. Cependant, les accidents de la route demeurent un problème majeur, selon les statistiques de l'Organisation Mondiale de la Santé (OMS), constituant l'une des principales causes de décès dans le monde et causant des dommages humains et économiques considérables. Malgré la mise en place de stratégies de sécurité routière pratiques et globales dans de nombreux pays comme la « Vision Zéro », visant à éliminer les décès et les blessures graves liés aux accidents de la route à long terme [63], le nombre d'accidents reste élevé et en augmentation.

En 2017, le Maroc a enregistré 89 375 accidents, dont 3 726 mortels et 130 011 corporels [64]. La figure 23 [65] compare les taux de mortalité dus aux accidents de la route entre le Maroc et d'autres pays motorisés sur une période de 1990 à 2019. Bien que le Maroc ait connu une amélioration de sa sécurité routière au fil des ans, avec un taux de mortalité en baisse, restant à 26,41 en 2019, la situation demeure préoccupante. D'évidence, cette comparaison avec d'autres pays permet orienter les mesures et initiatives visant à réduire le nombre de décès sur les routes. Un constat : les accidents de la circulation résultent souvent de facteurs humains, mais l'infrastructure et l'état des véhicules jouent également un rôle. Dans cette étude, nous avons examiné les données collectées tant dans les zones urbaines que périphériques, fournies par la Direction des Routes du Ministère de l'Équipement et de l'Eau, afin d'identifier les principales causes des accidents de la route.

De nombreuses études ont abordé les divers facteurs influençant la gravité des accidents de la route, incluant notamment le climat, le genre du conducteur et le niveau d'alcoolémie. Cependant, peu d'attention a été accordée à l'impact de l'état physique de la route sur la sévérité des accidents. Ce qui constitue un sujet sous-représenté dans la littérature. C'est dans ce contexte que notre recherche étude trouve sa principale source de motivation : examiner la relation entre l'état de la route et le taux de létalité des accidents, un aspect qui n'a pas été pleinement exploré. Les usagers de la route accordent généralement une grande importance à cet indicateur lors du choix de leur itinéraire et de la planification

de leurs déplacements. Précisons que notre approche méthodologique repose sur l'utilisation de l'analyse statistique, une technique bien établie permettant de donner un sens aux données numériques qui est largement utilisée dans ce domaine de recherche depuis des décennies [66]. Pour ce faire, nous utiliserons l'analyse de la variance (ANOVA) ; cette méthode permettant de décomposer la variabilité observée dans un ensemble de données en facteurs systématiques, exerçant une influence sur les résultats, et en facteurs aléatoires qui ne présentent pas de relation explicite avec les facteurs examinés. Elle permet ainsi de distinguer et d'évaluer la contribution respective des différents facteurs à la variation des données, offrant ainsi une compréhension plus approfondie des relations entre les variables sous investigation.

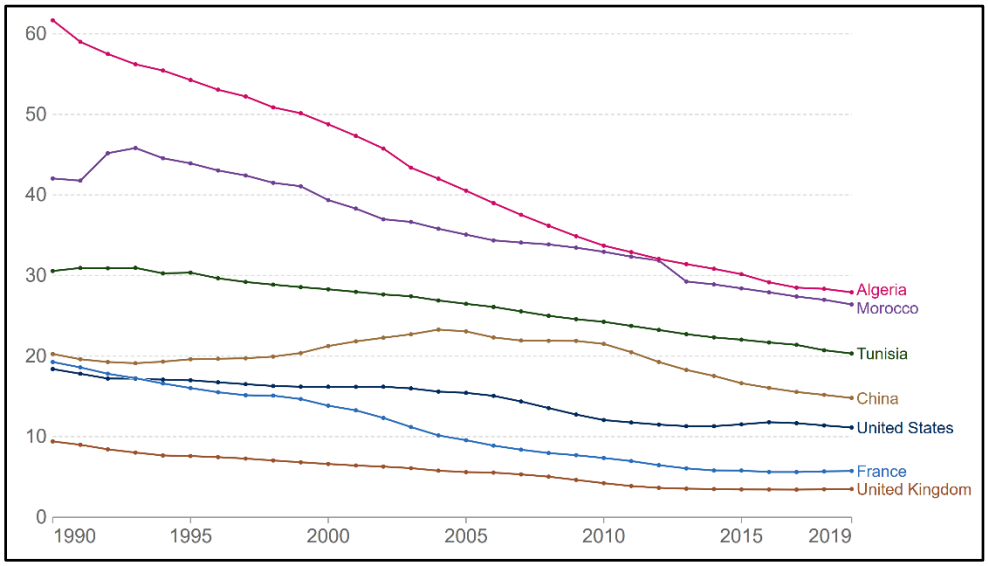


Figure 23. Taux de mortalité dû aux accidents de la route pour 100 000 personnes

Le reste de ce chapitre sera structuré de la manière suivante : une revue de la littérature sera présentée dans la deuxième partie, regroupant les études antérieures portant sur l'analyse des facteurs de risque influant sur la mortalité routière. La troisième section sera consacrée à la préparation des données relatives aux accidents routiers en vue d'une analyse approfondie. La quatrième partie détaillera la méthodologie adoptée, décrivant séquentiellement les étapes entreprises pour analyser les facteurs de risque, ainsi que les méthodes utilisées. L'analyse des facteurs d'infrastructure impactant la mortalité des accidents sera exposée dans la cinquième section, incluant la répartition des accidents selon les conditions routières, la distribution des taux de létalité, les résultats des tests statistiques ainsi qu'une évaluation de l'impact de chaque facteur sur la mortalité routière. Ensuite, une discussion des résultats obtenus sera entreprise, accompagnée de recommandations visant à améliorer la sécurité routière. Enfin, une conclusion synthétisera les principaux points abordés dans le chapitre, clôturant ainsi son développement.

2. Contexte théorique et historique

La présente section offre une synthèse des études antérieures se penchant sur les facteurs de risque liés à la mortalité routière. Cette partie vise à éclairer les recherches actuelles en identifiant les principaux éléments qui ont été étudiés dans ce domaine crucial de la sécurité routière. En examinant les travaux précédents, nous cherchons à situer notre propre analyse dans le contexte plus large des connaissances existantes, tout en mettant en évidence les lacunes et les perspectives émergentes dans ce domaine de recherche.

Un corpus substantiel d'études s'est intéressé au comportement global des conducteurs sur les voies de circulation, en mettant particulièrement en lumière la somnolence, identifiée comme l'un des principaux facteurs de risque d'accident [67]. Certains chercheurs ont également pris en considération les paramètres météorologiques et les conditions climatiques pour établir des analyses relatives à la sécurité routière [68]. Dans cette perspective, l'application de méthodes d'exploration de données s'est avérée pertinente pour anticiper les accidents à venir. Ce qui pourrait contribuer à l'élaboration de stratégies de prévention adaptées visant à renforcer la sécurité sur les routes [69]. D'autres investigations ont été axées sur l'optimisation et l'innovation des systèmes d'information existants, proposant ainsi de nouvelles approches pour la collecte, le traitement et l'analyse des données en lien avec la sécurité routière [70]. Malgré la richesse des études antérieures dans ce domaine, il convient de noter que l'incidence de l'état physique des voies de circulation sur la fatalité des accidents demeure un aspect sous-exploré dans la littérature académique.

Une étude récente réalisée par Gitelman et Doveh [71], avait pour objectif de comparer les performances en termes de sécurité des voies urbaines selon trois configurations d'itinéraires prioritaires pour les autobus (IPB) : les voies centrales séparées par des barrières, les voies centrales ouvertes et les itinéraires d'autobus en bordure de trottoir. Pour évaluer les fréquences d'accidents et examiner les différences de performance en matière de sécurité des voies avec différentes configurations IPB, des modèles de régression multivariés ont été ajustés, tenant compte d'autres caractéristiques de la route et du trafic. Les résultats des modèles ont révélé de manière consistante que des fréquences d'accidents plus élevées étaient prévues sur les tronçons routiers avec une configuration d'IPB en bordure de route par rapport aux IPB à voie centrale et à barrière séparée. En revanche, les différences de moyennes d'accidents entre les configurations à voie centrale ouverte et à barrière séparée n'étaient pas significatives. La configuration à voie centrale s'est avérée bien plus sûre sur les tronçons routiers que

les IPB en bordure de route ; la différence étant statistiquement significative. Toutefois, aux carrefours, ces deux configurations ont présenté les fréquences d'accidents les plus faibles pour certains types d'accidents, bien que les différences en termes de nombre d'accidents n'étaient pas statistiquement significatives. Par ailleurs, une étude menée par Salehian *et al.* [72] avait pour objectif d'analyser les facteurs influençant les blessures des piétons dans des scénarios d'intersection et hors intersection sur les routes de rase campagne, en utilisant un modèle Probit Ordonné. L'étude a distingué les accidents en intersection et hors intersection afin de mieux comprendre les schémas d'accidents impliquant les piétons. Il a été observé que l'impact de certaines variables pouvait varier entre l'analyse de l'ensemble des données et l'analyse en clusters. Parmi les facteurs contribuant à l'augmentation de la gravité des accidents, on retrouve des limitations de vitesse supérieures à 96 km/h, la présence de routes principales et de routes à chaussée unique. En outre, des variables telles que l'emplacement des piétons sur les voies piétonnes et les routes à sens unique se sont avérées statistiquement significatives uniquement dans la section des intersections.

Nous avons aussi relevé une étude ultérieure, réalisée par Ospina-Mateus *et al.* [73], qui avait pour objectif d'identifier les facteurs liés à la sécurité des motocyclistes aux intersections à trois et quatre voies non signalisées, en examinant le comportement des motocyclistes et en analysant les conflits de circulation potentiels. Les résultats ont révélé que les motocycles sont plus susceptibles de rencontrer des conflits de circulation graves aux intersections que les véhicules motorisés, et que l'état de l'infrastructure routière a une incidence significative sur le risque de conflits aux intersections. En outre, des comportements individuels tels que le non-respect des arrêts aux intersections et une conduite imprudente ont été associés à une augmentation du risque de conflits de circulation. Pour améliorer la sécurité des motocyclistes, cette recherche recommande des mesures visant à améliorer l'infrastructure, notamment la visibilité et la signalisation, ainsi que la mise en place de séparateurs pour réduire la vitesse et les conflits de circulation aux intersections. L'étude menée par Wan *et al.* [74] présente une approche combinée de la méthode de Grey Verhuls et de la méthode empirique de Bayes pour analyser les points noirs des accidents de la route en milieu urbain. Cette approche utilise le modèle de discrimination des points noirs pour prédire le nombre d'accidents dans les zones à risque, et pour dériver deux discriminateurs de points noirs, à savoir un espace d'amélioration de la sécurité et un indice de sécurité (IS), permettant de classer les points noirs en plusieurs niveaux en fonction de l'IS. Précisons que l'identification des points noirs repose principalement sur le nombre d'accidents et sur des facteurs tels que les caractéristiques géométriques des zones d'accident. Les résultats démontrent que cette approche

combinée permet d'identifier rapidement et précisément les points noirs en tenant compte des données d'accidents provenant de situations similaires.

Comme nous l'avons déjà évoqué, au cours de la dernière décennie, le Maroc a enregistré des avancées significatives dans la construction d'infrastructures routières, souvent guidées par des études intégrant divers paramètres tels que le volume de trafic et les répercussions socio-économiques régionales. Parallèlement, le pays s'est engagé dans un processus continu de réformes et d'analyses visant à renforcer son système de sécurité routière. Dans une étude récente réalisée par Ennajih *et al.* [75], les auteurs ont examiné l'impact des différentes variables de sécurité routière sur la fréquence des accidents de la route, en examinant notamment les possibilités d'amélioration de l'inspection des véhicules en s'inspirant du modèle allemand. Malgré ces initiatives, les statistiques relatives aux accidents routiers témoignent d'un maintien à un niveau élevé du taux de mortalité, signalant ainsi une persistance des défis à surmonter dans ce domaine.

3. Préparation des données

Les données relatives aux accidents de la circulation routière pour l'année 2017, fournies par la Direction des Routes (DR), se fondent sur les rapports statistiques établis par la sûreté nationale et la gendarmerie royale. Ce rapport englobe l'ensemble des catégories de routes, qu'elles soient situées en milieu urbain ou rural. Pour évaluer la sévérité des accidents, un ensemble de variables décrivant l'état physique et l'infrastructure des routes a été pris en considération. Parmi ces variables, certaines revêtent un caractère qualitatif comme le type de revêtement de la chaussée, l'état général de celle-ci, et le profil en long, tandis que d'autres sont de nature quantitative, comme la largeur de la chaussée et des accotements. Au total, treize variables détaillant l'état de la route et son infrastructure ont été recueillies, comme illustrées dans le tableau 9, récapitulant les variables ainsi que leurs modalités respectives.

Les données sont initialement filtrées pour exclure les enregistrements signalés comme inexacts ou non conservés dans l'ensemble de données initial. Pour mesurer la gravité de chaque accident en termes de pertes humaines, il s'avèrerait opportun d'utiliser une mesure spécifique. À cet effet, nous avons intégré l'indicateur de Taux de Létalité (TL) en tant que variable dépendante à modéliser. Ce paramètre est calculé par accident selon l'équation suivante :

$$TL = \frac{D}{D+B} \times 100 \tag{32}$$

Où D indique le nombre de décès dans un accident, et B le nombre de blessés. En épidémiologie, le taux de létalité est une mesure définie comme la proportion de décès au sein d'une population touchée par une maladie mortelle [76]. Dans notre contexte, ces caractéristiques se rapportent aux accidents entraînant des décès et/ou des blessures survenant pendant, après ou dans les 30 jours suivants l'événement. Cette mesure revêt une importance cruciale dans l'évaluation de la gravité des accidents de la route, permettant ainsi une analyse approfondie de l'impact de ces incidents sur la santé publique et la sécurité routière.

De nombreux indicateurs ont été utilisés dans la littérature pour mesurer la gravité des accidents. Ainsi, dans l'étude menée par Wang *et al.* [77], l'accent est mis sur les dommages humains moyens par accident, tandis qu'une autre étude de Hayakawa *et al.* [78], se penche sur le taux de mortalité ajusté par population, calculé pour un million d'habitants. De plus, dans l'étude de Oña *et al.* [79], deux indicateurs distincts sont utilisés pour chaque catégorie d'accident : les blessés légers et les tués ou blessés graves sont analysés en termes de fréquence pour évaluer la sévérité des incidents. Cette diversité d'approches reflète la complexité inhérente à la mesure de la gravité des accidents, soulignant ainsi l'importance d'une approche multidimensionnelle dans l'évaluation des risques et dans la formulation de stratégies de prévention efficaces.

4. Méthodologie de recherche

L'analyse statistique représente un outil essentiel pour interpréter les données numériques dans des contextes sans signification naturelle [66], et son utilisation est répandue depuis de nombreuses décennies dans des domaines de recherche divers. Ainsi, notre recherche adopte la technique ANOVA ; laquelle se distingue par sa capacité à partitionner la variabilité observée au sein d'un ensemble de données en deux composantes distinctes : les facteurs systématiques qui exercent une influence et les facteurs aléatoires qui n'en ont pas. L'ANOVA constitue la première étape de l'analyse des facteurs impactant un jeu de données spécifique, suivie de tests complémentaires portant sur les facteurs méthodiques contribuant de manière significative à l'incohérence des données.

L'ANOVA à un facteur, également connue sous le nom de *« one-way analysis of variance »,* se positionne comme l'une des méthodes les plus répandues en statistique. Cette approche vise à déterminer si des différences statistiquement significatives existent entre les moyennes de trois groupes indépendants ou plus. Dans le cadre de notre recherche, nous opterons pour le test de Kruskal-Wallis par rangs ; une alternative non paramétrique à l'ANOVA, spécialement adaptée à notre contexte de recherche. Ce choix

se justifie par le caractère ordinal de la variable indépendante et la non-normalité de la distribution de la variable dépendante dans nos données. L'objectif principal de ce test réside dans la comparaison des rangs des groupes d'observations indépendantes ; ce qui s'avère pertinent lorsque les hypothèses paramétriques de l'ANOVA ne sont pas satisfaites. Trois hypothèses majeures sous-tendent cette approche : la normalité de la variable dépendante au sein de chaque groupe de la variable indépendante, l'égalité de variance entre les groupes, ainsi que l'indépendance des observations.

L'évaluation de la distribution gaussienne a été réalisée au moyen du test de normalité de Shapiro-Wilk, l'un des principaux tests conçus pour identifier toute déviation par rapport à la normalité. Sous l'hypothèse nulle du test, les données sont supposées être distribuées selon une loi normale. Le rejet de l'hypothèse nulle intervient lorsque la valeur p est inférieure ou égale à 0,05. Un échec du test de normalité, avec un niveau de confiance de 95 %, indique que les données ne suivent pas une distribution gaussienne.

Dans de nombreuses analyses statistiques, il est supposé que la variance entre plusieurs groupes soit homogène. Une méthode appropriée pour évaluer cette hypothèse consiste à recourir au test de Levene, qui examine si la variance entre deux groupes ou plus est égale. Ce test générant une statistique de test ainsi qu'une valeur p associée. Si la valeur p est inférieure au niveau de signification prédéterminé. Cela permet de rejeter l'hypothèse nulle, conduisant ainsi à la conclusion qu'il existe des éléments insuffisants pour soutenir l'homogénéité de la variance.

Dans le cadre de notre recherche, l'analyse de la variance non paramétrique de Kruskal-Wallis est employée pour évaluer les données de chaque groupe de variables indépendantes en vue de déterminer leur impact sur la variable dépendante (le taux de létalité dans notre cas). Il convient de souligner que les facteurs sont examinés individuellement, un à un. En cas de résultat significatif du test de Kruskal-Wallis pour un facteur donné, cela indique que les groupes associés à ce facteur présentent des différences substantielles entre eux. Dans ce contexte, ledit facteur est considéré comme exerçant une influence mesurable sur la variable dépendante. Étant donné que notre modèle d'étude comporte un nombre considérable de groupes, il est primordial d'identifier ceux qui se distinguent les uns des autres. Ainsi, une analyse post-hoc est introduite dans le but de spécifier quelles paires de groupes démontrent des divergences significatives. Dans cette optique, le test de Dunn [80] est reconnu comme le choix optimal pour comparer les groupes au sein de chaque facteur. Ce test est une procédure post-hoc non paramétrique adaptée pour conduire des comparaisons multiples par paire lorsque l'hypothèse nulle du test de Kruskal-

Wallis est rejetée [81]. Il permet ainsi de réaliser k (k-1)/2 comparaisons multiples par paires pour chaque facteur, où k désigne le nombre de groupes considérés.

Pour déterminer la taille d'effet, les statisticiens ont introduit des formules pour calculer la taille de l'effet et d'interpréter les résultats [82]. L'Êta carré η_H^2 indique le pourcentage de variance de la variable dépendante expliquée par la variable indépendante [83]. Bien qu'il soit couramment utilisé dans l'ANOVA à un facteur, il est plus sujet à biais que l'epsilon-carré [84]. L'epsilon-carré, quant à lui, offre une estimation de l'ampleur de l'effet, calculée comme suit :

$$E_R^2 = \frac{H}{(n^2 - 1)/(n + 1)} \tag{33}$$

Où H est la valeur obtenue dans le test de Kruskal-Wallis, qui mesure la force de l'association entre les groupes dans un ensemble de données, et n est le nombre total d'observations dans l'ensemble de données.

En outre, les auteurs [85] ont proposé le Common Language Effect Size (CLES), une mesure utilisée pour évaluer la taille de l'effet entre deux groupes. Le CLES représente la probabilité qu'un individu sélectionné au hasard dans un groupe présente une valeur plus élevée qu'un individu sélectionné au hasard dans l'autre groupe :

$$P\left(Z < \frac{|M_2 - M_1|}{\sqrt{{S_1}^2 + {S_2}^2}}\right) \tag{34}$$

Où M1, M2 sont les moyennes des groupes et S1, S2 les écarts-types.

5. Facteurs de risque des accidents

Dans le cadre de notre recherche, le niveau de signification retenu est de 0,05, représentant la probabilité de rejeter l'hypothèse nulle lorsqu'elle est véridique. L'exécution de cette étude a été effectuée à l'aide de Python 3.6 sur une machine équipée d'un processeur Intel Core i5-7200U cadencé à 5 GHz. Par souci de clarté, les résultats seront présentés en utilisant le codage des groupes comme illustré dans le tableau 9.

5.1. Répartition des accidents sur les différentes conditions routières

Le tableau 9 montre les différents groupes de chaque facteur ainsi que le nombre d'accidents associé à chaque groupe. L'analyse de ces données révèle les conclusions suivantes :

- L'état de surface : la plupart des accidents (88 053) se produisent sur des routes sèches, suivies de près par les routes mouillées (976). Les accidents sur des routes enneigées ou verglacées et humides avec sel résiduel sont beaucoup moins fréquents (190 chacun). Cela souligne l'importance de maintenir des routes en bon état, surtout pendant les conditions météorologiques difficiles. En effet, l'état de surface de la route est un facteur clé dans la sécurité routière, car il peut influencer la distance de freinage et la tenue de route des véhicules. Une chaussée en mauvais état peut présenter des obstacles, des fissures ou des nids-de-poule qui peuvent causer des accidents. Par conséquent, il est essentiel de veiller à ce que les routes soient régulièrement entretenues et réparées pour garantir la sécurité des usagers de la route.
- L'état de chaussée : selon les données, la grande majorité des accidents surviennent sur des routes en bon état (86 835) en raison de leur plus grande fréquentation et vitesse. Cependant, les accidents sur des routes en mauvais ou très mauvais état, bien que moins fréquents (1 689 et 301 respectivement), peuvent entraîner des conséquences plus graves en raison des obstacles imprévus sur la route ou de la perte de contrôle du véhicule. Par conséquent, il est important de souligner l'importance de la maintenance et de la réparation des routes pour améliorer non seulement la sécurité routière, mais aussi le confort et la commodité des conducteurs et des passagers. En outre, une route en bon état peut également contribuer à réduire les émissions de gaz d'échappement en permettant une circulation plus fluide et une consommation de carburant plus efficace. Par conséquent, l'investissement dans l'entretien des routes peut avoir des avantages économiques, environnementaux et sociaux à long terme.
- Le type de chaussée : selon les données, il y a eu un nombre beaucoup plus élevé d'accidents sur les routes à chaussée unique (20 893) que sur celles qui sont séparées par un terre-plein (2 052). Cette tendance est préoccupante et indique clairement un besoin d'amélioration de la sécurité sur les routes à chaussée unique. Il pourrait être judicieux d'investir dans de nouvelles technologies de sécurité routière pour réduire le nombre d'accidents. Des mesures pourraient également être mises en place pour améliorer la signalisation, la visibilité et la qualité de la chaussée sur ces routes. En outre, il est important que les conducteurs soient informés de la

nécessité de conduire prudemment sur les routes à chaussée unique et soient régulièrement rappelés des règles de sécurité routière.

- La position de l'accotement : les accidents sont répartis de manière relativement égale entre les accotements au niveau de la chaussée et les accotements dénivelés par rapport à la chaussée, tant du côté droit que du côté gauche. Cela montre que des précautions doivent être prises pour la sécurité des conducteurs, quel que soit le type d'accotement. Ainsi, il est conseillé aux conducteurs de rester vigilants et de prendre des précautions supplémentaires lorsqu'ils conduisent sur des accotements dénivelés.

- La localisation : il est important de noter que la localisation des accidents de la route peut révéler des tendances importantes en matière de sécurité routière. En effet, selon les statistiques, les accidents en dehors des intersections sont plus fréquents (44 401) que ceux en intersection (19 985) ou en passage à niveau (26). Toutefois, cela ne signifie pas nécessairement que les intersections sont des lieux sûrs. En réalité, les intersections demeurent des zones critiques pour la sécurité routière, et il est important de prendre des mesures pour les rendre plus sûres. Par exemple, la mise en place de feux de signalisation ou de panneaux de signalisation clairs et facilement compréhensibles peut aider à réduire le nombre d'accidents en intersection. De plus, il est important de sensibiliser les conducteurs à l'importance de respecter les règles de la route et de conduire prudemment dans toutes les situations de conduite, y compris aux intersections.

- Le type de carrefour : il existe deux types de carrefours, les simples (CSSFS et CSAFS) et les multiples (CMSFS et CMAFS). Les carrefours simples ont tendance à avoir plus d'accidents (11 174 et 3 454 respectivement) que les carrefours multiples. Cela est souvent dû à des facteurs tels que le manque de signalisation ou une mauvaise gestion du trafic. Les carrefours multiples nécessitent également une attention particulière, car bien qu'ils aient moins d'accidents, ils peuvent être plus complexes et nécessitent une signalisation et une gestion du trafic plus pointues. La signalisation et la gestion du trafic sont des éléments clés dans la prévention des accidents aux carrefours, qu'ils soient simples ou multiples.

- La topographie : selon les statistiques, la plupart des accidents se produisent sur des routes plates (47 650 cas), suivies de celles en pente (1 674 cas) et des sommets de côte (86 cas). Cette information est cruciale pour améliorer la sécurité routière. En effet, il serait judicieux de porter une attention particulière aux routes en montagne ou en pente, où les risques d'accident sont plus élevés en raison des conditions de conduite plus difficiles. Par exemple, les routes en montagne

peuvent être plus étroites, plus sinueuses et plus escarpées, tandis que les routes en pente peuvent être glissantes et causer des dérapages. Il est donc important de prendre en compte la topographie des routes dans la planification de la sécurité routière pour réduire le nombre d'accidents et de victimes sur les routes.

- Le tracé du plan : les accidents sur des tronçons rectilignes sont les plus fréquents (40 776), suivis des courbes à gauche (2 654) et des courbes à droite (770). Cela suggère que les conducteurs peuvent être plus enclins à se relâcher sur les routes droites et que les courbes à gauche peuvent être plus difficiles à négocier que les courbes à droite. Il est donc important de prendre des mesures pour améliorer la sécurité sur les routes comportant des courbes, telles que la signalisation claire et la limitation de la vitesse.

- La majorité des accidents ont lieu sur des sections de route où la largeur d'accotement droit de 1 à 2 mètres (15 348 accidents). Cependant, il est également courant de trouver des routes avec une largeur d'accotement droit de moins de 1 mètre, qui ont causé 1 021 accidents, et des routes avec une largeur d'accotement droit de 2 à 3 mètres, qui ont été à l'origine de 4 130 accidents. Les routes avec une largeur d'accotement droit de 3 à 4 mètres ont causé 523 accidents, tandis que celles de 4 à 5 mètres et de 5 à 10 mètres, moins courantes, ont provoqué respectivement 120 et 143 accidents. Il est important de noter que les routes avec une largeur d'accotement droit plus large sont plus sûres pour les conducteurs et les piétons, car elles offrent plus d'espace pour manœuvrer en cas d'urgence.

- D'après les statistiques, la plupart des accidents ont lieu sur des sections de route où la largeur d'accotement gauche est comprise entre 1 et 2 mètres, avec un total de 15 113 cas recensés. Les routes avec des accotements de moins de 1 mètre (793 accidents) ou de 2 à 3 mètres (3 326 accidents) sont également souvent impliquées dans des accidents. En revanche, les routes avec une largeur d'accotement gauche de 3 à 4 mètres ne comptent que 434 accidents, tandis que celles avec une largeur d'accotement de 4 à 5 mètres ont causé 120 accidents et de 5 à 10 mètres ont été l'origine de 123 accidents. Cela souligne l'importance d'élargir les accotements de la route, en particulier dans les zones où les accidents sont plus fréquents, dans le but de réduire le nombre de victimes et de rendre les routes plus sûres pour tous les usagers.

- Les statistiques révèlent que la majorité des accidents se produisent sur des sections de route dont la largeur de chaussée est inférieure à 6 mètres, soit un total de 6 232 accidents. Cependant, il est également important de noter que les routes avec une largeur de chaussée de 6 à

7 mètres (7 486 accidents) et de 7 à 10,5 mètres (8 004 accidents) sont également des zones à risque élevé. Les routes avec une largeur de chaussée de 10,5 à 14 mètres (742 accidents) et de plus de 14 mètres (898 accidents) sont moins propices aux accidents, mais il est important de rester vigilant sur ces routes. Il est également important de noter que les routes de campagne peuvent avoir des largeurs de chaussée différentes, ce qui peut également affecter les risques d'accident.

Tableau 9. Répartition des accidents sur les conditions routières

Facteurs	Groupes	Description	Total
État de la surface	1	Sec	88 053
	2	Mouillé	976
	3	Enneigé ou verglacé	190
	4	Humide, sel résiduel	190
État de la chaussée	1	Revêtue bonne	86 835
	2	Revêtue mauvaise	1 689
	3	Revêtue très mauvaise	301
Type de chaussée	1	Route à chaussée unique	20 893
	2	Chaussée séparée par terre-plein	2 052
Position de l'accotement droit	1	LR : Au niveau de la chaussée	20 807
	2	NLR : Dénivelé par rapport à la chaussée	594
Position de l'accotement gauche	1	LR : Au niveau de la chaussée	19 975
	2	NLR : Dénivelé par rapport à la chaussée	580
Localisation	1	Hors intersection	44 401
	2	En intersection	19 985
	3	En passage à niveau	26
Type de carrefour	1	CSSFS : Carrefour simple sans feux ou stop	11 174
	2	CSAFS : Carrefour simple avec feux ou stop	3 454
	3	CMSFS : Carrefour multiple sans feux ou stop	211
	4	CMAFS : Carrefour multiple avec feux ou stop	159
Topographie	1	Pont ou radier	11 992
	2	Rétrécissement de la chaussée	236
	3	Obstacle naturel réduisant la visibilité	546
Profils en long	1	Plat	47 650
	2	Pente	1674
	3	Sommet de côte	86
Tracé plan	1	Partie rectiligne	40 776
	2	En courbe à gauche	2654

	3	En courbe à droite	770
	4	En S	513
Largeur de l'accotement droit	1	0< L <1	1 021
	2	1<= L <2	15 348
	3	2<= L <3	4 130
	4	3<= L <4	523
	5	4<= L <5	120
	6	5<= L <10	143
Largeur de l'accotement gauche	1	0< L <1	793
	2	1<= L <2	15 113
	3	2<= L <3	3 326
	4	3<= L <4	434
	5	4<= L <5	120
	6	5<= L <10	123
Largeur de chaussée	1	L <6	6 232
	2	6<= L <7	7 486
	3	7<= L <10.5	8 004
	4	10,5<= L <14	742
	5	14<= L	898

5.2. Répartition des taux de létalité

Le tableau 10 montre que la mortalité sur les surfaces enneigées ou verglacées est la plus élevée avec un pourcentage de 32,35 % et dépasse la moyenne avec un taux de 27,66 %. Ceci résulte du faible frottement sur la surface glacée de la chaussée. C'est également le cas des autres surfaces routières recouvertes de sable, d'huile ou autre, qui présentent un pourcentage de 28,15 %.

On constate également que les chaussées dont la largeur est inférieure à 6 mètres sont plus meurtrières que celles de plus grande largeur avec un TL de 21,79 % et sont suivies par les chaussées dont la largeur est de 14 mètres ou plus, avec un TL de 18,65 %. Les chaussées les moins dangereuses sont celles dont la largeur est comprise entre 10,5 et 14 mètres.

Nous observons aussi que la plupart des accidents de la route ont lieu sur des routes en bon état, qui sont les plus utilisées en raison de la qualité de l'infrastructure routière. De toute évidence, les routes en mauvais état sont moins régulièrement utilisées et à vitesse généralement limitées, et donc moins mortelles, avec un TL de 22,19 %. Bien qu'il existe plusieurs types de séparation des routes, nous ne

retenons au sein de cette étude que deux classes : les routes pleines et les routes séparées. Les routes pleines ont un TL de 18,91 %, et les routes séparées ont un TL de 12,28 %.

Tableau 10. Taux de létalité en fonction des facteurs infrastructurels

Description		**TL**	**Description**		**TL**
État de surface		25,34	**Largeur de chaussée**		18,42
1	Sec	25,33	1	L<6	21,79
2	Mouillé	25,53	2	6<=L<7	17,61
3	Enneige ou verglace	32,35	3	7<=L<10.5	16,78
4	Humide, sel résiduel	28,15	4	10,5<=L<14	15,58
État de chaussée		25,31	5	14<=L	18,65
1	Revêtue bonne	25,32	**Type de chaussée**		18,31
2	Revêtue mauvaise	25,58	1	Route à chaussée unique	18,91
3	Revêtue très mauvaise	22,19	2	Chaussée séparée par terre-plein	12,28
Largeur de l'accotement gauche		18,09	**Largeur de l'accotement droit**		17,78
1	0<L<1	17,95	1	0<L<1	20,13
2	1<=L<2	17,73	2	1<=L<2	17,67
3	2<=L<3	19,23	3	2<=L<3	17,01
4	3<=L<4	21,08	4	3<=L<4	18,17
5	4<=L<5	19,26	5	4<=L<5	21,25
6	5<=L<10	20,69	6	5<=L<10	30,08
Position de l'accotement droit		17,83	**Types de carrefours**		8,08
1	LR	17,8	1	CSSFS	8,68
2	NLR	18,72	2	STF	4,99
Position de l'accotement gauche		18,05	3	CMSFS	11,41
1	LR	18,01	4	CMAFS	28,62
2	NLR	19,42	**Tracé plan**		**23,23**
Localisation		**27,66**	1	Partie rectiligne	23,61
1	Hors intersection	36,41	2	En courbe à gauche	19,4
2	En intersection	8,24	3	En courbe à droite	11,82
3	En passage à niveau	28,84	4	En S	29,81
Profils en long		**23,63**	**Topographie**		**18,71**

1	Plat	23,62	1	Pont ou radier	18,72
2	Pente	24,66	2	Rétrécissement de la chaussée	19,15
3	Sommet de côte	12,86	3	Obstacle naturel réduisant la visibilité	18,11

Le tableau montre aussi que les routes rétrécies sont plus dangereuses que les routes avec des obstacles ou des ponts, avec une valeur TL de 19,15 %. Il apparaît également que les accidents sur les pentes sont plus mortels avec un TL de 24,66 %, suivis par les routes plates avec un TL de 23,62 %. En outre, les sommets ont un TL inférieur de 12,86 %. Cela s'explique par le fait que les conducteurs prennent plus de précautions sur ce type de tronçons routiers.

En ce qui concerne l'emplacement, les accidents qui se produisent en dehors des intersections sont plus désastreux que ceux qui se produisent dans les intersections avec un pourcentage de 36,41 % (contre 8,24 %). Cela doit être dû à la limitation de la vitesse sur ces tronçons.

Personne ne peut nier que le passage par une route à double virage est inconfortable et exige une plus grande vigilance de la part des conducteurs, car le risque d'accident est généralement plus élevé. Effectivement, les résultats montrent que ce type de route a un taux de létalité de 29,81 %, suivi par les routes droites avec un TL de 23,61 %. Néanmoins, les courbes à droite ont un TL faible de 11,82 %, qui est inférieur au TL moyen de 23,23 %.

Le tableau met aussi en évidence que la largeur de l'accotement droit est proportionnelle à la valeur de l'indice de mortalité. Les accidents qui se produisent sur des routes d'une largeur de l'accotement droit supérieure à 5 mètres ont une valeur TL de 30,08 %, ce qui est supérieur à la moyenne de 69,18 %. Les accotements qui ne sont pas au même niveau que la chaussée (élévation de l'accotement) sont un peu plus meurtriers que les accotements qui sont au même niveau que la chaussée. Cela s'applique aussi bien aux accotements de droite qu'à ceux de gauche.

En ce qui concerne les carrefours, il convient de souligner que les intersections multiples sont plus destructrices que les intersections simples. De plus, les accidents qui se produisent sur les routes contrôlées par des feux de circulation ont un taux de létalité de 28,62 %. Le service compétent devrait éventuellement en tenir compte pour les mesures nécessaires à prendre à l'avenir concernant ce type d'intersection.

5.3. Analyse des facteurs infrastructurels

Une analyse descriptive des divers facteurs examinés est présentée dans le tableau 11. Il est à noter que les groupes associés à un facteur donné peuvent ne pas avoir des tailles d'échantillon équivalentes. Afin d'assurer une puissance analytique suffisante, les groupes ont été égalisés en taille à celle possédant une plus petite taille, de manière aléatoire. Aucun des facteurs examinés précédemment n'a pu satisfaire aux hypothèses de normalité et d'homogénéité de la variance. Cette situation est fréquente dans l'analyse de données réelles, où il est rare d'observer une distribution normale parfaite. Face à cette réalité, certains statisticiens sont contraints de prétraiter les données en apportant des ajustements significatifs, tandis que d'autres optent pour des méthodes non paramétriques afin de contourner cette problématique. Les résultats de notre étude ont révélé que le taux de létalité le plus élevé est associé à l'état de surface, avec une valeur moyenne de 29,31 %, tandis que le taux le plus bas est observé pour le facteur du type d'intersection, s'établissant à 14,08 %.

Tableau 11. Statistique descriptive des facteurs infrastructurels

Facteurs	Taille d'échantillon	Moyenne TL	Écart-type	Shapiro-Wilk		Levene	
				F	p	F	p
État de surface	760	0,293 1	0,442 9	0,724 0	<0,001	6,134 7	<0,001
État de chaussée	903	0,240 2	0,415 5	0,638 8	<0,001	4,495 2	<0,01
Type de chaussée	400	0,160 7	0,352 5	0,593 9	<0,001	9,695 7	<0,05
Position de l'accotement droit	400	0,185 3	0,368 4	0,598 7	<0,001	4,536 6	<0,05
Position de l'accotement gauche	400	0,169 0	0,354 8	0,601 9	<0,001	8,078 3	<0,05
Localisation	300	0,210 8	0,388 9	0,751 9	<0,001	10,9	<0,001
Types de carrefours	636	0,140 8	0,343 8	0,645 0	<0,001	15,924 9	<0,001
Topographie	708	0,171 8	0,365 2	0,560 4	<0,001	3,589 1	<0,05
Profils en long	258	0,231 5	0,406 9	0,711 5	<0,001	4 978	<0,001
Tracé plan	2052	0,220 5	0,401 6	0,678 0	<0,001	18,36	<0,001
Largeur de l'accotement droit	720	0,202 4	0,382 4	0,666 5	<0,001	3,583 3	<0,01
Largeur de l'accotement gauche	720	0,210 9	0,382 7	0,664 5	<0,001	18,36	<0,001
Largeur de chaussée	1250	0,170 9	0,360 4	0,605 1	<0,001	5,397 2	<0,001

Les résultats de l'analyse, comprenant à la fois le test de Kruskal-Wallis et le test de Dunn, sont présentés de manière exhaustive dans le tableau 12. Ainsi, ce tableau illustre que le test de Kruskal-Wallis

démontre une significativité pour l'ensemble des facteurs étudiés, indiquée par des valeurs p inférieures au seuil prédéterminé de 0,05. En conséquence, l'hypothèse nulle est rejetée avec un niveau de confiance de 95 %. Parmi ces facteurs, le type d'intersection se distingue comme celui ayant l'effet le plus prononcé, suivi de près par la localisation des accidents, qui expliquent respectivement 7 % et 6,5 % de la variance du TL. Ces constats suggèrent que ces facteurs pourraient exercer une influence plus marquée sur la mortalité des accidents par rapport à ceux avec une moindre taille d'effet. La taille d'effet, un critère important d'évaluation, est souvent interprétée différemment selon les contextes de recherche : une taille d'effet inférieure à 0,08 étant considérée comme faible, entre 0,08 et 0,26 comme moyenne, et supérieure à 0,26 comme grande, bien que cette interprétation puisse varier selon le domaine d'étude. Dans notre analyse, l'objectif est de déterminer les associations entre les facteurs et le TL. Le test post-hoc de Dunn est employé pour identifier les différences significatives entre les groupes en termes de TL. Les résultats de ces tests, exposés dans la figure 24 avec leur niveau de signification, sont complétés dans le tableau 12 par leur taille d'effet respective. Il est notable que les tests de Dunn révèlent des différences significatives pour l'ensemble des facteurs analysés.

5.4. Évaluation de l'impact de chaque facteur sur la mortalité routière

En la matière, le tableau 12 fournit une analyse non paramétrique détaillée, réalisée à l'aide du test de Kruskal-Wallis, afin d'évaluer l'impact de divers facteurs sur la mortalité des accidents. En se basant sur la taille d'effet epsilon-carré, chaque facteur peut être interprété quant à son influence sur la mortalité des accidents. Selon cette interprétation, le facteur « État de chaussée » est associé à un effet négligeable. En revanche, les facteurs tels que « État de surface », « Type de chaussée », « Position de l'accotement droit », « Position de l'accotement gauche », « Topographie », « Profil en long », « Tracé du plan », « Largeur de l'accotement droit », « Largeur de l'accotement gauche » et « Largeur de chaussée » présentent une taille d'effet faible. En outre, les facteurs « Localisation » et « Type de carrefour » démontrent une taille d'effet modérée, suggérant ainsi une influence plus significative sur la mortalité des accidents.

Les résultats détaillés du test post-hoc de Dunn, accompagnés de la taille d'effet CLES, permettent d'appréhender les différences entre les groupes pour chaque facteur et l'ampleur de ces différences. Pour l'état de surface, la distinction entre le groupe 1 (Surface sèche) et le groupe 2 (Surface enneigée ou verglacée) est nette, avec un CLES de 31,72, indiquant une différence substantielle entre ces conditions. En ce qui concerne l'état de chaussée, la disparité entre le groupe 1 (Revêtue en bon état) et

le groupe 2 (Revêtue en mauvais état) est notable, avec un CLES de 25,18, témoignant d'une différence significative. Pour le facteur type de chaussée, la distinction entre le groupe 1 (Chaussée unique) et le groupe 2 (Chaussée séparée par terre-plein) est également notable, avec un CLES de 21,39. Les résultats pour la position de l'accotement droit et gauche révèlent des différences significatives entre les groupes 1 (au niveau de la chaussée) et 2 (dénivelé par rapport à la chaussée), avec des CLES respectifs de 23,39

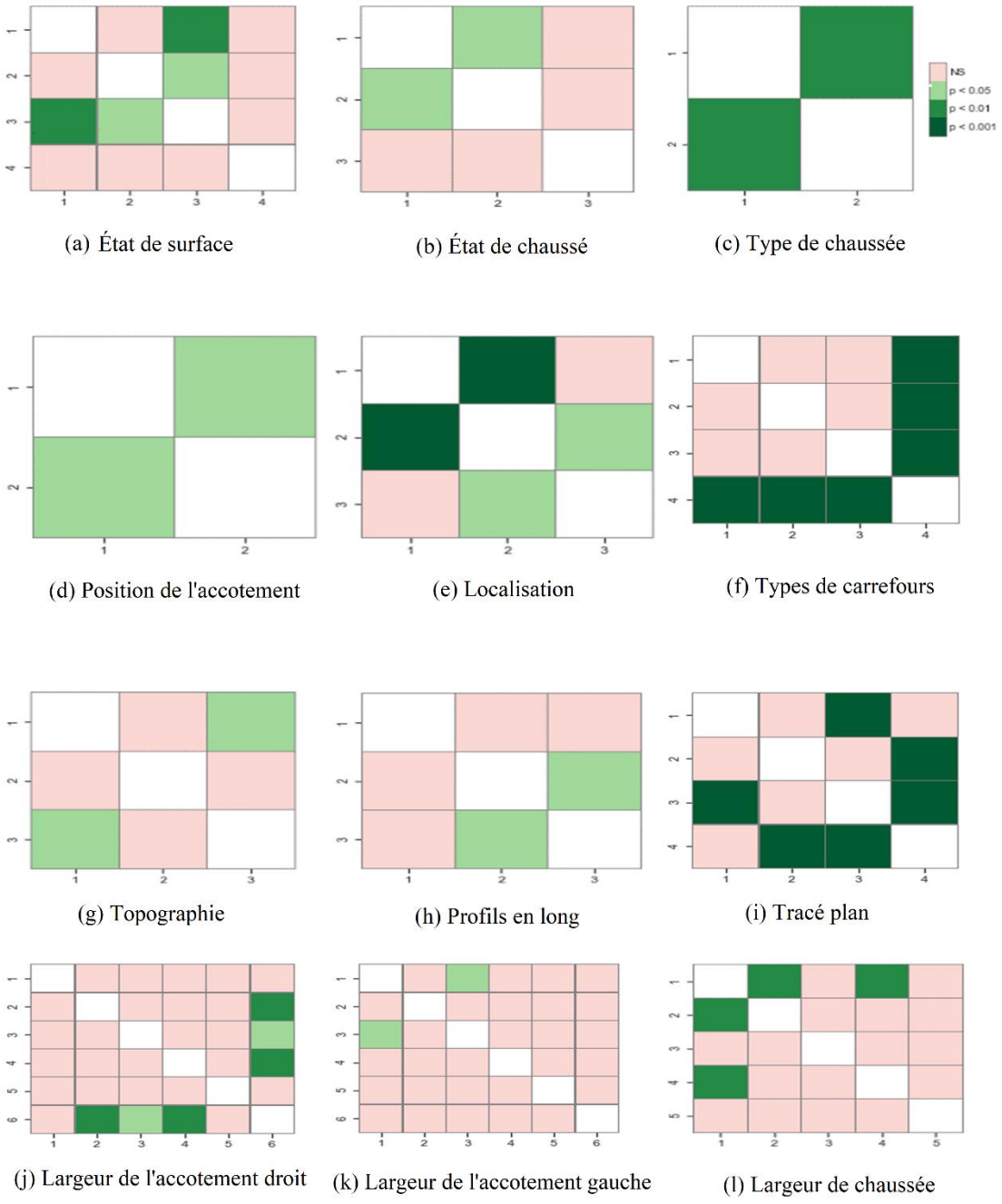


Figure 24. Résultats du test post-hoc avec le niveau de signification

(droit) et 23,66 (gauche). Pour le facteur de la localisation, la différence entre le groupe 1 (Hors intersection) et le groupe 2 (En intersection) est saisissante, avec un CLES de 32,74, dénotant une divergence substantielle. Concernant le type de carrefour, les divergences entre les groupes demeurent significatives, avec des CLES oscillant entre 26,03 et 28,62. L'analyse de la topographie indique une distinction notable entre le groupe 1 (Pont ou radier) et le groupe 2 (Obstacle naturel réduisant la visibilité), avec un CLES de 21,02. Pour le profil en long, une différence significative est observée entre le groupe 1 (Pente) et le groupe 2 (Sommet de côte), avec un CLES de 31,99. Les résultats pour le tracé du plan indiquent des divergences significatives entre les groupes, avec des CLES variant entre 25,33 et 29,14. Pour la largeur de l'accotement droit, des différences significatives sont notées entre les groupes, avec des CLES oscillant entre 30,88 et 32,32. Quant à la largeur de l'accotement gauche, une distinction notable est relevée entre le groupe 1 (0<L<1) et le groupe 2 (2<=L<3), avec un CLES de 31,85. Enfin, pour la largeur de chaussée, des différences significatives sont observées entre les groupes, avec des CLES allant de 24,66 à 24,75.

Tableau 12. Résultats du test de Krukal-Wallis et du test de post-hoc

Facteurs	**Kruskal-Wallis**		**Taille d'effet (Êta square)**	**Test de Dunn**		**Taille d'effet (CLES %)**
	F	***p***		**Groupe 1**	**Groupe 2**	
État de surface	15,537 6	<0,05	0,021	Sec	Enneige ou verglace	31,72
				Mouillé	Enneige ou verglace	30,08
État de chaussée	6,937 3	<0,05	0,008	Revêtue bonne	Revêtue mauvaise	25,18
Type de chaussée	6,925 7	<0,01	0,014	Route à chaussée unique	Chaussée séparée par terre-plein	21,39
Position de l'accotement droit	4,318 8	<0,05	0,011	Au niveau de la chaussée	Dénivelé par rapport à la chaussée	23,39
Position de l'accotement gauche	8 834	<0,05	0,02	Au niveau de la chaussée	Dénivelé par rapport à la chaussée	23,66
Localisation	19,378 6	<0,001	0,065	Hors intersection	En intersection	32,74
				En intersection	En passage à niveau	25,88
Type de carrefour	44 274	<0,001	0,07	CSSFS	CMAFS	26,03
				CSAFS	CMAFS	28,62
				CMSFS	CMAFS	26,19
Topographie	7,916 6	<0,05	0,011	Pont ou radier	Obstacle naturel réduisant la visibilité	21,02
Profil en long	8,323 9	<0,05	0,032	Pente	Sommet de côte	31,99

Tracé du plan	48,437 2	<0,001	0,024	Partie rectiligne	En courbe à droite	25,33
				En courbe à gauche	En S	26,74
				En courbe à droite	En S	29,14
Largeur de l'accotement droit	20,011 7	<0,01	0,028	1<=L<2	5<=L<10	31,31
				2<=L<3	5<=L<10	30,88
				3<=L<4	5<=L<10	32,32
Largeur de l'accotement gauche	14,028 4	<0,05	0,02	0<L<1	2<=L<3	31,85
Largeur de chaussée	18,252 5	<0,05	0,015	L<6	6<=L<7	24,75
				L<6	10,5<=L<14	24,66

6. Discussion des résultats et recommandations

L'analyse des données relatives aux accidents routiers de l'année 2017, portant sur un ensemble de 13 variables, met en exergue une corrélation significative entre l'état physique de la route et l'infrastructure, et la gravité des accidents. Par exemple, l'état de la surface de la route a révélé des différences marquées entre les surfaces sèches et enneigées, avec une probabilité de 31,72 % qu'un accident sur l'une soit plus fatal que sur l'autre. De même, la localisation des accidents, qu'ils se produisent dans ou hors d'une intersection, a montré une augmentation de la probabilité de mortalité de 32,74 %, suggérant que d'autres accidents peuvent également être graves dans environ un tiers des cas. Les accidents urbains, en particulier ceux survenant dans des intersections avec des feux de circulation ou des panneaux-stops, présentent une probabilité de mortalité accrue de 28,62 %. La mise en œuvre de mesures préventives telles que des campagnes de sensibilisation et l'application de sanctions restrictives peuvent contribuer à atténuer ce problème.

Par ailleurs, une différence significative a été observée entre les accidents survenant sur des routes en pente par rapport à ceux sur un sommet de côte, avec une probabilité de 32 %. Cela signifie qu'il y a une probabilité de 32 % qu'un accident qui se produit en pente soit plus grave qu'un accident qui se produit au sommet de côte. De même, la largeur des accotements s'est révélée être un facteur contribuant souvent à la gravité des accidents ; par exemple, une largeur de 5 mètres est associée à un TL de 30 %. L'analyse de la taille de l'effet (CLES) indique également qu'il existe une probabilité de 24 % que le TL soit supérieur pour un niveau d'accotement par rapport à un autre. Il est donc recommandé, par mesure de précaution, d'installer des barrières le long des voies piétonnes en dehors des zones urbaines ; une

mesure soutenue par des recherches antérieures [86] ayant montré une gravité accrue des accidents sur des accotements non protégés.

Enfin, les résultats de notre analyse concernant la largeur de l'accotement, la largeur et le type de chaussée corroborent les conclusions de [79] démontrant une corrélation entre ces variables et la sévérité des accidents. De plus, il a été observé que lorsque le tracé du plan est influencé par la topographie, la mortalité des accidents augmente, soulignant ainsi l'importance de prendre en compte ces facteurs dans la planification et la conception des routes.

Ces analyses suggèrent que de nombreux facteurs routiers ont un impact significatif sur la mortalité des accidents, avec des différences de taille d'effet allant de négligeable à significative. Voici une synthèse des résultats obtenus :

- L'état de la surface : le test de Kruskal-Wallis montre une différence significative, et le test post-hoc de Dunn indique une grande différence de mortalité entre les routes sèches et les routes enneigées ou glacées.
- L'état de la chaussée : il y a aussi une différence importante, et la différence de mortalité entre les routes en bon état et les routes en mauvais état est notable.
- Le type de chaussée : le test de Kruskal-Wallis montre une différence significative, et le test post-hoc de Dunn indique une différence significative de mortalité entre les routes à chaussée unique et les routes à chaussée séparée par une médiane.
- La position de l'accotement droit et de l'accotement gauche : les deux facteurs montrent des différences significatives, avec des différences notables de mortalité entre les accotements au niveau de la chaussée et celles dénivelées par rapport à la chaussée.
- La localisation : le test de Kruskal-Wallis affiche une différence significative, et le test post-hoc de Dunn indique une grande différence de mortalité entre les accidents hors intersection et les accidents aux intersections.
- Le type d'intersection : tous les groupes présentent des différences significatives, avec des différences notables de mortalité entre les différents types d'intersection.
- La topographie : le test de Kruskal-Wallis témoigne d'une différence significative, avec une différence notable dans les décès entre les accidents sur les ponts ou les radiers et ceux impliquant des obstacles naturels réduisant la visibilité.

- Le profil en long : le test de Kruskal-Wallis montre une différence significative, et le test post-hoc de Dunn indique une grande différence de mortalité entre les accidents sur les pentes et ceux sur les sommets des côtes.
- Le tracé du plan : tous les groupes affichent des différences importantes, avec des différences notables de mortalité selon le tracé des routes.
- La largeur des accotements droits : tous les groupes présentent des différences significatives, avec des différences notables de mortalité selon la largeur des accotements droits.
- La largeur des accotements gauches : il y a une différence significative, avec une différence notable de mortalité entre les groupes selon la largeur de l'accotement gauche.
- La largeur de la chaussée : les groupes indiquent des différences significatives, mais avec des différences de taille d'effet plus faibles pour la largeur de la chaussée.

Les recommandations formulées sur la base des résultats des tests de Kruskal-Wallis et du test post-hoc de Dunn dépendent des facteurs spécifiques qui ont montré des différences significatives dans leur impact sur la mortalité due aux accidents de la route. Les recommandations générales suivantes sont basées sur les facteurs qui ont montré des différences significatives :

- L'état de surface : étant donné la grande différence de mortalité entre les routes sèches et les routes enneigées ou verglacées, il est recommandé de mettre en œuvre des mesures plus efficaces de gestion du trafic et de déneigement en hiver. Il est bien connu que la conduite sur des routes enneigées ou verglacées est nettement plus dangereuse que la conduite sur des routes sèches. La différence entre les taux de létalité est particulièrement marquée, avec une probabilité beaucoup plus élevée d'accidents mortels dans des conditions hivernales. Pour remédier à ce problème, il est conseillé aux gouvernements et aux municipalités de mettre en œuvre des stratégies de gestion du trafic et des mesures de déneigement plus efficaces pendant les mois d'hiver. Il peut s'agir d'augmenter le nombre de chasse-neige sur les routes, d'utiliser des agents de déglaçage plus efficaces et de proposer des programmes de formation à la conduite plus complets afin d'aider les automobilistes à affronter les conditions hivernales dangereuses en toute sécurité. En prenant ces mesures proactives, il est possible de réduire le nombre d'accidents et de décès qui se produisent chaque année sur les routes, et de faire en sorte que chacun puisse voyager en toute sécurité et en toute confiance, quel que soit le temps qu'il fait à l'extérieur.
- L'état des chaussées : les routes dont la chaussée est en mauvais état de revêtement présentent une différence significative en termes de létalité. Il est donc recommandé de donner la

priorité à la réparation et à l'entretien des routes dont la chaussée est en mauvais état. Des études récentes ont montré que les routes en mauvais état présentent un risque important pour les conducteurs, entraînant un taux plus élevé d'accidents et de décès. Il est donc urgent d'accorder la priorité à la réparation et à l'entretien de ces routes, car il s'agit là d'une étape cruciale dans l'amélioration de la sécurité routière. En outre, l'investissement dans des technologies et des matériaux de construction routière modernes pourrait également contribuer à réduire le nombre d'accidents et de décès sur nos routes. Il est donc fortement recommandé aux décideurs politiques et aux autorités compétentes de prendre des mesures proactives pour résoudre ce problème et garantir la sécurité de tous les usagers de la route.

- La position des accotements droit et gauche : compte tenu des différences de mortalité entre les accotements situés au même niveau que la chaussée et ceux situés à des niveaux différents par rapport à la chaussée, il est important de maintenir une surface d'accotement sûre pour les conducteurs. La position de l'accotement droite et de l'accotement gauche est un facteur crucial dans le maintien d'une chaussée sûre. Aussi, il est important de prendre en compte les différences de taux de létalité entre les accotements situés au même niveau que la chaussée et ceux situés à des niveaux différents de la chaussée. En outre, il est important de maintenir une surface d'accotement sûre non seulement pour les conducteurs, mais aussi pour les piétons et les cyclistes qui peuvent utiliser l'accotement comme moyen de transport. En veillant à ce que l'accotement soit sûr et bien entretenu, nous pouvons réduire considérablement le risque d'accident et accroître la sécurité générale de la chaussée. Cet objectif peut être atteint grâce à des inspections régulières de la surface de l'accotement et la mise en œuvre de mesures telles que des barrières, des panneaux d'avertissement et un meilleur éclairage.
- La localisation : les accidents aux intersections ont montré une différence significative dans le nombre de décès. À ce titre, les autorités routières devraient examiner les intersections à haut risque afin d'identifier des mesures de sécurité supplémentaires. Les accidents aux intersections constituent une préoccupation majeure pour les autorités chargées de la sécurité routière. Il a été constaté que ces accidents entraînent un plus grand nombre de décès. Il est donc important d'examiner les intersections à haut risque et d'identifier les mesures de sécurité supplémentaires qui peuvent être mises en place pour prévenir ces accidents. Ces mesures de sécurité peuvent inclure l'installation de feux de circulation, de panneaux de limitation de vitesse et de passages pour piétons. En outre, la sensibilisation des conducteurs et des piétons à

l'importance du respect des règles de circulation et du code de la route pourrait également contribuer à réduire le nombre d'accidents à ces intersections à haut risque. Il est donc essentiel que les autorités chargées de la sécurité routière prennent des mesures proactives pour garantir la sécurité de tous les usagers de la route et éviter les pertes de vies inutiles.

- Le type d'intersection : les différents types d'intersections présentent des différences significatives. Il est conseillé de tenir compte de ces différences lors de la planification et de la conception des intersections routières. Les intersections sont un aspect essentiel des réseaux routiers, car elles permettent aux véhicules de changer de direction. Il existe différents types d'intersections routières, et chaque type présente des avantages et des défis uniques. Ainsi, les intersections peuvent être classées en fonction de leur forme, du nombre de branches, du nombre de voies et de la présence d'un terre-plein et de la signalisation. Il est donc essentiel de considérer ces différences lors de la planification et de la conception des intersections routières. Un rond-point, par exemple, est un type d'intersection de forme circulaire avec un îlot central. Il est conçu pour réduire la vitesse et améliorer la sécurité. En revanche, un carrefour en T n'a que trois branches et peut être difficile à gérer, en particulier pour les conducteurs qui ne connaissent pas bien la zone. Il importe de comprendre les différents types de carrefours pour garantir la sécurité et l'efficacité des réseaux routiers.
- La topographie : l'impact de la topographie sur la mortalité doit être pris en considération dans la conception des routes. Des mesures spécifiques peuvent être nécessaires pour améliorer la sécurité sur les ponts, les obstacles naturels réduisant la visibilité et le rétrécissement de la chaussée, par exemple. Une étude approfondie du terrain et du paysage doit être menée afin d'identifier tout obstacle naturel susceptible de réduire la visibilité. Ces informations peuvent ensuite être utilisées pour planifier des mesures spécifiques visant à améliorer la sécurité sur la route. Par exemple, les ponts et les routes étroites peuvent nécessiter des dispositifs de sécurité supplémentaires pour éviter les accidents. En outre, l'impact de la topographie sur les taux de létalité devrait être pris en compte lors de la conception des routes. En mettant en œuvre des mesures de sécurité appropriées, nous pouvons garantir que nos routes sont sûres pour tous les conducteurs et passagers.
- Le profil en long : les accidents sur les pentes ont montré une différence significative. Les autorités routières doivent accorder une attention particulière aux routes ascendantes et descendantes afin d'améliorer la sécurité. Le profil en long des routes est un facteur important à

considérer en matière de sécurité routière. Certaines études ont souligné que les accidents sur les pentes, tant en montée qu'en descente, sont plus fréquents que sur les routes plates. Il convient à ce titre que les autorités routières accordent une attention particulière à ces types de routes. Des améliorations telles qu'une meilleure conception des routes, un marquage au sol plus visible et un entretien adéquat peuvent réduire de manière significative le risque d'accident sur les pentes. En outre, l'éducation des conducteurs sur la manière de naviguer en toute sécurité sur les pentes peut également contribuer à l'amélioration de la sécurité routière.

- Le tracé plan : les différences significatives en fonction de la configuration de la route indiquent que des améliorations spécifiques de la sécurité peuvent être nécessaires sur les tronçons de route en courbe. À ce propos, l'étude souligne qu'il existe des différences significatives dans le nombre d'accidents sur les routes droites par rapport aux routes en courbe. Cela suggère qu'il pourrait être nécessaire d'apporter des améliorations ciblées en matière de sécurité sur les tronçons de route en courbe. En outre, les accidents sur les pentes sont plus fréquents que sur les routes plates. Il est donc essentiel que les autorités routières accordent une attention particulière à ces types de routes.
- La largeur des accotements gauche et droit : les variations de la largeur des accotements à droite et à gauche suggèrent que des accotements larges devraient être maintenus pour assurer la sécurité des conducteurs en cas d'urgence. En outre, des accotements plus larges offrent également plus d'espace aux conducteurs pour manœuvrer leur véhicule, en particulier dans les situations où des mouvements brusques sont nécessaires, comme lorsqu'un conducteur doit éviter un obstacle sur la route. Il est donc recommandé de prendre en considération la largeur des accotements gauche et droit lors de la conception des autoroutes et des autres routes afin de garantir la sécurité des conducteurs.
- La largeur de la chaussée : bien que la différence soit moins marquée, il est important de doter des chaussées d'une largeur adéquate pour assurer la sécurité routière. La largeur des chaussées est en effet un facteur crucial pour la sécurité routière. Un espace suffisant pour les piétons, les cyclistes et les autres usagers de la route peut contribuer à prévenir les accidents et à améliorer la fluidité du trafic. En outre, des trottoirs plus larges peuvent offrir davantage de possibilités d'aménagement paysager et de mobilier. Cependant, il est important de trouver un équilibre entre la nécessité d'avoir des trottoirs plus larges, la disponibilité de l'espace et d'autres

facteurs, tels que le coût. Trouver le bon équilibre conduit à des rues plus sûres et plus attrayantes pour tout le monde.

Toutes ces recommandations générales sont basées sur les différences significatives observées dans l'impact de chaque facteur sur la létalité des accidents. Les résultats de notre étude soulignent l'importance de prendre en compte les différents facteurs qui affectent la sécurité routière. En mettant en œuvre des mesures de sécurité appropriées, nous pouvons faire en sorte que les routes soient sûres pour tous les conducteurs et les passagers. Cependant, pour une prise de décision plus précise, des recherches supplémentaires et des évaluations approfondies de la sécurité routière peuvent être nécessaires pour chaque facteur spécifique.

7. Conclusion

Ce chapitre présente une analyse statistique visant à identifier les facteurs déterminants de la mortalité dans les accidents de la circulation. Les données ont été collectées concernant treize variables décrivant l'infrastructure routière et les conditions physiques de la route. Ensuite, le taux de létalité a été introduit afin de comparer les différents groupes pour chaque facteur à l'aide du test non paramétrique de l'ANOVA. Les résultats de ce test ont révélé des différences significatives entre tous les facteurs étudiés et le taux de létalité. Plus particulièrement, le type d'intersection et la localisation ont affiché une taille d'effet modérée, démontrant ainsi leur impact sur la gravité des accidents de la circulation. Ces conclusions corroborent l'hypothèse initiale selon laquelle l'infrastructure routière et les conditions de la route influent sur la mortalité des accidents. Ces résultats sont susceptibles de contribuer significativement à la littérature existante en matière de sécurité routière, en particulier en ce qui concerne l'amélioration de la prédiction des accidents et le développement de solutions d'ingénierie adaptées. Néanmoins, il est à noter que les interprétations tirées de cette analyse sont spécifiques aux données examinées. Des résultats différents pourraient être obtenus en utilisant d'autres ensembles de données ou en analysant des périodes différentes. Dans le cadre de futures recherches, il pourrait être envisagé d'étendre cette analyse à d'autres ensembles de données recueillis sur une période plus étendue, ainsi que d'explorer d'autres variables pouvant influencer le taux de létalité des accidents de la circulation. Une telle démarche permettrait une compréhension plus approfondie du problème et ouvrira de nouvelles perspectives de recherche dans le domaine de la sécurité routière.

Chapitre 5. Identification des points noirs d'accidents par la méthode d'indice de gravité pondéré

1. Introduction

Selon l'OMS, les accidents de la route sont actuellement considérés comme la 9e cause de décès dans le monde, causant la perte de 1,2 million de vies chaque année[6]. Au Maroc, les accidents de la route tuent environ 3500 personnes et causent environ 100 000 blessés chaque année. Entre 2009 et 2019, le nombre d'accidents a augmenté de 53,43 % alors que le parc automobile a augmenté de plus de 54 % [64]. Le rapport national sur les accidents de la route, lancé par la Direction des Routes en 2017, relève que les accidents urbains représentent une forte proportion des accidents de la circulation, 73 % causant seulement 35,43 % des décès sur la route, tandis que 64,57 % des décès sur la route se sont produits dans les zones rurales. Par ailleurs, ces accidents causent des pertes économiques importantes et affectent négativement le développement des zones rurales. Pour toutes ces raisons, le gouvernement marocain a mis en œuvre de nombreuses mesures visant à améliorer la sécurité routière et à réduire l'impact de la « guerre de la route » sur la société et l'économie. En effet, les accidents de la route coûtent encore chaque année environ 2 % de son Produit Intérieur Brut (PIB) du pays.

Dans cette perspective le Maroc a lancé sa première stratégie de sécurité routière en 2004-2013[7], qui a donné de bons résultats en termes de réduction du nombre de morts et de blessés graves. Cependant, le taux de mortalité par rapport à la taille de la population et au parc automobile reste préoccupant par rapport aux taux de mortalité dans la majorité des pays motorisés. A cet égard, la figure 25 montre le nombre annuel de décès dus à des accidents de la route pour 100 000 personnes. Elle établit que le Maroc a connu une situation grave entre les années 2000 et 2010. En effet, le taux de mortalité a augmenté légèrement à l'inverse des pays les plus motorisés comme la France et les États-Unis. Par la suite, le taux de mortalité a commencé à diminuer et à se stabiliser autour de 21 décès pour 100 000 personnes vers

[6] World Health Organization. (2022). Une nouvelle Déclaration politique historique fixe comme objectif de diviser par deux, d'ici à 2030, le nombre de morts et de blessés sur la route. https://www.who.int/fr/news/item/30-06-2022-new-political-declaration-to-halve-road-traffic-deaths-and-injuries-by-2030-is-a-milestone-achievement#:~:text=Au%20niveau%20mondial%2C%20les%20accidents%20de%20la%20route,dans%20des%20pays%20%C3%A0%20revenu%20faible%20ou%20interm%C3%A9diaire.

[7] Ministère de l'Équipement et de l'Eau. (2018). Plan Stratégique Intégré d'Urgence de Sécurité Routière. https://www.equipement.gov.ma/Transport-routier/Securite-routiere/Pages/PSIU.aspx.

l'année 2013. Cependant, la situation reste encore préoccupante, comparée aux pays développés. Face à cette situation alarmante, le gouvernement marocain doit accorder plus d'effort face à la guerre des routes en traitant les zones les plus accidentelles, plus précisément les sections les plus meurtrières. Dans ce contexte, une nouvelle stratégie a donc été lancée par les autorités marocain en 2017 avec pour objectif principal de réduire le taux de mortalité à 50 % d'ici 2026.

Ainsi, dans le contexte de la prévention des accidents, il revêt une importance cruciale de diriger les initiatives vers les voies présentant les taux d'accidents les plus élevés. C'est dans cette perspective que cette étude a été entreprise, visant à détecter les points noirs sur la route nationale 1 (N1) au Maroc. La N1 est une route principale reliant le nord au sud du Maroc. Elle représente 20,54 % de la totalité du réseau des routes nationales. Cependant, elle est également associée à une part significative des accidents mortels dans le pays, contribuant à hauteur de 29,28 % du total des décès, soit 28,97 % de la mortalité nationale. L'amélioration globale de cette route nécessite des investissements importants en termes de temps et de ressources financières. Afin d'accroître la sécurité sur cette route, il est impératif d'identifier les sections les plus dangereuses et de les soumettre à une analyse approfondie en vue de leur rénovation, en particulier si les accidents sont principalement attribuables à l'état de la route. Des améliorations supplémentaires doivent être envisagées le cas échéant. Ces considérations ont été la motivation

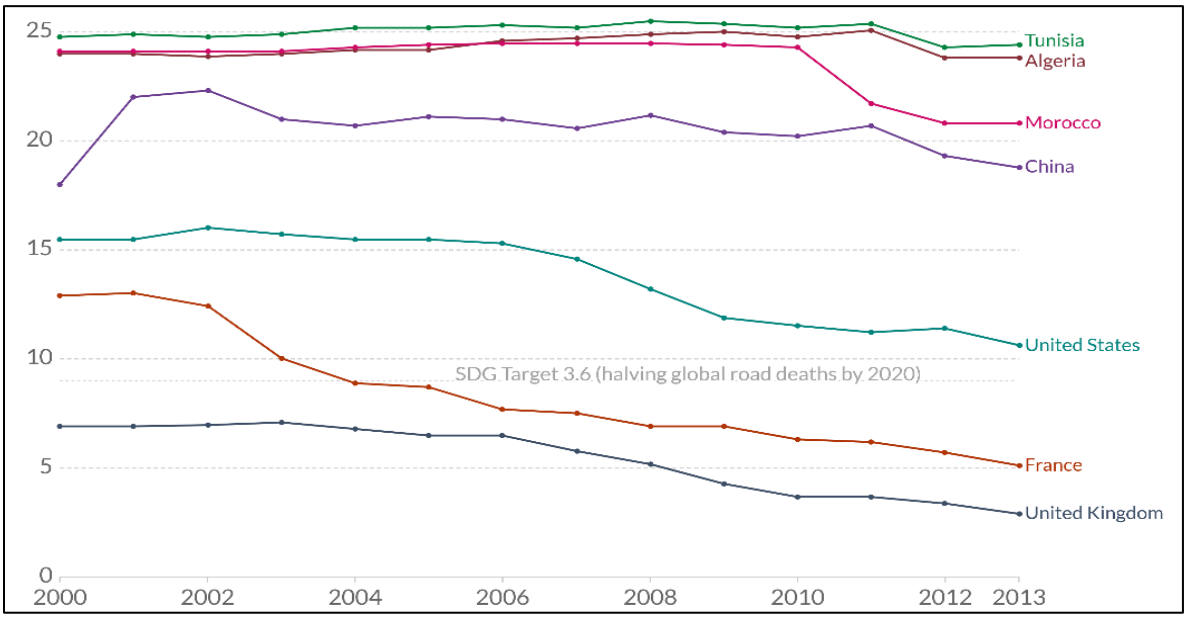


Figure 25. Taux de mortalité accidentelle pour 100 000 personnes[8]

[8] OurWorldInData. Death rate from road injuries, 2000 to 2013. https://ourworldindata.org/grapher/death-rate-from-road-accidents-ghe?tab=chart&time=2000..2013&facet=none&country=MAR~FRA~DZA~TUN~CHN~GBR~USA.

fondamentale de notre étude, qui vise à identifier les points dangereux de cette route spécifique, souvent considérée comme l'une des plus meurtrières du pays.

Afin d'atteindre ces objectifs, nous avons intégré la méthode de l'indice de gravité pondéré, visant à évaluer la gravité des sections routières en termes de dommages humains. Chaque catégorie de victime étant soumise à une pondération, de telle sorte que la plus sévère bénéficie d'un poids plus élevé. Cependant, le choix des pondérations doit être soigneusement établie. À cette fin, la méthode de l'arbre de décision a été introduite, permettant de générer des scores pour les variables d'entrée de manière à attribuer des scores élevés aux variables les plus importantes à la construction des arbres de décision. Cette approche a été appliquée à notre problème de pondération, démontrant son efficacité dans la génération des pondérations pour la méthode de l'indice de gravité pondéré. Celle-ci a permis d' identifier de manière fiable et efficace les points noirs d'accidents sur la route N1, les classant en trois groupes en fonction de l'indice de gravité pondéré. Le premier groupe englobe les points noirs présentant une gravité élevée, comprenant 74 sections dangereuses et le deuxième groupe les sections de gravité moyenne, totalisant 34 sections. Le troisième groupe indique une faible gravité, comprenant 65 sections.

2. Méthodes d'identification des points noirs

De multiples facteurs contribuent aux accidents, tels que l'état des routes, des véhicules et les conditions météorologiques, mais le comportement humain reste l'un des plus importants. Il est vrai que ces facteurs sont étroitement liés, mais leur combinaison peut considérablement augmenter le nombre d'accidents et accentuer la mortalité. Plusieurs études ont été menées pour analyser ces facteurs, en particulier dans les endroits où le nombre d'accidents est inhabituellement élevé. Dans [33], une analyse d'ANOVA a permis de montrer que l'infrastructure routière a un effet significatif sur les accidents de la route, à savoir que le type d'intersection et la localisation s'avèrent avoir un impact important sur la mortalité. Ces endroits sont appelés points noirs et définis comme des « segments de route qui ont fait l'objet d'un certain nombre d'accidents au cours d'une période « t » sur une distance « d » définie ». Ladi *et al.* [87], ont définis les points noirs comme « tout endroit où il y a eu au moins cinq accidents sur une période d'un an ». Au Maroc, un point noir désigne un tronçon de route où se sont produits au moins dix accidents corporels sur une distance d'un kilomètre, ayant entraîné au moins dix morts ou blessés graves. En général, leur détermination est basée sur un nombre d'accidents survenus sur un site ou une zone présentant les mêmes caractéristiques.

Dans ce cadre, l'Indice de Gravité Pondéré est l'une des méthodes utilisées pour identifier les points noirs d'une route. Elle consiste à pondérer les accidents corporels afin que les accidents les plus graves aient un poids plus important. Le choix de la méthode de pondération varie d'une approche à l'autre. Les auteurs de [88] ont conclu que IGP est la méthode la plus scientifique et la plus précise, car elle est calculée à partir de trois types d'accidents : mortels, corporels graves et corporels légers. Les poids de la méthode sont déterminés en fonction de l'analyse du coût des accidents mortels.

Les poids de l'Indice de Gravité Pondéré sont en effet pertinents, car leurs valeurs expriment l'importance des variables. Aussi, l'attribution des poids aux variables doit-il suivre un mécanisme spécifique afin de garantir leur fiabilité. Ces deux propriétés sont prises en charge par les méthodes d'ensemble d'arbres telles que les arbres à gradient boosté et les forêts aléatoires. Le XGBoost affecte des valeurs d'importance aux attributs d'entrée ; ces valeurs donnent un aperçu général du modèle prédictif. Le concept de calcul des valeurs d'importance est ainsi basé sur la réduction d'erreur apportée par chaque division dans chaque arbre, et l'importance de la caractéristique sera la somme des réductions d'erreur, appelée « gain », de toutes les divisions pour une caractéristique donnée.

Certaines études dans le domaine de la recherche sur la sécurité routière ont intégré l'Indice de Gravité Pondéré dans les Systèmes d'Information Géographique (SIG) en tant que nouvelle technique. Cette intégration vise à fournir un examen plus approfondi et plus précis des accidents de la circulation. Les auteurs de l'étude [89], ont utilisé l'IGP pour analyser les facteurs de risque susceptibles de contribuer aux accidents de la circulation. Ces facteurs sont pondérés sur une échelle de 1 à 10, de manière que celles augmentant la probabilité d'accident se voient attribuer des poids moins élevés. Cependant, les critères d'attribution des poids et le choix de leur intervalle ne sont pas explicitement justifiés. Dans l'étude citée en [90], les auteurs utilisent la méthode de rangement et l'indice de gravité pour identifier les points noirs sur les routes. De ce fait, leur méthode prend en compte la nature, la cause et la classification des accidents dans le processus d'identification des points noirs. Cependant, les sections identifiées comme points noirs sont souvent trop longues, ce qui rend difficile l'identification précise des zones de danger et la mise en place de mesures de sécurité adaptées. Ainsi que, leur analyse ne considère qu'un seul accident par segment de route, ce qui ne fournit pas une base adéquate pour élaborer un modèle représentatif de chaque section. Cette approche peut entraîner une compréhension incomplète des problèmes de sécurité routière, car elle ne tient pas compte de la fréquence ou de la gravité des accidents supplémentaires sur les sections examinées.

3. Modèle d'identification des points noirs d'accidents

3.1. Méthodologie utilisée

Dans le cadre de notre étude, nous utiliserons la méthodologie de l'Indice de Gravité Pondéré afin d'identifier les points noirs d'accidents. Le calcul du IGP implique à ce titre la considération de deux catégories de blessures et de quatre catégories de décès, plutôt que de se limiter à une seule modalité de décès. Le recours à quatre catégories de décès permet de distinguer les lieux de survenue de décès et de différencier les accidents en fonction de la période écoulée entre l'accident et le décès. Il est à noter que plus cette période est courte, plus la gravité de l'accident est élevée. À chaque type de victime, qu'il s'agisse de décès ou de blessures, est attribué un poids, de manière à attribuer un poids plus élevé aux blessures les plus graves. Précisons que les valeurs des poids pour l'IGP ont été soigneusement déterminées à l'aide de la fonction d'importance des caractéristiques fournie par l'algorithme XGBoost. La figure 26 résume la méthodologie adoptée au sein de cette étude, commençant par la préparation des données accidentogènes jusqu'à l'interprétation des résultats en passant par l'étape d'identification des points noirs d'accidents.

3.2. Préparation de données

Les données utilisées dans notre recherche proviennent du Ministère de l'Équipement et de l'Eau, se rapportant aux accidents de la route survenus entre 2013 et 2017. La première phase du processus d'analyse débute par la collecte des données, effectuée par les entités responsables de l'intervention en cas d'accidents. Il s'agit des agents de la sûreté nationale pour les accidents survenus en zones urbaines et des agents de la gendarmerie royale pour ceux survenus en zones rurales. Au sein du Ministère de l'Équipement et de l'Eau, ces deux sources de données sont fusionnées dans des tableaux comprenant les enregistrements bruts des accidents de la route, principalement destinés à une analyse descriptive. Ces tableaux contiennent une variété d'informations détaillées sur les accidents, comprenant le contexte temporel (date et heure) et le type d'accident (frontal, arrière, etc.), des détails sur les véhicules impliqués (type, année de fabrication, etc.), des informations précises sur les victimes (âge, sexe, profession, etc.) ainsi que des données détaillées sur l'infrastructure (chaussée, accotements, signalisation, etc.).

Ainsi, notre étude se concentre sur les données relatives à la localisation des accidents et aux caractéristiques des victimes impliquées. La localisation des accidents est exprimée en termes de points kilométriques (PK), tandis que les informations sur les victimes sont classées en trois catégories : les

conducteurs, les passagers et les piétons. Les victimes sont réparties en deux catégories principales : les décès et les blessés. Le groupe des décès est subdivisé en quatre niveaux, comprenant ceux survenus sur le champ de l'accident, pendant le transfert à l'hôpital, dans les 7 jours suivant l'accident, et dans les 30 jours suivant l'accident. Quant au groupe des blessés, ce dernier est scindé en deux catégories : les blessures graves nécessitant une hospitalisation et les blessures légères nécessitant des soins mineurs.

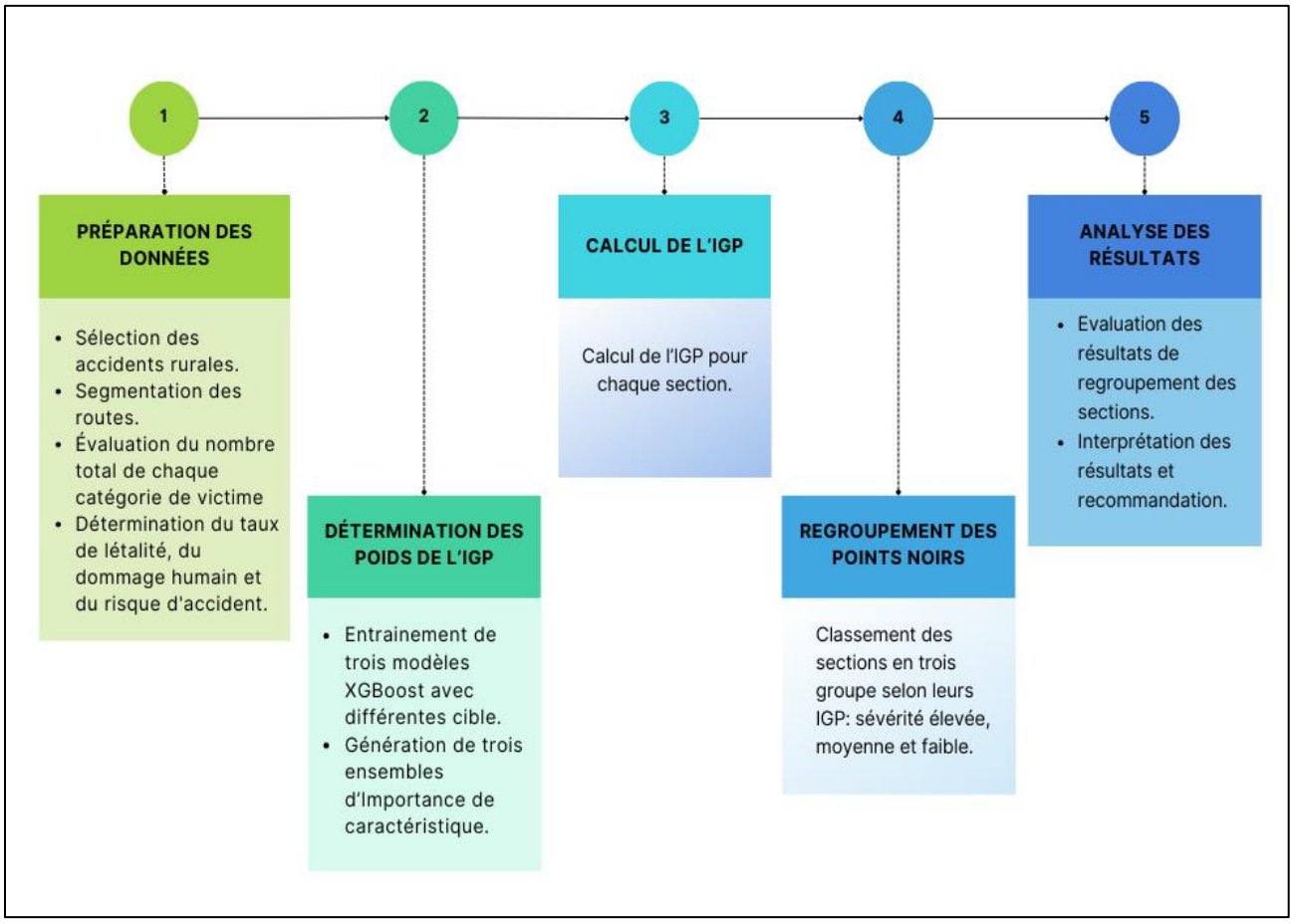


Figure 26. Méthodologie de recherche

À la suite d'une analyse initiale des données statistiques relatives à l'ensemble du réseau routier marocain, il a été décidé de nous focaliser sur la route nationale N1 (voir figure 27). Ce choix est justifié par plusieurs facteurs, notamment la longueur considérable de cette voie, totalisant 2339 km, la positionnant ainsi en tant que la route la plus étendue du pays. La sélection de la N1 repose également sur le constat du nombre significatif d'accidents mortels enregistrés annuellement sur cette route. Par ailleurs, il convient de souligner le rôle prépondérant de la N1 dans le contexte du développement socio-économique du Maroc, en tant que liaison vitale reliant les principales agglomérations du pays telles que

Tanger, Rabat, Casablanca, Agadir, et facilitant le lien avec le continent européen grâce à son port stratégique situé à Tanger.



Figure 27. Route nationale 1[9]

Pour ce faire, les accidents survenus sur les zones rurales de la route nationale N1 ont été recueillis et soumis à une analyse systématique. Les données ont été subdivisées en tronçons d'un kilomètre, avec le calcul du nombre de victimes de chaque type pour chaque segment. Au cours de ce processus, il a été observé que certaines sections ne présentaient aucun accident. Par conséquent, elles ont été exclues de l'étude et considérées comme des segments sécurisés. Par ailleurs, certains accidents ne fournissaient pas d'informations précises sur le point kilométrique en raison de leur survenue sur des routes non répertoriées par le Ministère de l'Équipement et de l'Eau. À la suite de cette manipulation des données, 9542 accidents, impliquant 15 873 véhicules, ont été répertoriés dans le cadre de notre étude. Par la suite,

[9] Maroc-Maps. (2016). Route Nationale Numéro 1 (RN1 ou N1). maroc-maps.blogspot.com.

les nombres de victimes et d'accidents a été utilisé comme des indicateurs pour évaluer la gravité de chaque section en termes de pertes humaines. À cet effet, trois indicateurs ont été calculés. Lesquels sont décrits comme suit :

- Le taux de létalité : le taux de létalité est défini comme le pourcentage de décès par rapport au nombre total de décès et de blessés dans une section donnée. Ce taux est calculé en divisant le nombre de décès par la somme du nombre de décès et de blessés. Cette métrique donne une indication de la gravité des incidents dans la section concernée. Un TL plus élevé indique une proportion plus importante de décès par rapport aux blessés, ce qui peut suggérer des conditions de sécurité routière moins favorables. Le TL est calculé comme suit :

$$\boldsymbol{TL = \frac{D}{D + B} \times 100} \tag{35}$$

- Les dommages humains : dans l'étude menée par Wei et al. [57], le dommage humain est défini comme la proportion de décès par rapport au nombre d'accidents. Cette mesure est définie comme le rapport entre le nombre total de victimes, y compris les morts et les blessés, et le nombre total de personnes impliquées dans des accidents sur une section spécifique. Cette mesure permet d'évaluer le nombre moyen de victimes par personne impliquée dans un accident. Ainsi, elle permet de mesurer l'impact global d'une situation d'accident sur la santé et la sécurité des personnes impliquées dans cette section particulière. Cette mesure est calculée comme suit :

$$\boldsymbol{DH = \frac{D + B}{P}} \tag{36}$$

- Le risque d'accident : ce dernier est défini comme le rapport entre le nombre de morts et le nombre total d'accidents. Cette mesure permet d'évaluer le taux de mortalité moyen par accident. Une valeur élevée de cette mesure indique un risque élevé de décès associé à un accident. Ainsi, une augmentation de cette valeur indique une augmentation de la gravité des accidents en termes de décès. Le risque d'accident est exprimé par la formule suivante :

$$\boldsymbol{RA = \frac{D}{A}} \tag{37}$$

Où D indique le nombre de morts dans l'accident, B le nombre de blessés, P le nombre des personnes impliquées et A le nombre d'accidents. Ces métriques seront utilisées, d'une part, pour évaluer

la gravité des sections routière en termes de décès, et serviront à la détermination des poids de l'Indice de Gravité Pondéré.

Le tableau 13 montre la structure finale des données à utiliser dans cette étude. Pour chaque section, le nombre total des différentes catégories de victimes a été calculé, de même que les trois taux de gravité ont été définis. Nul doute que cette approche de la présentation des données facilitera l'analyse et contribuera à une meilleure compréhension de la répartition des victimes d'accidents sur les sections routières. Contrairement à certaines études antérieures, qui ont choisi de ne pas différencier les types de décès, nous avons conservé les quatre modalités de décès. Nous considérons que cette distinction est cruciale, puisque le décès sur le lieu de l'accident indique une gravité élevée, tandis que le décès dans les 30 jours suggère que les blessures étaient graves, même si elles n'ont pas entraîné le décès sur le lieu de l'accident. Les travaux menés par les auteurs de [79][91], entre autres, considèrent la gravité des blessures comme une mesure de la gravité de l'accident, en classant les accidents en deux catégories : les blessures légères et les décès ou les blessures graves.

Tableau 13. Structure des données utilisées

PK	Tués sur le champ	Tués pendant le transfert à l'hôpital	Tués dans les 7 jours suivant l'accident	Tués dans les 30 jours suivant l'accident	Blessé grave	Blessé léger	TL	DH	RA
1	1	0	0	0	0	24	0,04	25	0,071 4
2	1	0	0	0	4	12	0.058 8	17	0,091
3	1	1	0	0	0	9	0,181 8	5,5	0,25
4	0	1	0	0	2	8	0,090 9	11	0,25
5	2	0	0	0	0	5	0,285 7	3,5	0,4

3.3. Détermination des poids de l'indice de gravité pondéré

L'Indice de Gravité Pondérée, tel que son nom l'indique, repose sur l'utilisation de poids qui permettent de pondérer les diverses catégories de victimes de manière que les catégories les plus graves, telles que les décès, soient attribuées à la pondération la plus élevée, tandis que les catégories moins graves, telles que les blessés légers, se voient attribuer une pondération plus faible. Dans ce cadre, nous avons opté pour la méthode largement utilisée qu'est XGBoost, reconnue pour ses performances élevées et sa capacité de généralisation. Soulignons qu'en général, les arbres de décision offrent une fonctionnalité permettant de générer des scores qui quantifient la contribution des caractéristiques à la

construction des arbres de décision. Cette approche, connue sous le nom d'Importance de Caractéristiques, est employée dans cette étude pour déterminer les poids de l'indice de gravité pondérée.

Le modèle XGBoost évalue l'importance des caractéristiques en fonction de leur contribution à la performance globale du modèle. Cette approche permet d'identifier les caractéristiques les plus informatives ou influentes pour la prédiction. La méthode utilisée par le modèle XGBoost pour calculer cette importance repose sur le concept de gain, déterminé par la mesure de la réduction d'erreur de la fonction de perte. Pendant la construction des arbres de décision, XGBoost utilise une caractéristique spécifique pour partitionner les données. Le gain est alors calculé pour chaque partition de caractéristique en tant que réduction moyenne de la fonction de perte. La caractéristique générant le gain optimal est sélectionnée pour la partition des données. Une fois tous les arbres construits, les gains individuels sont agrégés pour chaque caractéristique, aboutissant au score final d'importance de la caractéristique.

La structure du modèle XGBoost est affichée dans la figure 28. À la suite de l'entraînement du modèle XGBoost avec les données structurées présentes dans le tableau 13, les mesures d'importance des caractéristiques sont obtenues en sortie. Selon la cible de notre modèle, à savoir RA, DH et TL, XGBoost génère 3 ensembles différents d'importance de caractéristiques. Le choix de l'ensemble à considérer ultérieurement dans cette étude est déterminé par celui qui accorde une importance

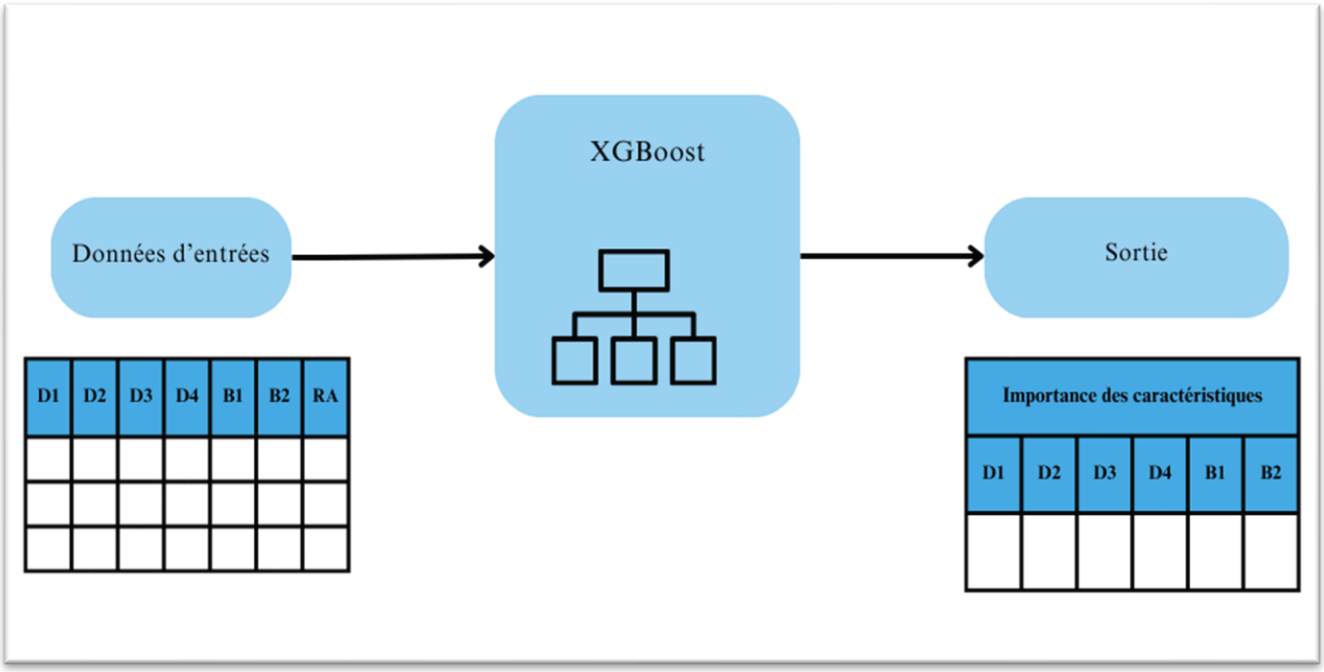


Figure 28. Modèle de génération de l'importance des caractéristiques

prépondérante aux événements graves tels que le décès et les blessures graves. Les valeurs ainsi sélectionnées seront utilisées comme des poids dans le calcul de l'indice de gravité pondéré.

3.4. Indice de gravité pondéré

Afin d'évaluer de manière précise la gravité des accidents de la circulation, nous avons opté pour l'utilisation de l'indice de gravité pondéré. Ce paramètre est calculé individuellement pour chaque accident, prenant en considération le nombre de décès et de blessés survenus. Il accorde une pondération différenciée aux décès survenus immédiatement par rapport à ceux survenus ultérieurement, et privilégie les blessés graves par rapport aux blessés légers. Pour accorder une importance significative aux décès sur le lieu de l'accident et aux blessés graves, l'indice de gravité pondéré intègre des poids soigneusement déterminés. Chaque catégorie de victime est ensuite pondérée en fonction du poids associé, et la somme de ces résultats délivre la valeur finale de l'indice de gravité pondéré. La formule générale est exprimée comme suit :

$$IGP = \sum_{i=1}^{n} P_i \times E_i \tag{38}$$

Où P_i est le poids de l'événement i, et E_i est l'événement i, avec $E_i = \{D1, D2, D3, D4, B1, B2\}$. Tel que, D1 désigne les tués sur le champ, D2 désigne les tués pendant le transfert à l'hôpital. D3 désigne les tués dans les 7 jours suivant l'accident, D4 désigne les tués dans les 30 jours suivant l'accident, B1 indique les blessés graves et B2 indique les blessés légers.

En prenant en compte les éléments de l'ensemble E_i, La formule de l'indice de gravité pondéré (5,1) devient :

$$IGP = \sum_{i=1}^{4} P_i \times D_i \; + \sum_{i=5}^{6} P_i \times B_{i-4} \tag{39}$$

4. Analyse de sections à haut risque d'accident

4.1. Classement des sections en fonction du nombre de victimes et d'accidents

Dans cette section, nous décrirons la répartition des points kilométriques en fonction du nombre de décès, de blessés graves, de blessés légers, de nombres d'accidents, ainsi que les trois métriques (TL,

DH, et RA). Certains PK se distinguent en tant que points présentant un risque plus élevé, avec des chiffres spécifiques de décès, de blessés graves et de blessés légers.

Le tableau 14 présente le classement des points kilométriques en fonction du nombre de morts, de blessés graves et de blessés légers. Les PK 1254, 787 et 420 se distinguent comme les plus meurtriers, avec respectivement 35, 14 et 11 décès (1414 décès au total), ce qui représente 2,47 %, 1 % et 0,78 % du total de décès. Les PK 245, 374 et 927 comptent le plus grand nombre de blessés graves, avec 21, 21 et 18 cas respectivement, constituant ainsi 0,96 %, 0,96 % et 0,82 % du total des blessés graves (2184 au total) sur l'ensemble de la route. En ce qui concerne les blessés légers, les PK 297, 244 et 253 sont les plus touchés, avec respectivement 192, 125 et 115 cas, représentant 1,43 %, 0,93 % et 0,85 % du total des blessures légères (13 462 au total) sur l'ensemble de la route.

Tableau 14. Classement des sections en fonction du nombre de tués, de blessés et d'accidents

PK	**Blessé grave**	**PK**	**Blessé léger**	**PK**	**Tués**	**PK**	**Nombre d'accidents**
245	21	297	192	1254	35	297	145
374	21	244	125	787	14	244	84
927	18	253	115	420	11	290	77
297	17	296	107	248	10	296	77
290	16	255	106	244	9	299	72
334	16	243	104	334	9	243	69
296	15	290	92	419	9	245	69
337	12	245	88	544	9	253	64
461	12	906	88	173	8	906	64
494	12	299	84	293	8	255	63
906	12	462	83	111	7	333	60
373	11	289	77	124	7	289	58
376	11	390	77	245	7	462	58
924	11	247	75	255	7	298	57
926	11	373	73	484	7	334	54
101	10	254	72	902	7	390	54
300	10	333	71	909	7	373	53
918	10	298	68	1122	7	288	49
1892	10	377	68	152	6	375	49
179	9	288	66	179	6	252	47

Le tableau 15 présente le classement des sections selon leur létalité, ordonné de manière décroissante. Les résultats du TL révèlent que la section 484 est la plus meurtrière, affichant une valeur de 14,29, Cela signifie qu'à ce point précis de la route, 14,29 % des personnes impliquées dans les accidents sont mortes de leurs blessures. Les sections 248, 638 et 334 présentent également une mortalité significative, avec des valeurs respectives de 13,89, 12,5 et 11,39. Des TL plus élevés indiquent une proportion plus importante de décès par rapport au nombre total de personnes touchées, ce qui peut indiquer des problèmes potentiels de sécurité routière à ces emplacements particuliers.

Les résultats du Dommage Humain (DH) indiquent que les sections 243, 297 et 333 ont entraîné le plus de pertes de vies humaines. Dans la section 243, le Dommage Humain est de 100, ce qui peut signifier qu'il y a un nombre relativement élevé de décès et de blessés par rapport au nombre total de personnes impliquées dans les accidents à ce point spécifique. Les sections 297 et 333 quant à elles affichent un DH de 100 et 76 respectivement. Ces valeurs suggèrent un niveau significatif de gravité des accidents à ces emplacements spécifiques, indiquant des préoccupations quant aux dommages humains associés à ces points kilométriques.

Les résultats de l'indice de risque d'accident, qui mesure la proportion de décès par rapport au nombre d'accidents, signalent que le point kilométrique 1254 présente le plus grand risque d'accident, avec une valeur de 11,67. Cela signifie qu'à ce point kilométrique particulier, le risque d'accident est relativement élevé, avec en moyenne 11,67 décès par accident. En d'autres termes, cela signifie qu'en moyenne, chaque accident survenant à cet emplacement entraîne environ 11,67 décès, un chiffre alarmant qui souligne la gravité de la situation. Les sections 787 et 1155, avec des valeurs de RA de 7 et 5 respectivement, présentent un risque d'accident inférieur à celui de la première section, mais néanmoins significatif.

En ce qui concerne la corrélation entre les conséquences morales et les indicateurs de gravité, nous observons que parmi les sections présentant le plus grand nombre de décès, le PK 1254 affiche un nombre de décès de 35, accompagné d'un risque d'accident de 11,57, lequel est le plus élevé. Le PK 787 suit avec 14 décès et un RA de 7. Le PK 420 enregistre 11 décès, présentant simultanément l'un des RA les plus élevés avec une valeur de 1,57. Le PK 248 affiche 10 décès ainsi qu'un TL remarquablement élevé, atteignant 13.89. Ainsi que le PK 484 présente un nombre de décès de 7, avec un TL le plus élevé, évalué à 14,29.

Tableau 15. Classement des sections selon DH, RA et TL

PK	DH	PK	RA	PK	TL
243	100	1254	11,67	484	14,29
297	100	787	7	248	13,89
333	76	1155	5	638	12,5
298	74	1637	4	334	11,39
288	72	1782	4	376	10,17
240	63	1675	3	241	8,7
249	62	510	2	257	8,51
296	62	696	2	378	8,16
291	57	1126	2	470	8
500	55	1417	2	927	8
242	53	1434	2	377	7,32
503	53	1653	2	466	7,27
299	47,5	1681	2	247	7,06
380	46	1741	2	390	6,7
502	46	1998	2	9	6,7
388	41	2057	2	295	6,52
335	39	2243	2	244	6,43
910	39	420	1,57	101	6,25
254	38	558	1,5	603	6,25
384	38	592	1,5	245	6,03
461	37,5	638	1,5	337	5,88
192	37	782	1,5	255	5,83
290	37	1060	1,5	906	5,66
432	35	1220	1,5	300	5,56
491	35	1498	1,5	374	5,33
198	34	1700	1,5	392	5,26
250	34	2312	1,5	256	5
387	34	1122	1,4	289	4,6
188	33	1171	1,33	252	4,35
439	33	1838	1,33	462	4,17

Parmi les sections comptabilisant un nombre substantiel de blessés, le PK 374 se distingue avec 21 Blessés Graves (BG) et concomitamment un TL parmi les plus élevés, affichant une valeur de 5,33. Le PK 243 présente un Dommage Humain plus élevé, évalué à 100, et recense 69 accidents mortels, ainsi qu'un nombre de Blessés Légers (BL) de 104. Le PK 297 présente un DH de 100, accompagné de 17 BG,

192 BL, et 145 accidents. Ainsi que le PK 333 affiche un DH de 76, 60 accidents mortels, et 71 blessés légers.

4.2. Résultats de l'importance des caractéristiques

La modélisation des caractéristiques des victimes et des trois taux de gravité (TL, DH et RA) a été effectuée à l'aide du modèle XGBoost. Les hyperparamètres ont été configurés afin d'optimiser la capacité du modèle à généraliser efficacement. Ce modèle prend en compte un ensemble de variables représentant les diverses caractéristiques des victimes en tant que données d'entrée, et génère en sortie un indice de gravité. Après le processus d'apprentissage, les importances de caractéristiques sont calculées pour chaque variable d'entrée (type de victime), reflétant leur « portée dans la construction de l'arbre de décision ». Trois modèles XGBoost sont employés pour modéliser les taux de gravité : TL, DH et RA, produisant ainsi trois ensembles de scores présentés dans le tableau 16. Il est à souligner que les scores sont relatifs aux indicateurs utilisés.

Selon les résultats d'importance de caractéristique du troisième modèle qui cible l'indicateur RA, les résultats concernant l'importance des caractéristiques indiquent que plus la durée écoulée entre l'accident et le décès est courte, plus le score attribué est élevé, indiquant une corrélation négative. En effet, une période plus courte entre l'accident et le décès est associée à un score attribué plus élevé, indiquant ainsi une gravité accrue de l'accident. Plus précisément, la variable de décès sur champs présente un poids de 1,219 9 qui plus grand au poids du décès pendant le transfert à l'hôpital qui est de 0,170 8. De manière similaire, le score attribué à la variable de blessures graves dépasse celui de la variable de blessures légères, avec un score de 0,328 pour les blessures graves, comparé à celui de 0,1934 pour les blessures légères. Par conséquent, les poids obtenus par le modèle basé sur le RA sont considérés comme les plus réalistes et les plus scientifiques. Ces poids seront ultérieurement appliqués dans le calcul de l'indice de gravité pondéré.

4.3. Classement des tronçons routiers selon l'indice de gravité pondéré

Après avoir établi les pondérations associées à chaque catégorie de victime, nous avons procédé au calcul de l'IGP pour chaque tronçon routier, en appliquant la somme des caractéristiques pondérées par les coefficients de pondération déterminés sur la base de l'indice de risque d'accident. Les résultats de nos calculs sont consignés dans le tableau 17. Notamment, les sections 1254, 297 et 244 se distinguent

comme les trois tronçons les plus critiques, affichant respectivement des valeurs d'IGP de 45,63, 42,92 et 33,97.

Tableau 16. Poids générés par les modèles XGBoost

Description (codification)	**TL**	**DH**	**RA**
Décès sur le champ (D1)	0,085 7	93,32	1,219 9
Décès pendant le transfert à l'hôpital (D2)	0,026 0	48,468	0,170 8
Décès dans les 7 jours suivant l'accident (D3)	0,017 1	125,957	0,104 3
Décès dans les 30 jours suivant l'accident (D4)	0,001 3	220,138	0,009 3
Blessure grave (B1)	0,007 8	79,569	0,328 0
Blessure légère (B2)	0,036 3	203,566	0,193 4

À ce propos, le graphique de la figure 29 représente la répartition des sections de la route par l'IGP. La figure révèle une tendance où le nombre de sections décroît en fonction de l'IGP. La courbe bleue initie son parcours à un point élevé sur l'axe des ordonnées, puis décline rapidement pour se rapprocher du bas de cet axe. Ce schéma traduit une diminution significative du nombre de sections à mesure que le IGP augmente, suggérant ainsi une forte corrélation négative entre ces deux variables. Ce constat met en évidence que les sections présentant un indice de gravité pondéré élevé sont moins nombreuses, tandis que celles affichant un IGP plus faible sont plus fréquentes. Cette observation établit que la formule utilisée pour calculer la gravité des sections ainsi que le choix des pondérations est précisé et appropriée. En effet, cette tendance confirme l'efficacité de la méthode employée pour évaluer la

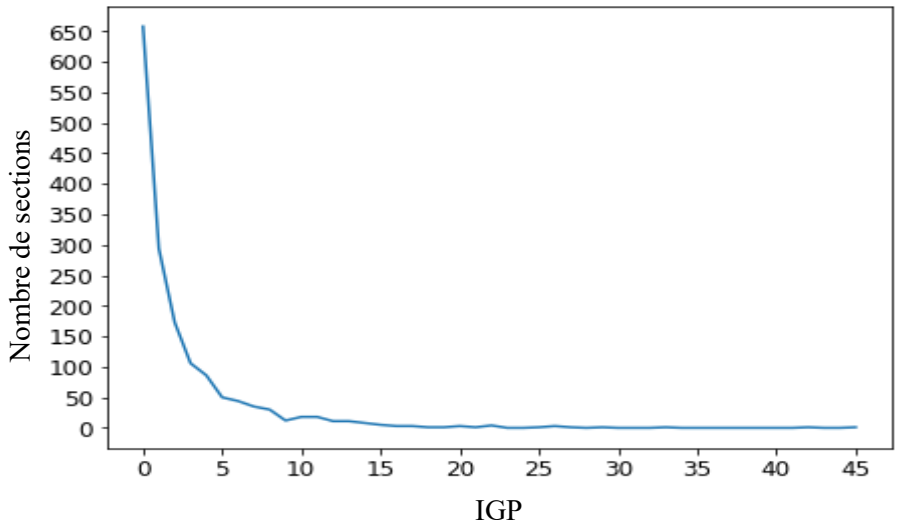


Figure 29. Répartition des sections de la route par IGP

gravité des sections routières, soulignant ainsi la pertinence des mesures prises pour prendre en compte l'impact des différents types de victimes sur la sécurité routière.

Tableau 17. Classement des sections selon l'IGP

PK	D1	D2	D3	D4	B1	B2	Nbr d'accidents	IGP
1254	35	0	0	0	6	5	3	45,631 5
297	0	0	2	0	17	192	145	42,917 4
244	6	3	0	0	6	125	84	33,974 8
245	4	3	0	0	21	88	69	29,299 2
255	3	3	1	0	7	106	63	27,072 8
296	1	0	0	1	15	107	77	26,843
253	1	2	2	0	7	115	64	26,307 1
906	4	0	2	0	12	88	64	26,043 4
290	2	0	1	0	16	92	77	25,584 9
390	4	1	1	0	7	77	54	22,342 5
243	1	0	0	0	3	104	69	22,317 5
247	5	1	0	0	4	75	44	22,087 3
377	5	1	0	0	8	68	46	22,045 5
334	4	3	2	0	16	54	54	21,292 2
289	3	0	1	0	6	77	58	20,623 8
299	1	1	0	0	9	84	72	20,588 3
374	3	0	1	0	21	50	47	20,322
462	0	2	2	0	9	83	58	19,554 4
248	4	6	0	0	3	59	45	18,299
373	0	0	0	0	11	73	53	17,726 2
293	8	0	0	0	6	30	19	17,529 2
461	1	0	1	0	12	61	43	17,057 6
420	9	2	0	0	6	17	7	16,576 5
333	1	0	0	0	4	71	60	16,263 3
927	4	0	0	0	18	28	39	16,198 8
544	8	1	0	0	6	21	8	15,959 4
254	1	0	1	0	2	72	42	15,905
376	3	3	0	0	11	42	47	15,902 9
288	1	0	0	0	5	66	49	15,624 3
375	0	0	3	0	9	63	49	15,449 1

4.4. Résultats du regroupement des points noirs

Après l'achèvement du calcul de l'IGP pour chaque segment routier, il devient impératif de déterminer quels segments présentent les niveaux de dangerosité les plus élevés en termes de décès et de blessures graves. Les segments de route affichant un IGP plus élevé seront désignés comme des « points noirs », tandis que ceux avec un IGP plus bas seront considérés comme sûrs et présentant moins de risques. Ces classifications sont détaillées dans le tableau 18, où sont présentés les seuils pour regrouper les segments routiers. Sur cette base, trois catégories de points noirs seront identifiées en utilisant la moyenne et l'écart-type du IGP [88].

Tableau 18. Classification des points noirs à partir de la moyenne et de l'écart-type de IGP

Niveau de gravité des accidents	IGP
Élevé	Moyenne + 2 E. T ≤ IGP
Moyen	Moyenne + 1,5 E.T ≤ IGP < Moyenne + 2 E. T
Faible	Moyenne + E.T ≤ IGP < Moyenne + 1,5 E.T

À ce titre, les résultats de la classification révèlent qu'un total de 74 sections routières sont identifiées comme présentant un niveau de gravité élevé, tandis que 34 sections sont classées comme ayant une gravité moyenne, et 65 sections comme ayant une gravité faible. Ces données sont synthétisées dans les tableaux 19, 20 et 21, qui fournissent des statistiques descriptives pour chaque niveau de gravité. Ainsi, nous remarquons que les sections classées avec un niveau de gravité élevé présentent des valeurs substantiellement plus élevées en termes de décès et de blessures graves par rapport aux niveaux inférieurs. Cette observation confirme la capacité de la méthode proposée à identifier de manière fiable les segments routiers les plus dangereux et les plus susceptibles d'être le théâtre d'accidents. Plus précisément, le niveau de gravité élevé est associé à un total de 335 décès, 4447 blessés graves, et 2803 accidents. Pour le niveau de gravité moyenne (respectivement faible), les chiffres sont de 280 (respectivement 223) décès, 2827 (respectivement 1694) blessés graves, et 1679 (respectivement 953) accidents.

Le tableau 17 présente le classement des 30 premiers segments routiers classés avec un niveau de gravité élevé. Ces sections sont jugées prioritaires et nécessitent une intervention immédiate. Une analyse détaillée a mis en évidence que les points kilométriques situés dans les plages suivantes : [240-249], [252-257], [288-291], [296-300], [373-378], [390-392], [460-462], et [905-909], sont particulièrement susceptibles de donner lieu à des accidents présentant un niveau de gravité élevé. Cette identification des

tronçons les plus à risque revêt une importance capitale pour orienter les efforts de sécurisation routière et allouer les ressources de manière stratégique afin de réduire les dangers potentiels sur ces tronçons spécifiques.

5. Discussion

Notre recherche démontre que l'utilisation de poids déterminés à partir de l'outil d'importance des caractéristiques de XGBoost, conjuguée à l'inclusion de différents types de décès supplémentaires, améliore la capacité d'identification des sections routières présentant un risque élevé d'accidents mortels. Alors que les études antérieures ont validé l'efficacité de la méthode IGP dans la détection des points noirs, l'attribution de poids n'a pas toujours été justifiée par les chercheurs et la diversité des types d'accidents mortels n'a pas été pleinement prise en considération. Ces constats ouvrent donc la voie à des analyses ultérieures visant à définir les interventions nécessaires pour accroître la sécurité sur ces tronçons routiers.

L'évaluation de l'importance des caractéristiques en nous basant sur le risque d'accident en tant que critère cible s'est avérée plus réaliste que celle qui repose sur les indicateurs taux de létalité et dommage humain. Cette réalité découle de la normalisation des données par la formule du RA en fonction du nombre total d'accidents, ce qui reflète de manière plus précise les conséquences potentielles des accidents. Le calcul de l'IGP attribue à chaque segment routier une valeur située entre 0 et 45,63. Les segments affichant des valeurs élevées d'IGP sont catégorisés comme points noirs, et ces derniers sont ensuite regroupés en trois niveaux : les points noirs de premier ordre. Ce groupe représente les sections routières à grand risque de létalité, nécessitant des interventions immédiates en termes d'amélioration de l'infrastructure, d'installation des signalisations pour informer les usagers de la route du danger du lieu. Par ailleurs, les points noirs de deuxième et troisième ordre appellent également à des améliorations, bien que dans un ordre de priorité moindre par rapport au premier niveau.

Nous avons également constaté que parmi les 74 sections identifiées comme les plus dangereuses, un total de 2803 accidents s'est produit, causant 335 décès, 522 blessures graves et 3925 blessures légères. Les points noirs classés de deuxième ordre ont été associés à 1679 accidents, dont : 280 ont été fatals, 411 ont entraîné des blessures graves, et 2416 ont résulté en des blessures légères. Enfin, les accidents classés de troisième ordre concernaient 953 incidents, avec 223 décès, 231 blessures graves et 1463 blessures légères. On observe que le nombre de décès, de blessures graves et d'accidents est décroissant allant du groupe de gravité élevé au groupe de gravité faible. Plus le risque de lieux est faible,

plus le nombre de morts et d'accidents mortels est faible. Ces résultats confirment la fiabilité de la méthode d'identification adoptée au sein de notre étude, ainsi que la méthode de regroupement des points noirs.

En outre, les résultats de notre analyse ont révélé l'existence de sections routières consécutives classées de premier ordre, où l'intervalle le plus notable est [240-249]. Le point kilométrique précédent cet intervalle est alors classé comme de second ordre. De manière similaire, l'intervalle de premier ordre [252-257] est précédé par le point kilométrique de second ordre PK 251. De plus, l'intervalle de second ordre [292-295] agit comme un élément intermédiaire entre les deux intervalles de premier ordre, à savoir [288-291] et [296-300]. Il est aussi à noter que l'intervalle de gravité élevée [373-378] est suivi de l'intervalle de gravité moyenne [379-383]. De même, l'intervalle de deuxième niveau [924-926] est suivi par l'intervalle de premier niveau [927-928]. La figure 30 montre un exemple d'une carte générée affichant les points noirs de premier ordre en rouge et ceux de deuxième ordre en orange. L'intégralité de la carte est consultable dans l'Annexe A. Enfin, en général, il a été observé que les points noirs sont fréquemment localisés à proximité des zones urbaines. Cette concentration pouvant être attribuée au fait que certaines sections traversent des agglomérations, des villages ou des zones agricoles, comportant parfois des intersections multiples et non réglementées. Cependant, d'autres facteurs, tels que les

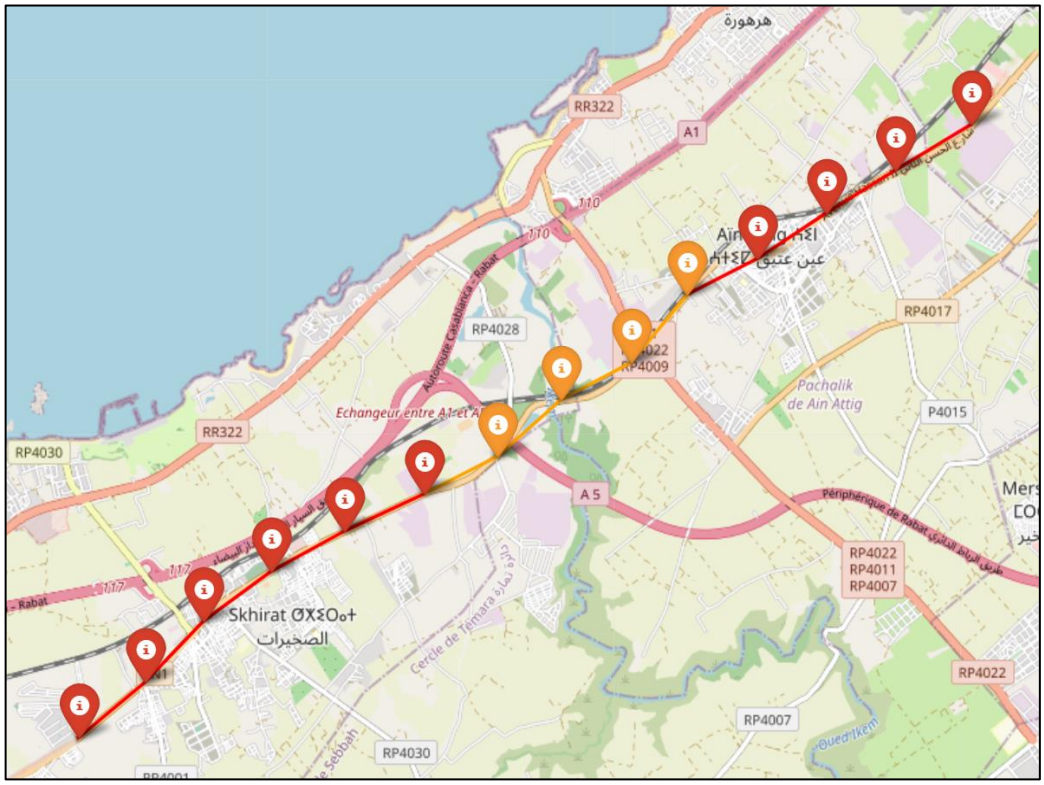


Figure 30. Carte générée affichant les points noirs

conditions humaines, météorologiques ou techniques, pourraient également contribuer à la survenue de ces accidents.

Tableau 19. Statistiques des points noirs de niveau de sévérité élevé

Niveau élevé	Tués	Blessés	D1	D2	D3	D4	B1	B2	Nbr d'accidents	TL	DH	RA	IGP
Comptage	335	4447	238	56	38	3	522	3925	2803	-	-	-	-
Moyenne	4,5	60,1	3,2	0,8	0,5	.04	7,1	53	37,9	1	23,3	0,4	16,6
E.T	4,6	32,4	4,5	1,1	0,8	0,2	4,4	30,6	23,4	0,1	23,9	1,6	6,8
Min	0	1	0	0	0	0	0	0	2	0	0	0	11,4
25 %	2	40,3	1	0	0	0	4	34	22,3	0,03	8,2	0,05	12,1
50 %	4	52,5	2	0	0	0	6	47	37	0,06	15	0,1	14,1
75 %	6	73	4	1	1	0	9	67,5	48,5	0,1	25,9	0,2	18,2
Max	35	209	35	6	3	1	21	192	145	0,9	100	11,7	45,6

Tableau 20. Statistiques des points noirs de niveau de sévérité moyen

Niveau moyen	Tués	Blessés	D1	D2	D3	D4	B1	B2	Nbr d'accidents	TL	DH	RA	IGP
Comptage	280	2827	207	39	33	2	411	2416	1679	-	-	-	-
Moyenne	2,83	28,56	2,1	0,4	0,3	0,02	4,15	24,4	16,9	0,12	13,5	0,27	8,7
E.T	1,78	10,9	1,6	0,7	0,6	0,14	2,5	10,7	8,1	0,12	12,1	0,32	1,2
Min	0	4	0	0	0	0	0	2	4	0	0	0	7
25 %	1	20,5	1	0	0	0	2	17	11,5	0	4,7	0,07	7,7
50 %	3	30	2	0	0	0	4	25	16	0,03	9	0,17	8,5
75 %	4	37	3	1	1	0	5,5	32	21	0,15	16	0,32	9,8
Max	7	54	6	3	3	1	12	52	39	0,56	46	1,5	11,1

Tableau 21. Statistiques des points noirs de niveau de sévérité faible

Niveau faible	Tués	Blessés	D1	D2	D3	D4	B1	B2	Nbr d'accidents	TL	DH	RA	IGP
Comptage	223	1694	173	22	27	1	231	1463	953	-	-	-	-
Moyenne	2,3	17,3	1,8	0,2	0,3	0,01	2,4	14,93	9,7	0,15	9,25	0,41	5,88
E.T	1. 3	7,4	1,1	0,6	0,6	0,1	1,9	7,47	4,7	0,15	7,67	0,63	0,58
Min	0	1	0	0	0	0	0	1	1	0	0	0	4,95
25 %	1	12	1	0	0	0	1	9,25	7	0,05	4	0,11	5,37
50 %	2	17	2	0	0	0	2	14	9	0,11	7,58	0,23	5,87
75 %	3	22	2	0	0	0	3	21,75	13	0,2	11,8	0,43	6,35
Max	6	34	4	3	2	1	10	34	23	0,8	29	5	6,95

6. Conclusion

Le Maroc est l'un des pays présentant chaque année un grand nombre d'accidents mortels, une problématique de plus en plus préoccupante en raison de ses conséquences causant des milliers de décès et de blessures graves. Dans cet environnement, la route nationale 1 occupe une position prépondérante au sein du réseau routier du pays du fait de son rôle crucial dans le processus de développement, en assurant la liaison entre les principales agglomérations. Cependant, son importance est entachée par son statut de l'une des voies les plus dangereuses, principalement en raison du nombre élevé d'accidents qui s'y produisent. L'objet de la présente étude était d'identifier les points noirs le long de la route N1 dans le but de réduire les accidents mortels ainsi que d'améliorer la sécurité routière sur cet axe. À cette fin, des métriques bien établies dans la littérature ont été utilisées pour analyser les données d'accidents sur une période de cinq ans, de 2013 à 2017. L'ensemble des données analysées comprend 9542 observations réparties sur 1584 sections routières. Pour ce faire, les sections routières ont été classées selon les indices de gravité : taux de létalité, dommage humain et risque d'accident.

La méthode que nous avons adoptée vise à examiner le nombre de victimes impliquées dans les accidents, en prenant en considération quatre types de décès : sur le lieu de l'accident, lors du transfert à l'hôpital, dans les sept jours suivant l'accident, ou dans les trente jours suivant l'accident. Les blessures sont également classées en deux niveaux : mineures ou graves. Par la suite, la méthode IGP a été appliquée pour identifier les points noirs et les classer en trois groupes en fonction du niveau de gravité. Trois modèles XGBoost ont été développés pour extraire les poids à utiliser lors du calcul du IGP, et il a été constaté que le RA est le plus approprié pour cette analyse. Nos résultats ont permis d'identifier 74 points noirs présentant un niveau de gravité élevé entre les points kilométriques 9 et 1254.

Cette approche pour calculer le IGP a prouvé son efficacité dans l'identification des points noirs. De plus, l'analyse des résultats de la classification des points noirs a démontré la fiabilité de l'approche proposée. Nul doute que ces conclusions peuvent guider les spécialistes de la sécurité routière dans l'élaboration de stratégies pour améliorer les sections routières étudiées, en priorisant les sections les plus dangereuses. Les résultats obtenus contribueront éventuellement à améliorer les processus de prise de décision visant à renforcer la sécurité routière dans le pays. En outre, la méthode proposée dans cette étude peut servir de base à d'autres recherches visant à améliorer la sécurité routière. Dans cette optique, nous encourageons les chercheurs futurs à adopter notre approche dans l'analyse du nombre de victimes en tenant compte des différents types de tués et de blessures.

Chapitre 6. Analyses des facteurs de risques et prédiction des points noirs des accidents sur les routes rurales

1. Introduction

Selon l'OMS, les accidents de la route constituent l'une des principales causes de décès dans le monde, entraînant des coûts importants en termes de vie et d'économie. Chaque jour, ces accidents font des milliers de morts et de blessés représentant 2 % de l'ensemble des décès. Selon les statistiques, les pays à revenu faible ou intermédiaire représentent la grande majorité des victimes de la route. Au Maroc, plus de 3000 personnes ont perdu la vie dans des accidents de la route chaque année au cours de la dernière décennie. Depuis 2008, ce taux d'accidents n'a cessé de croître, augmentant de plus de 38 % entre 2008 et 2017. Entre 2016 et 2017, le nombre d'accidents et de blessés a augmenté respectivement de 10,8 % et de 9,1 %, tandis que le nombre de morts est resté stable à environ 3500 [92]. En conséquence, le ministère chargé du transport du pays a établi une nouvelle stratégie en 2017 avec pour objectif de réduire le taux de mortalité à 50 % d'ici 2026. Cette stratégie consiste à renforcer l'ensemble du réseau routier, en particulier les sections les plus dangereuses. Celle-ci implique également d'enquêter sur les points noirs et de déterminer les causes fondamentales des accidents. Dans ce contexte, la motivation de la présente étude est d'être en conformité avec la stratégie du gouvernement en proposant un modèle précis pour l'identification des points noirs.

Face à l'accroissement des accidents de la route, le Maroc a été précurseur dans l'élaboration d'une stratégie nationale de sécurité routière dont l'objectif principal est de réduire le nombre de morts et de blessés graves de manière durable et continue. Il est à noter que jusqu'à la création de l'agence nationale de sécurité routière (NARSA) en 2018, le secteur de la sécurité routière du pays était géré par le Ministère de l'Équipement et de l'Eau. Compte tenu de l'importance des infrastructures routières et de leur rôle dans le secteur de la sécurité routière, le budget des infrastructures routières représente 15 % du total d'investissement public. En 2018, plus de 24 milliards de dirhams ont été investis dans les infrastructures de transport, les routes et autoroutes recevant 52 % des fonds. Selon les données de la même année, le réseau routier national comprend 1 800 km d'autoroutes et 57 334 km de routes, dont 44 180 km sont revêtus, soit 77 % de l'ensemble du réseau. L'indice de l'état de la chaussée (PCI) indique que 62,7 % du réseau routier serait dans un état acceptable ou bon. Malgré les investissements importants réalisés par le pays en matière de sécurité routière, le nombre d'accidents mortels continue d'augmenter,

entraînant d'importantes pertes humaines et économiques pour les victimes, leurs familles et le pays dans son ensemble.

Selon les statistiques du MEE, les accidents mortels sont enregistrés dans de bonnes conditions routières et météorologiques, c'est-à-dire des routes bien revêtues, des surfaces sèches avec une largeur comprise entre 6 et 10,5 mètres, et dans des conditions météorologiques normales. Des contre-mesures et des stratégies de prévention peuvent aussi être développées en étudiant les facteurs qui affectent directement les accidents mortels [93]. Au Maroc, comme ailleurs dans le monde, les accidents de la route tuent de plus en plus de personnes et sont principalement dus à la combinaison de trois facteurs : l'homme, la route et le véhicule. Les accidents de la circulation constituent un défi majeur dans la gestion du trafic routier, se produisant fréquemment dans les zones encombrées où les conducteurs essaient de maintenir une vitesse élevée pour éviter les embouteillages [94]. Outre les facteurs liés au comportement du conducteur ou à l'état du véhicule, les caractéristiques spatiales des routes sont considérées comme un élément clé à étudier afin de prévenir les incidents ou les embouteillages. Aussi, la prise en considération de l'aspect sécurité routière dans la conception et l'amélioration des infrastructures devient de plus en plus une priorité pour les acteurs de ce secteur. Face à ce constat, de nombreuses études ont montré que la composition spatiale de certaines sections routières peut avoir un lien étroit avec le type et le nombre d'accidents qui s'y produisent. Par conséquent, la combinaison de plusieurs caractéristiques peut contribuer à l'occurrence de certains accidents mortels de manière récurrente dans ces sections, souvent appelées « points noirs d'accidents ».

Un point noir est défini en tant que segment de route qui a connu un certain nombre d'accidents pendant une période donnée « t » et sur une distance donnée « d ». Au Maroc, une section de route d'un kilomètre de long est classée à haut risque si elle a connu au moins 10 accidents causant au moins 10 morts ou blessés graves. La mise en œuvre de mesures de sécurité sur ces sections est donc rentable, mais reste limitée en raison du manque de budget. Il convient donc d'accorder la priorité aux tronçons à haut risque lors de la planification des mesures de sécurité, en tenant compte des restrictions budgétaires. À ce propos, précisons que de nombreux pays prennent les mesures nécessaires pour réduire le nombre de points noirs, compte tenu de leur coût socio-économique, notamment en termes de pertes de vies humaines.

Il existe trois méthodes principales pour identifier les points noirs : les méthodes de filtrage, les méthodes de regroupement et les méthodes de prédiction des accidents [95] :

Une méthode de filtrage est une procédure d'examen d'un réseau routier visant à identifier et à classer les emplacements potentiels du plus probable au moins probable afin de réduire la fréquence des accidents sur la route. Elle se compose de nombreuses méthodes statistiques telles que la méthode des taux d'accidents [96][97], la méthode de la fréquence des accidents [98][99] ou la méthode du nombre d'accidents.

Quant aux méthodes de regroupement, celles-ci consistent à subdiviser un ensemble de données en plusieurs groupes, appelés clusters, de sorte que les éléments de chaque groupe aient des propriétés communes. Les clusters sont simplement des groupes de points de données disposés de sorte que la distance entre eux soit la plus courte possible. Ce type de recherche permet à la fois de détecter les points noirs et d'analyser les causes des accidents.

En ce qui concerne les approches prédictives, plusieurs techniques statistiques telles que l'apprentissage automatique, la modélisation prédictive et l'exploration de données sont explorées. Les approches prédictives permettent d'estimer ou de prévoir les résultats futurs à l'aide de données passées et présentes [94]. Les approches prédictives simples rencontrent toute fois des défis tels que l'augmentation de la variance des prédictions et des biais importants. Pour surmonter ces problèmes, les méthodes d'ensemble dans l'apprentissage automatique se présentent comme des systèmes multimodèles qui fusionnent plusieurs modèles en un modèle prédictif.

D'une manière générale, un modèle d'ensemble peut être classé en deux types : les méthodes séquentielles et les méthodes parallèles. Dans un ensemble parallèle, chaque prédicteur de base individuel est formé indépendamment des autres ; il est généré dans un ordre parallèle sans dépendance des données. Contrairement aux méthodes parallèles, qui tirent parti de l'indépendance de chaque prédicteur de base, les méthodes séquentielles tirent parti de la dépendance des prédicteurs de base. Plus précisément, les ensembles séquentiels forment un nouveau prédicteur de base au cours du processus d'apprentissage de manière à minimiser les erreurs causées par le prédicteur de base formé à l'étape précédente.

Dans la littérature, de nombreux travaux étudient l'impact des conditions météorologiques et de l'état des véhicules, ainsi que du comportement des usagers de la route, sur les points noirs des accidents. Cependant, à notre connaissance, aucune recherche sur l'utilisation de l'état des routes et des facteurs environnementaux dans la classification des points noirs d'accidents n'a été menée. Dans ce contexte, notre étude se concentre sur l'identification des facteurs de risque, et puis sur la classification des points noirs des accidents sur les routes rurales marocaines, à travers une analyse des données d'accidents

survenus en 2016 et 2017. Deux types de décès sont considérés : immédiat et suite aux blessures, ainsi que deux types de blessures : graves et légères. Quant à la partie classification des points noirs des accidents, celle-ci prend en compte des informations sur la localisation des accidents, telles que l'infrastructure routière et l'environnement. En effet, à notre connaissance, bien que certaines études aient été menées sur l'analyse des accidents dans le pays, aucune étude n'a été trouvée pour introduire une classification des points noirs des accidents en utilisant des données sur l'infrastructure et l'environnement. Cette constatation constitue la principale motivation de ce travail visant à proposer un classifieur des points noirs en fonction de leurs caractéristiques infrastructurelles.

2. Approches existantes d'identification et d'analyse des points noirs

Dans le domaine de la sécurité routière, plusieurs études ont porté sur les accidents de la route, et plusieurs d'entre elles ont évalué l'efficacité du réseau bayésien empirique pour la classification des points noirs en raison de leur flexibilité à classer des ensembles de données d'accidents sur la base d'informations de données antérieures. Pour leur part, Zhang *et al.* [100] ont proposé un réseau bayésien pour construire un modèle d'identification des points noirs, leur modèle de classification binaire montre la meilleure précision parmi les modèles examinés. En outre, les résultats de l'étude de corrélation montrent que le type de lieu de l'accident, le type d'accident, l'heure et la responsabilité ont un effet significatif sur les points noirs des accidents. Cependant, ces auteurs ont ignoré les variables spatiales qui pourraient avoir un impact significatif. Dans [101], la technique bayésienne empirique a été utilisée pour classer les sections collectées en fonction de leur degré de gravité. La technique des règles de décision a été utilisée pour déterminer les principaux attributs qui conduisent à des accidents dans des sections spécifiques. Les résultats montrent que les caractéristiques géométriques des sections de route ont un effet significatif sur la gravité de l'accident.

L'Indice de Gravité Pondéré ou la Pondération des Points d'Accident (PPA) est l'une des stratégies basées sur les statistiques les plus souvent utilisées pour identifier les points noirs en matière d'accidents. La procédure consistant à pondérer les incidents corporels pour donner plus de poids aux plus graves. Les chercheurs ont proposé de nombreuses stratégies de pondération pour le calcul de l'IGP. Toutefois, les arguments en faveur de la stratégie de pondération restent peu clairs et peu convaincants. Un nombre important et croissant de recherches sur les points noirs des accidents s'est concentré sur l'identification et l'évaluation de l'Indice de Gravité Pondéré. À ce propos, Iqbal *et al.* [102] ont utilisé l'PPA pour identifier les points noirs, qui prennent en compte les accidents mortels, majeurs, mineurs et

les dommages matériels. La principale conclusion de leur étude est que les erreurs humaines et les dysfonctionnements des véhicules, ainsi que les facteurs environnementaux, constituent les principaux responsables des accidents de la route. Calsavara *et al.* [103] ont utilisé l'IGP pour identifier les points noirs sur une route comportant plusieurs virages en S et fréquemment du brouillard. Les résultats ont montré que l'identification était fiable, contribuant ainsi à la sécurité des sections identifiées.

Par ailleurs, de nombreux chercheurs ont accordé une attention particulière à la distribution de Poisson-Tweedie dans l'analyse des accidents [104]. Ainsi, Debrabant *et al.* [105], ont utilisé les données historiques des hôpitaux pour identifier les points noirs. Ils utilisent un modèle autorégressif de Poisson-Tweedie pour modéliser le nombre attendu d'accidents, qui tient compte à la fois de l'inflation zéro et de la surdispersion et couvre un large éventail de distributions discrètes. Par ailleurs, les données utilisées manquaient d'informations crédibles sur les facteurs de risque locaux tels que la géométrie. De leur côté, Nguyen *et al.* [106] ont proposé un cadre basé sur les économies potentielles en coûts d'accidents ou en potentiel de sécurité (SAPO) en procédant à un test de Poisson pour les zones sélectionnées présentant une fréquence d'accidents élevée, puis en calculant le SAPO et classent les points noirs en fonction du SAPO.

Diverses approches alternatives ont été employées dans l'analyse et la hiérarchisation des points noirs en matière de sécurité routière, se basant sur l'utilisation du SIG. En ce sens, Vyas *et al.* [89] ont opté pour l'intégration des SIG dans leur étude visant à analyser et prioriser les points noirs détectés. De manière similaire, Zhu *et al.* [107] ont adopté une approche combinant un modèle mathématique et un modèle d'apprentissage automatique pour examiner la corrélation entre les accidents de la circulation et les caractéristiques spatiales des routes. Ces chercheurs ont élaboré une méthodologie reposant sur l'utilisation d'un SIG pour identifier de manière précise les points noirs associés aux accidents.

En outre, l'infrastructure et l'environnement routiers ont été étudiés par le biais de diverses approches afin d'extraire de nouvelles connaissances dans des cas spécifiques. Ainsi, l'étude de Babić *et al.* [108] a examiné les études précédentes sur les effets des marquages routiers sur le comportement des conducteurs. Ces auteurs ont constaté que les marquages larges réduisent les accidents et ont un effet positif sur la sécurité routière, en particulier lorsqu'ils sont bien entretenus et qu'ils ont une rétroréflectivité plus élevée. Une autre étude a montré que l'infrastructure routière améliore la sécurité routière et le comportement des usagers de la route [109]. Celle-ci a conclu que les avantages des lois sur le port du casque sont renforcés par l'amélioration de l'infrastructure routière. Ces lois étant associées

à une augmentation du port du casque chez les cyclistes et à une diminution des blessures à la tête et des décès chez les motocyclistes et les cyclistes. Dans une étude réalisée par Ambros *et al.* [110], l'impact de la largeur des accotements sur la fréquence des accidents a fait l'objet d'une analyse qui a établi que la largeur des accotements est associée à la fréquence des accidents sur les routes rurales.

Parallèlement, d'autres études se sont concentrées sur les accidents impliquant un seul véhicule, comme celle de Liu & Xia [111], qui ont étudié les accidents mortels impliquant un seul véhicule dans les zones métropolitaines. L'étude menée par Singh & Tiwari [112] a évalué l'impact des caractéristiques des bordures de route sur les accidents mortels sur les routes de campagne. En outre, certaines recherches ont analysé le comportement des conducteurs dans des situations spécifiques. C'est le cas de l'étude de Van Treese *et al.* [113] qui a étudié l'impact de la présence d'arbres le long des routes sur le comportement et la psychologie des conducteurs. L'étude de Ben-Bassat & Shinar [114] a analysé la perception et le comportement des conducteurs dans différentes situations d'accotement, de glissières de sécurité et de géométrie de la chaussée. Calvi *et al.* [115] ont également analysé le comportement des conducteurs sur des routes de type 2 +1. Les résultats montrent que le type de médian a un effet significatif sur la distance entre le véhicule et le médian, cette distance étant plus grande lorsque le médian est une barrière en câble.

Une autre recherche [116] a évalué l'impact de l'environnement de conduite sur les accidents impliquant des piétons ; les résultats ont montré que les facteurs de risque liés à l'environnement de conduite augmentent de manière significative la probabilité d'une blessure mortelle dans un accident impliquant un piéton. L'étude menée par Gherghina *et al.* [117], examine la relation entre l'infrastructure de transport routier et le PIB. Ces chercheurs ont établi que les infrastructures influencent positivement le PIB, et les investissements dans les infrastructures de transport ont un impact favorable sur la croissance économique.

L'apprentissage automatique est l'une des méthodes utilisées pour analyser la relation entre l'infrastructure routière et les accidents de la route. De cette façon, l'étude d'Ashraf *et al.* [118] a analysé les accidents de circulation à contresens à l'aide de méthodes d'apprentissage automatique et d'une technique d'apprentissage automatique interprétable. Par ailleurs, l'étude de Siregar *et al.* [119] a analysé l'effet de la géométrie des routes sur les accidents mortels en zones rurales. Les résultats ont montré que les caractéristiques des routes sont importantes pour expliquer les variations des accidents de la route. Enfin, la recherche menée par Siregar *et al.* [120] a montré que les caractéristiques géométriques des routes avaient des effets indirects sur le nombre d'accidents mortels, mais avaient un effet direct sur la

vitesse et les écarts de vitesse. En fait, une augmentation de la géométrie de la route entraîne une diminution de la vitesse.

Dans le domaine de la sécurité routière, les règles d'association sont des méthodes courantes et bien étudiées dans l'exploration de données. La plupart de ces méthodes se concentrant sur l'identification des facteurs qui affectent la gravité d'un accident. Les études [121][122] ont analysé les accidents mortels et ont démontré que l'infrastructure et l'environnement ont un impact sur les accidents. Ainsi, l'étude de Xu *et al.* [123] a examiné les facteurs de risque liés au véhicule, au conducteur, à la géométrie et à l'environnement routier dans les accidents mortels ayant fait plus de 10 morts en utilisant des règles d'association et des statistiques descriptives. Par ailleurs, l'étude menée par Das *et al.* [124] examine le comportement de conduite dans des conditions de brouillard, en particulier les compétences en matière de maintien de la voie. Ces chercheurs ont montré que l'infrastructure affecte négativement le comportement des conducteurs. Bien que l'infrastructure des intersections routières n'ait pas été examinée, Shahin *et al.* [125] ont étudié les accidents aux intersections et ont conclu que la construction de ralentisseurs peut minimiser le nombre d'accidents graves.

En raison des différentes méthodes et algorithmes proposés dans la littérature, l'extraction de connaissances à l'aide de techniques de règles d'association devient de plus en plus difficile et prend de plus en plus de temps. Pour surmonter ces difficultés, de nombreuses approches ont été proposées pour filtrer les règles d'association générées. Dans cette approche, l'étude menée par Zeinab *et al.* [122] a opté pour les méthodes Electre Tri et Prométhée pour sélectionner les règles intéressantes et pertinentes parmi le grand nombre de règles générées. De son côté, l'étude menée par Gu *et al.* [121] a utilisé les tests Lift Increase pour éliminer les règles redondantes. L'approche par Essai et Erreur consistant à itérer continuellement la combinaison des probabilités de soutien et de confiance et à déterminer un ensemble de seuils relativement élevés pour garantir la crédibilité des règles d'association. L'étude menée par Xu *et al.* [123] a choisi les seuils de soutien, de confiance et d'élévation sur la base d'études antérieures. Ils ont utilisé des valeurs de seuil relativement élevées dans le but d'augmenter la précision et la robustesse des règles d'association générées. Enfin, l'étude de Das *et al.* [124] s'est servi du Lift Increase Threshold pour sélectionner les règles dont les antécédents contribuent de manière significative à la robustesse de la règle. Ensuite, les règles avec plus d'antécédents ont été sélectionnées par rapport aux règles plus simples sur la base du critère d'augmentation de l'ascendance.

Après avoir examiné l'état de l'art en matière d'analyse des données d'accidents, force est de constater que les travaux antérieurs se concentraient sur la combinaison de plusieurs éléments, tels que les aspects liés au conducteur, au piéton et aux conditions météorologiques. Cependant, les variables liées à l'infrastructure et à l'environnement routiers ont reçu peu d'attention. Aussi, ce constat nous à encourage à effectuer une analyse plus approfondie en incluant des caractéristiques supplémentaires telles que les accotements gauche et droit, ainsi que leur largeur et leur position par rapport à la chaussée. Dans ce contexte, nous proposons une approche analytique pour extraire des informations importantes sur les caractéristiques des points noirs afin de prévenir les catastrophes routières. Notre approche étant basée sur l'Extraction de Règles d'Association (ERA), une technique d'extraction de données bien connue qui a fait ses preuves dans plusieurs domaines, y compris la sécurité routière. Cette méthode permettant d'identifier les corrélations entre les différents facteurs accidentogènes génère des règles sur les caractéristiques communes des points noirs. Les règles générées sont de ce fait faciles à interpréter et à comprendre, ce qui n'est pas le cas des autres méthodes d'intelligence artificielle.

3. Méthodologie de recherche

La méthodologie de recherche adoptée dans ce chapitre offre une analyse approfondie du risque d'accident routier, s'articulant autour d'une rigueur méthodologique (voir figure 31). Les données initiales proviennent de diverses sources, notamment les registres d'accidents, de véhicules et de piétons du Ministère de l'Équipement et de l'Eau. Un processus attentif de nettoyage et de préparation des données est entrepris, englobant l'élimination des doublons, la correction des erreurs et une mise en forme homogène. Afin de faciliter une analyse approfondie, une consolidation des différentes tables de données a été réalisée. Ainsi, la complexité des données a été réduite par une sélection judicieuse des variables pertinentes. De même, la segmentation des routes en segments de 1 kilomètre accroît la précision de l'analyse. Le calcul du risque d'accident pour chaque section permet d'identifier les zones présentant le plus grand danger. L'évaluation de l'importance des caractéristiques des variables sera suivie de l'identification de groupes de sections similaires. Enfin, l'élimination des données redondantes améliore la précision, tandis que la génération de sous-ensembles Bootstrap équilibre les données pour une robustesse améliorée. Pour ce faire, l'algorithme Apriori est déployé pour extraire des règles d'association de chaque sous-ensemble, et l'agrégation de ces ensembles de règles d'association est effectuée pour accroître la stabilité. Précisons que la sélection finale des règles est basée sur les niveaux de soutien, de confiance et de levée. Les résultats obtenus sont ensuite interprétés afin d'identifier les principaux contributeurs au risque d'accident.

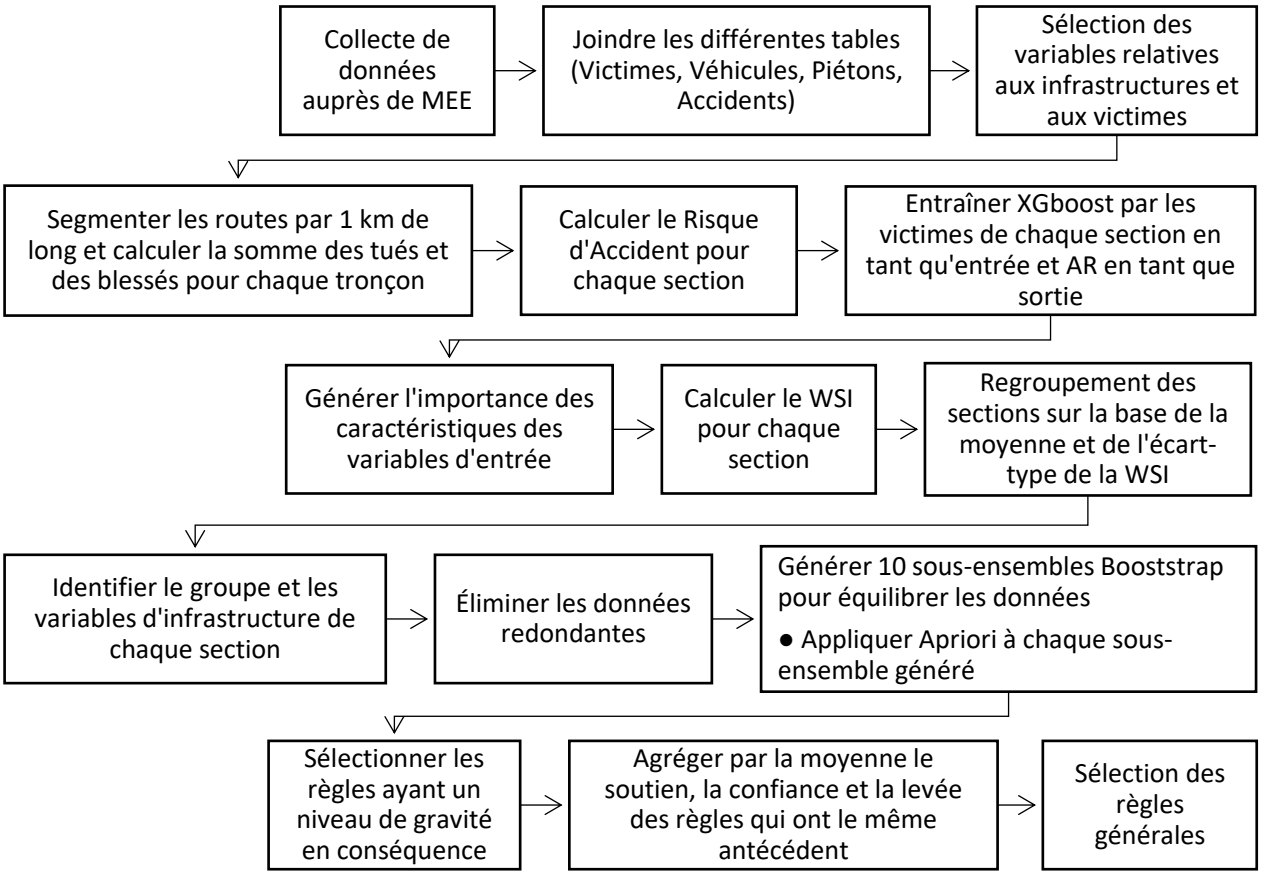


Figure 31. Méthodologie de recherche

4. Méthodes

4.1. Prétraitement de données

Les données utilisées dans notre étude proviennent de la DR, rattachée au ministère de l'Équipement et de l'Eau, du Maroc. Cette direction collectant un grand nombre de données sur les accidents de la circulation sur l'ensemble du réseau routier. Les ensembles de données étudiés sont constitués de nombreux fichiers distincts qui fournissent des statistiques sur les incidents de circulation. Chaque fichier contient des informations sur les accidents, ainsi que sur les conducteurs, les passagers et les piétons impliqués. Les métadonnées des fichiers regroupent des informations temporelles, spatiales, sur les véhicules et les usagers de la route.

Cette analyse menée spécifiquement sur la base des données d'accidents pour la période 2016 – 2017, porte sur les accidents mortels survenus en zones rurales. Notamment, les accidents véhicule-véhicule, à un seul véhicule et véhicule-piéton. Les composantes de notre recherche consistent en une

enquête de santé sur les victimes et un examen sur place de la section routière qui est effectué par des personnels qualifiés. L'enquête sur la santé a été menée par le biais d'entretiens personnels, et l'examen de la section routière effectué par des mesures directes du champ de l'accident.

In fine, les données doivent être traitées afin d'en tirer des informations. Pour commencer, les tables individuelles ont été concaténées et fusionnées en une seule qui contient toutes les propriétés pertinentes pour l'étude. Plus précisément, pour chaque accident, des variables spatiales sont collectées, telles que le numéro de la route et le point kilométrique, ainsi que des informations sur l'état de la route et l'environnement. De même, des variables statistiques telles que le nombre de morts et de blessés ont été collectées. Des enquêtes telles que celle menée par [126][127] ont mis en évidence que les méthodes qui appliquent la fréquence des accidents graves et mortels sont plus performantes que celles qui appliquent la fréquence totale des accidents. Celles-ci ont mis l'accent sur le fait que la gravité des accidents est le point clé du processus d'évaluation des points noirs des accidents.

Les statistiques sur les décès à la suite d'un accident ont été divisées en quatre catégories en fonction du moment du décès, à savoir les décès sur les lieux de l'accident, pendant le transfert à l'hôpital, dans les 7 jours ou dans les 30 jours suivant l'accident. Les blessures sont divisées en deux groupes selon leur gravité : les blessures légères et les blessures graves. L'étude a aussi pris en compte 30 576 observations, soit 3 604 décès et 54 359 blessés. Par la suite, les données ont été regroupées par numéro de route afin de pouvoir les analyser indépendamment les unes des autres. Finalement, 652 routes différentes ont été répertoriées. Puis, chaque route a été divisée en segments équidistants d'un kilomètre de long. Le choix d'une distance fixe s'est avéré efficace sur les routes à grande vitesse [95]. L'un des avantages de l'analyse à distance fixe pour les routes à grande vitesse étant qu'elle évite le problème des faux positifs (c'est-à-dire des sections sûres incluses dans l'étude des points noirs) ou des faux négatifs (c'est-à-dire des sections de route non incluses dans l'étude des points noirs alors qu'elles sont requises). Enfin, les facteurs pertinents et leur description sont présentés dans le tableau 22.

L'indicateur de risque d'accident, qui est la variable de sortie, est ensuite ajouté pour quantifier la gravité en termes de pertes de vies humaines. Il est destiné à calculer la fréquence des décès par rapport au nombre d'accidents dans chaque section. La formule de RA est la suivante :

$$RA = \frac{D}{A} \tag{40}$$

Où D représente le nombre de morts et A le nombre d'accidents sur la section. Les sections routières présentant un RA élevé sont considérées comme plus dangereuses en termes de décès. Ce qui signifie qu'il y a beaucoup plus de décès par accident sur ces sections que sur d'autres sections du même type.

Tableau 22. Description des facteurs étudiés

Facteurs	**Description**
Section routière	Point kilométrique
État de chaussée	Bon
	Mauvais
	Très mauvais
	Non revêtue
État de surface	Sec
	Mouillé
	Enneigée ou glacée
	Humide, salée ou sablée
Largeur de chaussée (RP)	[0,2]
	] 2,5]
	] 5,8]
	] 8,12]
	] 12,90]
Type de chaussée	Unique
	Séparé par médian
Largeur de l'accotement droit (RS) (resp. largeur de l'accotement gauche [LS])	[0 ; 0,5]
	] 0,5 ; 1]
	] 1 ; 1,5]
	] 1,5 ; 4]
	] 4 ; 10]
Position de l'accotement droit (resp. position de l'accotement gauche)	Au niveau de la chaussée
	Dénivelé par rapport à la chaussée
Topographie	Pont
	Rétrécissement

	Obstacle
Profils en long	Plat
	Pente
	Sommet de cote
Tracé plan	Partie rectiligne
	En courbe à gauche
	En courbe à droite
	En S

4.2. Indice de gravité pondéré

Pour évaluer le risque d'accident lié aux points noirs, l'Indice de Gravité Pondéré a été utilisé, tel que révisé en 2021 [128]. Le IGP a été conçu comme une représentation simple du risque de points noirs. Le principe de l'IGP est le suivant : si l'indice IGP est inférieur à un certain seuil, le risque est relativement faible et l'endroit peut donc être considéré comme sûr. En revanche, si le IGP dépasse un seuil spécifique, le risque est fortement accru et l'endroit peut être classé comme excessivement dangereux. L'IGP a été calculé sur la base des décès immédiats, des décès suite aux blessures, des blessures légères et des blessures graves. Quatre catégories de risque de points noirs ont été déterminées, comme le montre le tableau 23.

Le modèle XGBoost a par ailleurs été utilisé pour modéliser les caractéristiques des victimes et le RA en utilisant les quatre types de victimes comme données d'entrée et l'indice de RA comme données de sortie. Après le processus d'apprentissage, le score de chaque type de victime a été calculé à l'aide de l'outil d'évaluation de l'importance des caractéristiques. Le score de chaque caractéristique indique son importance dans la construction des arbres. Ainsi, le modèle XGBoost simule la gravité du taux d'accidents. Ce qui permet d'obtenir un ensemble de poids à utiliser dans la formule de l'IGP :

$$IGP(j) = \sum_{i=1}^{4} W_i \times A_i \tag{41}$$

Où $IGP(j)$ est le IGP du j^{ème} section, W_i est le poids du i^{ème} type de victime, et A_i est le i^{ème} type de victime.

Pour identifier le niveau de gravité des points noirs, la moyenne et l'écart-type du IGP ont été calculés. Les quatre catégories des points noirs d'accidents ont été déterminées à l'aide des valeurs du IGP, de la moyenne et de l'écart-type, comme le montre le tableau 23. Le concept est simple : si la valeur IGP est inférieure à une valeur spécifique, le nombre attendu d'accidents mortels à cet endroit est relativement faible, et l'endroit peut donc être considéré comme sûr. En revanche, si le IGP dépasse un seuil spécifique, la probabilité d'accidents mortels sur le site augmente considérablement et le site peut être considéré comme excessivement risqué [129].

Tableau 23. Classification des points noirs sur la base de la moyenne et de l'écart-type du IGP

Niveau de gravité des sections routières	**IGP**
Élevé	Moyenne + 2 É.T ≤ IGP
Moyen	Moyenne + 1,5 É.T ≤ IGP < Moyenne + 2 É.T
Faible	Moyenne + É.T ≤ IGP < Moyenne + 1,5 É.T
Sûr	IGP < Moyenne + É.T

4.3. Extraction de règles d'association

L'extraction de règles d'association a été introduite pour la première fois par [130]. Elle se compose de deux étapes principales. La première consiste à trouver tous les ensembles fréquents qui satisfont au support minimum. La deuxième étape nécessite de générer des règles d'association à partir des ensembles fréquents trouvés à l'étape précédente. Une règle d'association est de la forme A → B, où A ∩ B = Ø, « A » représente l'antécédent et « B » le conséquent. Précisons que l'algorithme Apriori est l'algorithme d'extraction de règles d'association le plus connu qui appartient à l'apprentissage automatique non supervisé, qui quantifie la corrélation des occurrences les plus fréquentes d'éléments avec un soutien, une confiance, une levée et une conviction, montrant que chacun se produit fréquemment au sein d'un groupe donné [131]. Cet algorithme a été conçu pour fonctionner dans cette étude sur un ensemble de données contenant des caractéristiques spatiales des points noirs des accidents.

Soit « D » un ensemble de données sur les accidents comprenant n sections routières. Soit « C » un ensemble de caractéristiques routières $C = \{c_1, c_2, \dots, c_m\}$ et $S = \{\text{Élevé}, Moyen, Faible, \text{Sûr}\}$ un ensemble de niveaux de gravité des sections routières, soit « A » un ensemble de caractéristiques où A ⊂ C et « B » un élément des niveaux de gravité des points noirs où B ⊂ S et A ∩ B = Ø. Trois mesures ont été utilisées pour évaluer les règles d'association générées :

Le support de la règle (A → B) est la probabilité que les deux phénomènes soient plus susceptibles de se produire ensemble, où A est l'antécédent et B est le conséquent. Le support est calculé comme suit :

$$Support\ (A\ \rightarrow\ B) =\ P(A\ \cap\ B) = \frac{Nbr\ de\ sections\ comportant\ A\ et\ B}{Nombre\ de\ sections} \qquad (42)$$

La confiance pour la règle (A → B) est la probabilité conditionnelle d'avoir un niveau de gravité B étant donné qu'une section routière a un élément A. La confiance est de 1 pour une règle (A→B) si le conséquent et l'antécédent se produisent toujours ensemble. Elle se calcule comme suit :

$$Confiance\ (A \rightarrow B) = \frac{P(A\ \cap\ B)}{Support(A)} \qquad (43)$$

La levée de la règle (A → B) est la confiance de la règle divisée par le support du conséquent B. Une valeur de levée, de deux variables, plus élevée signifie une plus grande corrélation entre eux. Notons que levée > 1 signifie que l'antécédent et le conséquent sont plus susceptibles d'apparaître ensemble, tandis que levée < 1 signifie qu'ils sont plus susceptibles d'apparaître séparément. Enfin, levée = 1 signifie qu'il n'y a pas d'association entre l'antécédent et le conséquent. La levée est déterminée par la formule suivante :

$$\text{Levée}\ (A\ \rightarrow\ B) = \frac{P(A\ \cap\ B)}{Support(A)\ \times\ Support(B)} = \frac{Confiance\ (A\ \rightarrow\ B)}{Support\ (B)} \qquad (44)$$

4.4. Méthode de rééchantillonnage Bootstrap

Les données obtenues à l'étape précédente présentaient une distribution déséquilibrée, avec la classe de niveau faible largement dominante par rapport aux autres classes supérieures. Ce résultat était attendu étant donné que les sections considérées comme « sûres » sont plus nombreuses que celles considérées comme risquées. Pour résoudre ce problème, la méthode de rééchantillonnage Bootstrap a été utilisée pour équilibrer les données. Cette méthode permis de générer des sous-ensembles à partir de la classe de risque des points noirs, en veillant à ce que toutes les classes soient de taille égale à la classe la plus petite. Le sous-ensemble Bootstrap a été extrait par un processus de rééchantillonnage aléatoire

avec remplacement. Ce processus a été répété plusieurs fois pour générer des sous-ensembles de taille égale.

Concrètement, un ensemble de 10 sous-ensembles ont été créé, lesquels ont été utilisés dans le processus de génération de règles d'association. Pour garantir une représentation égale de toutes les classes de gravité, le nombre d'enregistrements pour chaque classe a été fixé à celui de la plus petite classe (Niveau élevé, avec 71 instances). L'algorithme Apriori a ensuite été appliqué à chaque sous-ensemble en utilisant les paramètres suivants : un soutien minimum de 0,01, une levée minimum de 1,1, une confiance minimum de 0,75 et une longueur minimum de 2 pour les règles d'association conséquentes. Les résultats combinés de ces dix sous-ensembles ont produit 373 402 règles d'association, dont plusieurs étaient dupliquées. Pour remédier à ce problème, les doublons ont été agrégés en utilisant la moyenne, ce qui a permis d'obtenir 151 671 règles uniques. Pour sélectionner les règles les plus pertinentes, celles qui présentent une levée minimale de 3 et un niveau de confiance de 0,8 ont été retenues. Cette démarche a conduit à la génération de 468 règles.

4.5. Sélection des règles générales

La sélection des règles d'association constitue une étape cruciale du processus de découverte des connaissances dans l'exploration des données. L'objectif de cette méthode étant de découvrir les relations entre les éléments d'un ensemble de données transactionnelles. Pour déterminer les règles d'association les plus générales, les critères d'évaluation du soutien, de la confiance et de la levée sont couramment utilisés. Effectivement, les règles d'association présentant les valeurs les plus élevées pour ces mesures sont considérées comme les plus fiables et sont donc choisies pour la découverte ultérieure de connaissances.

Soit « D » l'ensemble des règles d'association. Les règles d'association à supprimer de D supposent que pour deux règles : $R_1 : A_1 \rightarrow C$ et $R_2 : A_2 \rightarrow C$, telles que $A_2 \subset A_1$ et Levée R_1 <= Levée R_2, donc supprimer la règle R_1. L'ensemble « S » ainsi défini contient les règles d'association de l'ensemble D qui n'ont pas de règle de Levée inférieure dans l'ensemble D. Ainsi, S contient les règles d'association les plus générales de l'ensemble D. L'ensemble des règles générales est défini par :

$$S = \{A \rightarrow C \;/\; \nexists\, A_1 \in D \; telle\; que\; A_1 \subset A \; et \; Levée\; A \leq Levée\; A_1\} \qquad (45)$$

5. Résultats d'identification des facteurs de risque

Dans cette section, nous résumerons et discuterons les principaux résultats de notre étude dans l'identification des facteurs de risque des accidents. Pour anticiper l'état des tronçons routiers, plusieurs étapes ont été menées et, à chaque étape, des résultats ont été obtenus que les détaillons ci-dessous. Les résultats de la simulation ont été obtenus sur un ordinateur Intel Core i7-8550U cadencé à 2,00 GHz et utilisant Python 3.8.0.

5.1. Résultats des poids de la méthode de l'indice de gravité pondéré

L'apprentissage du modèle XGBoost est basé sur les types de victimes en entrée et le RA en sortie. Ensuite, l'importance des caractéristiques a été explicitement estimée pour chaque entrée. Cela nous donne les poids des victimes qui seront utilisés dans la méthode IGP. À ce titre, le tableau 24 présente les résultats obtenus à partir de l'analyse préliminaire de l'importance des caractéristiques de XGBoost. Selon les données présentées dans le tableau, nous observons que le poids attribué au décès immédiat est significativement supérieur à celui attribué aux décès suite à des blessures. Par ailleurs, le poids relatif aux individus gravement blessés s'avère être plus élevé que celui des individus légèrement blessés.

Tableau 24. Poids des caractéristiques obtenus par le modèle XGBoost

Tués sur le champ	Tués suite aux blessures	Blessures graves	Blessures légères
21,88	11,79	5,45	3,15

5.2. Résultats de la méthode de l'indice de gravité pondéré

Le IGP a été calculé pour chaque section de route en utilisant le nombre total de caractéristiques pondérées. Les sections de la route ayant un score IGP plus élevé seront considérées comme ayant un niveau de gravité élevé, tandis que celles ayant une valeur plus faible sont estimées comme étant plus sécurisées. De manière générale, si la valeur IGP d'une section dépasse un seuil spécifique, le tronçon est considéré comme un point noir. Pour déterminer le niveau de gravité de chaque section, la moyenne et l'écart-type ont été calculés à partir de la valeur de l'IGP. Ainsi, les résultats de la classification ont permis d'identifier 506 sections avec un niveau de gravité élevé, 692 sections avec un niveau de gravité moyen, 760 sections avec un niveau de gravité faible, et 9744 sections qui représentent des endroits sécurisés, car elles ont un IGP très faible. D'une manière générale, le regroupement des sections de route

indique un problème clairement déséquilibré avec des groupes significativement déséquilibrés. Ce résultat était attendu puisque les accidents de la route sont imprévisibles et discrets par nature.

5.3. Résultats d'extraction des règles d'association

Les résultats de la sélection des règles générales sont présentés dans les tableaux 25, 26, 27 et 28. Ces règles nous permettront de mieux appréhender la relation entre les caractéristiques de la route et le niveau de gravité des sections routières. Les règles d'association générées ont révélé divers facteurs associés aux points noirs des accidents. Les règles d'association générées pour les sections de niveau de gravité élevé (tableau 25) ont révélé que les sections sèches, d'une largeur de 8 à 12 mètres, comportant un pont et ayant un accotement droit d'une largeur de 0,5 à 1 mètre sont communes à toutes les règles générées. Ainsi, les sections présentant de telles caractéristiques sont dangereuses en termes de mortalité. La valeur de support de 0,021 1 indique que dans 2,11 % des accidents enregistrés, ces caractéristiques et le niveau élevé de gravité se produisent concomitamment. La valeur de confiance de 1 signifie que dans 100 % des cas d'accident, lorsque les caractéristiques de la section sont présentes, des accidents de grande gravité se produisent également. Plus précisément, cette valeur indique qu'il existe 100 % de chances d'avoir des accidents mortels si la topographie est un pont, si l'état de la route est bon, si la route est sèche et rectiligne, si sa largeur est comprise entre 8 et 12 mètres et si l'accotement droit est d'une largeur comprise entre 0,5 et 1 mètre. La valeur de levée de 4 met aussi en évidence que les caractéristiques présentes et le niveau de gravité élevé sont fortement corrélés entre eux ; c'est-à-dire qu'il y a plus de 4 fois plus de chances qu'une section classée au niveau de gravité élevé et ces caractéristiques se produisent ensemble que ce à quoi on pourrait s'attendre par simple hasard.

Les règles générées par la méthode de sélection des règles générales englobent des facteurs de risque tels que les ponts, la largeur de l'accotement droit comprise entre 0,5 et 1 mètre, la largeur de la chaussée comprise entre 8 et 12 mètres et la surface sèche de la chaussée. Un résultat similaire a été démontré par [132], selon lequel la largeur de la chaussée, la largeur de l'accotement droit et la présence de ponts ont un effet significatif sur le niveau de gravité des points noirs. Tous ces résultats sont également cohérents avec la stratégie de la direction des routes du Maroc pour l'identification des points noirs. De ce fait, cette stratégie a permis d'identifier des tronçons de bonne qualité, avec une largeur de chaussée sans séparation qui varie entre 7 et 16 mètres, et des accotements qui sont au même niveau que la chaussée, plats et droits. Les sections identifiées présentent également une surface sèche à 95 % et des ponts à 48 %. Ashraf *et al.* [118] confirment également nos conclusions en établissant que la longueur de

la chaussée et la largeur des voies étaient positivement associées aux segments d'accidents liés à la conduite à contresens ; ce qui a un impact négatif sur la sécurité routière et augmente le nombre d'accidents mortels. Dans notre analyser, nous avons constaté que l'absence de séparation médiane dans de bonnes conditions routières (bonne chaussée, sèche et rectiligne) est associée à des points noirs de niveau de gravité élevé. Cette découverte a pareillement été rapportée par Calvi *et al.* [115], qui affirment que le type de séparation médiane a un effet négatif sur le comportement sécuritaire du conducteur. ces auteurs ont montré que le type de médian a un effet significatif sur la distance entre le véhicule et le médian, à savoir une grande distance lorsque la séparation médiane est une barrière de câbles. De même, nous avons constaté que les sections droites étaient associées à des points noirs. Ce résultat est cohérent avec celui de Siregar *et al.* [119] qui ont démontré que les segments droits étaient plus susceptibles d'augmenter le nombre d'accidents mortels que les courbes. De même, les résultats de Xu *et al.* [123] ont montré que le segment de route rectiligne est un facteur de risque d'accident mortel.

Les tronçons de niveau de gravité moyen (voir tableau 26) concernent principalement des routes mouillées et en mauvais état, d'une largeur de 5 à 8 mètres, ce qui inférieure à celle du niveau de gravité élevé. Ce résultat est cohérent avec ceux de [111][122], qui ont établi que parmi les attributs qui influencent la gravité des accidents de la route dans les points noirs, il y a les routes mouillées, en particulier dans le cas d'accidents impliquant un seul véhicule. Les résultats de Das *et al.* [124] confirment également que les caractéristiques de l'infrastructure, y compris l'état des routes mouillées, ont un effet néfaste sur le comportement des conducteurs. D'autre part, nos résultats soulignent que les routes mouillées et mal entretenues, avec une largeur d'accotement gauche comprise entre 1,5 et 4 mètres et soit avec un accotement gauche, soit avec un accotement droit au niveau de la route, ou une largeur de chaussée comprise entre 5 et 8 mètres, sont plus dangereuses. En outre, les sections plates, sèches et séparées, avec une largeur de chaussée comprise entre 5 et 8 mètres et contenant un pont, sont associées aux points noirs de gravité moyenne (Sup = 0,020 2 ; Conf = 0,802 0 ; Levée = 3,21). De plus, si ces sections sont caractérisées par un accotement gauche au niveau de la route, la probabilité qu'elles soient classées au niveau de gravité moyen augmente (Sup = 0,015 8 ; Conf = 0,828 6 ; Levée = 3,31).

Contre toute attente, les ponts se sont révélés être un facteur de risque dans les tronçons routiers les plus meurtriers (niveaux de gravité élevé et moyen). Cela peut s'expliquer par le fait que les ponts routiers qui traversent des rivières connaissent souvent une certaine période de brouillard, qui réduit la visibilité du conducteur et augmente la distance de freinage du véhicule [133]. Une autre explication possible est que les ponts sont parfois soumis au vent, ce qui entraîne un mauvais contrôle de la voie et

donc une perte de contrôle du véhicule. Pour remédier à ce problème, les parties prenantes recommandent d'installer des pare-vent, qui sont couramment adoptés pour améliorer la sécurité routière sur les ponts [134]. Par ailleurs, les ponts figurent également parmi les tronçons les plus meurtriers dans les zones urbaines [135].

Tableau 25. Règles générales générées sur les sections de niveau de gravité élevé

Antécédents (Niveau élevé)	**Supp**	**Conf**	**Levée**
Plat, Sec, Aco_Dr_0.5_1, Lar_Cha_8_12, Bonne, Pont	0,021 13	1	4
Rectiligne, Aco_Dro_Niv_Cha, Sec, Aco_Dr_0.5_1, Aco_Gau_Niv_Cha, Lar_Cha_8_12, Pont	0,021 13	1	4
Rectiligne, Aco_Dro_Niv_Cha, Sec, Lar_Cha_8_12, Aco_Dr_0.5_1, Pont, Unique	0,021 13	1	4
Rectiligne, Sec, Aco_Dr_0.5_1, Lar_Cha_8_12, Bonne, Pont	0,021 13	1	4
Rectiligne, Sec, Aco_Dr_0.5_1, Lar_Cha_8_12, Pont	0,021 13	0,9	3,62
Rectiligne, Sec, Aco_Gau_Niv_Cha, Lar_Cha_8_12, Aco_Dr_0.5_1, Pont, Unique	0,021 13	1	4
Rectiligne, Sec, Pont, Lar_Cha_8_12, Aco_Dro_Niv_Cha, Plat, Aco_Dr_0.5_1	0,021 13	1	4
Rectiligne, Sec, Pont, Lar_Cha_8_12, Aco_Gau_Niv_Cha, Plat, Aco_Dr_0.5_1	0,021 13	1	4
Rectiligne, Sec, Pont, Lar_Cha_8_12, Unique, Plat, Aco_Dr_0.5_1	0,021 13	1	4
Sec, Pont, Lar_Cha_8_12, Aco_Gau_Niv_Cha, Aco_Dro_Niv_Cha, Plat, Aco_Dr_0.5_1	0,021 13	1	4
Sec, Pont, Lar_Cha_8_12, Unique, Aco_Dro_Niv_Cha, Plat, Aco_Dr_0.5_1	0,021 13	1	4
Sec, Pont, Lar_Cha_8_12, Unique, Aco_Gau_Niv_Cha, Plat, Aco_Dr_0.5_1	0,021 13	1	4

Tableau 26. Règles générales générées sur les sections de niveau de gravité moyen

Antécédents (Niveau moyen)	**Supp**	**Conf**	**Levée**
Aco_Dro_Niv_Cha, Sec, Aco_Dr_0.5_1, Pont, Lar_Cha _5_8, Bonne, Unique	0,015 8	0,833	3,33
Aco_Dro_Niv_Cha, Sec, Aco_Gau_Niv_Cha, Aco_Dr_0.5_1, Lar_Cha_5_8, Bonne, Pont, Unique	0,017 6	1	4
Aco_Gau_1.5_4, Aco_Dro_Niv_Cha, Mouillé, Mauvaise	0,034 0	0,936	3,745
Aco_Gau_1.5_4, Aco_Gau_Niv_Cha, Mouillé, Mauvaise	0,034 0	0,936	3,745
Aco_Gau_1.5_4, Lar_Cha_5_8, Mouillé, Mauvaise	0,030 5	0,929	3,718
Courbe_Gauche, Sec, Aco_Gau_Niv_Cha, Aco_Dr_0.5_1, Bonne, Pont	0,017 6	1	4
Sec, Pont, Aco_Gau_1.5_4, Séparé, Aco_Gau_Niv_Cha, Plat, Lar_Cha_5_8	0,015 8	0,828	3,314
Sec, Pont, Aco_Gau _1,5_4, Séparé, Plat, Lar_Cha_5_8	0,020 2	0,802	3,208

Les résultats de l'approche de sélection des règles générales ont donné lieu à une seule règle qui associe les caractéristiques routières aux sections de niveau de gravité faible, comme l'illustre le

tableau 27. Ces sections étant caractérisées par une chaussée unique et un très mauvais état de surface de la route. Les critères de mesure indiquent que la probabilité qu'une section présentant ces caractéristiques soit classée au niveau faible est de 0,8. En outre, il y a trois fois plus de chances qu'une section classée au niveau faible et une section présentant ces caractéristiques se produisent ensemble que prévu (Sup=0,0152 ; Conf=0,8079 ; Levée=3,23). Ces résultats sont conformes aux conclusions de Siregar *et al.* [116], qui ont montré que l'état des routes et le statut sont statistiquement significatifs. En particulier, le mauvais état des routes augmente le risque relatif d'accidents mortels.

Tableau 27. Règles générales générées sur les sections de niveau de gravité faible

Antécédents (Niveau faible)	**Support**	**Confidence**	**Levée**
Unique, Très_Mauvaise	0,015 25	0,807 93	3,231

Dans les sections sûres (tableau 28), les règles générées indiquent que celles-ci présentent des caractéristiques qui n'ont pas été identifiées dans les autres niveaux de gravité supérieures, notamment le rétrécissement de la route, les courbes à gauche et à droite, les pentes et les routes humides, salées ou sablonneuses, avec une largeur de route comprise entre 2 et 8 mètres, et non inférieure à 2 mètres ou supérieure à 8 mètres, une largeur de l'accotement gauche comprise entre 0,5 et 1 mètre, la position de l'accotement droite (resp. gauche) par rapport à la chaussée.

Nos résultats ont ainsi nous permis d'identifier une seule règle contenant le facteur de risque lié aux courbes parmi les trois niveaux de gravité (niveau élevé, niveau moyen, niveau faible). En revanche, de nombreuses règles associent les courbes et la classe « Sûr ». Ce résultat est conforme à l'étude de Siregar *et al.* [120] qui établit que l'augmentation des courbes dans une route entraîne une diminution de la vitesse du trafic et, par conséquent, du taux de mortalité et d'accident. Cela peut s'expliquer par le fait que les courbes dans les zones rurales peuvent contribuer à la sécurité routière en ayant un impact positif sur le comportement des conducteurs (réduction de la vitesse, amélioration de l'anticipation des dangers potentiels) et en augmentant la visibilité pour les conducteurs (création de vues claires et élimination des angles morts). Toutefois, ce résultat reste en contradiction avec ceux de [116], qui suggèrent que les courbes présentent un risque accru de blessures graves ou mortelles dans un accident impliquant un piéton.

D'autre part, nos résultats montrent que la pente de la route est un facteur insignifiant dans les points noirs d'accidents. C'est-à-dire que les pentes et les zones sûres sont associées. Cela correspond aux conclusions de menées par Siregar *et al.* [120], selon lesquelles un plus grand nombre de pentes sur

une route réduit la vitesse de circulation, ce qui entraîne des vitesses plus faibles et convergentes, et donc des taux de mortalité et d'accident plus faibles. En ce qui concerne l'impact de la largeur des accotements sur la sécurité routière, nos résultats concordent avec ceux d'Ambros *et al.* [110], qui ont démontré que la largeur des accotements a un effet sur la sécurité routière. En effet, ces deniers ont constaté que des accotements plus larges étaient associés à la gravité des accidents. Ils concordent également avec les résultats de Ben-Bassat & Shinar [114] qui ont étudié l'impact de la largeur des accotements sur la sécurité routière. Ces auteurs ont constaté que les accotements ont un effet significatif sur la vitesse réelle, la position dans la voie et la vitesse de sécurité perçue en présence de glissières de sécurité. En outre, les réglementations sur la largeur des accotements ont un impact positif sur la sécurité routière en termes de contrôle de la vitesse et de maintien de la position dans la voie. Les résultats de l'étude de Singh & Tiwari [112] concordent également avec ceux de notre étude. Effectivement, ils indiquent que les accotements larges présentent un risque plus élevé d'accidents mortels que les accotements d'une largeur comprise entre 1,0 et 1,5 mètre.

Tableau 28. Règles générales générées sur les sections sûres

Antécédents (Sûr)	**Supp**	**Conf**	**Levée**
Aco_Dr_0.5_1, Aco_Dro_Niv_Cha, Plat, Courbe_Gauche, Rétrécissement	0,017 60	1	4
Aco_Dr_0.5_1, Plat, Sec, Aco_Gau_Niv_Cha, Rétrécissement, Lar_Cha_5_8, Bonne	0,017 60	1	4
Aco_Dro_Niv_Cha, Humide_Salée _ Sablée	0,017 60	1	4
Aco_Gau_0.5_1, Aco_Dr_0.5_1, Aco_Dro_Niv_Cha, Plat, Rétrécissement, Lar_Cha_5_8	0,017 60	1	4
Aco_Gau_0.5_1, Aco_Dr_0.5_1, Aco_Dro_Niv_Cha, Rétrécissement, Lar_Cha_5_8, Bonne	0,017 60	1	4
Aco_Gau_0.5_1, Aco_Dr_0.5_1, Aco_Dro_Niv_Cha, Sec, Courbe_Gauche, Rétrécissement	0,021 12	1	4
Aco_Gau_0.5_1, Aco_Dro_Denivelé, Rectiligne, Aco_Gau_Dénivelé	0,017 60	1	4
Aco_Gau_0.5_1, Lar_Cha_2_5, Aco_Dr_0.5_1, Aco_Dro_Niv_Cha, Sec, Plat	0,015 84	0,87	3,5
Aco_Gau_0.5_1, Lar_Cha_2_5, Aco_Dr_0.5_1, Aco_Gau_Niv_Cha, Plat	0,016 43	0,87	3,47

Un résultat inattendu est que le rétrécissement de la route est associé à des sections sûres, que l'état de la chaussée soit bon ou mauvais. Cela peut être attribué au fait que le rétrécissement des routes a un impact positif sur le comportement des conducteurs en les incitant à la prudence et à la prise de conscience et ce en limitant leur vitesse de conduite. En outre, les rétrécissements de route peuvent calmer le trafic et améliorer la visibilité aux intersections. Ce résultat est cohérent avec celui de Sołowczuk & Kacprzak [136], qui ont montré que l'installation de rétrécissements de chaussée avec d'autres dispositifs de sécurité peut améliorer la sécurité routière aux intersections. Cependant, il est en contradiction avec

les conclusions de l'étude de Sharma *et al.* [137], qui a indiqué que les pistes cyclables étroites étaient un obstacle à la satisfaction de l'utilisation de ces pistes, avec un impact négatif sur la sécurité des cyclistes.

5.4. Analyse des règles d'association par diagramme de dispersion

Les diagrammes de dispersion présentés dans la figure 32 montrent la distribution des règles d'association générales à l'aide des valeurs de Support et de Confiance. À ce titre, il nous permet de visualiser rapidement les règles d'association qui ont une valeur de Support et de Confiance élevée. Les points situés en haut à droite du diagramme représentent les règles d'association dont les valeurs de Support et de Confiance sont élevées ; ce qui indique une forte association entre l'antécédent et le conséquent. Les points situés en bas à gauche du diagramme représentent des règles d'association dont le Support et la Confiance sont faibles, montrant ainsi une faible association. La taille des points change en fonction de la valeur de Levée, une grande taille indiquant une grande valeur de Levée, et vice versa.

La figure montre également les sections caractérisées par : d'un côté, le rétrécissement de la route, l'accotement droit entre 0,5 et 1 mètre, et la largeur de la route entre 5 et 8 mètres. De l'autre côté, en plus des caractéristiques précédentes, une surface sèche et un accotement gauche entre 0,5 et 1 mètre sont fortement associés aux sections sûres. Par ailleurs, la figure montre que les sections caractérisées par : une surface humide, une surface de chaussée médiocre, un accotement gauche entre 1,5 et 4 mètres, et soit avec des accotements au même niveau que la chaussée, soit avec une largeur de chaussée entre 5 et 8 mètres ; ces sections sont fortement associées aux points noirs du deuxième niveau de sévérité.

6. Résultats de prédiction des points noirs

Dans les problèmes de classification, l'inconvénient fondamental des données déséquilibrées est de conduire à des résultats incorrects, voire à des erreurs de classification. Il existe à ce titre plusieurs approches de l'apprentissage supervisé pour réduire l'impact des données déséquilibrées. Le rééchantillonnage des données, les méthodes sensibles aux coûts et les méthodes d'ensemble sont les trois principales catégories de ces techniques. Les performances du modèle ELM ont été évaluées sur des données d'accidents de la route multiclasses. Dans ce dispositif, le tableau 29 présente les données, les caractéristiques et les classes disponibles. Les données de l'ensemble d'apprentissages et de l'ensemble des tests ont été générées de manière aléatoire, dont 20 % ont été affectés à l'ensemble des tests.

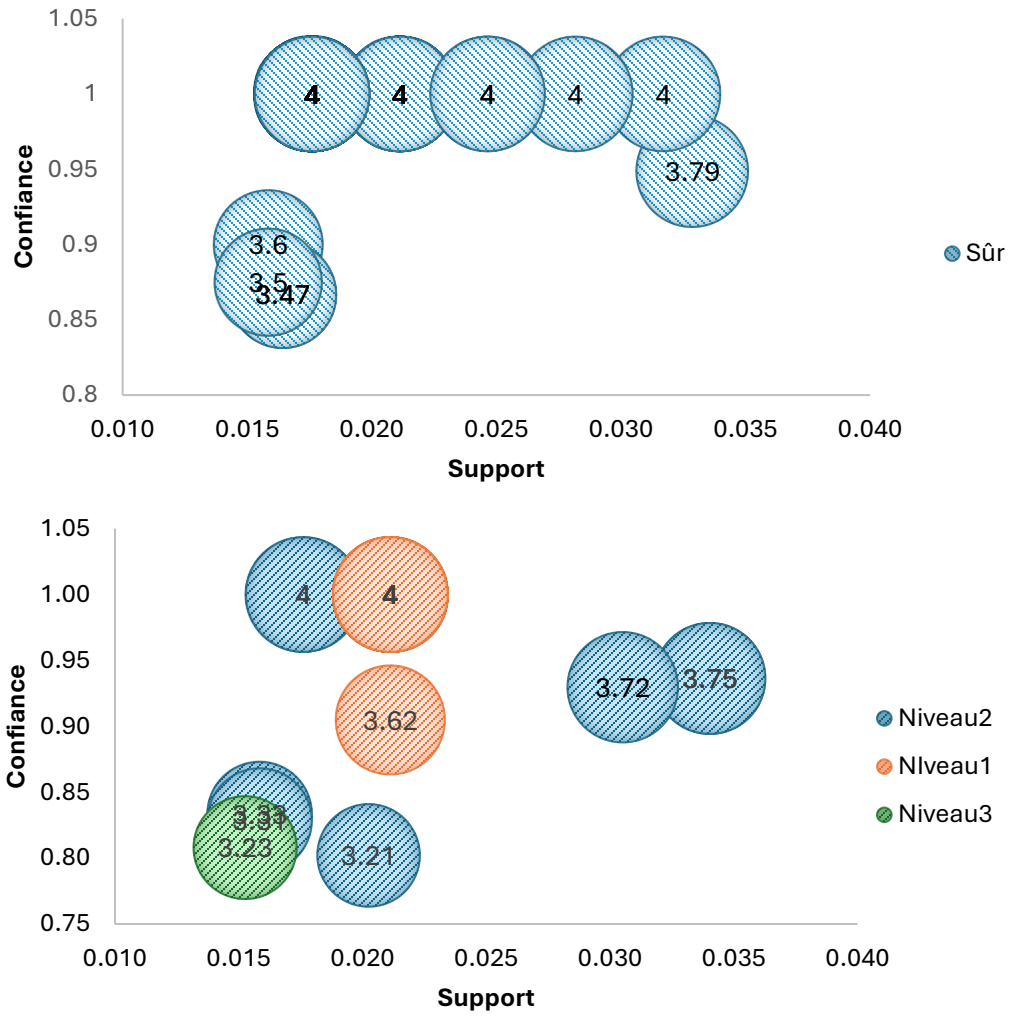


Figure 32. Visualisation par graphique à bulle des règles générées

Tableau 29. Propriétés de l'ensemble de données final

Taille des données	Nombre de caractéristiques	Nombre de classes
11 702	11	4

Les résultats de la précision de la classification préliminaire des points noirs sont présentés dans le tableau 30. Comme le montre le tableau, le modèle ELM a un temps d'exécution rapide, ce qui est un avantage. Cependant, sa précision d'apprentissage de 81,6 % et sa précision de test de 90 % indiquent certaines difficultés à capturer des modèles complexes dans les données. De plus, la disparité entre la précision d'apprentissage et celle de test peut indiquer un surajustement du modèle aux données d'apprentissage. Cela signifie que le modèle peut s'adapter correctement aux données d'apprentissage,

mais qu'il peut ne pas se généraliser correctement sur de nouvelles données. Dans cette optique, le modèle Bagging ELM montre une amélioration significative de la précision par rapport au modèle ELM de base.

En effet, sa précision d'apprentissage de 97,4 % et sa précision de test de 98,6 % suggèrent qu'il est capable de mieux se généraliser sur de nouvelles données. Cela peut être attribué à l'utilisation de l'approche du Bagging, qui entraîne plusieurs instances du modèle ELM sur des sous-ensembles aléatoires de données et combine leurs prédictions. Cela réduit le risque de surapprentissage et améliore la fiabilité du modèle. Pour sa part, le modèle SVM a un temps d'exécution plus élevé que les modèles ELM et Bagging ELM. Sa précision d'apprentissage de 80,9 % et sa précision de test de 77,2 % suggèrent qu'il peut également être sujet au surapprentissage, bien que la différence entre les deux précisions soit inférieure à celle du modèle ELM. Toutefois, ses performances en termes de précision sont inférieures à celles du modèle Bagging ELM. Enfin, le modèle MLP présente des performances comparables à celles du modèle ELM en termes de précision. Son temps d'exécution est le plus élevé de tous les modèles évalués. Toutefois, l'écart entre la précision de l'apprentissage et celle du test est relativement faible, ce qui indique qu'il se généralise mieux que le modèle ELM.

Tableau 30. Comparaison des performances entre B-ELM, ELM, SVM et MLP

Modèle	**Temps (s)**	**Précision de l'apprentissage**	**Précision du test**
ELM	0,03	81,6	90
Bagging ELM	0,27	97,4	98,6
SVM	1,14	80,9	77,2
MLP	1,85	83,6	89,8

Le modèle Bagging ELM se distingue dans notre analyse avec une précision de test de 98,6 %. Ce qui en fait le modèle le plus performant parmi ceux qui ont été évalués. Il est également plus robuste que le modèle ELM de base et le modèle SVM, car il évite le surapprentissage et se généralise mieux sur de nouvelles données. De plus, la différence entre la précision d'apprentissage et la précision de test est très faible (1,2 %), ce qui indique que le modèle Bagging ELM présente une meilleure fiabilité, car il est moins sensible aux variations des données. Toutefois, il est important de noter que la fiabilité et la robustesse d'un modèle dépendent également de l'ensemble de données spécifique utilisé pour l'apprentissage et le test. Des expériences supplémentaires avec différents ensembles de données pourraient permettre de mieux comprendre les performances de ces modèles dans différents contextes.

7. Discussions et recommandations

La stratégie de la Direction des Routes visant à identifier les points noirs des accidents de la route a donné lieu à la localisation de 21 points noirs à l'échelle du réseau routier national. Ces emplacements concordent avec les résultats de notre propre identification, étant donné qu'ils affichent un niveau de gravité élevé. Toutes ces sections présentent des caractéristiques favorables, telles qu'une chaussée en bon état, des accotements au même niveau que la chaussée, une topographie plate et une trajectoire rectiligne. De plus, 95 % de ces points noirs disposent d'une surface sèche, 48 % sont situés sur des ponts et 38 % correspondent à des rétrécissements de chaussée. Par ailleurs, leur largeur varie entre 7 et 16 mètres. Ces points noirs ont été associés à un total de 45 décès, 216 blessures graves et 520 accidents.

Des études antérieures [111][138][139] ont montré que les routes courbes et rectilignes sont les plus susceptibles d'être des points noirs lorsque l'on considère le tracé de la route ; les mêmes résultats ont été obtenus par Katre *et al.* [140]. Ces auteurs ont en effet identifié un nombre maximal d'accidents survenant spécifiquement sur les tronçons de route rectiligne ainsi que dans les virages légers et serrés. Un autre constat significatif réside dans l'impact substantiel de la largeur de l'accotement sur la gravité des points noirs ; un résultat également corroboré par Yasmin *et al.* [141] qui ont mis en évidence l'effet de la largeur de la voie et de la largeur de l'accotement sur la proportion d'accidents en fonction de leur gravité. Par ailleurs, des recherches antérieures ont démontré que l'accroissement de la largeur des accotements est associé à une intensification du niveau de gravité des accidents routiers. Les résultats de l'étude menée par [142] mettent aussi en évidence l'effet favorable de la largeur de l'accotement sur la fréquence des accidents, une conclusion partagée par Gitelman *et al.* [143], qui a également trouvé une corrélation positive entre la gravité des accidents et la largeur de l'accotement. En effet, il a été constaté qu'une augmentation de la largeur de l'accotement, jusqu'à 2 mètres, augmentait le risque d'accident.

Enfin, des études précédentes ont mis en évidence que les routes dont la chaussée est plus large sont généralement associées à des accidents plus graves [141][144][145]. À ce propos, les résultats de Dadashova *et al.* [146] ont montré que la courbure horizontale et verticale et la largeur de la voie sont associées à la gravité de l'accident. En revanche, le profil en long de la route a été identifié comme un facteur non significatif du niveau de gravité des points noirs. Ces résultats reflètent ceux de [111] qui ont également constaté que les pentes des routes n'avaient pas d'impact significatif sur les accidents mortels. L'expérience met dans ces conditions en évidence l'effet de la position de l'accotement droit sur la gravité

des points noirs, ce qui est également cohérent avec les résultats de [147] qui soulignent que le niveau d'élévation de l'accotement a un impact sur la détermination de l'emplacement des points noirs.

8. Conclusion

Dans notre approche, nous proposons une méthode analytique utilisant l'extraction de règles d'association pour identifier la corrélation entre les caractéristiques spatiales des routes et les points noirs sur les routes rurales du Maroc. Dans ce cadre, la méthodologie a consisté à calculer l'indicateur de risque d'accident pour chaque section routière d'un kilomètre, à entraîner un modèle XGBoost, à extraire les poids associés à chaque entrée, à calculer l'Indice de Gravité Pondéré, et à regrouper les sections en quatre niveaux sur la base des valeurs du IGP. Ces données ont ensuite été transformées en règle d'association et les règles les plus générales ont été sélectionnées grâce à l'approche proposée. L'étape finale consistaient à identifier les caractéristiques fréquentes prédominantes pour chaque niveau de gravité des points noirs.

Nos résultats ont montré que plusieurs caractéristiques spatiales de la route, telles que la largeur de la route, le revêtement de la route, le tracé rectiligne et la présence d'un pont, contribuaient à l'occurrence d'accidents mortels dans un type spécifique de section routière. Ils démontrent également que diverses caractéristiques routières ont un impact significatif sur la gravité des points noirs. À ce titre, la contribution de notre thèse réside dans l'identification des caractéristiques routières qui contribuent aux points noirs des accidents. Ces résultats pouvant être utilisés par les autorités routières pour concevoir et construire des routes plus sûres, hiérarchiser les interventions de sécurité routière et développer des stratégies efficaces.

Notre étude se concentre principalement sur les caractéristiques spatiales des routes, mais d'autres facteurs tels que le comportement des conducteurs, les conditions météorologiques et les caractéristiques des véhicules peuvent également influencer la survenue des accidents routiers. D'autres études sont par ailleurs recommandées pour valider ces résultats et examiner d'autres facteurs susceptibles de contribuer aux accidents de la route. De plus, notre méthodologie s'appuie sur des modèles prédictifs et des techniques d'apprentissage automatique qui peuvent être sensibles aux biais des données et nécessitent une validation rigoureuse sur des ensembles de données indépendants pour garantir leur généralisabilité.

D'autre part, dans cette étude, nous proposons un modèle pour l'identification et la classification des points noirs sur les routes rurales du Maroc en utilisant une méthode d'ensemble basée sur l'Extreme

Learning Machine. Pour ce faire, le modèle de classification nommé Bagging ELM, utilise les caractéristiques spatiales comme l'état de la route et les conditions environnantes pour la prédiction des points noirs des accidents. Ainsi, le modèle proposé montre la meilleure généralisation en termes de précision et fiabilité surpassant les modèles réputés en atteignant une précision de 98,6 %.

De toute évidence, le modèle d'analyse des points noirs proposé est de nature à aider les spécialistes de la sécurité routière à prendre de bonnes décisions et à apporter des améliorations sur les tronçons routiers prioritaires. Dans les travaux futurs, nous prévoyons de prouver le plein potentiel de notre approche en étudiant l'ensemble du réseau routier marocain sur une longue période afin d'extraire les propriétés communes des points noirs. Effectivement, ces méthodes d'analyse et de prédiction pourraient être affinées en incluant également des variables météorologiques et temporelles afin d'augmenter la fiabilité et d'améliorer la précision de prédiction, contribuant de manière considérable au renforcement du système de transport intelligent dans le pays.

Conclusion générale

Les objectifs de cette recherche sont axés sur la sécurité routière, un élément fondamental de la santé publique, en abordant deux dimensions : le trafic routier et les accidents de la circulation. Dans le cadre du trafic routier, notre étude se concentre sur deux aspects principaux. Tout d'abord, elle examine la gestion des données manquantes générées par les compteurs du trafic routier en dysfonctionnement temporaire. Ensuite, elle propose un modèle intelligent visant à renforcer les interventions proactives et à guider les décideurs dans la gestion du trafic, en fournissant des prévisions à court terme sur l'état futur du trafic routier. Quant à la composante dédiée à la sécurité routière, elle est divisée en deux volets distincts. D'une part, elle vise à identifier les facteurs de risque associés aux accidents mortels, en mettant l'accent sur les caractéristiques infrastructurelles ayant un impact sur leur occurrence. D'autre part, elle se concentre sur l'identification des caractéristiques communes des points noirs des accidents, en proposant un modèle reposant sur l'extraction des règles d'association. La seconde partie de cette composante se focalise sur la prédiction des points noirs d'accidents en se basant sur les caractéristiques infrastructurelles, dans le but d'identifier les sections routières présentant un risque élevé en fonction de leur configuration.

Parmi les modèles proposés pour la génération des données manquantes du trafic routier, le modèle SARIMA a démontré une précision remarquable dans la reconstitution du trafic routier manquant, surpassant ainsi les autres modèles comparés en termes de précision et robustesse. De plus, le modèle de prédiction à court terme du trafic routier, basé sur l'Extreme Learning Machine et une méthode d'ensemble, a fait preuve d'une grande précision et d'une capacité de généralisation élevée, en surpassant les autres modèles comparés. Concernant l'identification des facteurs de risque associés aux accidents mortels, les résultats ont mis en évidence l'impact significatif des facteurs infrastructurels sur la gravité de ces accidents. Plus précisément, le type de carrefour et la localisation ont montré un effet modéré sur la létalité des accidents. Par ailleurs, les résultats issus des règles d'association ont confirmé une forte association entre les caractéristiques infrastructurelles telles que la largeur de chaussée et des accotements, le tracé rectiligne, et la présence de pont, avec les points noirs d'accidents de niveau de gravité élevé.

Notre thèse présente plusieurs bénéfices et applications pratiques des résultats obtenus dans le domaine de la sécurité routière. En effet, la prédiction du trafic routier manquant va remédier au problème

de perte d'information fréquemment apparu à cause de dysfonctionnement des dispositifs électroniques. Cela va augmenter la robustesse et la fiabilité des résultats des analyses sur le trafic routier. Le but principal est de proposer une solution fiable et précise aux données manquantes du trafic routier. Ainsi que d'aider les décideurs et les entités responsables du trafic routier à bien exploiter la masse des données du trafic routier sans aucune perte d'information. D'autre part, la prédiction à court terme du trafic routier va permettre aux décideurs de bien gérer le trafic routier de manière proactive, afin d'optimiser la gestion du trafic et d'améliorer l'efficacité des interventions en temps réel. Cette approche anticipative contribuera à améliorer la planification opérationnelle et la réactivité face aux éventuelles perturbations, renforçant ainsi la pertinence et l'efficacité globale du système de gestion du trafic routier. Quant à l'analyse des accidents mortels en termes des facteurs de risques infrastructurels, l'identification des points noirs des accidents et la prédiction des points noirs des accidents sur la base des facteurs de risques, l'objectif sous-jacent est d'orienter judicieusement les investissements dans l'infrastructure routière en concentrant les ressources sur les sections à forte létalité, ainsi que de réduire de manière significative la létalité liée aux accidents de la route. Cette démarche favorise une utilisation rationnelle des ressources allouées, optimisant ainsi l'impact des investissements sur la sécurité routière. La prédiction des points noirs sur la base des infrastructures et conditions de la route apporte plusieurs avantages dans le domaine de la sécurité routière. Notamment la prévention des accidents en identifiant les zones susceptibles de devenir des points noirs d'accidents, la planification des infrastructures à long terme pour améliorer l'infrastructure routière, l'allocation des ressources en dirigeant les efforts vers les zones prévues comme étant à haut risque et l'évaluation de l'efficacité des interventions effectuées sur des zones jugées à haut risque. Cette approche intégrée de la prédiction du trafic routier et de l'analyse des accidents offre des perspectives prometteuses pour une gestion plus efficace et plus sûre des réseaux routiers, contribuant à réduire la congestion, les risques d'accident et les pertes humaines.

La prédiction des données manquantes du trafic routier sur un compteur en se basant sur les données historiques des compteurs voisins, peut ne pas capturer les variations temporelles et spatiales, en particulier dans des sections où le trafic routier est influencé uniquement par des facteurs locaux. Ainsi, il serait judicieux de développer des modèles de prédiction incluant d'autres variables pertinentes telles que les conditions météorologiques, les événements locaux et les caractéristiques spécifiques de l'infrastructure routière. Parfois, les compteurs voisins peuvent être soumis à des erreurs de mesure, ce qui entraîne des prédictions erronées pour le compteur en question. Afin de pallier cette problématique, il est impératif de déployer des mécanismes de correction d'erreur pour filtrer les données incorrectes

provenant des compteurs voisins. De plus, il convient d'explorer l'utilisation de capteurs supplémentaires ou de technologies alternatives, telles que les caméras, pour compléter les données des compteurs voisins. D'autre part, la prédiction à court terme du trafic routier sur la base du trafic historique et d'autres facteurs temporels a montré sa fiabilité dans de nombreuses études empiriques. De plus, cette approche s'est avérée être un outil utile pour la gestion proactive du trafic. Cependant d'autres facteurs peuvent intervenir dans la dynamique du trafic routier, comme les conditions météorologiques, les événements locaux et les caractéristiques spécifiques des infrastructures, ainsi que les comportements des conducteurs et les événements imprévus. Une étude profonde comportant ces facteurs va permettre une modélisation plus précise et complète du trafic routier, offrant ainsi des prédictions plus fiables et une gestion plus efficace des flux de circulation.

En ce qui concerne les accidents de la circulation, ils sont souvent résultats de plusieurs facteurs tels que le facteur humain, physique et météorologique. Une étude profonde comportant plusieurs variantes de facteurs est indispensable pour une modélisation complète des accidents. D'un autre part, les infrastructures peuvent connaître des évolutions au fil du temps, ce qui peut rendre les résultats obtenus obsolètes et nécessiter des mises à jour régulières afin de maintenir leur pertinence. Par conséquent, toute identification des points noirs basée uniquement sur les caractéristiques des infrastructures doit être régulièrement révisée pour tenir compte des améliorations apportées aux routes par les autorités compétentes. De même, les points noirs des accidents sont fréquemment le produit d'une interaction complexe entre divers facteurs, incluant notamment les conditions météorologiques, les aspects temporels et les événements locaux. Ainsi, pour développer un modèle de prédiction plus fiable, il est impératif d'intégrer non seulement les facteurs liés aux infrastructures, mais également d'autres variables telles que le flux du trafic, les conditions météorologiques, les événements et les caractéristiques temporelles.

Bibliographies

[1] D. Luo and K. Chen, "A comparative study of statistical ensemble methods on mismatch conditions," in *Proceedings of the International Joint Conference on Neural Networks*, 2002, vol. 1, pp. 59–64. doi: 10.1109/ijcnn.2002.1005442.

[2] P. Poonia, V. K. Jain, and A. Kumar, "Short term traffic flow prediction methodologies: a review," *Mody Univ. Int. J. Comput. Eng. Res.*, vol. 2, no. 1, pp. 37–39, 2018.

[3] A. M. Nagy and V. Simon, "Survey on traffic prediction in smart cities," *Pervasive and Mobile Computing*, vol. 50. Elsevier B.V., pp. 148–163, Oct. 01, 2018. doi: 10.1016/j.pmcj.2018.07.004.

[4] S. Benkirane *et al.*, "Adapted Speed System in a Road Bend Situation in VANET Environment," *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 3781–3794, 2023, doi: 10.32604/CMC.2023.033119.

[5] H. Khalfaoui, A. Azmani, A. Farchane, and S. Safi, "Symbiotic Combination of a Bayesian Network and Fuzzy Logic to Quantify the QoS in a VANET: Application in Logistic 4.0," *Computers*, vol. 12, no. 2, Feb. 2023, doi: 10.3390/COMPUTERS12020040.

[6] A. Barodi, A. Zemmouri, A. Bajit, M. Benbrahim, and A. Tamtaoui, "Intelligent Transportation System Based on Smart Soft-Sensors to Analyze Road Traffic and Assist Driver Behavior Applicable to Smart Cities," *Microprocess. Microsyst.*, vol. 100, p. 104830, Jul. 2023, doi: 10.1016/j.micpro.2023.104830.

[7] A. Kherraki and R. El Ouazzani, "Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring," *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, pp. 110–120, Mar. 2022, doi: 10.11591/IJAI.V11.I1.PP110-120.

[8] H. Derrouz *et al.*, "End-to-end quantum-inspired method for vehicle classification based on video stream," *Neural Comput. Appl.*, vol. 34, no. 7, pp. 5561–5576, Apr. 2022, doi: 10.1007/s00521-021-06718-9.

[9] M. Akallouch, O. Akallouch, K. Fardousse, A. Bouhoute, and I. Berrada, "Prediction and Privacy Scheme for Traffic Flow Estimation on the Highway Road Network," *Inf.*, vol. 13, no. 8, Aug. 2022, doi: 10.3390/INFO13080381.

[10] A. Barodi, A. Bajit, T. El Harrouti, A. Tamtaoui, and M. Benbrahim, "An Enhanced Artificial Intelligence-Based Approach Applied to Vehicular Traffic Signs Detection and Road Safety Enhancement," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 672–683, Feb. 2021, doi: 10.25046/aj060173.

[11] A. Kherraki, S. S. Warraich, M. Maqbool, and R. El Ouazzani, "Residual balanced attention network for real-time traffic scene semantic segmentation," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 3281–3289, Jun. 2023, doi: 10.11591/IJECE.V13I3.PP3281-3289.

[12] A. Charroud, K. El Moutaouakil, V. Palade, and A. Yahyaouy, "XDLL: Explained Deep Learning LiDAR-Based Localization and Mapping Method for Self-Driving Vehicles," *Electronics*, vol. 12, no. 3, p. 567, Jan. 2023, doi: 10.3390/electronics12030567.

[13] K. Mohammed, M. Abdelhafid, K. Kamal, N. Ismail, and A. Ilias, "Intelligent driver monitoring system: An Internet of Things-based system for tracking and identifying the driving behavior,"

Comput. Stand. Interfaces, vol. 84, p. 103704, Mar. 2023, doi: 10.1016/j.csi.2022.103704.

[14] Y. Benmessaoud, L. Cherrat, and M. Ezziyyani, "Real-Time Self-Adaptive Traffic Management System for Optimal Vehicular Navigation in Modern Cities," *Computers*, vol. 12, no. 4, Apr. 2023, doi: 10.3390/COMPUTERS12040080.

[15] Y. Karouani and M. Elgarej, "Milk-run collection monitoring system using the internet of things based on swarm intelligence," *Int. J. Inf. Syst. Supply Chain Manag.*, vol. 15, no. 3, Jul. 2022, doi: 10.4018/IJISSCM.290018.

[16] Y. Regragui and N. Moussa, "A real-time path planning for reducing vehicles traveling time in cooperative-intelligent transportation systems," *Simul. Model. Pract. Theory*, vol. 123, Feb. 2023, doi: 10.1016/J.SIMPAT.2022.102710.

[17] Y. Zhang and Y. Liu, "Comparison of parametric and nonparametric techniques for non-peak traffic forecasting," *World Acad. Sci. Eng. Technol.*, vol. 39, no. 3, pp. 242–248, 2009, doi: doi.org/10.5281/zenodo.1329472.

[18] B. M. Williams, M. Asce, L. A. Hoel, and F. Asce, "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results", doi: 10.1061/ASCE0733-947X2003129:6664.

[19] H. Moeeni and H. Bonakdari, "Impact of Normalization and Input on ARMAX-ANN Model Performance in Suspended Sediment Load Prediction," *Water Resour. Manag.*, vol. 32, no. 3, pp. 845–863, Feb. 2018, doi: 10.1007/s11269-017-1842-z.

[20] H. Moeeni and H. Bonakdari, "Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model," *Stoch. Environ. Res. Risk Assess.*, vol. 31, no. 8, pp. 1997–2010, Oct. 2017, doi: 10.1007/s00477-016-1273-z.

[21] M. Awad, R. Khanna, M. Awad, and R. Khanna, "Support Vector Regression," in *Efficient Learning Machines*, Apress, 2015, pp. 67–80. doi: 10.1007/978-1-4302-5990-9_4.

[22] B. Çetiner, M. Sari, and O. Borat, "A Neural Network Based Traffic-Flow Prediction Model," *Math. Comput. Appl.*, vol. 15, no. 2, pp. 269–278, Apr. 2010, doi: 10.3390/mca15020269.

[23] R. Li and H. Lu, "Combined neural network approach for short-term urban freeway traffic flow prediction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5553 LNCS, no. PART 3, pp. 1017–1025. doi: 10.1007/978-3-642-01513-7_112.

[24] M. Jiber, I. Lamouik, Y. Ali, and M. A. Sabri, "Traffic flow prediction using neural network," *2018 Int. Conf. Intell. Syst. Comput. Vision, ISCV 2018*, vol. 2018-May, pp. 1–4, 2018, doi: 10.1109/ISACV.2018.8354066.

[25] B. Hicham, A. Ahmed, and M. Mohammed, "Vehicle Type Classification Using Convolutional Neural Network," *Colloq. Inf. Sci. Technol. Cist*, vol. 2018-October, pp. 313–316, Dec. 2018, doi: 10.1109/CIST.2018.8596500.

[26] I. El Mallahi, A. Dlia, J. Riffi, M. A. Mahraz, and H. Tairi, "Prediction of Traffic Accidents using Random Forest Model," in *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, May 2022, pp. 1–7. doi: 10.1109/ISCV54655.2022.9806099.

[27] Z. Elamrani Abou Elassad, H. Mousannif, and H. Al Moatassime, “A real-time crash prediction fusion framework: An imbalance-aware strategy for collision avoidance systems,” *Transp. Res. Part C Emerg. Technol.*, vol. 118, p. 102708, Sep. 2020, doi: 10.1016/j.trc.2020.102708.

[28] E. A. E. Zouhair, H. Mousannif, and H. Al Moatassime, “Towards analyzing crash events for novice drivers under reduced-visibility settings: A simulator study,” *ACM Int. Conf. Proceeding Ser.*, Mar. 2020, doi: 10.1145/3386723.3387849.

[29] B. Marzak, K. El Guemmat, E. H. Benlahmar, and M. Talea, “Clustering in Vehicular Ad-Hoc Network Using Artificial Neural Network,” *Int. Rev. Comput. Softw.*, vol. 11, no. 6, p. 548, Jun. 2016, doi: 10.15866/irecos.v11i6.9328.

[30] B. Ben Elallid, N. Benamar, N. Mrani, and T. Rachidi, “DQN-based Reinforcement Learning for Vehicle Control of Autonomous Vehicles Interacting With Pedestrians,” *2022 Int. Conf. Innov. Intell. Informatics, Comput. Technol. 3ICT 2022*, pp. 489–493, 2022, doi: 10.1109/3ICT56508.2022.9990801.

[31] Z. Charouh, A. Ezzouhri, M. Ghogho, and Z. Guennoun, “Video Analysis and Rule-Based Reasoning for Driving Maneuver Classification at Intersections,” *IEEE Access*, vol. 10, pp. 45102–45111, 2022, doi: 10.1109/ACCESS.2022.3169140.

[32] M. Jiber, A. Mbarek, A. Yahyaouy, M. A. Sabri, and J. Boumhidi, “Road traffic prediction model using extreme learning machine: the case study of Tangier, Morocco,” *Information*, vol. 11, no. 12, Nov. 2020, doi: 10.3390/info11120542.

[33] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, “Road traffic mortality in Morocco: Analysis of statistical data,” *2020 Int. Conf. Intell. Syst. Comput. Vis.*, pp. 1–7, Jun. 2020, doi: 10.1109/ISCV49265.2020.9204325.

[34] L. Guillaume, “Road Traffic Data: Collection Methods and Applications.” 2008.

[35] T. V. Mathew, “Automated Traffic Measurement.” https://www.civil.iitb.ac.in/~vmtom/nptel/524_AutoMer/web/web.html (accessed Aug. 30, 2023).

[36] “The Traffic Data Collection Renaissance - UrbanLogiq.” https://urbanlogiq.com/the-traffic-data-renaissance/ (accessed Aug. 30, 2023).

[37] S. M. Turner, W. L. Eisele, and R. J. Benz, “TRAVEL TIME DATA COLLECTION HANDBOOK,” vol. 7, pp. 8–72, 1700, doi: 10.21949/1404545.

[38] L.-E. Y. Mimbela, L. A. Klein, and U. S. J. P. O. for I. T. Systems, “Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems,” United States. Joint Program Office for Intelligent Transportation Systems, Aug. 2007. doi: 10.21949/1503647.

[39] K. O. Kusimo and F. O. Okafor, “Comparative Analysis of Mechanical and Manual Mode of Traffic Survey for Traffic Load Determination,” *Niger. J. Technol.*, vol. 35, no. 2, pp. 226–233, Mar. 2016, doi: 10.4314/NJT.V35I2.1.

[40] S. M. Lundberg, P. G. Allen, and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[41] M. Tahifa, J. Boumhidi, and A. Yahyaouy, “Multi-agent reinforcement learning-based approach

for controlling signals through adaptation," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 11, no. 2, p. 129, 2018, doi: 10.1504/IJAACS.2018.092019.

[42] G. D'Angelo, R. Pilla, J. B. Dean, and S. Rampone, "Toward a soft computing-based correlation between oxygen toxicity seizures and hyperoxic hyperpnea," *Soft Comput.*, vol. 22, no. 7, pp. 2421–2427, 2018, doi: 10.1007/s00500-017-2512-z.

[43] G. D'Angelo, R. Pilla, C. Tascini, and S. Rampone, "A proposal for distinguishing between bacterial and viral meningitis using genetic programming and decision trees," *Soft Comput.*, vol. 23, no. 22, pp. 11775–11791, 2019, doi: 10.1007/s00500-018-03729-y.

[44] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, Sep. 2015, doi: 10.1007/s12544-015-0170-8.

[45] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 42, no. 2, pp. 513–529, Apr. 2012, doi: 10.1109/TSMCB.2011.2168604.

[46] H. F. Yang, T. S. Dillon, E. Chang, and Y. P. P. Chen, "Optimized Configuration of Exponential Smoothing and Extreme Learning Machine for Traffic Flow Forecasting," *IEEE Trans. Ind. Informatics*, vol. 15, no. 1, pp. 23–34, Jan. 2019, doi: 10.1109/TII.2018.2876907.

[47] N. Chiang, L. Tam, K. Lai, K. Wong, and W. Si Tou, "Floating Car Data-Based Real-Time Road Traffic Prediction System and Its Application in Macau Grand Prix Event," in *Intelligent Transport Systems for Everyone's Mobility*, Springer Singapore, 2019, pp. 377–392. doi: 10.1007/978-981-13-7434-0_21.

[48] Y. Xing, X. Ban, X. Liu, and Q. Shen, "Large-Scale Traffic Congestion Prediction Based on the Symmetric Extreme Learning Machine Cluster Fast Learning Method," *Symmetry 2019, Vol. 11, Page 730*, vol. 11, no. 6, p. 730, May 2019, doi: 10.3390/SYM11060730.

[49] Z. Ma, G. Luo, and D. Huang, "Short term traffic flow prediction based on on-line sequential extreme learning machine," *Proc. 8th Int. Conf. Adv. Comput. Intell. ICACI 2016*, vol. 2, no. 2, pp. 143–149, 2016, doi: 10.1109/ICACI.2016.7449818.

[50] W. Feng, H. Chen, and Z. Zhang, "Short-term traffic flow prediction based on wavelet function and extreme learning machine," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, Apr. 2018, vol. 2017-November, pp. 531–535. doi: 10.1109/ICSESS.2017.8342971.

[51] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, no. 2, pp. 114–121, Feb. 2006, doi: 10.1061/(ASCE)0733-947X(2006)132:2(114).

[52] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. Part C Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017, doi: 10.1016/j.trc.2017.02.024.

[53] M. Zeynoddin, H. Bonakdari, A. Azari, I. Ebtehaj, B. Gharabaghi, and H. Riahi Madavar, "Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate," *J. Environ. Manage.*, vol. 222, pp. 190–206, Sep. 2018, doi: 10.1016/j.jenvman.2018.05.072.

[54] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.

[55] S. Poornima, M. Pushpalatha, and U. Aglawe, "Predictive analytics using extreme learning machine," *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 7, pp. 1959–1966, 2018.

[56] Lawrence A. Klein, M. K. Mills, and D. R. P. Gibson, "Traffic Detector Handbook: Third Edition—Volume II," United States. Federal Highway Administration, Oct. 2006.

[57] Y. Wei, H. Huang, B. Chen, B. Zheng, and Y. Wang, "Application of Extreme Learning Machine for Predicting Chlorophyll-A Concentration Inartificial Upwelling Processes," *Math. Probl. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/8719387.

[58] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–44, 2006, doi: 10.1109/MCAS.2006.1688199.

[59] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 415–425, Mar. 2002, doi: 10.1109/72.991427.

[60] J. Cao and Z. Lin, "Extreme Learning Machines on High Dimensional and Large Data Applications: A Survey," *Math. Probl. Eng.*, vol. 2015, pp. 16–18, 2015, doi: 10.1155/2015/103796.

[61] M. Zeynoddin, H. Bonakdari, I. Ebtehaj, F. Esmaeilbeiki, B. Gharabaghi, and D. Zare Haghi, "A reliable linear stochastic daily soil temperature forecast model," *Soil Tillage Res.*, vol. 189, pp. 73–87, Jun. 2019, doi: 10.1016/j.still.2018.12.023.

[62] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, Dec. 1974, doi: 10.1109/TAC.1974.1100705.

[63] J. Whitelegg and G. Haq, "Vision Zero : Adopting a Target of Zero for Road Traffic Fatalities and Serious Injuries," 2006.

[64] "Statistiques des accidents." http://www.equipement.gov.ma/Infrastructures-Routieres/Reseau-Routier-du-Royaume/Pages/Statistiques-des-accidents.aspx (accessed Apr. 11, 2022).

[65] Institute for Health Metrics and Evaluation, "Death rate from road injuries, 1990 to 2019," 2019. https://ourworldindata.org/grapher/death-rates-road-incidents?tab=chart&country=MAR~DZA~TUN~FRA~CHN~USA~GBR (accessed Sep. 09, 2020).

[66] Z. Ali and S. B. Bhaskar, "Basic statistical tools in research and data analysis," *Indian Journal of Anaesthesia*, vol. 60, no. 9. Indian Society of Anaesthetists, pp. 662–669, Sep. 01, 2016. doi: 10.4103/0019-5049.190623.

[67] A. S. Houssaini, M. A. Sabri, H. Qjidaa, and A. Aarab, "Real-time driver's hypovigilance detection using facial landmarks," *2019 Int. Conf. Wirel. Technol. Embed. Intell. Syst. WITS 2019*, Apr. 2019, doi: 10.1109/WITS.2019.8723768.

[68] S. Allach, B. Benamrou, M. Ben Ahmed, A. A. Boudhir, and M. Ouardouz, "A new architecture based on ARIMA models for the safety classification of inter-city routes using meteorological metrics," in *Proceedings of the 4th International Conference on Smart City Applications - SCA '19*, Oct. 2019, pp. 1–9. doi: 10.1145/3368756.3369067.

[69] A. M. Addi, A. Tarik, and G. Fatima, “An approach based on association rules mining to improve road safety in Morocco,” in *2016 International Conference on Information Technology for Organizations Development, IT4OD 2016*, May 2016, pp. 1–6. doi: 10.1109/IT4OD.2016.7479311.

[70] N. Laaraj, S. Boutahari, and F. Jawab, “The Road Safety Information Systems Appropriate to the Systemic Approach: The Case of Morocco,” in *Colloquium in Information Science and Technology, CIST*, Dec. 2018, vol. 2018-October, pp. 74–79. doi: 10.1109/CIST.2018.8596590.

[71] V. Gitelman and E. Doveh, “A comparative evaluation of the safety performance of bus priority route configurations,” *Eur. Transp. Res. Rev.*, vol. 15, no. 1, pp. 1–16, Dec. 2023, doi: 10.1186/S12544-023-00589-Y/FIGURES/4.

[72] A. Salehian, K. Aghabayk, M. A. Seyfi, and N. Shiwakoti, “Comparative analysis of pedestrian crash severity at United Kingdom rural road intersections and Non-Intersections using latent class clustering and ordered probit model,” *Accid. Anal. Prev.*, vol. 192, p. 107231, Nov. 2023, doi: 10.1016/J.AAP.2023.107231.

[73] H. Ospina-Mateus, L. Quintana Jiménez, and F. J. López-Valdés, “Analyzing traffic conflicts and the behavior of motorcyclists at unsignalized three-legged and four-legged intersections in Cartagena, Colombia,” *Accid. Anal. Prev.*, vol. 191, p. 107222, Oct. 2023, doi: 10.1016/J.AAP.2023.107222.

[74] Y. Wan, W. He, and J. Zhou, “Urban road accident black spot identification and classification approach: A novel grey Verhuls–empirical Bayesian combination method,” *Sustain.*, vol. 13, no. 20, Oct. 2021, doi: 10.3390/SU132011198.

[75] D. Ennajih, M. Elgameh, A. Salik, A. Echchelh, and A. Chaouch, “Study of the Influence of Parameters of Road Safety on the Road Accidents,” *Int. J. Res. Stud. Sci. Eng. Technol.*, vol. 2, no. 4, pp. 22–27, 2015.

[76] L. Liu, “Biostatistical Basis of Inference in Heart Failure Study,” in *Heart Failure: Epidemiology and Research Methods*, Elsevier, 2018, pp. 43–82. doi: 10.1016/b978-0-323-48558-6.00004-9.

[77] D. Wang, Q. Liu, L. Ma, Y. Zhang, and H. Cong, “Road traffic accident severity analysis: A census-based study in China,” *J. Safety Res.*, vol. 70, pp. 135–147, 2019, doi: 10.1016/j.jsr.2019.06.002.

[78] H. Hayakawa, P. S. Fischbeck, and B. Fischhoff, “Traffic accident statistics and risk perceptions in Japan and the United States,” *Accid. Anal. Prev.*, vol. 32, no. 6, pp. 827–835, Nov. 2000, doi: 10.1016/S0001-4575(00)00007-5.

[79] J. De Oña, G. López, R. Mujalli, and F. J. Calvo, “Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks,” *Accid. Anal. Prev.*, vol. 51, pp. 1–10, 2013, doi: 10.1016/j.aap.2012.10.016.

[80] O. J. Dunn, “Multiple Comparisons among Means,” *J. Am. Stat. Assoc.*, vol. 56, no. 293, pp. 52–64, 1961, doi: 10.1080/01621459.1961.10482090.

[81] A. Dinno, “Nonparametric pairwise multiple comparisons in independent groups using Dunn’s test,” *Stata J.*, vol. 15, no. 1, pp. 292–300, Apr. 2015, doi: 10.1177/1536867x1501500117.

[82] D. Lakens, “Calculating and reporting effect sizes to facilitate cumulative science: A practical

primer for t-tests and ANOVAs," *Front. Psychol.*, vol. 4, no. NOV, 2013, doi: 10.3389/fpsyg.2013.00863.

[83] M. Tomczak and E. Tomczak, "The need to report effect size estimates revisited. An overview of some recommended measures of effect size," 2014.

[84] "The 20% Statistician: Why you should use omega-squared instead of eta-squared." https://daniellakens.blogspot.com/2015/06/why-you-should-use-omega-squared.html#comment-form (accessed May 04, 2020).

[85] K. O. McGraw and S. P. Wong, "A common language effect size statistic.," *Psychol. Bull.*, vol. 111, no. 2, pp. 361–365, Mar. 1992, doi: 10.1037/0033-2909.111.2.361.

[86] J. M. Pardillo-Mayora, C. A. Domínguez-Lira, and R. Jurado-Piña, "Empirical calibration of a roadside hazardousness index for Spanish two-lane rural roads," *Accid. Anal. Prev.*, vol. 42, no. 6, pp. 2018–2023, Nov. 2010, doi: 10.1016/j.aap.2010.06.012.

[87] S. Ladi, D. C. Wijeyesekera, A. J. Brimicombe, and Y. Li, "Road traffic accident hotspot identification using modified Voronoi process," in *Proceedings of Advances in Computing and Technology*, 2009, pp. 189–198.

[88] N. Srinivasan, V. S. Iyer, M. Chand, and K. Srinath, "Scientific identification and improvement of accident prone locations on national highways in Kerala," *J. Indian Roads Congr.*, vol. 48, no. 3, pp. 477–525, 1987.

[89] P. R. Vyas, M. L. Honnappanavar, and B. H B, "Identification of black spots for safe commuting using weighted severity index and GIS," *Int. J. Adv. Struct. Geotech. Eng.*, vol. 04, no. 01, pp. 2319–5347, 2015.

[90] B. Snehal U and P. Jalindar R, "Accidental Study on Pune - Solapur National Highway," *Int. J. Sci. Res.*, vol. 4, no. 6, pp. 2224–2228, 2015.

[91] J. De Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accid. Anal. Prev.*, vol. 43, no. 1, pp. 402–411, 2011, doi: 10.1016/j.aap.2010.09.010.

[92] "Ministère de l'Equipement et de l'Eau," 2023. http://www.equipement.gov.ma/Infrastructures-Routieres/Actualites/Pages/Actualites.aspx (accessed Jan. 12, 2023).

[93] "Ministère de l'Equipement et de l'Eau." https://www.equipement.gov.ma/Pages/accueil.aspx (accessed Nov. 20, 2021).

[94] M. Jiber, A. Mbarek, A. Yahyaouy, M. A. Sabri, and J. Boumhidi, "Road traffic prediction model using Extreme Learning Machine: the case study of Tangier, Morocco," *Inf. 2020, Vol. 11, Page 542*, vol. 11, no. 12, p. 542, Nov. 2020, doi: 10.3390/INFO11120542.

[95] M. Ghadi and Á. Török, "Comparison different black spot identification methods," *Transp. Res. Procedia*, vol. 27, pp. 1105–1112, 2017, doi: 10.1016/j.trpro.2017.12.104.

[96] H. Xin-min, L. Tao, Z. Tian-hua, H. Yong, X. Hai-wei, and Y. Tao, "Highway black spot recognition and improvement," *J. Traffic Transp. Eng.*, vol. 4, no. 1, pp. 106–109, 2004.

[97] W. Cheng and S. Washington, "New Criteria for Evaluating Methods of Identifying Hot Spots,"

Transp. Res. Rec., vol. 2083, no. 1, pp. 76–85, Jan. 2008, doi: 10.3141/2083-09.

[98] V. Ratanavaraha and D. Watthanaklang, "Road safety audit: identification of bus hazardous location in thailand," *Indian J. Sci. Technol.*, vol. 6, no. 8, pp. 5126–5133, 2013, doi: 10.17485/ijst/2013/v6i8.6.

[99] I. W. Agustin, "Analysis of Car Accident at the Location of Black-Spot and Rating for Accident-Prone Roads in Surabaya," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 328, no. 1, Oct. 2019, doi: 10.1088/1755-1315/328/1/012023.

[100] C. Zhang, Y. Shu, and L. Yan, "A novel identification model for road traffic accident black spots: a case study in Ningbo, China," *IEEE Access*, vol. 7, pp. 140197–140205, 2019, doi: 10.1109/ACCESS.2019.2942647.

[101] M. Q. Ghadi and Á. Török, "Evaluation of the Impact of Spatial and Environmental Accident Factors on Severity Patterns of Road Segments," *Period. Polytech. Transp. Eng.*, vol. 49, no. 2, pp. 146–155, 2021, doi: 10.3311/PPTR.14692.

[102] A. Iqbal, Z. U. Rehman, S. Ali, K. Ullah, and U. Ghani, "Road traffic accident analysis and identification of black spot locations on highway," *Civ. Eng. J.*, vol. 6, no. 12, pp. 2448–2456, Dec. 2020, doi: 10.28991/cej-2020-03091629.

[103] F. Calsavara, F. I. Kabbach, and A. P. C. Larocca, "Effects of fog in a brazilian road segment analyzed by a driving simulator for sustainable transport: drivers' speed profile under in-vehicle warning systems," *Sustainability*, vol. 13, no. 19, Sep. 2021, doi: 10.3390/su131910501.

[104] D. Saha, P. Alluri, E. Dumbaugh, and A. Gan, "Application of the Poisson-Tweedie distribution in analyzing crash frequency data," *Accid. Anal. Prev.*, vol. 137, Mar. 2020, doi: 10.1016/J.AAP.2020.105456.

[105] B. Debrabant, U. Halekoh, W. H. Bonat, D. L. Hansen, J. Hjelmborg, and J. Lauritsen, "Identifying traffic accident black spots with Poisson-Tweedie models," *Accid. Anal. Prev.*, vol. 111, pp. 147–154, Feb. 2018, doi: 10.1016/J.AAP.2017.11.021.

[106] H. H. Nguyen, P. Taneerananon, and P. Luathep, "Approach to Identifying Black Spots Based on Potential Saving in Accident Costs," *Eng. J.*, vol. 20, no. 2, pp. 109–122, May 2016, doi: 10.4186/ej.2016.20.2.109.

[107] H. Zhu, Y. Zhou, and Y. Chen, "Identification of potential traffic accident hot spots based on accident data and GIS," *MATEC Web Conf.*, vol. 325, p. 01005, 2020, doi: 10.1051/MATECCONF/202032501005.

[108] D. Babić, M. Fiolić, D. Babić, and T. Gates, "Road Markings and Their Impact on Driver Behaviour and Road Safety: A Systematic Review of Current Findings," *J. Adv. Transp.*, vol. 2020, 2020, doi: 10.1155/2020/7843743.

[109] B. Du *et al.*, "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 972–985, 2020, doi: 10.1109/TITS.2019.2900481.

[110] J. Ambros, Z. Křivánková, R. Zůvala, K. Bucsuházy, and J. Frič, "ANALYSIS OF SAFETY IMPACT OF PAVED SHOULDER WIDTH ON CZECH SECONDARY ROADS," *Arch. Transp.*, vol. 60, no. 4, pp. 125–136, 2021, doi: 10.5604/01.3001.0015.6148.

[111] X. Liu and J. (Cecilia) Xia, "Locally analysing the risk factors for fatal single vehicle crashes hot spots in Western Australia," *Int. J. Crashworthiness*, vol. 20, no. 6, pp. 524–534, Nov. 2015, doi: 10.1080/13588265.2015.1055649.

[112] L. Singh Bisht and G. Tiwari, "Assessing the Effect of Paved Shoulder Width on Road Traffic Crash Risk on a Six-Lane Rural Highway in India," *J. Transp. Eng. Part A Syst.*, vol. 149, no. 7, Jul. 2023, doi: 10.1061/JTEPBS.TEENG-7296.

[113] J. W. Van Treese, A. K. Koeser, G. E. Fitzpatrick, M. T. Olexa, and E. J. Allen, "A review of the impact of roadway vegetation on drivers' health and well-being and the risks associated with single-vehicle crashes," *Arboric. J.*, vol. 39, no. 3, pp. 179–193, Jul. 2017, doi: 10.1080/03071375.2017.1374591.

[114] T. Ben-Bassat and D. Shinar, "Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior," *Accid. Anal. Prev.*, vol. 43, no. 6, pp. 2142–2152, Nov. 2011, doi: 10.1016/J.AAP.2011.06.004.

[115] A. Calvi, S. D. Cafiso, C. D'Agostino, M. Kieć, and G. Petrucci, "A driving simulator study to evaluate the effects of different types of median separation on driving behavior on 2 + 1 roads," *Accid. Anal. Prev.*, vol. 180, Feb. 2023, doi: 10.1016/J.AAP.2022.106922.

[116] M. L. Siregar, R. Jachrizal Sumabrata, A. Kusuma, O. B. Samosir, and S. N. Rudrokasworo, "Analyzing Driving Environment Factors in Pedestrian Crashes Injury Levels in Jakarta and the Surrounding Cities," *J. Appl. Eng. Sci.*, vol. 17, no. 4, pp. 482–489, Dec. 2019, doi: 10.5937/JAES17-22121.

[117] Ş. C. Gherghina, M. Onofrei, G. Vintilă, and D. Ş. Armeanu, "Empirical Evidence from EU-28 Countries on Resilient Transport Infrastructure Systems and Sustainable Economic Growth," *Sustain. 2018, Vol. 10, Page 2900*, vol. 10, no. 8, p. 2900, Aug. 2018, doi: 10.3390/SU10082900.

[118] M. T. Ashraf, K. Dey, and S. Mishra, "Identification of high-risk roadway segments for wrong-way driving crash using rare event modeling and data augmentation techniques," *Accid. Anal. Prev.*, vol. 181, Mar. 2023, doi: 10.1016/J.AAP.2022.106933.

[119] M. L. Siregar, T. Tjahjono, and N. Yusuf, "Predicting the Segment-Based Effects of Heterogeneous Traffic and Road Geometric Features on Fatal Accidents," *Int. J. Technol.*, vol. 13, no. 1, pp. 92–102, 2022, doi: 10.14716/IJTECH.V13I1.4450.

[120] M. L. Siregar, T. Tjahjono, and Nahry, "Endogenous relationship of accident occurrence with speed, traffic heterogeneity and driving environment on inter-urban roads in Indonesia," *J. Appl. Eng. Sci.*, vol. 18, no. 4, pp. 475–484, 2020, doi: 10.5937/JAES0-25837.

[121] C. Gu, J. Xu, C. Gao, M. Mu, E. Guangxun, and Y. Ma, "Multivariate analysis of roadway multi-fatality crashes using association rules mining and rules graph structures: A case study in China," *PLoS One*, vol. 17, no. 10 October, Oct. 2022, doi: 10.1371/JOURNAL.PONE.0276817.

[122] F. Zeinab, K. Ali, D. Bassam, and C. Pierre, "An Improved Approach to Analyze Accidents and Promote Road Safety using Association Rule Mining and Multi-Criteria Decision Analysis Methods," *Recent Adv. Comput. Sci. Commun.*, vol. 13, no. 4, pp. 731–746, Aug. 2019, doi: 10.2174/2213275912666190807113914.

[123] C. Xu, J. Bao, C. Wang, and P. Liu, "Association rule analysis of factors contributing to

extraordinarily severe traffic crashes in China," *J. Safety Res.*, vol. 67, pp. 65–75, Dec. 2018, doi: 10.1016/J.JSR.2018.09.013.

[124] A. Das, M. M. Ahmed, and A. Ghasemzadeh, "Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach," *Accid. Anal. Prev.*, vol. 129, pp. 250–262, Aug. 2019, doi: 10.1016/J.AAP.2019.05.024.

[125] M. Shahin *et al.*, "Cluster-Based Association Rule Mining for an Intersection Accident Dataset," *2021 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2021 - Proc.*, 2021, doi: 10.1109/ICECUBE53880.2021.9628206.

[126] X. Qu and Q. Meng, "A note on hotspot identification for urban expressways," *Saf. Sci.*, vol. 66, pp. 87–91, Jul. 2014, doi: 10.1016/J.SSCI.2014.02.006.

[127] A. Montella, "A comparative analysis of hotspot identification methods," *Accid. Anal. Prev.*, vol. 42, no. 2, pp. 571–581, Mar. 2010, doi: 10.1016/J.AAP.2009.09.025.

[128] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "A new model for black spots identification using Weighted Severity Index," in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Oct. 2021, pp. 1–7. doi: 10.1109/ICDS53782.2021.9626758.

[129] M. Ghadi and Á. Török, "Comparison different black spot identification methods," *Transp. Res. Procedia*, vol. 27, pp. 1105–1112, Jan. 2017, doi: 10.1016/J.TRPRO.2017.12.104.

[130] R. Agrawal, T. Imieliński, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207–216, Jan. 1993, doi: 10.1145/170036.170072.

[131] M. Harahap, A. M. Husein, S. Aisyah, F. R. Lubis, and B. A. Wijaya, "Mining association rule based on the diseases population for recommendation of medicine need," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, Apr. 2018, doi: 10.1088/1742-6596/1007/1/012017.

[132] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Black spots identification on rural roads based on extreme learning machine," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 3149–3160, Jun. 2023, doi: 10.11591/IJECE.V13I3.PP3149-3160.

[133] Y. Dai, Y. Bian, Y. Wu, X. Zhao, and J. Zhang, "A variable speed limit model for freeway bridges considering the influence of both traffic service level and visibility in fog," *Phys. A Stat. Mech. its Appl.*, vol. 602, Sep. 2022, doi: 10.1016/J.PHYSA.2022.127590.

[134] Y. Zou *et al.*, "Analysis of the effects of wind barrier on driving safety and comfort of vehicles on long-span bridges under crosswinds," *Structures*, vol. 42, pp. 367–385, Aug. 2022, doi: 10.1016/J.ISTRUC.2022.05.098.

[135] X. Lyu and Q. Wang, "Analysis of Traffic Accident Severity on Circular Expressway Based on Logistic," *Wuhan Ligong Daxue Xuebao (Jiaotong Kexue Yu Gongcheng Ban)/Journal Wuhan Univ. Technol. (Transportation Sci. Eng.*, vol. 46, no. 3, pp. 411–415, Jun. 2022, doi: 10.3963/J.ISSN.2095-3844.2022.03.006.

[136] A. B. Sołowczuk and D. Kacprzak, "Effect of Lane Narrowing Implemented on an Urban Dual Carriageway on Speed Reduction and Acoustic Environment," *Build. 2022, Vol. 12, Page 31*, vol. 12, no. 1, p. 31, Dec. 2021, doi: 10.3390/BUILDINGS12010031.

[137] B. Sharma, H. K. Nam, W. Yan, and H. Y. Kim, “Barriers and Enabling Factors Affecting Satisfaction and Safety Perception with Use of Bicycle Roads in Seoul, South Korea,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 5, Mar. 2019, doi: 10.3390/IJERPH16050773.

[138] S. Reed and A. Morris, “Characteristics of fatal single-vehicle crashes in Europe,” *Int. J. Crashworthiness*, vol. 17, no. 6, pp. 655–664, Dec. 2012, doi: 10.1080/13588265.2012.709394.

[139] S. Barman and R. Bandyopadhyaya, “Crash severity analysis for low-speed roads using structural equation modeling considering shoulder- and pavement-distress conditions,” *J. Transp. Eng. Part A Syst.*, vol. 146, no. 7, Jul. 2020, doi: 10.1061/JTEPBS.0000373.

[140] N. Katre, N. H. Pitale, and S. Bobade, “Analysis of Black Spots on NH-3 and Its Rectification,” *J. Transp. Syst.* , vol. 4, no. 2, pp. 16–32, May 2019, doi: 10.5281/ZENODO.2908866.

[141] S. Yasmin, N. Eluru, J. Lee, and M. Abdel-Aty, “Ordered fractional split approach for aggregate injury severity modeling,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2583, no. 1, pp. 119–126, Jan. 2016, doi: 10.3141/2583-15.

[142] N. Stamatiadis, J. G. Pigman, D. Lord, J. Sacksteder, and W. Ruff, *Impact of Shoulder Width and Median Width on Safety*. Washington, D.C.: National Academies Press, 2009. doi: 10.17226/14252.

[143] V. Gitelman, E. Doveh, R. Carmel, and S. Hakkert, “The influence of shoulder characteristics on the safety level of two-lane roads: A case-study,” *Accid. Anal. Prev.*, vol. 122, pp. 108–118, Jan. 2019, doi: 10.1016/j.aap.2018.10.003.

[144] G. Sugiyanto, Y. A. Prasetyo, E. W. Indriyati, Y. Yanto, and M. Y. Santi, “Determining of black spot location in Purbalingga Regency using road geometric approach,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 847, no. 1, Sep. 2021, doi: 10.1088/1755-1315/847/1/012019.

[145] S. Vadlamani, E. Chen, S. Ahn, and S. Washington, “Identifying Large Truck Hot Spots Using Crash Counts and PDOEs,” *J. Transp. Eng.*, vol. 137, no. 1, pp. 11–21, Jun. 2010, doi: 10.1061/(ASCE)TE.1943-5436.0000183.

[146] B. Dadashova, C. Silvestri-Dobrovolny, J. Chauhan, M. Perez, and R. Bligh, “Hot-spot analysis of motorcyclist crashes involving fixed objects using multinomial logit and data mining tools,” *J. Transp. Saf. Secur.*, 2021, doi: 10.1080/19439962.2021.1898070.

[147] E. de la Peña González, “Design of a model for the identification and analysis of road sections without accidents: a new vision of road safety (in Spanish: Diseño de un modelo para la identificación y análisis de tramos de carreteras sin accidentes : una nueva visión de la seguri,” Universidad Politecnica de Madrid - University Library, 2015. doi: 10.20868/UPM.thesis.38774.

Liste des publications

Revues internationales :

A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Accident black spots identification based on association rule mining," Bull. Electr. Eng. Informatics, vol. 13, no. 3, pp. 2075–2085, Jun. 2024, doi: 10.11591/EEI.V13I3.6135.

A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Black spots identification on rural roads based on extreme learning machine," Int. J. Electr. Comput. Eng., vol. 13, no. 3, pp. 3149–3160, Jun. 2023, doi: 10.11591/IJECE.V13I3.PP3149-3160.

M. Jiber, A. Mbarek, A. Yahyaouy, M. A. Sabri, and J. Boumhidi, "Road traffic prediction model using Extreme Learning Machine: the case study of Tangier, Morocco," *Inf. 2020, Vol. 11, Page 542*, vol. 11, no. 12, p. 542, Nov. 2020, doi: 10.3390/INFO11120542.

Conférences internationales :

A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "A new model for black spots identification using Weighted Severity Index," in 2021 Fifth International Conference on Intelligent Computing in Data Sciences (ICDS), Oct. 2021, pp. 1–7. doi: 10.1109/ICDS53782.2021.9626758.

A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Road traffic mortality in Morocco: Analysis of statistical data," in 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Jun. 2020, pp. 1–7. doi: 10.1109/ISCV49265.2020.9204325.

Chapitre :

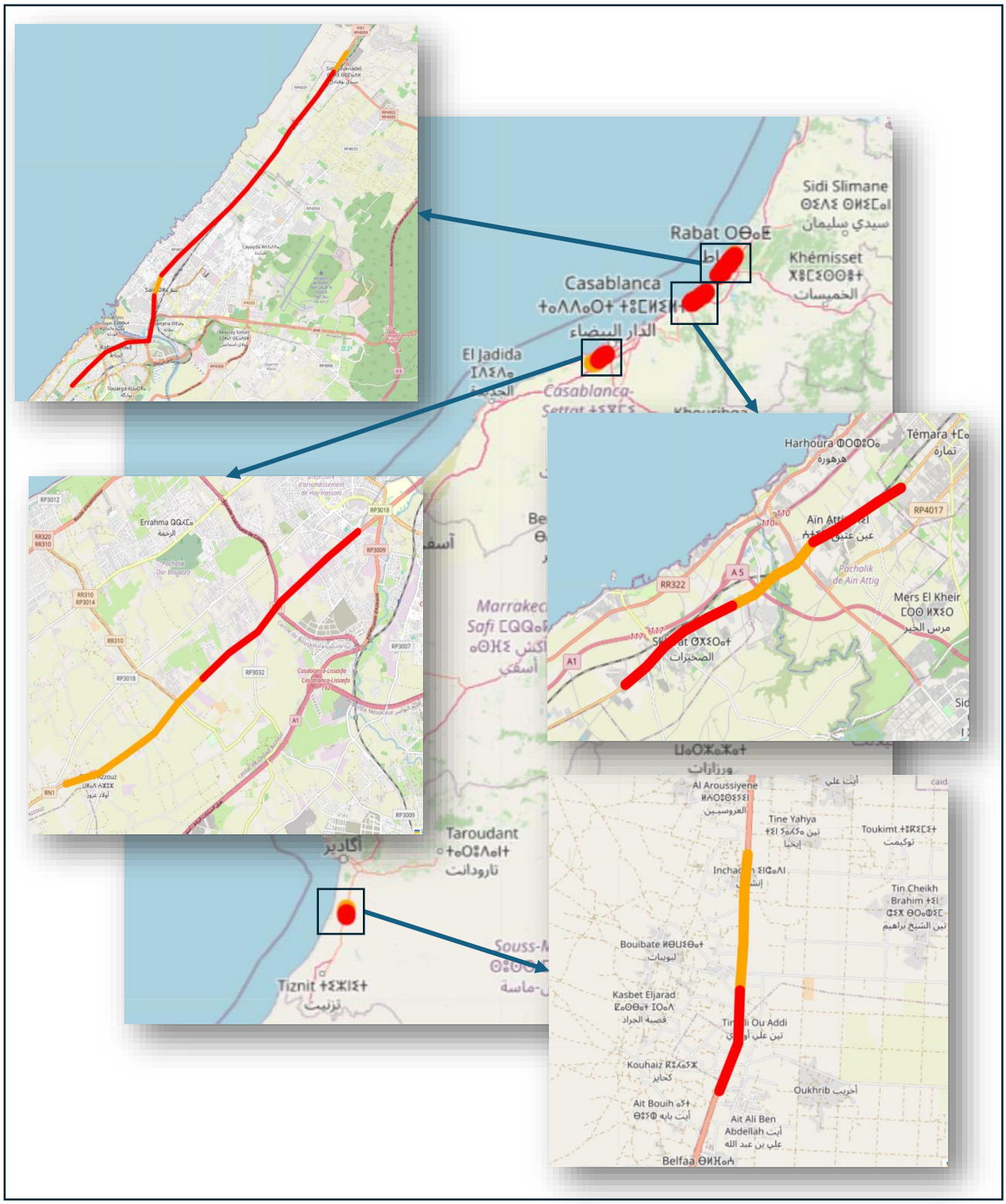
A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Road-Traffic Data Collection: Handling Missing Data," in Explainable Artificial Intelligence for Intelligent Transportation Systems, 1st ed., A. Adadi and A. Bouhoute, Eds. CRC Press, 2023, pp. 119–134. doi: 10.1201/9781003324140-6.

Présentation en conférence :

A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Identification of black spots in rural areas using extreme learning machine," in jun, 2022. 1st National Doctoral Conference, FLSH Meknès.

Annexes

Annexe A. Carte affichant l'ensemble des points noirs consécutifs : les sections de niveau de gravité élevée sont en rouge et les sections de gravité moyenne sont en orange.



Annexe B. Exemple d'un fichier généré par un compteur permanant.

Site identifier	000000000137											
Site reference	00000137											
Title	RN7 PK 132 MARRAKECH											
Start	Thu 08 Dec 2016 12:00											
End	Mon 02 Jan 2017 11:00											
Interval	01:00											
		#1										
Date	Time	LT	Tot	<52	<65	<105	105+	#2 Tot	<52	<65	<105	105+
				0.1m	0.1m	0.1m	0.1m		0.1m	0.1m	0.1m	0.1m
Thu 08 Dec 2016	12:00	6187.50	321	267	12	14	28	259	239	9	4	7
	13:00	6187.54	321	267	12	14	28	259	239	9	4	7
	14:00	6187.58	323	276	10	17	20	255	240	6	6	3
	15:00	6187.63	327	276	10	14	27	228	209	5	9	5
	16:00	6187.67	326	276	11	16	23	296	281	4	5	6
	17:00	6187.71	292	248	8	18	18	429	390	12	13	14
	18:00	6187.75	283	243	8	10	22	482	457	8	8	9
	19:00	6187.79	236	202	7	8	19	320	297	13	3	7
	20:00	6187.83	214	182	9	11	12	254	238	6	1	9
	21:00	6187.88	146	137	3	3	3	180	176	2	0	2
	22:00	6187.92	127	112	6	6	3	122	121	1	0	0