

UNIVERSITÉ DE NGAOUNDÉRE

FACULTÉ DES SCIENCES

*DÉPARTEMENT DE
MATHÉMATIQUES ET
INFORMATIQUE*



THE UNIVERSITY OF
NGAOUNDERE

*FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE*

**ÉCOLE DOCTORALE : SCIENCES, TECHNOLOGIE ET
INGÉNIERIE (STI)**

**UFD : MATHÉMATIQUES, INFORMATIQUE, INGÉNIERIE ET
APPLICATION (M2IAP)**

Deep Analysis and prediction of chikungunya using ensemble regression approach

Mémoire en vue d'obtention du Master II

Filière : Informatique

Spécialité : Systèmes et Logiciels en Environnements distribués

Par

MOHAMED EL BACHIR BOUBA NGANADAKOUA

(Master I Systèmes et Logiciels en Environnements distribués)

Matricule : 19A666FS

Sous la direction de :

PR. DR. ING DAYANG PAUL

Maître de Conférence

Faculté des Sciences

Université de Ngaoundéré

DR. ABBOUBAKAR HAMADJAM

Chargé de Cours

Institut Universitaire de Technologie

Université de Ngaoundéré

Année académique 2023-2024

Dédicace

*à mon père BOUBA Nana Dekwa,
à ma mère HABIBA Kouyessi,
à mes sœurs.*

Remerciements

La réalisation de ce mémoire a été un long voyage, parsemé de défis et d'apprentissages, et n'aurait pas été possible sans le soutien et l'assistance de nombreuses personnes. C'est avec une profonde gratitude que je tiens à exprimer mes sincères remerciements à tous ceux qui ont contribué, de près ou de loin, à la concrétisation de ce travail.

- Tout d'abord, je remercie Allah le Tout-Puissant qui m'a donné la volonté et la force nécessaires pour parfaire ce travail et le mener à terme ;
- Le Doyen de la Faculté des Sciences de l'Université de Ngaoundéré, le Professeur NGAMENI Emmanuel, pour son soutien et ses encouragements ;
- Le Chef du Département de Mathématiques et Informatique de la Faculté des Sciences de l'Université de Ngaoundéré : le Pr. Dr. Ing DAYANG Paul pour le suivi de notre formation et la supervision de ce travail ;
- Mon directeur de mémoire, Dr. ABOUBAKAR Hamadjam, Chargé de Cours à l'Université de Ngaoundéré, au Département de Génie Informatique à l'Institut Universitaire de Technologie, pour m'avoir proposé ce sujet et pour ses précieux conseils ;
- Les enseignants du Département de Mathématiques et Informatique de la Faculté des Sciences de l'Université de Ngaoundéré, ainsi que les membres du jury, pour avoir accepté d'examiner ce travail ;
- Ma famille, pour leurs nombreux conseils, encouragements et soutien tout au long de mes études ;
- Mes camarades de promotion, pour la convivialité, le partage de connaissances et l'entraide dont nous avons bénéficié ensemble.

À tous, un grand merci.

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste de figures	vi
Liste des tableaux	vii
Notation et abréviations	viii
Résumé	ix
Abstract	x
General Introduction	0
1 Épidémiologie de la Chikungunya	1
1.1 Origine de la Chikungunya	1
1.2 Agent Pathogène	2
1.3 Mode de Transmission	2
1.3.1 Vecteurs : Les moustiques Aedes	3
1.3.2 La transmission	3
1.4 Symptômes et Diagnostic	4
1.4.1 Symptômes	4
1.4.2 Méthodes de diagnostic	4
1.5 Méthodes de Contrôle et Traitement	4
1.5.1 Méthodes de contrôle	4
1.5.2 Options de traitement	5
1.6 Cas du Tchad	5
1.6.1 Point saillants	6
1.6.2 Contexte	6

2	Revue de la Littérature et Concepts de Base	8
2.1	Introduction	8
2.2	Apprentissage Automatique	8
2.2.1	Techniques d'Apprentissage Automatique	9
2.2.1.1	Apprentissage Supervisé	9
2.2.1.2	Apprentissage Non Supervisé	10
2.2.1.3	Apprentissage Semi-Supervisé	10
2.2.1.4	Apprentissage par Ensemble	10
2.2.2	Applications de l'Apprentissage Automatique à l'Épidémiologie	11
2.2.3	Avantages et Limites de l'Apprentissage Automatique	11
2.3	Deep Learning	11
2.3.1	Définition et Concepts Fondamentaux	11
2.3.2	Réseaux de Neurones Artificiels	11
2.3.2.1	Perceptron Multicouche (MLP)	11
2.3.2.2	Réseaux de Neurones Convolutifs (CNN)	11
2.3.2.3	Réseaux de Neurones Récurrents (RNN)	11
2.3.3	Applications du Deep Learning à l'Épidémiologie	11
2.3.4	Avantages et Limites du Deep Learning	11
2.4	Ensemble Regression Learning	11
2.4.1	Définition et Concepts Fondamentaux	11
2.4.2	Techniques d'Ensemble Learning	11
2.4.2.1	Bagging	11
2.4.2.2	Boosting	11
2.4.2.3	Stacking	11
2.4.3	Applications de l'Ensemble Regression Learning	11
2.4.4	Avantages et Limites de l'Ensemble Regression Learning	11
2.5	Conclusion	11
3	Implémentation des Modèles	12
3.1	Introduction	12
3.2	Présentation des Outils Utilisés	12
3.2.1	Langages et bibliothèques	12
3.2.2	Infrastructure matérielle	12
3.3	Méthodes de l'Apprentissage Automatique	12
3.3.1	Le jeu de données	12
3.3.2	Préparation et nettoyage des données	12
3.3.3	Implémentation des modèles de machine learning	12
3.4	Méthodes de l'Apprentissage Profond	12
3.4.1	Le jeu de données	13
3.4.2	Préparation des données pour le deep learning	13
3.4.3	Implémentation des modèles de deep learning	13

General Conclusion	15
Bibliographie	16

Liste de figures

1.1	CHIKV in the Western Hemisphere.	2
1.2	Electron microscopic image of chikungunya virus	2
1.3	Aedes aegypti mosquito full of blood	3
1.4	Transmission mode Chikungunya	3
1.5	Region infecté au Tchad	6
1.6	Evolution journalière des cas et décès du Chikungunya	7
2.1	9
2.2	Flux de travail de l'apprentissage supervisé	9
2.3	Apprentissage Non Supervisé	10

Liste des tableaux

Notation et abréviations

CHIKV	Chikungunya virus
WHO	World Health Organization
PAHO	Pan American Health Organization
OMS	Organisation Mondiale de la Sante
CDC	Centers for Disease Control and Prevention
Ae	Aedes

Résumé

Le virus du chikungunya, transmis principalement par les moustiques **Aedes aegypti** et **Aedes albopictus**, représente une menace croissante pour la santé publique mondiale en raison de sa propagation rapide et de ses effets débilissants. Récemment, des épidémies ont été signalées non seulement en Afrique centrale et orientale, mais aussi en Amérique du Sud et en Asie du Sud-Est. Prédire ces épidémies reste un défi majeur en raison de l'interaction complexe entre les facteurs environnementaux, climatiques et biologiques. Les approches traditionnelles de surveillance épidémiologique se révèlent souvent insuffisantes pour anticiper les épidémies de manière proactive.

Cette recherche se concentre sur le **Tchad**, le **Brésil**, et le **Paraguay**, trois pays où le chikungunya a émergé comme un problème significatif de santé publique. En raison de la disponibilité limitée de données au Tchad, des données supplémentaires provenant du Brésil et du Paraguay sont intégrées pour renforcer l'analyse. Le but de cette thèse est de développer et d'évaluer des modèles prédictifs pour les épidémies de chikungunya en utilisant des techniques avancées d'apprentissage automatique, en particulier la régression d'ensemble.

Les modèles choisis pour cette étude incluent le **Random Forest Regressor**, le **XGBoost Regressor** optimisé via **Grid Search**, ainsi qu'un modèle d'ensemble (**Voting Regressor**) combinant **Linear Regression**, **Decision Tree Regressor**, et le **XGBoost Regressor** optimisé. Ces modèles seront formés et validés à partir de données épidémiologiques et climatiques.

Cette thèse commence par une analyse approfondie de l'épidémiologie du chikungunya et de ses dynamiques de transmission, suivie d'une revue des applications actuelles de l'intelligence artificielle dans la prédiction des maladies. Les modèles sont ensuite testés et évalués sur la base de critères de performance tels que la sensibilité, la spécificité, et la valeur prédictive.

Les résultats de cette recherche montrent l'efficacité des techniques de régression d'ensemble pour améliorer la précision des prévisions épidémiologiques, et offrent des perspectives nouvelles pour l'intervention en santé publique dans les régions affectées.

Mots-clés : *Chikungunya, Régression d'ensemble, Random Forest, XGBoost, Apprentissage Automatique, Données Climatiques, Prévision Épidémiologique*

Abstract

The chikungunya virus, transmitted mainly by the mosquitoes **Aedes aegypti** and **Aedes albopictus**, represents a growing threat to global public health due to its rapid spread and debilitating effects. Recently, epidemics have been reported not only in Central and East Africa, but also in South America and Southeast Asia. Predicting these epidemics remains a major challenge due to the complex interplay between environmental, climatic and biological factors. Traditional epidemiological surveillance approaches often prove insufficient to proactively anticipate epidemics.

This research focuses on **Chad, Brazil, and Paraguay**, three countries where chikungunya has emerged as a significant public health problem. Due to limited data availability in Chad, additional data from Brazil and Paraguay are incorporated to strengthen the analysis. The aim of this thesis is to develop and evaluate predictive models for chikungunya epidemics using advanced machine learning techniques, in particular ensemble regression.

The models chosen for this study include the **Random Forest Regressor**, the **XGBoost Regressor** optimized via **Grid Search**, as well as an ensemble model (**Voting Regressor**) combining **Linear Regression**, **Decision Tree Regressor**, and the optimized **XGBoost Regressor**. These models will be trained and validated using epidemiological and climatic data.

This thesis begins with an in-depth analysis of the epidemiology of chikungunya and its transmission dynamics, followed by a review of current applications of artificial intelligence in disease prediction. The models are then tested and evaluated on the basis of performance criteria such as sensitivity, specificity and predictive value.

The results of this research demonstrate the effectiveness of ensemble regression techniques in improving the accuracy of epidemiological forecasts, and offer new perspectives for public health intervention in affected regions.

Keywords : *Chikungunya, Ensemble regression, Random Forest, XGBoost, Machine learning, Climatic data, Epidemiological forecast*

Introduction Générale

Le chikungunya est une maladie virale transmise par la piqûre de moustiques infectés, principalement des espèces *Aedes aegypti* et *Aedes albopictus*. Endémique dans certaines régions d’Afrique de l’Est, d’Asie et plus récemment d’Amérique du Sud, le chikungunya se manifeste par des symptômes tels que fièvre, douleurs articulaires, maux de tête, douleurs musculaires, gonflements articulaires, et éruptions cutanées. Bien que les décès dus au chikungunya soient rares, le virus peut provoquer de graves complications, en particulier chez les personnes âgées ou celles souffrant de maladies chroniques. Une détection précoce est cruciale pour prévenir la propagation de la maladie.

L’objectif de ce mémoire est de développer un modèle prédictif du chikungunya, en utilisant des approches de régression d’ensemble issues de l’intelligence artificielle. Compte tenu du manque de données suffisantes pour le Tchad, nous avons étendu notre étude aux données du Brésil et du Paraguay, où des cas de chikungunya ont également été signalés. Cette approche nous permet de tirer parti d’un ensemble de données plus vaste et diversifié pour améliorer la précision de nos prévisions.

Nous nous posons donc la question suivante : comment utiliser les techniques d’apprentissage automatique, en particulier l’approche ensembliste, pour prédire et analyser les épidémies de chikungunya ?

Les objectifs spécifiques de ce travail sont les suivants :

- Comprendre les concepts liés à la régression d’ensemble et au chikungunya ;
- Prédire les épidémies de chikungunya en se basant sur les données climatiques et les cas rapportés au Tchad, au Brésil, et au Paraguay ;
- Analyser la dynamique du chikungunya dans ces régions et proposer des stratégies d’intervention.

Le reste du document est structuré comme suit : le premier chapitre est dédié à l’épidémiologie du chikungunya avec un accent particulier sur les cas au Tchad, au Brésil et au Paraguay. Le deuxième chapitre introduit l’apprentissage automatique dans le contexte des maladies infectieuses. Le troisième chapitre présente la conception générale et détaillée du modèle proposé. Dans le quatrième chapitre, nous illustrons l’efficacité de notre modèle à travers des résultats de simulation. Enfin, le mémoire se conclut par une synthèse des résultats et des perspectives pour des recherches futures.

Chapitre 1

Épidémiologie de la Chikungunya

Introduction

Le chikungunya est une maladie virale transmise à l'homme par des moustiques infectés par le virus du chikungunya. Les moustiques impliqués dans la transmission sont *Aedes aegypti* et *Aedes albopictus* [10].

1.1 Origine de la Chikungunya

La fièvre **Chikungunya** est une maladie virale décrite pour la première fois en 1952 lors d'une épidémie dans le sud de la **Tanzanie**. Le nom vient d'un mot de la langue *Makonde*, parlée dans le sud-est de la Tanzanie et le nord du Mozambique, qui signifie "*devenir contorsionné*" ou "*ce qui se plie*". Le virus a été isolé pour la première fois en Thaïlande en 1958.[13]

Évolution géographique et épidémiologique

En **avril 2005**, il a été confirmé que le CHIKV était à l'origine d'une épidémie de maladie ressemblant à la dengue sur les îles Comores, situées au large de la côte est du Mozambique septentrional ; il s'agissait de la première émergence connue du CHIKV dans la région du sud-ouest de l'océan Indien. En raison de similitudes cliniques, cette épidémie a d'abord été suspectée d'être causée par le virus de la dengue, soulignant le fait que la maladie CHIKV est souvent mal diagnostiquée et que le nombre réel de cas dans une région donnée peut être sous-estimé. Peu après, les premiers cas de CHIKV ont été signalés à Mayotte, à Maurice et sur l'île française de La Réunion. Le nombre de cas dans ces régions a rapidement augmenté, notamment en raison de taux d'attaque atteignant **35% à 75%**. À la fin de l'année **2005**, après une période apparente d'environ 32 ans pendant laquelle le CHIKV n'a pas été détecté, l'Inde a signalé des cas de maladie à CHIKV dans de nombreux États, le nombre officiel de cas suspects atteignant finalement plus de 1,3 million. L'épidémie de CHIKV a continué à se propager, provoquant d'importantes flambées au Sri Lanka et dans de nombreux autres pays d'Asie du Sud-Est. Au cours de cette épidémie, le CHIKV a été introduit dans des pays où il n'est pas endémique par des voyageurs virémiques, et la transmission autochtone du CHIKV a été observée pour la première fois dans de nombreux pays, dont l'Italie, la France, la Nouvelle-Calédonie, la Papouasie-Nouvelle-Guinée, le Bhoutan et le Yémen. La propagation rapide et explosive du CHIKV a incité l'Organisation panaméricaine de la santé

(OPS) et les Centers for Disease Control and Prevention (CDC) à publier un guide de préparation qui prévoyait de futures épidémies potentielles de CHIKV dans les Amériques. Cette prédiction s’est maintenant concrétisée, puisqu’en décembre 2013, l’Organisation mondiale de la santé (OMS) a signalé la première transmission locale du CHIKV dans l’hémisphère occidental, sur l’île caribéenne de Saint-Martin. Le 18 juillet 2014, le CHIKV avait provoqué plus de 440 000 cas de maladie dans plus de 20 pays des Caraïbes, d’Amérique centrale et d’Amérique du Sud (Fig. 1). En outre, les CDC ont signalé plus de 230 cas importés d’infection à CHIKV sur le territoire continental des États-Unis, ainsi que des cas acquis localement en Floride. Ainsi, en moins de 10 ans, le CHIKV s’est propagé depuis les côtes du Kenya dans l’océan Indien, le Pacifique et les Caraïbes, provoquant des millions de cas de maladie dans plus de 50 pays. En d’autres termes, le CHIKV est redevenu un véritable agent pathogène mondial.



FIGURE 1.1 – CHIKV in the Western Hemisphere.

1.2 Agent Pathogène

Le virus du chikungunya (CHIKV), un alphavirus transmis par les moustiques, dont le génome est constitué d’un ARN monocaténaire à sens positif de ~ 12 kb [9].

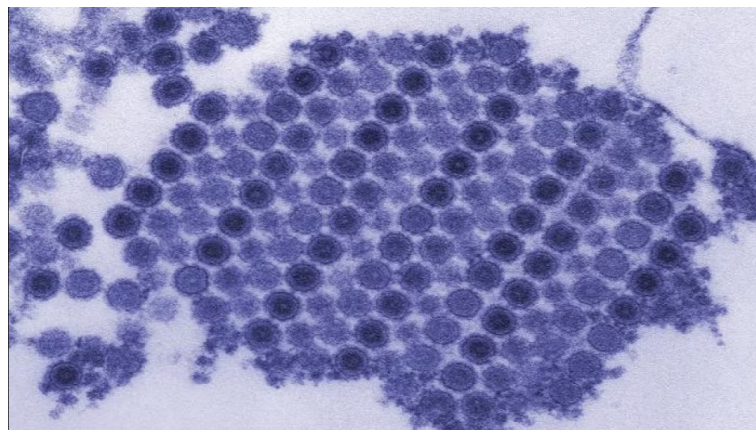


FIGURE 1.2 – Electron microscopic image of chikungunya virus

1.3 Mode de Transmission

Comprendre les mécanismes de transmission du virus chikungunya est essentiel pour développer des stratégies de prévention efficaces. Cette section examine les principaux vecteurs de la maladie, les conditions environne-

mentales qui favorisent la propagation du virus et les dynamiques de transmission entre les hôtes humains et animaux.

1.3.1 Vecteurs : Les moustiques Aedes

Le virus du chikungunya est un alphavirus, similaire aux virus de Mayaro et de Ross River, appartenant à la famille *Togaviridae*, au genre *Alphavirus*. Le virus du chikungunya est principalement transmis à l'homme par la piqûre d'un moustique infecté, principalement *Aedes aegypti* (Voir Figure 1.3) et *Ae. albopictus*.



FIGURE 1.3 – *Aedes aegypti* mosquito full of blood

1.3.2 La transmission

Le **CHIKV** se transmet selon deux cycles différents :

- **Cycle urbain** : transmission de l'homme au moustique.
- **Cycle sylvatique** : transmission de l'animal au moustique, puis à l'homme [5].

Le cycle sylvatique est la principale forme de transmission en Afrique [5]. Ailleurs, dans les zones plus densément peuplées, le CHIKV se maintient principalement dans un cycle urbain, dans lequel les humains sont les principaux hôtes et les moustiques du genre *Aedes* les vecteurs [5] (voir figure 1.4). bien que *Ae. aegypti* continue d'être un vecteur viral important, comme on l'a vu lors de la flambée épidémique dans les Caraïbes en 2013 [5]. La transmission verticale de la mère à l'enfant a été postulée pour expliquer les incidences postérieures à 2005 [5], étant particulièrement délétère lorsque :

- la mère est infectée jusqu'à quatre jours après l'accouchement [5],

bien que cette hypothèse ait été contestée [5].

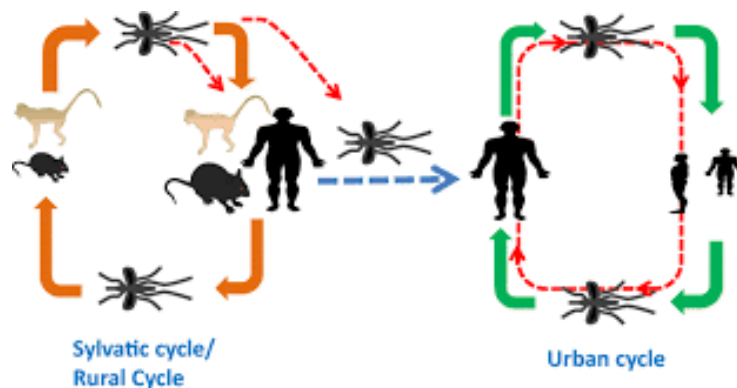


FIGURE 1.4 – Transmission mode Chikungunya

1.4 Symptômes et Diagnostic

La reconnaissance des symptômes et l'établissement d'un diagnostic précis sont essentiels pour la gestion efficace des cas de chikungunya. Cette section examine les manifestations cliniques typiques de l'infection par le virus chikungunya ainsi que les méthodes diagnostiques utilisées pour identifier la maladie.

1.4.1 Symptômes

Chez les patients symptomatiques, la maladie à **CHIKV** se déclare généralement 4 à 8 jours (entre 2 et 12 jours) après la piqûre d'un moustique infecté. Elle se caractérise par une brusque poussée de fièvre, souvent accompagnée de fortes douleurs articulaires. Les douleurs articulaires sont souvent invalidantes et durent généralement quelques jours, mais peuvent être prolongées et durer des semaines, des mois, voire des années. D'autres signes et symptômes courants sont le gonflement des articulations, les douleurs musculaires, les maux de tête, les nausées, la fatigue et les éruptions cutanées. Comme ces symptômes se confondent avec ceux d'autres infections, notamment celles dues aux virus de la dengue et du Zika, les cas peuvent être mal diagnostiqués. En l'absence de douleurs articulaires importantes, les symptômes des personnes infectées sont généralement légers et l'infection peut passer inaperçue.

La plupart des patients se rétablissent complètement de l'infection ; toutefois, des cas occasionnels de complications oculaires, cardiaques et neurologiques ont été signalés dans le cadre d'infections par le **CHIKV**. Les patients situés aux extrémités du spectre d'âge sont plus exposés à une maladie grave. Les nouveau-nés infectés pendant l'accouchement et les personnes âgées souffrant de pathologies sous-jacentes peuvent devenir gravement malades et l'infection par le **CHIKV** peut augmenter le risque de décès [14].

Une fois qu'une personne est guérie, les données disponibles suggèrent qu'elle est probablement immunisée contre les infections futures [1].

1.4.2 Méthodes de diagnostic

Le diagnostic peut être retardé en raison de la confusion possible des symptômes avec ceux de la dengue ou du Zika. Les tests *immuno-enzymatiques* (ELISA) peuvent être utilisés pour confirmer la présence d'anticorps **anti-CHIKV**, les niveaux d'anticorps IgM étant les plus élevés trois à cinq semaines après l'infection et persistant jusqu'à deux mois. La PCR peut également être utilisée pour génotyper le virus [9].

1.5 Méthodes de Contrôle et Traitement

La gestion efficace de l'épidémie de chikungunya repose sur une combinaison de stratégies de contrôle des vecteurs et d'interventions médicales. Cette section explore les diverses approches utilisées pour prévenir la transmission du virus et traiter les symptômes chez les patients infectés.

1.5.1 Méthodes de contrôle

La prévention de l'infection en évitant les piqûres de moustiques est la **meilleure protection**. Les patients suspectés d'être infectés par le **CHIKV** doivent éviter les piqûres de moustiques pendant la première semaine de la maladie afin d'empêcher la transmission aux moustiques, qui peuvent à leur tour infecter d'autres personnes.

La principale méthode pour réduire la transmission du **CHIKV** consiste à contrôler les moustiques vecteurs. Pour ce faire, il faut mobiliser les communautés, qui jouent un rôle essentiel dans la réduction des sites de reproduction des moustiques en :

- vidant et en nettoyant chaque semaine les récipients contenant de l’eau,
- éliminant les déchets,
- soutenant les programmes locaux de lutte contre les moustiques.

Pendant les épidémies, des insecticides peuvent être :

- pulvérisés pour tuer les moustiques adultes volants,
- appliqués sur les surfaces à l’intérieur et autour des conteneurs où les moustiques se posent,
- utilisés pour traiter l’eau dans les conteneurs afin de tuer les larves immatures.

Les autorités sanitaires peuvent également prendre des mesures d’urgence pour contrôler la population de moustiques.

Pour se protéger pendant les épidémies de **chikungunya**, il est conseillé de :

- porter des vêtements qui minimisent l’exposition de la peau aux vecteurs qui piquent pendant la journée,
- utiliser des moustiquaires aux fenêtres et aux portes pour empêcher les moustiques de pénétrer dans les maisons,
- appliquer des répulsifs sur la peau exposée ou sur les vêtements en respectant scrupuleusement les instructions figurant sur l’étiquette du produit.

Les répulsifs doivent contenir du **DEET**, de l’**IR3535** ou de l’**icaridine** [?].

1.5.2 Options de traitement

Le traitement actuel vise à atténuer la gravité des symptômes plutôt qu’à guérir la maladie. Le traitement repose principalement sur l’utilisation d’**antipyrétiques** et d’**AINS**. Cependant, aucune étude n’a évalué systématiquement l’efficacité de ces traitements, et les symptômes peuvent disparaître sans intervention. L’utilisation de **corticostéroïdes** pour le traitement de la phase aiguë a connu un succès mitigé et est utilisée avec hésitation en raison de la possibilité d’aggravation des symptômes après le traitement. Il est particulièrement important de maintenir des niveaux de liquide adéquats.

Il existe également des preuves émergentes que les médicaments qui entravent le transport du cholestérol, tels que les composés **amphiphiles cationiques de classe II U18666A** et l’**imipramine**, peuvent être efficaces contre la fusion membranaire du **CHIKV**, et ont un potentiel d’action contre d’autres arbovirus.

Pour les arthralgies chroniques graves, des **antirhumatismaux modificateurs de la maladie (ARMM)**, notamment le **méthotrexate**, l’**hydroxychloroquine** ou la **sulfasalazine**, ont été proposés. Comme pour les traitements aigus, l’efficacité systématique des **DMARD** pour le traitement chronique est inconnue, bien que des rapports décrivent des résultats positifs avec une cessation des symptômes dans les 4 à 6 mois [5].

1.6 Cas du Tchad

L’analyse spécifique des cas de chikungunya au Tchad permet de comprendre l’impact de cette maladie dans un contexte régional spécifique. Cette section examine les Points saillant, la mise en contexte du problème dans ce pays et ainsi que les mesures prises par l’état pour la gestion de l’infection par le virus chikungunya dans ce

pays d'Afrique centrale.

1.6.1 Point saillants



FIGURE 1.5 – Region infecté au Tchad

1.6.2 Contexte

Au Tchad, à partir du mois de juillet 2020, le Médecin Chef du District d'Abéché a été alerté par le Responsable du Centre de Santé Samalat de la survenue des cas d'une pathologie surnommée localement Kourgnalé, caractérisée par une forte fièvre, des céphalées, des douleurs articulaires intenses et invalidantes et parfois associées aux vomissements et traités comme cas de paludisme. A partir du mois d'août l'augmentation des cas enregistrés a retenu l'attention des autorités sanitaires locales. Les prélèvements biologiques envoyés au Laboratoire Mobile de Ndjaména ont mis en évidence le virus de Chikungunya le 12 août 2020 et confirmé par l'Institut Pasteur de Yaoundé le 26 août 2020. Selon les investigations épidémiologiques et entomologiques réalisées (243 patients ont été enquêtés dans 138 ménages) par l'équipe d'appui du niveau central, il a été mis en évidence la présence d'*Aedes Egyptis*. Du 14/08 au 03/09/2020, 13 488 cas ont été enregistrés avec zéro décès ; La tranche d'âge la plus touchée est celle de 15 ans et plus et le sexe féminin est prédominant ; Plus de trois quarts des patients développent des fortes fièvres, des céphalées, des douleurs articulaires atroces et invalidantes et un tiers de ces malades ont développé des éruptions cutanées maculo-papuleux.

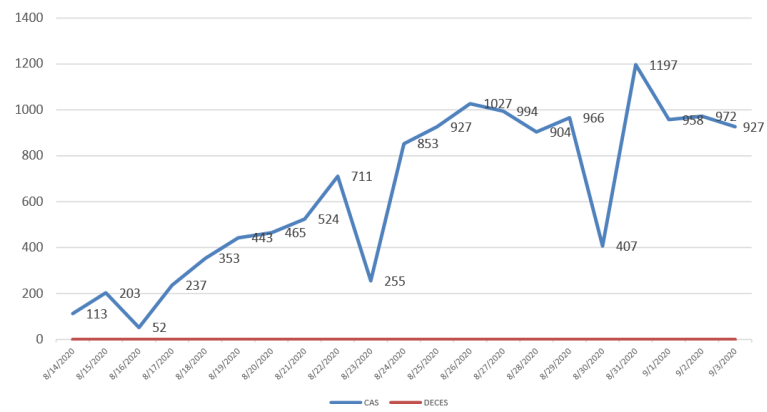


FIGURE 1.6 – Evolution journalière des cas et décès du Chikungunya

Chapitre 2

Revue de la Littérature et Concepts de Base

2.1 Introduction

Dès **1997**, le potentiel du **data mining** pour améliorer les problèmes dans le domaine médical avait été identifié par l'Organisation mondiale de la santé (**OMS**) (Gulbinat, 1997). L'utilité de la détection de connaissances à partir des dépôts de données médicales a été soulignée par l'OMS, car elle bénéficie au diagnostic médical et à la prédiction. Le data mining est un processus de découverte de connaissances utiles à partir de bases de données pour construire une structure (c'est-à-dire un modèle ou un schéma) qui peut interpréter de manière significative les données. Le data mining est le processus de découverte de schémas et de connaissances intéressants à partir d'une grande quantité de données (Han et al., 2001). Le data mining utilise de nombreuses techniques d'apprentissage automatique pour découvrir des schémas cachés dans les données. Ces techniques peuvent être réparties en trois catégories principales : les techniques d'**apprentissage supervisé**, les techniques d'**apprentissage non supervisé** et les techniques d'**apprentissage semi-supervisé** (Huang al., 2014). Voir figure 2.1. Les systèmes experts développés par des techniques d'apprentissage automatique peuvent être utilisés pour aider les médecins dans le **diagnostic** et la **prédiction des maladies** (Kononenko, 2001). En raison de l'importance du diagnostic des maladies pour l'humanité, plusieurs études ont été menées sur le développement de méthodes pour leur classification [7].

2.2 Apprentissage Automatique

Depuis leur évolution, les humains ont utilisé de nombreux types d'outils pour accomplir diverses tâches de manière plus simple. La créativité du cerveau humain a conduit à l'invention de différentes machines. Ces machines ont facilité la vie humaine en permettant aux gens de répondre à divers besoins de la vie, y compris le voyage, les industries et l'informatique. Et l'apprentissage automatique en fait partie. Selon Arthur Samuel, l'**apprentissage automatique** est défini comme le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. *Arthur Samuel* était célèbre pour son programme de jeu de dames. L'apprentissage automatique (**ML**) est utilisé pour apprendre aux machines comment gérer les données

de manière plus efficace. Parfois, après avoir examiné les données, nous ne pouvons pas interpréter les informations extraites des données. Dans ce cas, nous appliquons l'apprentissage automatique. Avec l'abondance des ensembles de données disponibles, la demande pour l'apprentissage automatique est en hausse. De nombreuses industries appliquent l'apprentissage automatique pour extraire des données pertinentes. Le but de l'apprentissage automatique est d'*apprendre à partir des données* [6].

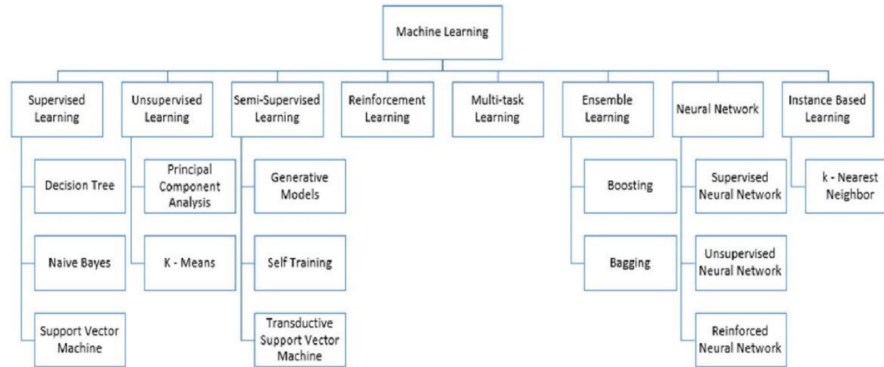


FIGURE 2.1

2.2.1 Techniques d'Apprentissage Automatique

Voici un aperçu rapide de certains des algorithmes couramment utilisés en apprentissage automatique (ML)

2.2.1.1 Apprentissage Supervisé

Les algorithmes d'**apprentissage supervisé** sont ceux qui nécessitent une assistance externe. L'ensemble de données d'entrée est divisé en ensemble d'entraînement et ensemble de test. L'ensemble d'entraînement a une variable de sortie qui doit être prédite ou classée. Tous les algorithmes apprennent un certain type de schémas à partir de l'ensemble d'entraînement et les appliquent à l'ensemble de test pour la prédiction ou la classification. Le flux de travail des algorithmes d'apprentissage supervisé est donné dans la figure 2.2 ci-dessous [6].

Exemple d'apprentissage supervisé :

- **Arbre de Décision** : est un graphe pour représenter les choix et leurs résultats sous forme d'arbre.
- **Naïve Bayes** : C'est une technique de classification basée sur le théorème de Bayes avec une hypothèse d'indépendance entre les prédicteurs. En termes simples, un classificateur Naïve Bayes suppose que la présence d'une caractéristique particulière dans une classe est indépendante de la présence de toute autre caractéristique.

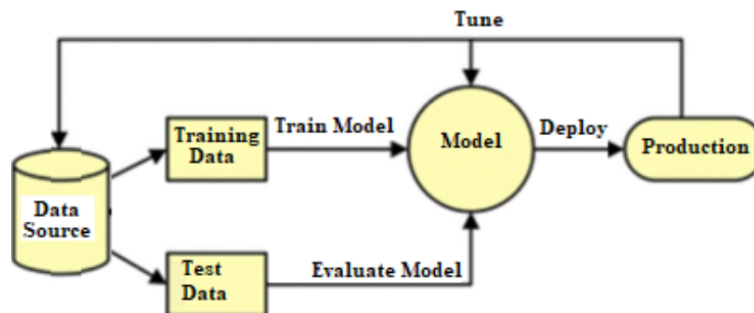


FIGURE 2.2 – Flux de travail de l'apprentissage supervisé

2.2.1.2 Apprentissage Non Supervisé

Contrairement à l'apprentissage supervisé ci-dessus, il n'y a pas de réponses correctes et il n'y a pas de professeur. Les algorithmes sont laissés à leurs propres dispositifs pour découvrir et présenter la structure intéressante dans les données. Les algorithmes d'apprentissage non supervisé apprennent quelques caractéristiques à partir des données. Lorsque de nouvelles données sont introduites, elles utilisent les caractéristiques apprises précédemment pour reconnaître la classe des données. Il est principalement utilisé pour le *clustering* et la réduction des *features* [6].

Exemple d'apprentissage non supervisé :

- **Clustering K-means** : est l'un des algorithmes d'apprentissage non supervisé les plus simples qui résout le problème bien connu du clustering. La procédure suit une manière simple et facile de classer un ensemble de données donné par un certain nombre de clusters.

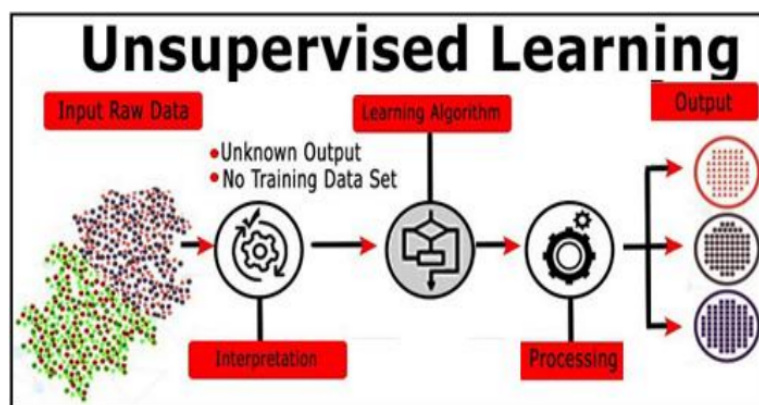


FIGURE 2.3 – Apprentissage Non Supervisé

2.2.1.3 Apprentissage Semi-Supervisé

L'apprentissage **semi-supervisé** est une combinaison des méthodes d'apprentissage *supervisé* et *non supervisé*. Il peut être fructueux dans ces domaines de l'apprentissage automatique et du data mining où les données non étiquetées sont déjà présentes et obtenir les données étiquetées est un processus fastidieux [6].

2.2.1.4 Apprentissage par Ensemble

L'**apprentissage par ensemble** est le processus par lequel plusieurs **modèles**, sont générés et combinés stratégiquement pour résoudre un problème particulier d'intelligence computationnelle. L'apprentissage par ensemble est principalement utilisé pour améliorer les performances d'un modèle ou réduire la probabilité de sélectionner un modèle médiocre [6].

Exemple d'apprentissage par ensemble :

- **Boosting** : Le terme "*Boosting*" fait référence à une famille d'algorithmes qui convertit les apprenants faibles en apprenants forts. Le boosting est une technique d'apprentissage par ensemble utilisée pour réduire le biais et la variance.
- **Bagging** : est appliqué lorsque la précision et la stabilité d'un algorithme d'apprentissage automatique doivent être augmentées. Il est applicable en classification et en régression. Le bagging réduit également la variance et aide à gérer le surapprentissage.

2.2.2 Applications de l'Apprentissage Automatique à l'Épidémiologie

2.2.3 Avantages et Limites de l'Apprentissage Automatique

2.3 Deep Learning

2.3.1 Définition et Concepts Fondamentaux

2.3.2 Réseaux de Neurones Artificiels

2.3.2.1 Perceptron Multicouche (MLP)

2.3.2.2 Réseaux de Neurones Convolutifs (CNN)

2.3.2.3 Réseaux de Neurones Récurrents (RNN)

2.3.3 Applications du Deep Learning à l'Épidémiologie

2.3.4 Avantages et Limites du Deep Learning

2.4 Ensemble Regression Learning

2.4.1 Définition et Concepts Fondamentaux

2.4.2 Techniques d'Ensemble Learning

2.4.2.1 Bagging

2.4.2.2 Boosting

2.4.2.3 Stacking

2.4.3 Applications de l'Ensemble Regression Learning

2.4.4 Avantages et Limites de l'Ensemble Regression Learning

2.5 Conclusion

Chapitre 3

Implémentation des Modèles

3.1 Introduction

3.2 Présentation des Outils Utilisés

Bla

3.2.1 Langages et librairies

Bla

3.2.2 Infrastructure matérielle

Bla

3.3 Méthodes de l'Apprentissage Automatique

Bla

3.3.1 Le jeu de données

Bla

3.3.2 Préparation et nettoyage des données

jijj

3.3.3 Implémentation des modèles de machine learning

Bla

3.4 Méthodes de l'Apprentissage Profond

Bla

3.4.1 Le jeu de données

Bla

3.4.2 Préparation des données pour le deep learning

Bla

3.4.3 Implémentation des modèles de deep learning

Bla

Conclusion

General Conclusion

Bibliographie

- [1] Auerswald, H., Boussioux, C., In, S., et al. : Broad and long-lasting immune protection against various chikungunya genotypes demonstrated by participants in a cross-sectional study in a cambodian rural community. *Emerging Microbes & Infections* **7**(1), 13 (2018). DOI 10.1038/s41426-018-0013-0
- [2] CDC : Transmission of chikungunya virus. <https://www.cdc.gov/ncezid> Visité le 10/juin/2024
- [3] cdc — US centers for disease control, prevention : National center for emerging and zoonotic infectious diseases (ncezid). <https://www.cdc.gov/ncezid> Visité le 12/juin/2024
- [4] Dr Jean Bosco NDIHOKUBWAYO, D.B.H.e.a. : Rapport de la situation Épidémiologique chikungunya
- [5] Ganesan, V., Duan, B., Reid, S. : Chikungunya virus : Pathophysiology, mechanism, and modeling. *Viruses* **9**(12), 368 (2017). DOI 10.3390/v9120368
- [6] Mahesh, B. : Machine learning algorithms - a review **9**. DOI 10.21275/ART20203995
- [7] Mehrbakhsh Nilashi Othman bin Ibrahim, H.A.L.S. : An analytical method for diseases prediction using machine learning techniques DOI <http://dx.doi.org/doi:10.1016/j.compchemeng.2017.06.011>
- [8] Moreira, J., Soares, C., Jorge, A., Sousa, J. : Ensemble approaches for regression : A survey. *ACM Computing Surveys* **45**, 10 :1–10 :40 (2012). DOI 10.1145/2379776.2379786
- [9] Morrison, T.E. : Reemergence of chikungunya **88**(20), 11,644 – 11,647. DOI <https://doi.org/10.1128/jvi.01432-14>
- [10] PAHO : Chikungunya - paho/who — pan american health organization. <https://www.paho.org/en/topics/chikungunya> Visité le 12/juin/2024
- [11] Reddy, B.V.K., Patel, S., Singh, A., Singh, N., Ranjit, S., Dantkale, S. : Data science system of predicting and detecting the chikungunya virus using ensemble techniques. In : 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1420–1423 (2022). DOI 10.1109/ICIRCA54612.2022.9985708
- [12] Rougeron, V., Sam, I.C., Caron, M., Nkoghe, D., Leroy, E., Roques, P. : Chikungunya, a paradigm of neglected tropical disease that emerged to be a new health global risk. *Journal of clinical Virology* **64**, 144–152 (2015)
- [13] WHO : Chikungunya — world health organization. https://www.who.int/health-topics/chikungunya#tab=tab_1 Visité le 10/juin/2024
- [14] WHO : Chikungunya fact sheet. <https://www.who.int/news-room/fact-sheets/detail> Visité le 12/juin/2024

[15] WHO : Disease outbreak news - tchad. <https://www.who.int/fr/emergencies/disease-outbreak-news/item/chikungunya-chad> Visité le 10/juin/2024