

UNIVERSITÉ DE NGAOUNDÉRE
FACULTÉ DES SCIENCES

THE UNIVERSITY OF NGAOUNDERE
FACULTY OF SCIENCE



***DEPARTEMENT DE MATHEMATIQUES ET
INFORMATIQUE
DEPARTMENT OF MATHEMATICS AND COMPUTER
SCIENCE***

**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 :

DR.ABBOUBAKAR HAMADJAM

Chargé de cours IUT

Université de Ngaoundéré

Superviseur

PR.DR.ING DAYANG PAUL

Maitre de conférences

Université de Ngaoundéré

Directeur

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	v
Liste des tableaux	vi
Notation et abréviations	vii
Abstract	viii
Résumé	ix
General Introduction	1
1 Épidémiologie de la Chikungunya	2
1.1 Origine de la Chikungunya	2
1.2 Agent Pathogène	2
1.2.1 Le virus Chikungunya	2
1.3 Mode de Transmission	3
1.3.1 Vecteurs : Les moustiques Aedes	3
1.3.2 La transmission	4
1.4 Symptômes et Diagnostic	5
1.4.1 Symptômes	5
1.4.2 Méthodes de diagnostic	5
1.5 Méthodes de Contrôle et Traitement	5
1.5.1 Méthodes de contrôle	5
1.5.2 Options de traitement	6
1.6 Cas du Tchad	6
1.6.1 Facteurs climatiques influençant la propagation	6
1.6.2 Études de cas et données climatiques	6

2	Revue de la littérature et concepts de base	7
2.1	Introduction	7
2.2	Apprentissage Automatique (Machine Learning)	7
2.2.1	Concepts de base	7
2.2.2	La régression linéaire	7
2.2.3	Forêt aléatoire (Random Forest)	7
2.3	Apprentissage Profond (Deep Learning)	7
2.3.1	Réseaux de neurones profonds	7
2.3.2	Algorithmes spécifiques (e.g., LSTM, CNN)	7
2.4	Régression par Ensemble	8
2.4.1	Concepts de régression par ensemble	8
2.4.2	Avantages des méthodes d'ensemble	8
2.4.3	Algorithmes d'ensemble (e.g., Gradient Boosting, XGBoost)	8
3	Implémentation des Modèles	9
3.1	Introduction	9
3.2	Présentation des Outils Utilisés	9
3.2.1	Langages et bibliothèques (e.g., Python, Scikit-learn, TensorFlow)	9
3.2.2	Infrastructure matérielle (e.g., GPU, serveurs)	9
3.3	Méthodes de l'Apprentissage Automatique	9
3.3.1	Le jeu de données	9
3.3.2	Préparation et nettoyage des données	9
3.3.3	Implémentation des modèles de machine learning	9
3.4	Méthodes de l'Apprentissage Profond	9
3.4.1	Le jeu de données	10
3.4.2	Préparation des données pour le deep learning	10
3.4.3	Implémentation des modèles de deep learning	10
	General Conclusion	12
	Bibliographie	13

Liste de figures

1.1	Aedes aegypti mosquito full of blood	2
1.2	Electron microscopic image of chikungunya virus	3
1.3	Tools to Develop our Model	3
1.4	Tools to Develop our Model	4
1.5	Tools to Develop Model	4
1.6	Tools to Develop our Model	4

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

Abstract

Chikungunya virus, transmitted by **Aedes aegypti** and **Aedes albopictus** mosquitoes, poses a significant public health threat globally due to its rapid spread and debilitating impact on affected populations. In recent years, outbreaks have afflicted populations in East and Central Africa, South America and Southeast Asia [7]. Predicting chikungunya epidemics is a challenge, due to the complex interaction of environmental, social and biological factors. Traditional epidemiological surveillance methods, while essential, may be insufficient to proactively predict epidemics. This research focuses specifically on **Chad**, a country where chikungunya has recently emerged as a major public health concern. With the increase in chikungunya cases in Chad, it is crucial to develop accurate predictive models to anticipate epidemics and implement effective control measures. The advent of artificial intelligence (AI) and machine learning techniques offers new opportunities to improve epidemiological forecasting.

This thesis aims to develop and evaluate a predictive model for chikungunya epidemics using ensemble regression approaches. Ensemble methods combine several learning algorithms to improve predictive accuracy over individual models. The research exploits comprehensive datasets including epidemiological, climatic and environmental variables to train and validate predictive models. Leading machine learning algorithms such as Random Forest, Gradient Boosting and XGBoost are implemented and compared to identify the most effective approach.

The study begins with a detailed exploration of chikungunya epidemiology, transmission dynamics, clinical manifestations and current control strategies. It reviews the existing literature on AI applications in disease prediction, highlighting the potential of ensemble regression techniques to improve the accuracy and reliability of epidemiological forecasts.

Through a rigorous methodology and an in-depth analysis of the results, this thesis evaluates the performance of ensemble regression models in predicting chikungunya epidemics. Performance metrics such as sensitivity, specificity and predictive value are used to assess model efficiency and robustness. In addition, the study examines the interpretability of model results and discusses implications for public health interventions.

The findings of this research contribute to advancing the field of infectious disease prediction by demonstrating the effectiveness of ensemble regression techniques for improving the prediction of chikungunya epidemics.

Résumé

Le virus du chikungunya, transmis par les moustiques **Aedes aegypti** et **Aedes albopictus**, constitue une menace importante pour la santé publique mondiale en raison de sa propagation rapide et de son impact débilisant sur les populations touchées. Ces dernières années, des épidémies ont touché des populations d’Afrique centrale et orientale, d’Amérique du Sud et d’Asie du Sud-Est [7]. Prédire les épidémies de chikungunya est un défi, en raison de l’interaction complexe de facteurs environnementaux, sociaux et biologiques. Les méthodes traditionnelles de surveillance épidémiologique, bien qu’essentielles, peuvent s’avérer insuffisantes pour prévoir les épidémies de manière proactive. Cette recherche se concentre spécifiquement sur le **Tchad**, un pays où le chikungunya a récemment émergé comme un problème majeur de santé publique. Avec l’augmentation des cas de chikungunya au Tchad, il est crucial de développer des modèles prédictifs précis pour anticiper les épidémies et mettre en œuvre des mesures de contrôle efficaces. L’avènement de l’intelligence artificielle (IA) et des techniques d’apprentissage automatique offre de nouvelles opportunités pour améliorer les prévisions épidémiologiques.

Cette thèse vise à développer et évaluer un modèle prédictif pour les épidémies de chikungunya en utilisant des approches de régression d’ensemble. Les méthodes d’ensemble combinent plusieurs algorithmes d’apprentissage pour améliorer la précision prédictive par rapport aux modèles individuels. La recherche exploite des ensembles de données complets comprenant des variables épidémiologiques, climatiques et environnementales pour former et valider des modèles prédictifs. Les principaux algorithmes d’apprentissage automatique tels que Random Forest, Gradient Boosting et XGBoost sont mis en œuvre et comparés afin d’identifier l’approche la plus efficace.

L’étude commence par une exploration détaillée de l’épidémiologie du chikungunya, de la dynamique de transmission, des manifestations cliniques et des stratégies de contrôle actuelles. Elle passe en revue la littérature existante sur les applications de l’IA dans la prédiction des maladies, en soulignant le potentiel des techniques de régression d’ensemble pour améliorer la précision et la fiabilité des prévisions épidémiologiques.

Grâce à une méthodologie rigoureuse et à une analyse approfondie des résultats, cette thèse évalue la performance des modèles de régression d’ensemble dans la prévision des épidémies de chikungunya. Des mesures de performance telles que la sensibilité, la spécificité et la valeur prédictive sont utilisées pour évaluer l’efficacité et la robustesse des modèles. En outre, l’étude examine l’interprétabilité des résultats des modèles et discute des implications pour les interventions de santé publique.

Les résultats de cette recherche contribuent à faire progresser le domaine de la prédiction des maladies infectieuses en démontrant l’efficacité des techniques de régression d’ensemble pour améliorer la prédiction des épidémies de chikungunya.

Introduction Générale

Le chikungunya est une maladie virale transmise par la piqure d'un moustique infecté et endémique en Afrique de l'Est et dans certaines régions d'Asie. Les symptômes du chikungunya sont la fièvre, des douleurs articulaires, des maux de tête, des douleurs musculaires, un gonflement des articulations ou une éruption cutanée. Les décès dus au chikungunya sont rares. Le virus peut entraîner de graves problèmes chez certaines personnes, en particulier les personnes âgées et les personnes souffrant d'autres maladies chroniques. Une détection précoce du chikungunya peut empêcher la maladie de se propager. Notre projet vise à prédire si une personne est atteinte ou non du chikungunya et à prendre des mesures de précaution en fonction du résultat. Il n'existe pas de traitement antiviral spécifique pour le chikungunya. La prise en charge clinique se concentre sur le soulagement des symptômes, comme le soulagement des douleurs articulaires à l'aide d'antipyrétiques, d'analgésiques appropriés, d'une consommation suffisante de liquides et d'une relaxation générale [8]. Les progrès récents dans le domaine de l'intelligence artificielle (IA) ont permis d'améliorer considérablement les prédictions grâce à des algorithmes qui se sont révélés capables de saisir les relations non linéaires entre les données d'entrée et de sortie. Ces algorithmes sont généralement connus sous le nom d'algorithmes d'apprentissage automatique qui sont basés sur des données historiques et actuelles.

L'objectif de ce mémoire de master est de développer un modèle de prédiction du chikungunya. Pour ce faire, nous avons opté pour une méthode d'apprentissage automatique issue de l'intelligence artificielle et basée sur l'approche de régression d'ensemble.

Dans ce travail, nous nous intéressons à répondre à la problématique : Comment utiliser l'apprentissage automatique par l'approche ensembliste pour prédire et analyser le chikungunya ?

Les objectifs spécifiques sont donc les suivants.

- Comprendre les concepts liés à la régression d'ensemble et au chikungunya ;
- Prédire le chikungunya au Tchad ;
- Analyser la situation du chikungunya au Tchad.

Le reste du manuscrit est organisé comme suit. Le premier chapitre est consacré à l'épidémiologie du chikungunya et à un focus sur le cas du Tchad. Le deuxième chapitre donne une introduction à l'apprentissage automatique pour les maladies. Dans le troisième chapitre, nous présentons la conception générale et détaillée du modèle proposé. Dans le quatrième chapitre, nous illustrons l'efficacité de notre modèle par des résultats de simulation. Le travail se termine par une conclusion générale et quelques perspectives.

Chapitre 1

Épidémiologie de la Chikungunya

Introduction

1.1 Origine de la Chikungunya

Chikungunya fever (CHIKF) is a viral disease that was first described in **1952** during an outbreak in southern Tanzania. The name comes from a word in the *Makonde* language, spoken in southeast Tanzania and northern Mozambique, that means "*to become contorted*" or "*that which bends up*". The virus was first isolated in Thailand in 1958.[10]

1.2 Agent Pathogène

Blajxckc

1.2.1 Le virus Chikungunya



FIGURE 1.1 – *Aedes aegypti* mosquito full of blood

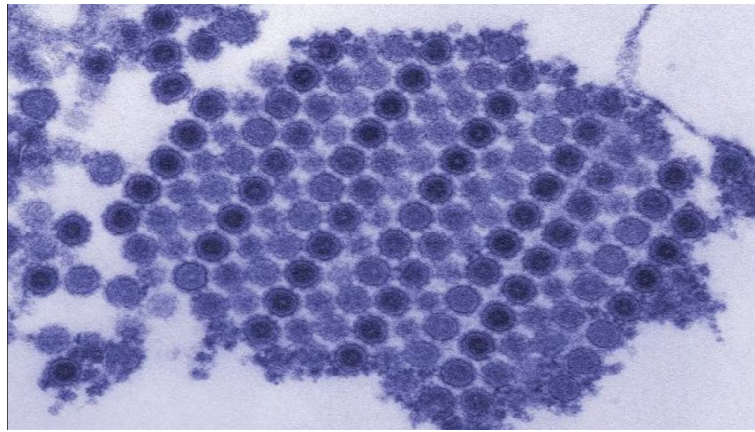


FIGURE 1.2 – Electron microscopic image of chikungunya virus



FIGURE 1.3 – Tools to Develop our Model

1.3 Mode de Transmission

Bla

1.3.1 Vecteurs : Les moustiques Aedes

Bla

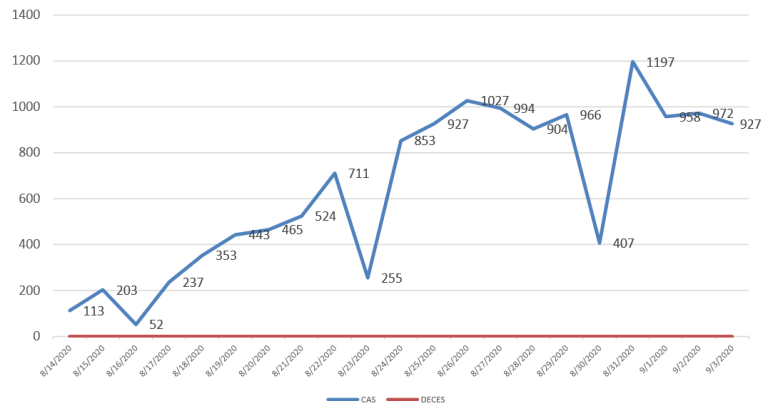


FIGURE 1.4 – Tools to Develop our Model

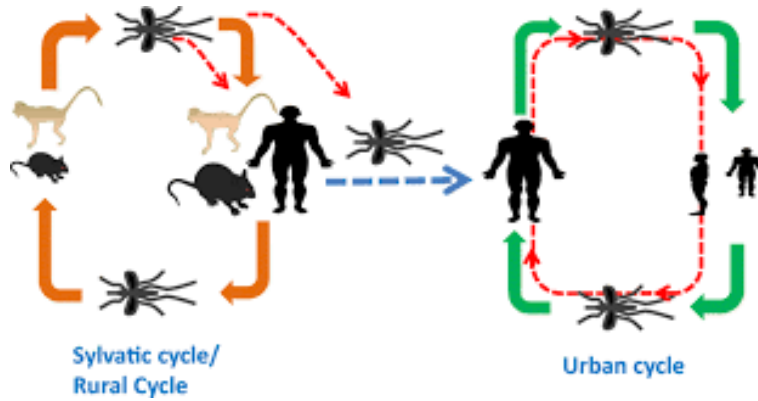


FIGURE 1.5 – Tools to Develop Model



FIGURE 1.6 – Tools to Develop our Model

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??). 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].

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 [11].

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

ccxc Bla

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

Bla

1.5.2 Options de traitement

Bla

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 caractéristiques épidémiologiques, les stratégies de contrôle et les défis rencontrés dans la gestion de l'infection par le virus chikungunya dans ce pays d'Afrique centrale.

1.6.1 Facteurs climatiques influençant la propagation

Bla

1.6.2 Études de cas et données climatiques

Chapitre 2

Revue de la littérature et concepts de base

2.1 Introduction

2.2 Apprentissage Automatique (Machine Learning)

Bla

2.2.1 Concepts de base

Bla

2.2.2 La régression linéaire

Bla

2.2.3 Forêt aléatoire (Random Forest)

Bla

2.3 Apprentissage Profond (Deep Learning)

Bla

2.3.1 Réseaux de neurones profonds

Bla

2.3.2 Algorithmes spécifiques (e.g., LSTM, CNN)

Bla

2.4 Régression par Ensemble

Bla

2.4.1 Concepts de régression par ensemble

Bla

2.4.2 Avantages des méthodes d'ensemble

Bla

2.4.3 Algorithmes d'ensemble (e.g., Gradient Boosting, XGBoost)

Bla Bla

Chapitre 3

Implémentation des Modèles

3.1 Introduction

3.2 Présentation des Outils Utilisés

Bla

3.2.1 Langages et librairies (e.g., Python, Scikit-learn, TensorFlow)

Bla

3.2.2 Infrastructure matérielle (e.g., GPU, serveurs)

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

Bla

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

Roads face several problems in managing traffic, reducing pollution and helping to improve road safety. ITS issues in making roads smarter, more efficient and better managed. Prediction of road traffic speeds plays a crucial role in intelligent transport systems. Prediction helps to avoid traffic congestion over time, to achieve intelligent recall of traffic congestion.

For accurate prediction of traffic speeds, we have proposed in this work a machine learning method called KnnLSTM-KF. This method, predicts the characteristics of traffic speeds over time.

In order to validate our work, we conducted experiments on four hybrid machine learning models : Simple KNN, KNN-KF, KNN-LSTM and KNN-LSTM-KF. Dataset of traffic flow of New York City in the United States was used for simulations. Comparative study shown that, the proposed KnnLSTM-KF model outperformed other hybrid models in terms of MAPE and MSE.

This research can be improved as follows :

- Taken into account other parameters such as road conditions and weather ;
- Used GPS positions (longitude and latitude) taken every five minutes on road links ;
- Extended these experiments to other types of vehicles.

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] Morrison, T.E. : Reemergence of chikungunya **88**(20), 11,644 – 11,647. DOI <https://doi.org/10.1128/jvi.01432-14>
- [7] PAHO : Chikungunya - paho/who — pan american health organization. <https://www.paho.org/en/topics/chikungunya> Visité le 12/juin/2024
- [8] 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
- [9] 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)
- [10] WHO : Chikungunya — world health organization. https://www.who.int/health-topics/chikungunya#tab=tab_1 Visité le 10/juin/2024
- [11] WHO : Chikungunya fact sheet. <https://www.who.int/news-room/fact-sheets/detail> Visité le 12/juin/2024
- [12] WHO : Disease outbreak news - tchad. <https://www.who.int/fr/emergencies/disease-outbreak-news/item/chikungunya-chad> Visité le 10/juin/2024