Letters in Animal Biology

Journal homepage: www.liabjournal.com

Advancing vector-borne disease prediction through functional classifier integration: A novel approach for enhanced modeling

Mokammel Hossain Tito *[1], Most Hoor E. Jannat [1], Mst Tachhlima Aktar [2], Barshon Saha [1], Puja Das [1], Md. Kawser [1], Md. Arafat Hossain [1], Md. Neyamul Islam [1], Muhammad Shahzad Chohan [3], Shahzad Khan [3]

[1] *Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh*
[2] *Bangladesh Agricultural University, Mymensingh, Bangladesh*
[3] *Department of Biomedical Sciences, King Faisal University, Al Hafuf, Saudi Arabia*

Article info

Received: 06 January 2024
Received in revised form: 19 February 2024
Accepted: 24 February 2024
Published online: 26 February 2024

Keywords
Machine learning models
Vector borne disease
Accuracy
Efficacy
Simple logistic

* *Corresponding author:*
M.H. Tito
Email: mokammel.17asvm014@bsmrstu.edu.bd

Reviewed by:
Md. Arifuzzaman
Kind Faisal University, Saudi Arabia

Abstract

This paper evaluates various machine learning models for predicting vector-borne diseases, focusing on performance metrics that reveal insights into their efficacy. The Multilayer Perceptron (MLP) model demonstrated the highest accuracy at 92%, surpassing the Simple Logistic (SL) and Support Vector Machine (SVM) models, which achieved 88% and 90.87% accuracy, respectively. Notably, the MLP model excelled in precision, recall, and F-Measure, indicating superior classification accuracy. Conversely, the SVM model exhibited noteworthy computational efficiency with the lowest processing time at 0.3 seconds, emphasizing its potential for real-time applications in public health interventions. In contrast, the Radial Basis Function Network (RBFN) lagged in accuracy and other metrics. The results underscore the trade-offs between accuracy and computational efficiency, emphasizing the need for a nuanced model selection. Considering the holistic evaluation, the SVM model emerged as a compelling choice, balancing high accuracy and efficient processing, making it promising for real-time public health applications. This study contributes valuable insights into machine learning model performance, emphasizing the importance of selecting models tailored to the specific needs of vector-borne disease prediction. As we confront emerging infectious diseases, the SVM model stands as an indispensable tool, supporting a proactive and data-driven approach to mitigate the global health impact of vector-borne diseases.

1. Introduction

Vector-borne diseases, fueled by the intricate interplay between pathogens, vectors, and hosts, constitute a significant global health challenge. These infectious illnesses are primarily transmitted through the bites of arthropod vectors, such as mosquitoes, ticks, sandflies, and fleas, acting as intermediary hosts for pathogens like bacteria, parasites, and viruses (WHO 2014; Knudsen and Slooff 1992). Diseases like malaria, dengue fever, Lyme disease, and Zika virus have led to substantial morbidity and mortality worldwide, with changing environmental conditions, urbanization, globalization, and ecological transformations altering the distribution and abundance of vectors, thereby intensifying the global threat of these diseases (WHO 2014; Gubler 2009). The dynamic nature of vector-borne diseases necessitates advanced predictive tools to anticipate their emergence, spread, and impact on human populations (Wilson et al. 2020). In recent years, machine learning approaches, specifically function classifiers, have emerged as powerful tools in epidemiology, providing the potential to enhance our understanding of the complex interactions between environmental factors, vectors, and pathogens (Basu et al. 2020; Uddin et al. 2019).

This study explores the application of function classifiers in predicting vector-borne diseases, emphasizing their capacity to discern intricate patterns within diverse datasets. Function classifiers, a subset of machine learning algorithms, leverage mathematical functions to map input features to specific outputs, allowing for the identification of complex relationships within multidimensional data (Almustafa 2020). These classifiers hold promise for predicting the occurrence and spread of vector-borne diseases by analyzing diverse sets of variables, including climatic conditions, land use patterns, and socio-economic factors. In contrast to traditional statistical methods, function classifiers excel at capturing non-linear and interactive effects, providing a more nuanced understanding of the intricate dynamics underlying disease transmission (Kaur et al. 2022; Raizada et al. 2021; Shaikh et al. 2023). This study aims to predict vector-borne diseases by employing function classifiers, leveraging advanced computational techniques. It focuses on identifying the most effective classifier within the function classifiers domain to enhance disease prediction accuracy and inform targeted interventions. Through rigorous

analysis, it seeks to optimize predictive models for better public health outcomes in combating vector-borne illnesses.

2. Materials and methods

2.1 Data collection and Preprocessing

The investigation into vector-borne diseases in this study relied on meticulously curated datasets obtained from reputable sources, including Kaggle competitions, research databases, and public health organizations (https://www.kaggle.com/datasets/richardbernat/vector-borne-disease-prediction). A heartfelt appreciation is extended to the contributors who have devoted their efforts to compile and share invaluable data related to vector-borne diseases, enriching the scope and depth of this study. This study focuses on 11 vector-borne diseases: Chikungunya, Dengue, Zika, Yellow Fever, Rift Valley Fever, West Nile Fever, Malaria, Tungiasis, Japanese Encephalitis, Plague, and Lyme Disease. Information for this research was diligently gathered from diverse sources such as public health records, surveillance databases, and relevant literature. The dataset, a comprehensive compilation of environmental factors, demographic data, and historical disease occurrences form the bedrock of the analytical endeavours of this study.

Before embarking on model training, the amassed data underwent an extensive preprocessing phase to ensure impeccable quality and consistency. This phase encompassed a range of tasks, including rigorous cleaning processes, addressing missing values, and standardizing data formats. Categorical variables were appropriately encoded, and numerical features underwent normalization to achieve uniformity in scale. Additionally, the dataset underwent meticulous curation using WEKA 3.9.6, where outliers and extreme values were carefully identified and removed. In pursuit of identifying the most influential features for accurate disease prediction, a robust feature selection process unfolded. This involved the strategic application of statistical techniques and domain expertise to discern and filter out less informative variables. The objective was to ensure that the input features for machine learning models used in this study were not only meaningful but also directly relevant to the intricate landscape of vector-borne diseases. This comprehensive approach underscored the commitment of this study to excellence in data handling and analysis, guaranteeing the reliability and significance of subsequent findings for the broader field of vector-borne disease research.

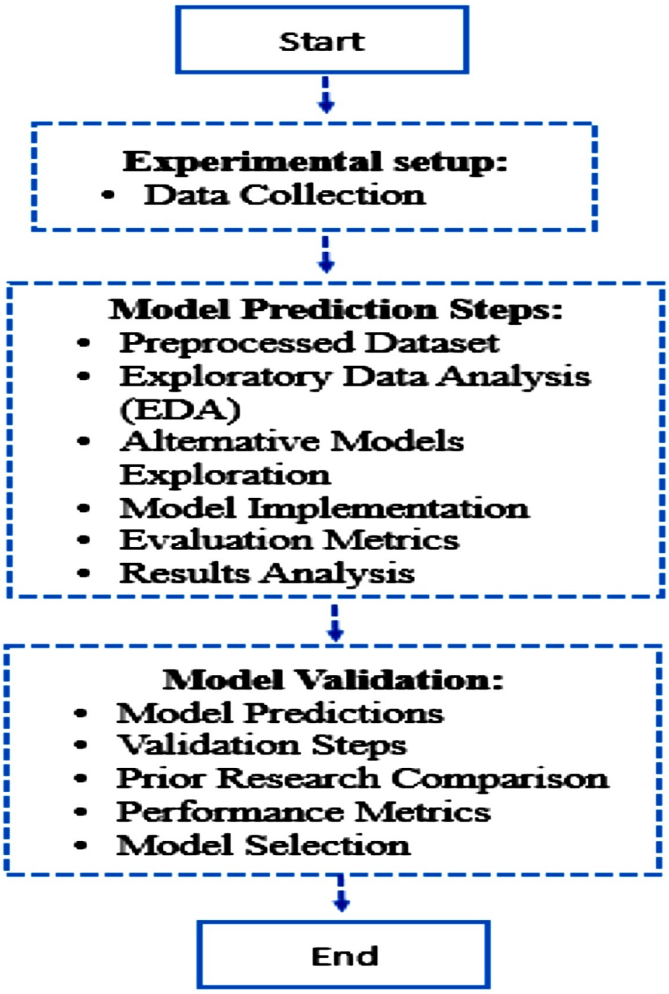

***Fig. 1**: Operational workflow the study*

2.2 Description of utilized models

Total of four machine learning models were used in this study including Multilayer Perceptron (MLP), Radial Basis Function Network (RBFN), Simple Logistic (SL), and Support Vector Machine (SVM) for predicting vector-borne diseases. This approach allowed for diverse exploration, comparison, and evaluation of methodologies, enhancing robustness and aiding in model selection for accurate predictions.

2.2.1 Multilayer Perceptron (MLP)

A Multilayer Perceptron (MLP) model, a type of feedforward neural network, is a potent tool in the prediction of vector-borne diseases (Kumar et al. 2024). Inspired by the human brain, an MLP consists of interconnected nodes organized into layers, including an input layer, one or more hidden layers, and an output layer (Fig 2). In the context of vector-borne diseases, the MLP model processes input data related to environmental conditions, demographic factors, and historical disease occurrences through the input layer. The hidden layers, characterized by nodes employing weighted connections and activation functions, enable the network to discern intricate patterns and non-linear relationships within the data (Javaid et al. 2023; Kofidou et al. 2021). The output layer provides predictions or classifications, such as the likelihood of vector-borne disease occurrence in a specific region or population. Activation functions in the output layer depend on the nature of the prediction task, employing sigmoid functions for binary classification and softmax functions for multiclass classification. Training an MLP involves adjusting weights and biases to minimize the difference between predicted outputs and actual outcomes, typically utilizing optimization algorithms like gradient descent (Javaid et al. 2023).

MLP models excel in capturing complex patterns and relationships within datasets, especially in scenarios where traditional linear models may be inadequate (Tito et al. 2023). Their ability to handle non-linearities and interactions between

features makes them well-suited for predicting the occurrence and spread of vector-borne diseases. However, successful implementation requires careful tuning of hyperparameters such as the number of hidden layers, nodes, and learning rates (Kumar et al. 2024). In the context of vector-borne diseases, MLP models contribute to more accurate predictions, assisting public health officials in proactive measures and interventions. By leveraging the strengths of artificial neural networks, MLP models enhance the understanding of complex disease dynamics, ultimately aiding in the development of targeted strategies to mitigate the impact of vector-borne diseases on communities (Erraguntla et al. 2019).

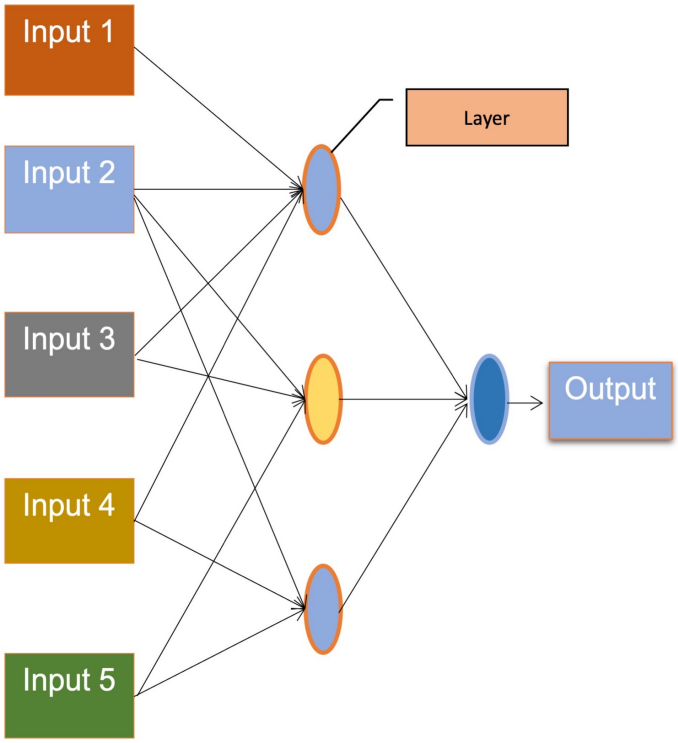

***Fig. 2**: Visual representation of Multilayer Perceptron (MLP) Model*

2.2.2 Radial Basis Function Network (RBFN)

The Radial Basis Function Network (RBFN) model emerges as a potent tool in the context of predicting and understanding vector-borne diseases. The RBFN model, a type of artificial neural network, excels in capturing complex non-linear relationships within datasets (Kesorn et al. 2015), a feature crucial in deciphering the intricate dynamics of vector-borne diseases. In the realm of disease prediction, the RBFN model proves particularly adept at handling the multifaceted interplay between environmental factors, host characteristics, and pathogen behaviour. By leveraging radial basis functions, this model can effectively map the input features onto a high-dimensional space, allowing it to discern subtle patterns and interactions that may elude traditional linear models. The RBFN's ability to adapt and generalize makes it a valuable asset for forecasting the spatial and temporal spread of vector-borne diseases, contributing significantly to the development of robust predictive models and informed public health interventions (Alfred and Obit 2021; da Silva et al. 2022).

2.2.3 Simple Logistic (SL)

The Simple Logistic model, a fundamental component in predictive modeling, offers a straightforward yet powerful approach when applied to the study of vector-borne diseases. In the context of these diseases, the Simple Logistic model aims to predict the likelihood of occurrence or spread based on a set of input features. Typically used when the outcome variable is binary (e.g., presence or absence of a particular vector-borne disease), this model estimates the probability of the event occurring (Eisen and Eisen 2011). It employs the logistic function to transform a linear combination of predictor variables into a probability score, which is then thresholded to make predictions. The Simple Logistic model is particularly useful in vector-borne disease research, where understanding the factors influencing disease presence is essential for effective public health interventions. It provides a foundational framework for analyzing and interpreting the relationships between various predictors, contributing valuable insights into the dynamics of vector-borne diseases and aiding in the identification of key factors influencing their occurrence (Brownstein et al. 2002).

2.2.4 Support Vector Machines (SVM)

Support Vector Machines (SVM) serve as a formidable tool in the realm of vector-borne disease prediction. SVM, a supervised machine learning algorithm (Fig. 3), is particularly well-suited for complex datasets characterized by non-linear relationships. In the context of vector-borne diseases, where diverse factors such as climate, demographics, and historical occurrences intricately influence transmission dynamics, SVM excels at discerning patterns that may elude traditional methods (Raizada et al. 2020). By transforming the input data into a higher-dimensional space and identifying an optimal hyperplane that maximally separates different classes of vector-borne diseases, SVM achieves a remarkable ability to classify and predict instances. The flexibility of SVM allows it to adapt

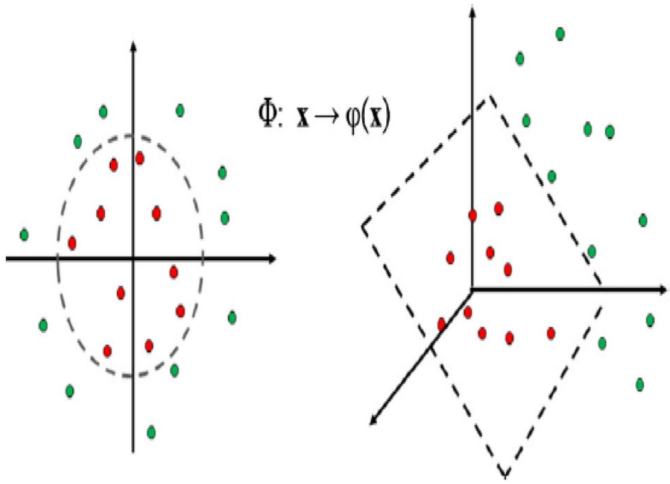

***Fig. 3**: Visual representation of Support Vector Machines (SVM) model (Arifuzzaman et al. 2021)*

Table 1 Summarized results of four Machine Learning Models

	Model	MLP	RBFN	SL	SVM
Evaluation Criteria	Accuracy	92.0%	82.0%	88.0%	90.9%
	Root mean squared error	0.1120	0.1784	0.1321	0.2642
	Mean absolute error	0.0255	0.0327	0.0245	0.1492
	Relative absolute error	15.32%	19.70%	14.73%	90.24%
	Root relative squared error	38.75%	61.73%	45.70%	91.86%
	Kappa statistic	0.9105	0.7987	0.8658	0.8996
	Precision	0.934	0.888	0.912	0.913
	Recall	0.920	0.820	0.880	0.909
	F-Measure	0.917	0.814	0.873	0.908
	Processing *Time (s)*	3.37	0.75	0.43	0.30

MLP - Multilayer Perceptron; RBFN - Radial Basis Function Network; SL - Simple Logistic; SVM - Support Vector Machine

to the intricacies of disease spread, providing a robust framework for understanding the multifaceted interactions between vectors, hosts, and environmental variables. Leveraging SVM in the prediction of vector-borne diseases not only enhances accuracy but also offers valuable insights into the intricate dynamics governing the spatial and temporal patterns of these diseases, ultimately contributing to more effective public health interventions (Fuchida et al. 2017; Munirathinam et al. 2023).

3. Result and Discussions

From the evaluation of various machine learning models for predicting vector-borne diseases, the performance metrics reveal intriguing insights. The Multilayer Perceptron (MLP) model exhibited the highest accuracy at 92%, closely followed by the Simple Logistic (SL) and Support Vector Machine (SVM) models at 88% and 90.87%, respectively. The MLP model also outshined others in terms of precision, recall, and F-Measure, indicating its superior ability to correctly classify instances and balance between false positives and false negatives (Table 1). However, the SVM model demonstrated noteworthy efficiency with the lowest processing time at 0.3 seconds, suggesting its computational advantage. On the other hand, the RBFN lags behind in terms of accuracy, precision, and recall.

The results showcased the trade-offs between accuracy, computational efficiency, and other performance metrics, emphasizing the need for a nuanced decision in selecting the most suitable model for vector-borne disease prediction. Considering the holistic evaluation, the SVM model emerged as a compelling choice (Fig 4), striking a balance between high accuracy and efficient processing time, making it a promising candidate for real-time applications in public health interventions. The application of machine learning (ML) in predicting vector-borne diseases represents a pivotal advancement in the field of public health. The results presented in this study underscore the efficacy of ML models, specifically Multilayer Perceptron (MLP), Radial Basis Function Network (RBFN), Simple Logistic (SL), and Support Vector Machine (SVM), in discerning complex patterns inherent in the transmission dynamics of vector-borne diseases.

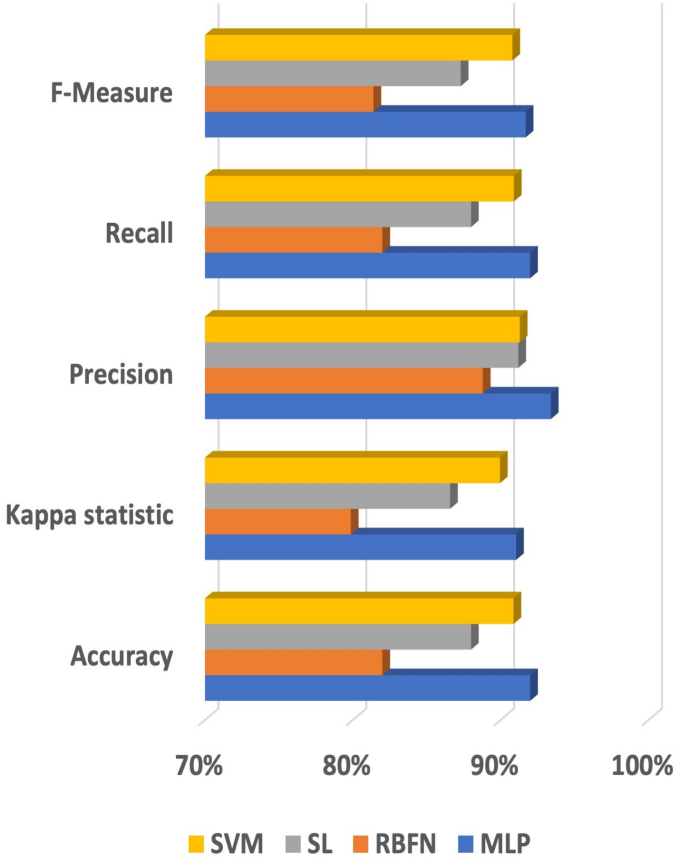

***Fig. 4**: Visual representation showcasing the efficacy of different ML models*

The high accuracy achieved by these models, particularly the MLP and SVM, suggests their potential as robust tools for early detection and proactive management of outbreaks. The comprehensive evaluation metrics, including precision, recall, and F-Measure, provide a holistic understanding of the models' performance, highlighting their ability to balance true positives, true negatives, false positives, and false negatives. Kappa Statistics provide insights into the reliability and validity of the predictive models by quantifying the agreement between predicted and observed outcomes. This metric helped for the evaluation of the model's performance while considering the possibility of chance agreement, thus offered a more nuanced evaluation of the predictive accuracy and robustness of the disease prediction models. Furthermore, the consideration of processing time unveils practical implications for real-time applications, with the SVM model demonstrating notable efficiency. The study not only contributes valuable insights into the comparative performance of ML models but also underscores the significance of selecting an appropriate model that aligns with the specific needs of vector-borne disease prediction. As this study navigates the landscape of emerging infectious diseases, ML-driven predictive models stand as indispensable tools, fostering a proactive and data-driven approach to mitigate the impact of vector-borne diseases on global health.

4. Conclusions

Utilization of machine learning (ML) function classifiers did indeed contribute to more accurate prediction of outbreaks and identification of areas at higher risk for vector-borne diseases. By analyzing various data inputs such as climate conditions, environmental factors, host population dynamics, vector abundance, and historical disease incidence, ML models could detect patterns and relationships that humans might have overlooked, thus enhancing predictive capabilities. After the evaluation of function classifier models for predicting vector-borne diseases, Support Vector Machine (SVM) showed the best result based on different evaluation criteria. With an impressive accuracy of 90.87% and notable efficiency reflected in the lowest processing time at 0.3 seconds, the SVM model emerges as a compelling choice for real-time applications in public health interventions. Its ability to strike a balance between high accuracy and efficient processing highlights its potential as a robust tool for early detection and proactive management of vector-borne disease outbreaks. The comprehensive evaluation metrics, including precision, recall, and F-Measure, further validate the SVM model's efficacy in achieving a nuanced understanding of the complex dynamics underlying disease transmission.

Recommendations and future work

While recognizing the potential for further enhancement in this dataset, its current utility is constrained by specific limitations. Machine learning holds promise for predicting various diseases, including but not limited to Brucellosis, Subclinical Mastitis, Anthrax, and LSD.

Declarations

Funding: None

Conflict of interest: None

Ethical approval: Not applicable

Acknowledgements: None

References

Alfred R, Obit JH. (2021). The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. Heliyon 7(6): e07371. https://doi.org/10.1016/j.heliyon.2021.e07371

Almustafa KM. (2020). Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinformatics 21(1): 278. https://doi.org/10.1186/s12859-020-03626-y

Arifuzzaman M, Islam M, Hossain M, Tito MH, Anwar M, Fuhaid AA. (2021). Application of AI on moisture damage of modified asphalt binders. 4th Smart Cities Symposium (SCS), Online Conference, Bahrain, 21-23 November 2021, pp. 307–311. https://doi.org/10.1049/icp.2022.0361

Basu S, Johnson KT, Berkowitz SA. (2020). Use of machine learning approaches in clinical epidemiological research of diabetes. Current Diabetes Reports 20(12): 80. https://doi.org/10.1007/s11892-020-01353-5

Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D. (2002). Spatial analysis of West Nile virus: Rapid risk assessment of an introduced vector-borne zoonosis. Vector-Borne and Zoonotic Diseases 2(3): 157–164. https://doi.org/10.1089/15303660260613729

da Silva CC, de Lima CL, da Silva ACG, Moreno GMM, Musah A, Aldosery A, Dutra L, Ambrizzi T, Borges IVG, Tunali M, Basibuyuk S, Yenigun O, Massoni TL, Jones K, Campos L, Kostkova P, da Silva Filho AG, dos Santos WP. (2022). Spatiotemporal forecasting for dengue, chikungunya fever and Zika using machine learning and artificial expert committees based on meta-heuristics. Research on Biomedical Engineering 38(2): 499–537. https://doi.org/10.1007/s42600-022-00202-6

Eisen L, Eisen RJ. (2011). Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. Annual Review of Entomology 56(1): 41–61. https://doi.org/10.1146/annurev-ento-120709-144847

Erraguntla M, Zapletal J, Lawley M. (2019). Framework for infectious disease analysis: A comprehensive and integrative multi-modeling approach to disease prediction and management. Health Informatics Journal 25(4): 1170 – 1187. https://doi.org/10.1177/1460458217747112

Fuchida M, Pathmakumar T, Mohan RE, Tan N, Nakamura A. (2017). Vision-based perception and classification of mosquitoes using support vector machine. Applied Sciences 7(1): 51. https://doi.org/10.3390/app7010051

Gubler DJ. (2009). Vector-borne diseases. Revue Scientifique et Technique (International Office of Epizootics) 28(2): 583 – 588. https://doi.org/10.20506/rst.28.2.1904

Javaid M, Sarfraz MS, Aftab MU, Zaman Q. uz, Rauf HT, Alnowibet KA. (2023). Web GIS-based real-time surveillance and response system for vector-borne infectious diseases. International Journal of Environmental Research and Public Health 20(4): 4. https://doi.org/10.3390/ijerph20043740

Kaur I, Sandhu AK, Kumar Y. (2022). Artificial intelligence techniques for predictive modeling of vector-borne diseases and its pathogens: A systematic review. Archives of Computational Methods in

Engineering 29(6): 3741–3771. https://doi.org/10.1007/s11831-022-09724-9

Kesorn K, Ongruk P, Chompoosri J, Phumee A, Thavara U, Tawatsin A, Siriyasatien P. (2015). Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the *Aedes aegypti* infection rate in similar climates and geographical areas. PLOS One 10(5): e0125049. https://doi.org/10.1371/journal.pone.0125049

Knudsen AB, Slooff R. (1992). Vector-borne disease problems in rapid urbanization: New approaches to vector control. Bulletin of the World Health Organization 70(1): 1–6.

Kofidou M, de Courcy Williams M, Nearchou A, Veletza S, Gemitzi A, Karakasiliotis I. (2021). Applying remotely sensed environmental information to model mosquito populations. Sustainability 13(14): 7655. https://doi.org/10.3390/su13147655

Kumar S, Srivastava A, Maity R. (2024). Modeling climate change impacts on vector-borne disease using machine learning models: Case study of Visceral leishmaniasis (Kala-azar) from Indian state of Bihar. Expert Systems with Applications 237: 121490. https://doi.org/10.1016/j.eswa.2023.121490

Raizada S, Mala S, Shankar A. (2020). Vector borne disease outbreak prediction by machine learning. 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, pp. 213–218. https://doi.org/10.1109/ICSTCEE49637.2020.9277286

Raizada S, Mala S, Shankar A. (2021). Vector-borne disease outbreak prediction using machine learning techniques. In: Prakash KB, Kannan R, Alexander SA, Kanagachidambaresan GR, editors, Advanced deep learning for engineers and scientists. EAI/Springer innovations in communication and computing. Springer Cham. Pp. 227–241. https://doi.org/10.1007/978-3-030-66519-7_9

Shaikh SG, Kumar BS, Narang G, Pachpor NN. (2023). Diagnosis of Vector borne disease using various machine learning techniques. International Journal of Intelligent Systems and Applications in Engineering 11(4s): 517-526. https://ijisae.org/index.php/IJISAE/article/view/2721

Tito MH, Arifuzzaman M, Jannat MHE, Rahman MS, Sharmy ST, Nasrin A, Asaduzzaman M, Ashrafuzzaman M, Prince DB, Asif AH. (2023). A comparative study of ensemble machine learning algorithms for brucellosis disease prediction. Letters In Animal Biology 3(2): 23-27. https://doi.org/10.62310/liab.v3i2.119

Uddin S, Khan A, Hossain ME, Moni MA. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making 19(1): 281. https://doi.org/10.1186/s12911-019-1004-8

WHO (2014). World Health Organization, Regional Office for South-East Asia. Vector-borne diseases. https://iris.who.int/handle/10665/206531

Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, Lindsay SW. (2020). The importance of vector control for the control and elimination of vector-borne diseases. PLOS Neglected Tropical Diseases 14(1): e0007831. https://doi.org/10.1371/journal.pntd.0007831

Citation

Tito MH, Jannat MHE, Aktar MT, Saha B, Das P, Kaiser M, Hossain MA, Islam MN, Chohan MS, Khan S. (2024). Advancing vector-borne disease prediction through functional classifier integration: A novel approach for enhanced modeming. Letters in Animal Biology 04(1): 17 – 22.