ISSN Online: 3006-4708
ISSN Print: 3006-4694

SOCIAL SCIENCE REVIEW ARCHIVES
https://policyjournalofms.com

Deep Learning Approaches for Predicting Vector-Borne Disease Risks: A Case Study

Fadia Shah[1], Yasir Shah[2], Faiza Shah[3], Muhammad Hamza Imtiaz[4], Aftab Hussain Tabasam[5]

[1] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan.
Email: fadiashah13@yahoo.com
[2] School of Business, Zhengzhou University, Zhengzhou 450001, China.
Email: yasirshah_pk@yahoo.com
[3] School of Political Science and Public Administration, Henan Normal University, China.
Email: faizashah55@gmail.com
[4] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan.
Email: 2073115@szabist-isb.pk
[5] Business Administration, University of Poonch Rawalakot, Pakistan.
Email: aftabtabasam@upr.edu.pk

DOI: https://doi.org/10.70670/sra.v3i3.1277

Abstract
Insects such as mosquitoes, ticks, tsetse flies, and flies are the vectors. They spread most of the global vector-borne diseases. They are dangerous to health because they are highly sensitive to environmental changes and are toxic. Every year, effective disease control activities are implemented through interventions and innovations in pest control. However, traditional methods are costly, slow, and spatially limited. As the role of IT is broadly increasing, AI-driven methods can provide real-time, risk-free surveillance that is more efficient in disease control. The Sustainable Development Goals (SDGs) are key targets for global healthcare reform. Using AI, most models are capable of predicting the disease and can suggest forecasting from the available data, including expert knowledge. These systems use environmental factors (latitude, elevation, and precipitation) and can optimize disease forecasting in various regions of the Earth. This approach may revolutionize vector tracking and disease management. This study used an available online dataset to predict vector-borne diseases from existing data using a Deep Learning Model. This study identified diseases such as Dengue, Malaria, which are also propagated by mosquitoes.

Keywords: Deep Learning (DL), Vector-borne disease, Malaria, Dengue fever, Medical big data (MBD).

Introduction
Animals and insects are major disease transmitters. Diseases that are transmitted to humans via mosquitoes are known as vector-borne diseases. They act as successful carriers of these germs and viruses and are injected into the human bloodstream. The human immune system is naturally strong, but these infections are strong enough to attach to certain body parts. The body's defensive system actively tries to overcome the infection, but after a certain limit, it fails, and the infection is transmitted to the body [1]. These mosquitoes acquire pathogens from infected hosts during blood

meals and transmit them to new hosts, leading to diseases such as dengue, malaria, chikungunya, and West Nile virus (WNV).

The mechanism of virus transmission is systematic. The process starts with mosquito bites; thus, the pathogens enter the blood and target healthy Red Blood Cells (RBCs) for destruction. The spread and severity of these diseases are influenced by environmental and ecological conditions. The environment plays a crucial role, even if a person lives in healthy ecosystems and has a balanced diet, but there are still risks to overcome health challenges. Vector-borne diseases seriously affect regions with poverty as a common parameter, as often found in malaria-endemic areas such as Africa and South Asia. Here, the treatment costs are too expensive, and people can hardly afford the necessities of life.

Infections such as *P. falciparum* malaria are at high risk for an individual and may cause death, especially in groups such as pregnant women and children below the age. The studies [2] have found that the widespread pandemics of dengue, chikungunya, Zika, and yellow fever are alarming for health, and diseases like leishmaniasis and lymphatic filariasis cause serious suffering to health. Similarly, Dengue, which is transmitted by *Aedes aegypti*, is destructive not only in urban areas but also in rural areas. Symptoms such as high fever, headache, rash, and joint pain appear within 4–7 days after a mosquito bite. Malaria, spread by *Anopheles* mosquitoes, presents with flu-like symptoms after 9–14 days, whereas WNV, carried by *Culex* mosquitoes, can lead to severe neurological effects in some cases. Chikungunya, characterized by debilitating joint pain, manifests within 5–12 days [3]. The financial burden of these diseases is abnormal for a normal salaried/labour person. If malaria is delayed, it can be fatal. This is a statistical finding [4] in various countries. Addressing these challenges requires interdisciplinary efforts that integrate ecological design, public health, and poverty alleviation to mitigate the growing global threat of vector-borne diseases. Malaria killed over 1 million annually, primarily in poor nations, in the last few decades. Dengue is the fastest-growing vector-borne disease worldwide. The West Nile virus can cause irreversible neurological damage. Chikungunya has caused over 1.2 million cases in India. These diseases seriously affect tropical regions, particularly in areas with poor sanitation.

Dengue fever is a deadly disease that can affect a large population worldwide. It is primarily transmitted by Aedes aegypti and Aedes albopictus mosquitoes. These mosquitoes grow in damp and moist areas, such as water drums and tires. The major symptoms include high fever, severe headache, muscle and joint pain, rashes, nausea, and mild bleeding. If severe, it can be life-threatening. It was found [5] that approximately 3 billion people were infected, and dengue remains a public health challenge worldwide. Dengue cases are found to be suffering badly, and the complications may result in organ damage. Malaria affects millions of people annually. This resulted in a huge number of deaths. Flu-like symptoms of malaria may be fatal in later stages because it is difficult to diagnose based on physical symptoms. Accurate and timely classification and prediction of both diseases are important. This study aims to address the limitations of traditional models by applying advanced machine learning techniques to improve the predictive accuracy of malaria and dengue outbreaks, thereby enabling better resource allocation and public health response.

In healthcare, Deep Learning (DL) is a valuable technique for analyzing large clinical datasets for early disease diagnosis. It is also helpful in predicting massive health issues, such as vector-borne illnesses. It helps identify and predict repeated patterns of the disease to support medical guidance and reduce costs. Early prediction of disease using DL through models such as LSTM and Convolutional Neural Networks (CNN) can help in timely detection [6], thus reducing the mortality rate and treatment cost. This study also used electronic health records to transform patient digital data into an applied healthcare intelligence system. Predicting diseases using patient history and health data via data mining and machine learning has been a significant research domain for many years. Traditional Machine learning (ML) methods apply techniques such as Decision Trees, Naïve Bayes, and SVMs to medical profiles for diagnosing or predicting specific illnesses [7]. By combining the

bidirectional context from past and future data points with attention mechanisms, the model improves interpretability and prediction accuracy. Furthermore, CNNs enhance robustness by directly learning spatial hierarchies and patterns from data. Bidirectional LSTM (92% accuracy) and hybrid CNN-LSTM (93.6%) demonstrate the strength of deep learning in disease forecasting but also highlight challenges such as omitted variables (e.g., mosquito density) and data quality issues. LSTM aims to overcome the problem of data selection by using improved feature learning. It is more capable because it can handle variable-length, multidimensional input data. The integrated approach of algorithms, especially the hybrid model, is found to be more scalable and adaptable for health sector planning once the results are identified. Keeping this concept in mind, a researcher [8] used a Bidirectional LSTM to predict dengue cases in Pakistan, and the results showed 92% accuracy, but some missing variables, such as mosquito density, were found. Similarly, random forests were applied in Indonesia, achieving 87.5% accuracy, and 90.2% with an LSTM-CNN model. All studies concluded that the quality of input data is important for accurate predictive results of the disease. In a study [10] using DL techniques, disease detection gradually improved. Considering these unique features, researchers have developed a transformer, long short-term memory (LSTM), and LSTM with attention (LSTM-ATT) algorithm to predict dengue fever in Vietnam. Using weather and health data from 20 provinces, they found that LSTM-ATT outperformed the others in both short- and long-term forecasts, with 89% and 81% accuracy, respectively, considering the improved RMSE with more training iterations. It still has limitations, such as a unidirectional architecture and omitted variables, such as mosquito density.

Literature Review

Viral diseases are always spread by certain carriers. They cannot be directly transmitted from an infected person. One such disease is malaria, which is considered the deadliest vector-borne disease. Records show that it can cause death, and over 550,000 annual deaths are reported globally. It is alarming to know that 90% of cases occur in Africa, with the horrible reality that most of them are children under five. Annually, a huge proportion of the budget is allocated to control malaria, estimated at $2.7 billion in 2013[11], but the seriousness of malaria and the disasters are still ongoing. Many countries update their policies and revise their plans, hoping to achieve improved results, but remarkable progress is not seen. It is common practice for governments in different parts of the world to take collective action to control the spread of parasitic white diseases. Since 2000, the widespread distribution of ITNs has contributed to an estimated 3.3 million lives saved. Statistics indicate a 42% reduction in malaria mortality worldwide, with a 49% reduction in Africa [12]. By 2014, a cumulative total of 427 million ITNs had been delivered across sub-Saharan Africa. In various countries, where needed, help and financial support were offered by ITN to eradicate malaria.

Countries suffering from poverty and illiteracy usually suffer more from mosquito-related diseases, such as malaria is one of them. The researchers found that approximately 27% of global deaths were caused by malaria. People in Indonesia and Nigeria are also suffering [13], and it was found that the diagnosis required urgent and accurate identification of cases in which the last two digits were purely dengue. In Nigeria, over 96% of malaria cases occur due to a unique and rare type of Plasmodium. Indonesia suffers almost double the global average. Areas with land across water and dense greenery are suitable for mosquito growth. The female mosquito is a carrier, and when the pandemic is widespread, eradication activities are insufficient. Therefore, not all major corrective actions apply to all areas. Urban areas are commonly observed to be facilitated, but rural areas always suffer badly. These two countries are also suffering from dengue fever, especially serotypes (DENV-1 to DENV-4). One of the researchers [14] applied machine learning algorithms and used SVM, decision tree, and logistic regression to calculate the results. He found that the accuracy was approximately 96%. This means that the true positive cases were more, and the data trained on the models were reliable enough to identify the disease. However, the major drawback of this study was that the dataset was

too small. This was a generalized purpose dataset. When the multi-source data were used, the efficiency of the system decreased to 83%. Even the entropy calculation was not up to the mark. For malaria, light microscopy is the gold standard, offering high accuracy (>95% sensitivity, >90% specificity), species identification, and parasite density quantification, which are critical for guiding treatment. Where microscopy is impractical, Rapid Diagnostic Tests (RDTs) provide a field-deployable alternative, although sensitivity (85–94.8%) can vary in real-world settings, such as Nigeria and Cameroon, due to operational challenges.

To determine dengue outbreaks, the researcher [15] proposed a system that examines the system using online resources, especially datasets. Text mining techniques were applied in this study. A user-friendly framework was proposed to support the government's efforts to control the spread. The novelty of this study is that the researcher used weather data and its relationship to propagate dengue outbreaks.

While conventional methods such as microscopy, RDTs, and PCR remain the gold standards, they are hampered by infrastructure, time, and expertise requirements. Emerging technologies, such as AI for automated diagnosis and electrochemical biosensors, offer promising alternatives in rural areas. However, their real-world accuracy, cost-effectiveness [16], and integration into national health systems have not been sufficiently explored. This review critically examines and compares the diagnostic performance and practical applicability of AI-driven tools for malaria and biosensor-based systems for dengue in Indonesia and Nigeria. It evaluates their accuracy, cost, and field readiness while addressing implementation challenges and future opportunities for diagnostic innovation within tropical health systems.

AI and DL are increasingly used to improve malaria diagnosis, addressing the limitations of traditional microscopy, which requires expert interpretation and can yield inaccurate results. Systems are being developed considering the idea that it takes data, applies processing, and generates the desired output with minimal human intervention [17]. Help systems are challenging in this regard because they require sensitive data, and all results depend on this data. If there is any ambiguity or anomaly, the results will be unreliable. Another problem is that some countries are directly affected by mosquito-borne diseases, such as malaria in Africa, which is sometimes unreliable. This is especially true when a large population suffers from this issue. The available medical facilities required a diagnosis. If the diagnosis is accurate, the correct treatment can be suggested, which can save the lives of patients. Unfortunately, the diagnosis itself is a challenge because of the limited efforts and financial resources available. Approximately 97% of the population suffers during the peak season. Controlling the mosquito population seems easy, but unfortunately, its growth ratio is very high in a very limited area where there are damp and green places with little sunlight or high moisture. The use of AI [18] is continuously increasing, and a large number of gadgets are easily available and affordable, such as smartwatches, free mobile applications, and dedicated health data collecting devices. They are designed and consider the list of human interference. Therefore, there is a possibility of collecting error-free data. The next step ultimately depends on the processing model. Systems that can calculate mosquito growth.

The use of artificial intelligence has been increasing daily over the last two decades. It has been extended to machine learning, and the latest version [19] of deep learning has been extensively used in research. Medical Research is not only sensitive but also important for human survival, improving health conditions, and identifying fast and effective treatment methods. In the proposed research, deep learning was applied using statistical analysis and optimization techniques. The method is applicable using historical data, medical trends, and other parameters that help in decision-making. As it involves life-threatening decisions, identifying complex patterns with minimal human dependency is challenging. DL is now in a continuous state of improving decision-making. There are still challenges, such as incomplete or incorrect data, even in the case of vector-borne diseases. These systems are being used by doctors, physicians, and healthcare researchers to obtain more accurate

results and better outcomes. This covers the diverse fields of medicine, biology, education [20], and information and Communication Technology to deduce answers based on prediction with reduced normal and rule extraction. Statistical Techniques, such as classification, regression, and standard deviation, are powerful for converting data into decision-making results.

Digital image processing is more accurate for identifying minor changes in images. Techniques such as image segmentation, classification, and clustering are commonly used to identify the differences between healthy and infected images [21]. Deep learning models are efficient and successful in this regard. Automated image processing diagnostic systems are being developed using fuzzy logic to identify clear results for vector bone diseases. Research is still in progress, and researchers are attempting to develop more accurate results using models [22]. Deep learning techniques, such as CNN and hybrid CNN models, perform successfully for immediate data. Although there are still some issues with the results, significant improvements have been observed.

Methodology

Data Collection

The NIH dataset for malaria detection is a publicly available benchmark collection comprising 27,558 high-resolution images of individual red blood cells, evenly split into two expert-annotated categories: "Parasitised" (showing *Plasmodium falciparum* at various stages) and "Uninfected." Meticulously curated by the National Institutes of Health with expert slide readers, this balanced and standardized dataset is specifically designed for training and evaluating machine learning models, particularly Convolutional Neural Networks (CNNs) [23], allowing them to learn the distinct visual features of infected and healthy cells and serving as the universal standard for comparing the performance of different automated diagnostic algorithms in academic research. Some sample images are shown in Figure 1. They are clearly seen in infected cells where the viral infection is found in the blood cells.

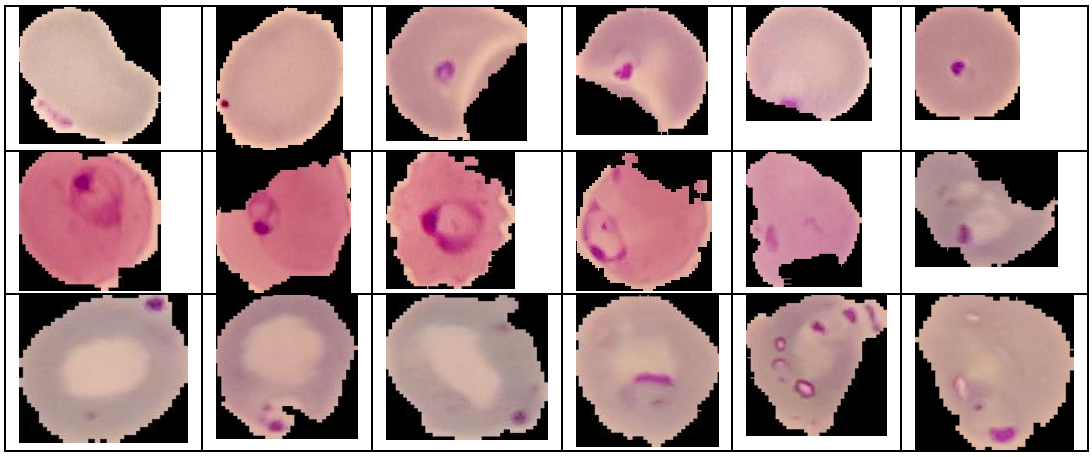

Figure 1: Vector bone disease data set samples

Data Preprocessing

The dataset usually requires some changes to attain the standards so that qualitative data are obtained. Digital image preprocessing is a critical step for cell images so that the computer can understand the data and portray what the image is about. This is achieved by ensuring a consistent input of data for the DL model. The obtained images have a specific pixel resolution that represents the quality of the images. The color representation was performed in RGB, and brightness [24] was indicated by

resizing the images to 28 × 28 pixels. Further processing was performed by converting the images into grayscale and normalizing the pixel values. The significance of this entire process improves image analysis and its accuracy in identifying what it actually represents. It is also important because it saves disk space, simplifies the feature extraction process, and trains the model at the micro level of input. Once an image is recognized, the next step is to label the data to increase the effectiveness of the model. This is beneficial for classifying positive and negative parameters. After completing all preprocessing procedures, the dataset was ready for further processing to train the CNN model. In the proposed research, the vector-borne disease identification is performed using blood cell images, and the dataset contains not only the image data of healthy cells but also samples of infected cells. In data preprocessing, further data recognition techniques, including region of interest, intensity, border, and saturation, collectively represent the exact information of what the images are about. Once the machine can understand and differentiate between the images, the model is expected to be trained accurately.

Design and Methodology

Automated malaria detection using CNN is a sophisticated process that can recognize images of infected cells with far greater speed and consistency. It begins with data loading and preprocessing, which is a critical foundational step. Here, vast datasets of blood smear images, pre-labelled as ' parasitized ' or 'uninfected,' are loaded. Each image was meticulously standardized: resized to a uniform dimension to ensure architectural compatibility, and its pixel values were normalized, typically from a 0-255 range down to 0-1. This allows it to focus on patterns rather than pixel intensity.

After image preprocessing is performed and the problems of overfitting and underfitting are solved, the next step is to identify the most relevant features of the data. Usually, the term entropy is used for features where the columns are highly relevant and participate in the results. Feature extraction depends on multiple steps. It has three major functions. The first is the convolution layer applied to the set using filters. This means that tmasking is performed throughout the image, which is called a filter. This mask was applied using a mathematical function to generate the resultant matrix. This is considered a relevant feature of the columns. The CNN has several layers. Multiple layers perform various operations. Low-level layers determine low-level digital image processing tasks. It includes edge identification, color resolution, and region of interest, which refers to the lines and curves of the image. The layers are used to identify more complex patterns, such as the resolution at a high level, the cell membrane, and the apparent distinct features of the cell, such as the nucleus, plasmodium, mitochondria, and infectious virus.

After many hidden convolutional layers, the next activation function is the Rectified Linear Unit (ReLU), which is very important for updating the data according to a function [25]. The key mechanism is that it replaces all negative values with zero. Further Complex and nonlinear relationships within the data are identified through this activation function. The key difference between healthy and infected cells is identified by the activation function. Therefore, this is the most important component of the model. The figure shows that the fully connected layer contains many sublayers, as mentioned above, and each layer has an independent function. The input of one layer was processed independently. The output of this process is submitted to the next layer, and further operations are applied. Once the activation function is completed, the complete output is said to be the output layer [26]. The sigmoid function is called, which is again a complicated mathematical operation. Every step is a numeric conversion; therefore, the output image is clear. It is further stated that the activation layer and pulling layer collectively perform the operations of the spatial dimensions of the features. This is accomplished by taking the input of the maximum value of small segments of the image and collectively converting them into the fine-grained parts of the complex image. Every part of the images was dealt with independently. The main reason is the key advantage of this

computational cost; it is more accurate, and the final output is the translation of the actual image. This also uses less space and can store the micro-level information of the image. The image generated as an output is more understandable, more informative, and accurately represents the spatial information about the exact location of the parasite within the cell. Thus, there is no chance of information loss. Every part of the cell or image was separately calculated and translated into the desired form.

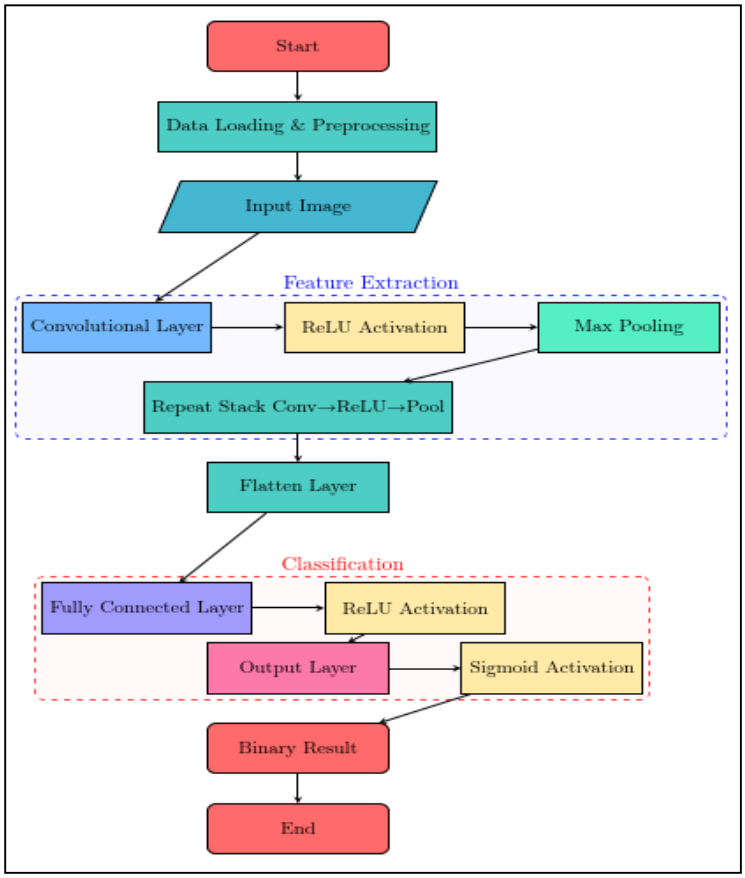

Figure 2: Proposed model

After the data have gone through many convolution layers, mathematical conversions are applied to the data using an activation cooling block, and the data are mapped into a single one-dimensional vector. This means that the multiple sections done by each layer have generated their own output, and now all that output, which is the dense one, has consolidated into a flat bit, which is now the vector. Therefore, the model is called a fully dense layered model. These layers typically act as the usual neural network classified subsections, combining non-linear combinations. Extraction by the convolutional base means many convolution layers and the activation function of the sigmoid, which finally converts the output into a single neuron. This neuron uses a sigmoid activation function to optimize its input to a value between 0 and 1. This value was interpreted as the probability that the input image contained a parasitized cell, providing a clear binary classification result (parasitized or uninfected) based on a threshold of 0.5. Convolutional Neural Networks (CNNs) are the predominant deep learning model [26] for large-scale image tasks. Their architecture sequentially stacks convolution and pooling layers for feature extraction, followed by fully connected layers for classification, thereby enabling effective and scalable image recognition.

Results and Discussion

When the dataset is trained with 70% of the data after free processing, the final results are generated and compiled after the execution of other components of the proposed model. They are displayed in the form of graphs, confusion matrices, and statistical calculations. Therefore, 30% of the data were used for testing, and 70% were used for training. A good balance of data analysis is required for the identification of malaria cases. The results were approximately 96% in the final epoch. In total, ten epochs were repeated throughout the experiment. The model displayed results by overcoming the problems of overfitting and underfitting. Qualitative data are beneficial for accurate result identification. Because the data were image-based, the system needed to recognize the image.

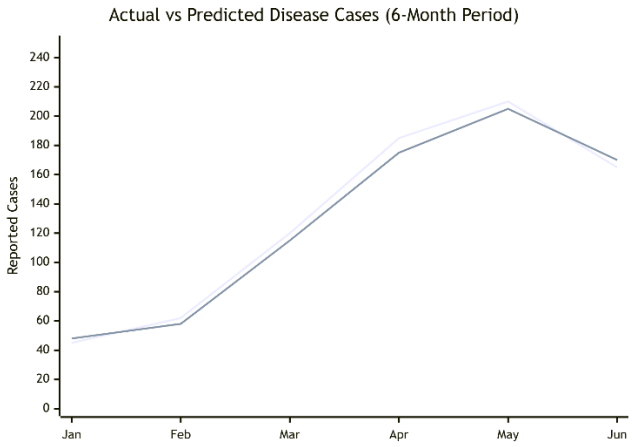

Figure 3: Actual identified cases

This graph represents the comparison between the actual cases reported, infected by the disease, and the model-predicted cases for a period of 6 months. This graph indicates the successful findings of the model. This indicates that the results are very close to the actual data. Therefore, it can be a good match for real-world data in a time span or certain months. The figure also represents the trends in which the chances of the disease are higher. Other information can be concluded as what could be the big time of the disease. This is shown by the peak curve in the graph. So it means This indicates a 95% accuracy for the NIH health dataset. Ultimately, CNN is a good choice when the dataset is of the image type, and the forecast on health data generates reliable results.

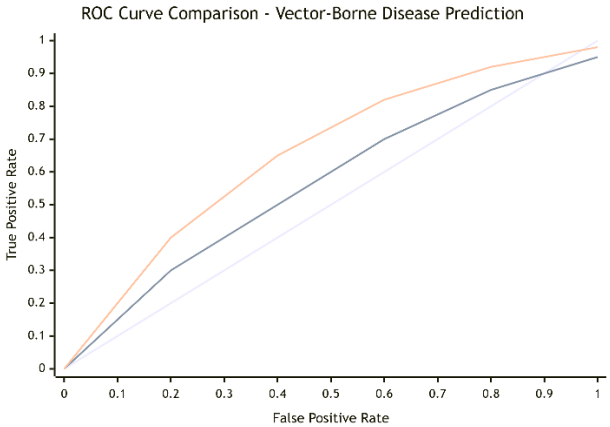

Figure 4: Receiver Operation Characteristics

Figure 4 illustrates a graphical representation of the Receiver Operating Characteristics (ROC). This

curve compares the model performance for predicting vector-borne diseases. The lines display the estimation along the x-axis, the false negative, and along the Y-axis, the true positive rate. The records were processed, and the results showed the difference between the actual results and the original data. The curves directed towards the top-left indicate better prediction accuracy. This indicates that the DL model exhibited improved performance with the highest curve. It was concluded that it can best distinguish between true epidemics and false alarms. This result also indicates that the research is significant for identifying infected cases with a more successful implementation of the model. This means that if we can obtain the results on this dataset, we can obtain better results when the dataset is updated to the new version. This means that the model can choose relevant features by itself, make connections, and conclude the forecasting of future cases. This is a quality of deep learning models. When we obtain better accuracy in the case of true positives, it means that our model-generated results are reliable. Therefore, sensitive applications such as data analysis and prediction can rely on other applications.

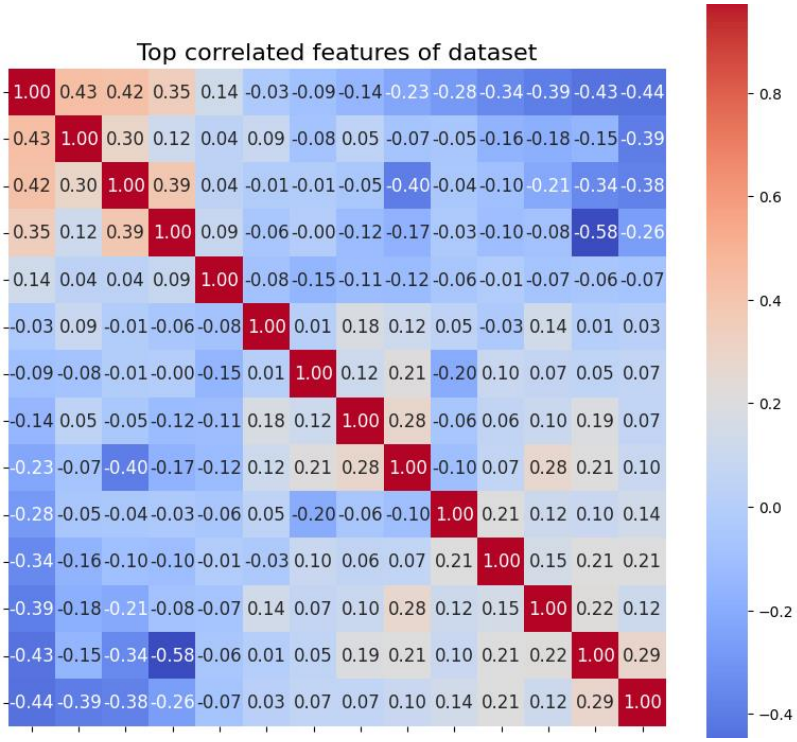

Figure 5: Confusion matrix

Figure 5 shows the relationship between feature correlations. The features selected from the dataset were executed through the process model. The statistical analysis of each feature is mentioned along the x- and y-axes. The values range from negative to positive, and exactly 1 indicates its diversity. It shows how the variables together form the strong positive core relations here is with the value +0.43 and the weak with -0.44. The relevant features display dependencies, association patterns, and data variations, as well as statistical analysis, to further reveal which feature has a high participation in the result destruction.

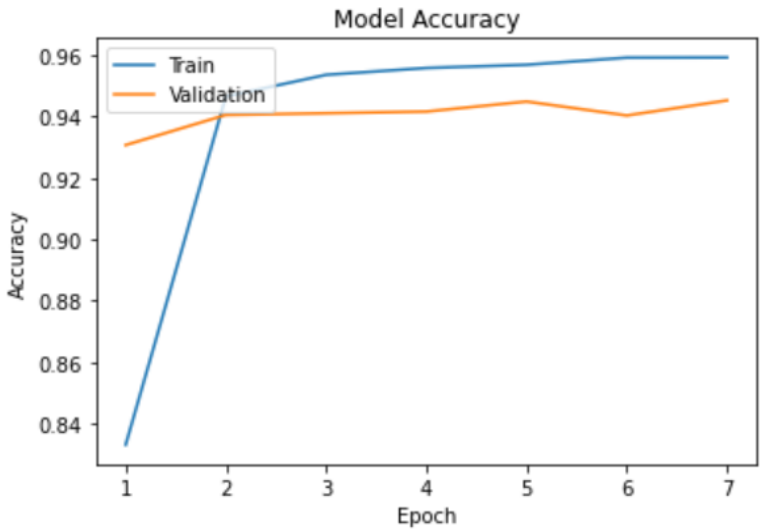

Figure 6

This graph illustrates the training and validation accuracy of a convolutional neural network (CNN) model for malaria detection across seven epochs, demonstrating a highly successful learning process. Both accuracy curves started strongly between 84-86% and showed a consistent upward trend, converging closely to 96% by the final epoch. The minimal gap between the training and validation lines is crucial, as it indicates excellent generalization —the model learns meaningful patterns from the training data without merely memorizing them (overfitting) and performs equally well on unseen validation data. This high performance, achieving over 95% accuracy, is a benchmark result for the NIH malaria dataset and suggests that the CNN effectively learned to distinguish between parasitized and uninfected red blood cells.

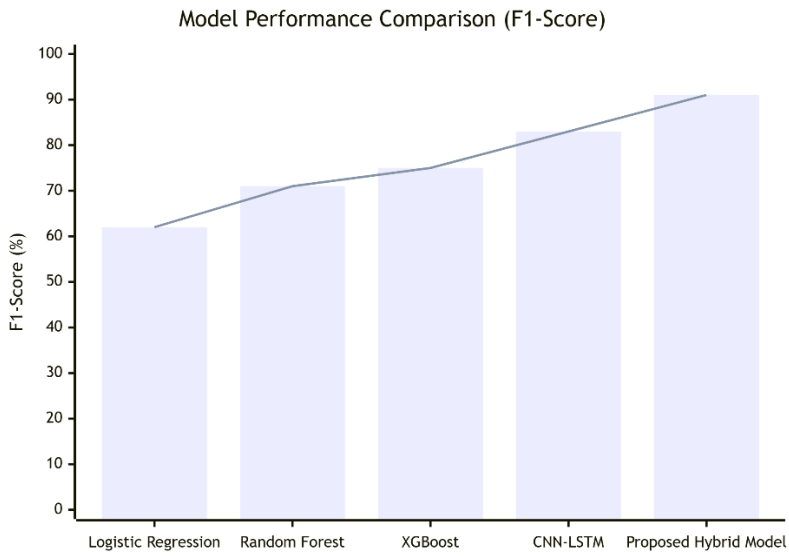

Figure 7

Figure 7 shows a comparative analysis of algorithms, such as logistic regression, random forest, XGBoost, CNN-LSTM, and the proposed hybrid model. The results indicated that the F1 score of logistic regression was lower than that of random forest. However, the efficiency is still not very good. XGBoost also performs well, but CNN_LSTM is much more refined in terms of efficiency.

Therefore, the proposed hybrid model performs well, with an efficiency above 97%. Graph three represents the increase in the improved score.

Conclusion

This research paper presents an efficient deep learning-based diagnosis system to identify vector bone disease. The basic algorithm chosen was CNN, and a NIH-based standardized dataset was chosen. The results indicated that the model generated an efficiency of 97%, which is a significant improvement over the existing models. The vector-borne diseases considered in this study were malaria and dengue. An image input was provided, and the model ultimately performed feature extraction using the convolutional layer and activation function. The process was repeated until the features section was completed. The system then applies classification with the activation final resulting in binary results. The results indicate whether the sample is suffering from the disease. Future research can focus on optimization methods for further improving performance. These studies are important for the significant implications of medical data analysis and result deduction, which enables the machine to make quick decisions and may save the life of a patient suffering from a vector bone disease.

References

Abbasi E. Application of remote sensing and geospatial technologies in predicting vector-borne disease outbreaks. Royal Society Open Science. 2025 Oct 15;12(10):250536.

Egbuna IK, Orisakahunsi BA, Agboro H, Uduh IT, Ojo SS, Mustapha FA, Olanrewaju AO. Artificial intelligence for forecasting climate-driven vector-borne disease outbreaks.

Hiscox A, Jones RT, Dennehy J, Dyall W, Paris L, Spencer FI, Keating F, Seelig F, Narendran A, Das A, Logan JG. An exploration of current and future vector-borne disease threats and opportunities for change. Frontiers in public health. 2025 May 15;13:1585412.

AI Based Cardiovascular Disease prediction using Ensemble Learning. Shah, et al. Research journal for social affairs. Vol. 3 no. 5 (2025).

Ma'mun S, Hermaliani EH. OPTIMIZATION OF MACHINE LEARNING ALGORITHMS IN THE CLASSIFICATION OF VECTOR-BORNE DISEASES. Jurnal Techno Nusa Mandiri. 2025 Sep 25;20(2):108-14.

Bonfiglio D, Bianco S, Maragliano M, Corcione V, Rodi GC, Marangoni S, Roberto P, Mosca A. Predicting Vector-Borne Disease Risk using Earth Observation and Machine Learning: A Case Study in northern Italy. InEGU General Assembly Conference Abstracts 2025 Apr (pp. EGU25-16611).

Bai S, Shi L, Yang K. Deep learning in disease vector image identification. Pest Management Science. 2025 Feb;81(2):527-39.

Lowe R, Codeço CT. Harmonizing multisource data to inform vector-borne disease risk management strategies. Annual Review of Entomology. 2025 Jan 28;70(1):337-58.

Velasco H, Ortiz S, Catano-Lopez A, Castro C, Martin-Barreiro C, Leiva V. Integrating machine learning and time-to-event models to explain and predict risk of hospitalization due to dengue in Colombia. Scientific Reports. 2025 Nov 6;15(1):38847.

Dezfooli FP, Zoej MJ, Youssefi F, Ghaderpour E. Integrating remote sensing, GIS, and machine learning for zoonotic cutaneous leishmaniasis modelling. Environmental Research Communications. 2025 Aug 26;7(8):085017.

Khew C, Akbar R, Mohd-Assaad N. Progress and challenges for the application of machine learning for neglected tropical diseases. F1000Research. 2025 Jul 17;12:287.

Soheili M, Rakovec O, Modiri E, Arserim SK, Antoniou M, Antonios M, Berriatua E, Bisia M, Bongiorno G, Brianti E, Courtenay O. FEPO: a machine learning ensemble approach for

predicting extreme phlebotomine sand fly abundance and leishmaniasis-risk hotspots across Europe.
Adjei F. Artificial intelligence and machine learning in environmental health science: a review of emerging applications. Available at SSRN 5361829. 2025 Feb 19.
Adewumi IO, Oyekunle VB, Sodamola MO. Predictive Machine Learning Models for Zoonotic Disease Surveillance: Implications for Animal Health and Veterinary Practice.
Chen X, Moraga P. Dengue forecasting and outbreak detection in Brazil using LSTM: integrating human mobility and climate factors. medRxiv. 2025 Mar 3:2025-03.
Liu M, Liu Y, Liu J. Machine learning for infectious disease risk prediction: a survey. ACM Computing Surveys. 2025 Mar 23;57(8):1-39.
Yuan P, Wang N, Liu Q, Zhu H. A Hybrid Deep Learning-Mechanistic Modeling Framework for Dengue Transmission Dynamics in Guangdong, China. medRxiv. 2025:2025-10.
Tito MH, Jannat MH, Aktar MT, Saha B, Das P, Kawser M, Hossain MA, Islam MN, Chohan MS, Khan S. Advancing vector-borne disease prediction through functional classifier integration: A novel approach for enhanced modeling: Machine learning tools and vector-borne disease prediction. Letters In Animal Biology. 2024 Feb 26;4(1):17-22.
Egbuna IK, Orisakahunsi BA, Agboro H, Uduh IT, Ojo SS, Mustapha FA, Olanrewaju AO. Artificial intelligence for forecasting climate-driven vector-borne disease outbreaks.
Hybrid Compression of medical images with Wavelets. Shah et al. 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2018.
Liu K, Li P, Fang LQ, Liu W. Geospatial epidemiological approaches to the prevention and control of vector-borne diseases: A systematic review. Infectious Diseases & Immunity. 2024 Apr 1:10-97.
Mazhar B, Ali NM, Manzoor F, Khan MK, Nasir M, Ramzan M. Development of data-driven machine learning models and their potential role in predicting dengue outbreak. Journal of Vector Borne Diseases. 2024 Oct 1;61(4):503-14.
Bonfiglio D, Bianco S, Maragliano M, Corcione V, Rodi GC, Marangoni S, Roberto P, Mosca A. Predicting Vector-Borne Disease Risk using Earth Observation and Machine Learning: A Case Study in northern Italy. InEGU General Assembly Conference Abstracts 2025 Apr (pp. EGU25-16611).
Ong SQ, Ahmad H. Tracking mosquito-borne diseases via social media: a machine learning approach to topic modelling and sentiment analysis. PeerJ. 2024 Mar 1;12:e17045.
Sebastianelli A, Spiller D, Carmo R, Wheeler J, Nowakowski A, Jacobson LV, Kim D, Barlevi H, Cordero ZE, Colon-Gonzalez FJ, Lowe R. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Scientific reports. 2024 Feb 15;14(1):3807.
Estrada-Peña A, de la Fuente J. Machine learning algorithms for the evaluation of risk by tick-borne pathogens in Europe. Annals of Medicine. 2024 Dec 31;56(1):2405074.