WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Diagnosing Malaria and Jaundice Using Selected Machine Learning Methods

[a]*Olajide Blessing Olajide, [b]Soneye Olufemi Sobowale, [c]Ogunniyi Olufunke Kemi, [c]Adeogun Adetoyese Elijah, [d]Temitope Omotayo Olayinka, [e]Ogunjinmi Bamidele Oluwabunmi, [f]Adeosun Titilayo Helen

[a] Department of Computer Engineering, Federal University Wukari, Wukari, Nigeria.
[b] Department of Computer Engineering, Lagos State University of Science and Technology, Ikorodu, Lagos State, Nigeria.
[c] Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
[d] Department of Electrical and Computer Engineering, University of Maine, Orono, United State.
[e] Department of Computer Science, Ajayi Crowther University, Oyo, Nigeria.
[f] Department of Accounting, University of Ilesa, Ilesa, Nigeria.

A B S T R A C T

Malaria and jaundice are critical diseases caused by the bite of infected female Anopheles mosquitoes and by viral infections of the liver (including hepatitis A, B, C, D, and E) or other factors, respectively. The primary method for diagnosing these conditions involves analyzing microscopic images of blood samples. This process necessitates the expertise of a trained microscopist, which can be time-intensive and may yield suboptimal results for large-scale screenings. Consequently, this study presents the development of four machine learning predictive models aimed at diagnosing malaria and jaundice. The models utilize Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), and Naïve Bayes (NB) algorithms. The research further evaluates the performance of each model based on accuracy and recall metrics. For the prediction of malaria, jaundice, and unaffected samples, the SVM achieved accuracies of 86.46%, 87.50%, and 87.78%, respectively; KNN recorded accuracies of 82.03%, 81.60%, and 83.05%; DT achieved 71.61%, 67.71%, and 71.94%; while NB had accuracies of 75.78%, 73.39%, and 76.39%. In terms of recall for malaria, jaundice, and unaffected samples, SVM demonstrated values of 88.33%, 90.00%, and 89.78%, respectively; KNN had recalls of 84.38%, 84.72%, and 85.56%; DT recorded recalls of 73.96%, 70.83%, and 71.16%; and NB achieved recalls of 78.12%, 76.39%, and 78.89%. The findings indicate that the SVM model outperformed the other three models. The study concludes that the application of machine learning algorithms significantly improves early detection, optimizes resource allocation for interventions, and ultimately mitigates the incidence and effects of malaria on human health.

Keywords: malaria, jaundice, diagnosis, machine learning, accuracy, recall.

1. INTRODUCTION

Malaria is an infectious condition that occur following the bite of an infected female anopheles' mosquito, which transmits protists of the Plasmodium genus into the blood of the infected humans. The mosquito injects its saliva having protist into the circulatory system of the human, which then transit to the liver where they reproduce and multiply [1]. The most critical form of malaria is attributed to Plasmodium falciparum. Other species affecting humans, such as P. vivax, P. ovale, P. malariae, and occasionally P. knowlesi, can lead to acute and severe illness, although they generally present lower mortality rates. Malaria is one of the foremost infectious diseases in tropical and subtropical regions, continuing to pose a substantial global health issue, with more than 40% of the world's population facing varying levels of malaria risk across around 100 countries. The diagnosis of malaria entails the identification of parasites or their antigens/products in the blood of affected individuals. Despite the apparent simplicity of this process, its effectiveness is contingent upon various factors, such as the distinct forms of the five malaria species, the stages of erythrocytic schizogony, the endemic characteristics of each species, the relationship between transmission levels, population dynamics, parasitemia, immunity, and clinical symptoms [2]. The exigency of malaria necessitates prompt and appropriate treatment. Delays in both diagnosis and treatment are among the factors that account for mortality in numerous countries [3]. The diagnosis of malaria can pose challenges, particularly in regions where the disease is no longer endemic, as healthcare providers may lack familiarity with its presentation. Clinicians might not pay close attention to malaria as a potential diagnosis for certain patients, resulting in the omission of essential diagnostic tests. Additionally, laboratory technicians may not possess adequate experience with malaria, which can lead to missed detections of parasites during blood smear examinations. In certain regions, such as Africa, the intensity of malaria transmission can result in a substantial portion of the population being infected yet asymptomatic. These individuals have developed enough immunity to prevent the manifestation of malarial illness, although they remain susceptible to infection. Consequently, the presence of malaria parasites in a symptomatic individual does not necessarily

indicate that the illness is attributable to these parasites. Furthermore, healthcare personnel often face challenges such as inadequate training, insufficient resources, overwhelming patient load and low compensation [3].

Jaundice is characterized by a yellowish discoloration of the skin, sclera, and mucous membranes due to the accumulation of bilirubin, a yellow-orange pigment found in bile. Bilirubin is produced as a result of the breakdown of heme rings, primarily from the metabolism of red blood cells. Clinically, this discoloration is usually observed when the serum bilirubin concentration exceeds 3 mg per dL (51.3 μmol per L). In adult patients, jaundice may arise from a diverse range of both benign and potentially life-threatening conditions [4]. Jaundice results from the buildup of bilirubin in the bloodstream. This condition may arise from an excessive production of metabolites required for the metabolism and elimination of bilirubin. The occurrence of jaundice in infants is estimated to be between 1 in 2,500 and 5,000 live births, with likely underlying conditions varying from benign breast milk jaundice to severe, life-threatening disorders such as biliary atresia (BA) and liver failure. While the clinical manifestations of certain diseases are readily apparent, others may present with more subtle signs, requiring a heightened level of suspicion for accurate diagnosis [4]. In situations with limited resources, the diagnosis of acute bilirubin encephalopathy (ABE) is often made without the aid of magnetic resonance imaging (MRI) or auditory brainstem response (ABR) confirmation. In such contexts, seasoned pediatricians possess the skills necessary to identify the specific signs associated with this condition. Nevertheless, efforts to assess the prevalence and severity of ABE across various communities and among physicians with differing levels of expertise have historically faced challenges due to the absence of a standardized diagnostic algorithm [4]. In 1999, the Bilirubin-induced Neurologic Dysfunction (BIND) scoring system was introduced, which allocates 1, 2, or 3 points to reflect mild, moderate, or severe abnormalities in an infant's mental status, muscle tone, or cry. Elevated scores signify an increase in acute neurotoxicity linked to excessive hyperbilirubinemia, thereby offering a shared descriptive framework for both clinicians and researchers [4].

To assess the severity of neonatal jaundice, the BIND scoring system offers a straightforward approach, akin to the Apgar score, making it a valuable diagnostic tool for identifying acute bilirubin encephalopathy (ABE). The Modified BIND scoring algorithm includes the assessment of upward gaze abnormalities, a recognized indicator of bilirubin encephalopathy, along with additional qualifiers. The objective was to create a more detailed scoring system that could more effectively differentiate between varying degrees of BIND severity. In this revised scoring framework, the maximum achievable score for BIND-M is 12. A score ranging from 1 to 4 is anticipated to signify mild ABE, which is typically reversible with prompt and aggressive treatment. A score of 5 to 6 is expected to indicate moderate ABE, which may also be reversible with urgent bilirubin reduction. Scores of 7 and above are likely to reflect severe or very severe ABE, often associated with irreversible brain damage in most affected infants. The integration of medical diagnostic systems with bioinformatics-inspired methodologies can yield valuable insights from medical data, under the guidance of a human expert. Such insights can aid medical professionals in categorizing diseases in patients and facilitate timely treatment interventions. In light of this, the research focused on developing and comparing the efficacy of selected machine learning models to predict malaria and jaundice conditions. To assess the severity of jaundice, the BIND scoring system offers a straightforward approach, akin to the Apgar score, making it a valuable diagnostic tool for identifying acute bilirubin encephalopathy (ABE) [4]. The Modified BIND scoring algorithm includes the assessment of upward gaze abnormalities, a recognized indicator of bilirubin encephalopathy, along with additional qualifiers. The objective was to create a more detailed scoring system that could more effectively differentiate between varying degrees of BIND severity. In this revised scoring framework, the maximum achievable score for BIND-M is 12. A score ranging from 1 to 4 is anticipated to signify mild ABE, which is typically reversible with prompt and aggressive treatment. A score of 5 to 6 is expected to indicate moderate ABE, which may also be reversible with urgent bilirubin reduction. Scores of 7 and above are likely to reflect severe or very severe ABE, often associated with irreversible brain damage in most affected infants. The integration of medical diagnostic systems with bioinformatics-inspired methodologies can yield valuable insights from medical data, under the guidance of a human expert. Such insights can aid medical professionals in categorizing diseases in patients and facilitate timely treatment interventions [5]. In light of this, the research focused on developing and comparing the efficacy of selected machine learning models to predict malaria and jaundice conditions.

2. LITERATURE REVIEW

Mosabbir and Islam [6] in their study created an ensemble deep learning model to identify malaria parasites in red blood cell imagery. This model integrates three different architectures: VGG16, VGG19, and DenseNet201, to form an adaptive weighted average ensemble. To mitigate the variability in predictions, a max voting ensemble approach was implemented in conjunction with the adaptive weighted average ensemble models. The study also incorporated data augmentation methods to expand the dataset and tackle the overfitting challenge faced by the model. The ensemble learning model proposed in this research achieved remarkable performance, with an accuracy rate of 97.92% in classifying parasitized versus uninfected cells. Dennis *et al.,* [7] conducted a study to examine twelve predictor variables in relation to the outcome variable, which consisted of malaria test results. Five machine learning models were utilized in this analysis: k-nearest neighbors, support vector machines, random forest, tree bagging, and boosting. Upon evaluating the models, the random forest model was identified as the most effective for classifying and predicting the final malaria test results. It achieved a classification accuracy of 97.33%, a sensitivity of 71.1%, a specificity of 98.4%, a balanced accuracy of 84.7%, and an area under the curve of 98.3%. The final model indicated that the presence of Plasmodium falciparum was the most significant feature, followed by region, endemic zone, and anemia level. The feature with the least importance in predicting final malaria test results was having mosquito nets. Dhrgam et al.,[8] assessed the performance of Convolutional Neural Networks (CNNs) in the identification of malaria by analyzing cell images. The dataset comprises 27,558 images obtained from the prominent Malaria Cell Images Dataset on Kaggle, which includes various types of cells. To mitigate the challenges associated with vanishing gradients and to promote convergence, the activation function known as Rectified Linear Unit (ReLU) is effectively utilized. The evaluation of the model's performance through a perplexity grid reveals remarkable results: a precision rate of 99.59%, specificity of 99.69%, sensitivity of 99.40%, an F1 score of 99.44%, and a precision of 99.48%. These findings illustrate the model's ability to accurately distinguish between cells affected by malaria and those that are not. The study highlights the considerable promise of CNNs in the automated detection of malaria through image analysis.

Ahmed et al.,[9] developed machine learning model for detecting utilizing machine learning algorithms, which were formulated from a dataset containing 767 images of infants, captured using a computer device and a USB webcam. The research assessed the efficacy of four machine learning algorithms: support vector machine (SVM), k nearest neighbor (k-NN), random forest (RF), and extreme gradient boost (XGBoost), all based on the aforementioned dataset. This approach allows for jaundice detection in a patient with minimal effort, simply by capturing the patient's image through a USB webcam. The initial results from the evaluation of the machine learning algorithms revealed that XGBoost achieved the highest accuracy at 99.63%, surpassing the other algorithms. The RF algorithm followed closely with an accuracy of 98.99%, while the k-NN algorithm attained an accuracy of 98.25%. The SVM algorithm exhibited the lowest performance among the group, recording an accuracy of 96.22%. Oguntimilehin *et al*., [10] proposed an innovative machine learning technique for diagnosing jaundice in clinical settings. Researchers compiled a labeled dataset reflecting the severity levels of jaundice, as assessed by healthcare experts. This dataset includes diagnostic variables and classifies severity into five distinct categories: very low, low, moderate, high, and very high, to formulate effective guidelines for improving diagnostic precision. The researchers indicated that the implementation of this system could result in a reduction in both mortality rates and patient wait times. Nonetheless, a significant limitation identified in the research was the issue of rule extraction, which, if resolved, could enhance diagnostic accuracy. Samuel and Omisore [11] presented a combination of fuzzy logic and neural networks for the effective diagnosis of jaundice. This hybrid model enables the neural network component to autonomously refine the diagnostic process by formulating diagnostic rules for the fuzzy inference system. Reports indicate that this combined approach delivers a reliable diagnosis that is both time-efficient and cost-effective. Nevertheless, the proposed hybrid model may result in computational burdens stemming from the unverified concept of weight adjustment within neural networks. Fatumo et al., [12] A robust computer-simulated medical expert was developed, utilizing input diagnosis variables as rules stored within the inference engine to identify various types of malaria-related issues. This medical expert system demonstrates effectiveness and accessibility; however, a limited number of rules and symptoms in the knowledge base may hinder its overall efficacy. Djam et al. [13] created a fuzzy expert system for diagnosing and treating malaria, which relies on the degree of participation of each diagnostic variable, employing the root sum square and center of gravity for reasoning and decision-making, respectively. This fuzzy expert system successfully provided reasonable diagnoses for malaria with a certain level of confidence. The authors highlighted the user-friendly nature of the system, which aims to facilitate medical consultations. Nonetheless, a significant drawback of the system lies in the challenges of knowledge representation that are common to many rule-based systems. Aminu et al. [14] introduced a predictive symptoms-based system that utilizes binary classification through Support Vector Machines (SVM) to improve the joint classification of malaria and jaundice. The researchers indicated that this proposed system serves as a dependable alternative for disease diagnosis, although the evaluation results revealed a low level of classification accuracy. Boruah and Kakoty [15] conducted a comparative study of various data mining techniques aimed at predicting and diagnosing malaria. Their findings suggested that ensemble data mining techniques may offer greater efficiency in malaria prediction and diagnosis compared to individual predictive models. Femi et al. [16] created a decision support system for diagnosing malaria and fever, employing a bioinformatics approach. This technology originated from the concept of developing computer-based tools to identify functional patterns within biological sequences, such as the locations of functional structures in Deoxyribonucleic Acid (DNA). The researchers noted that many existing studies on disease diagnosis systems often neglect to assess their systems in terms of detection accuracy, simplicity, and accessibility. Consequently, there is a pressing need to establish a multi-target disease diagnosis system capable of identifying two or more disease conditions in patients.

3. METHODOLOGY

The environmental setup for developing the SVM and LSTM models for analyzing the computer network user's activities from a dataset sourced from the Kaggle machine learning repository for this study are; a 64-bit Windows Operating System, with an Intel(R) Corel (TM) i5-3630QM CPU @2.40GHZ with 4.00 GB of RAM, Anaconda programming environment using the Python 3.8 software development kit as the programming language, Sklearn API application programming interface, Python dependencies such as NumPy for vector operations, pandas for reading files, TensorFlow API (Application Programming Interface), and the Mat-plot library for data visualization operations. This research was divided into three stages which are; data pre-processing state, model implementation stage, and model evaluation stage. In the data pre-processing stage, the techniques used consist of reading dataset gotten from Kaggle repository. The dataset comprises of microscopic blood cell images of infected and non-infected tested patient for malaria, and jaundice. Data filtration was conducted on the dataset to treat missing values, scaling was done to convert non numeric data to numeric. The images in the dataset were cropped by cropping tools to trim off the excessive edges in order to give the precised image that was needed. Image resizing was done and contrast adjustment for images to be clear for its contents to be viewed and read. Further to this, Local Binary Pattern (LBP) was also employed to carry out feature extraction from the predictors in the dataset. The model implementation stage of the methodology involved splitting the pre-processed dataset into 70:30 ratio for training and test set respectively. Also, the pre-processed dataset was fed into the four proposed machine learning algorithms namely; Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT) and Naïve Bayes (NB) Algorithms. The methodology described is as illustrated in Figure 1.

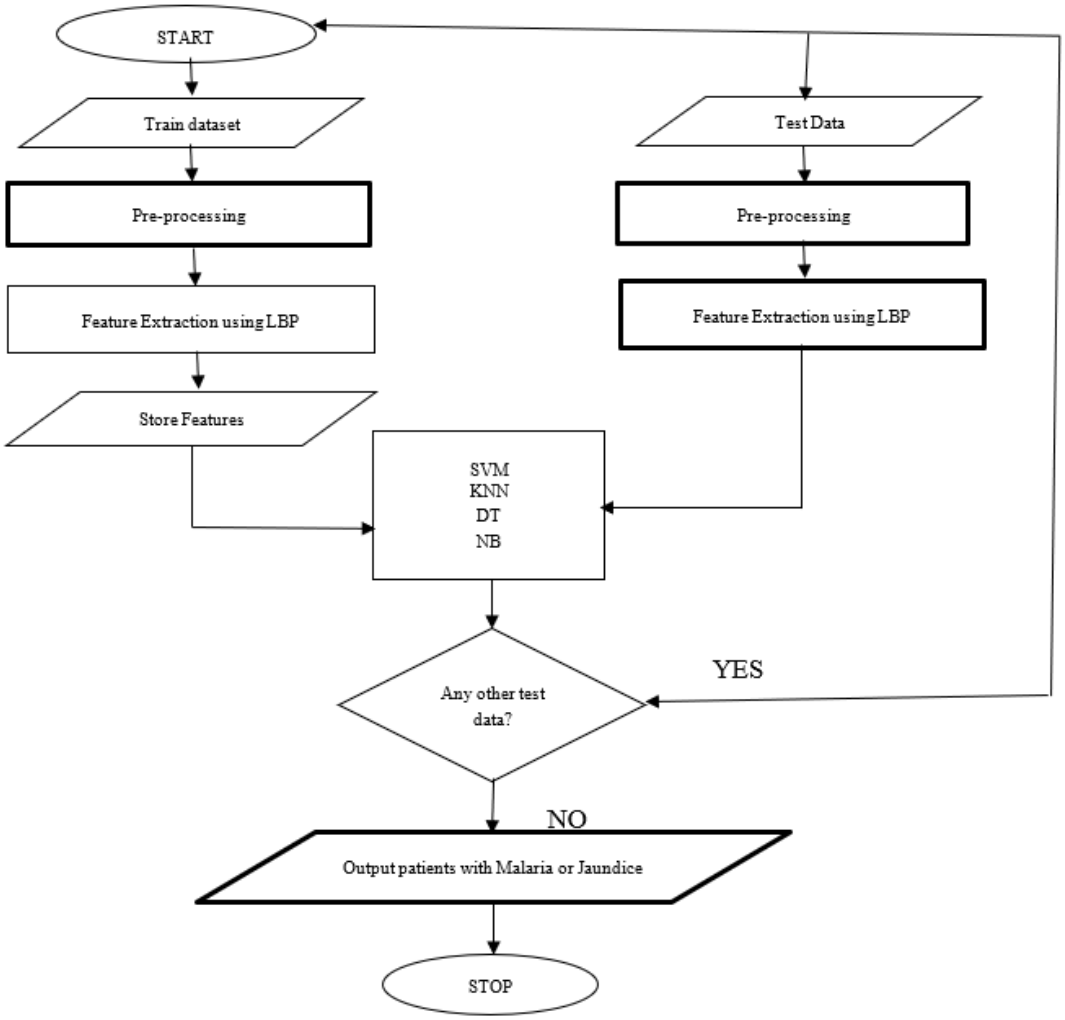

Figure 1: Flowchart for the Development of Machine Learning Predictive Model for Malaria and Jaundice

Finaly, performance evaluation was done using the following metrics; false positive, recall, specification, precision, accuracy and precision time. The operation of the four selected machine learning algorithms for the models are described as follow;

3.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) employs methodologies for addressing two-group classification problems through various classification techniques. The foundation of the SVM algorithm lies in the concept of decision planes, utilizing hyperplanes to classify a specific collection of objects. Upon receiving a labeled dataset, SVM is capable of classifying new data [17]. The primary objective of the SVM approach is to establish a line or decision boundary that separates n-dimensional space into distinct classes, facilitating the swift assignment of future data points to the correct category. The hyperplane serves as the optimal decision boundary, selecting the extreme vectors and points that contribute to its formation [18]. This is illustrated in the Figure 2:

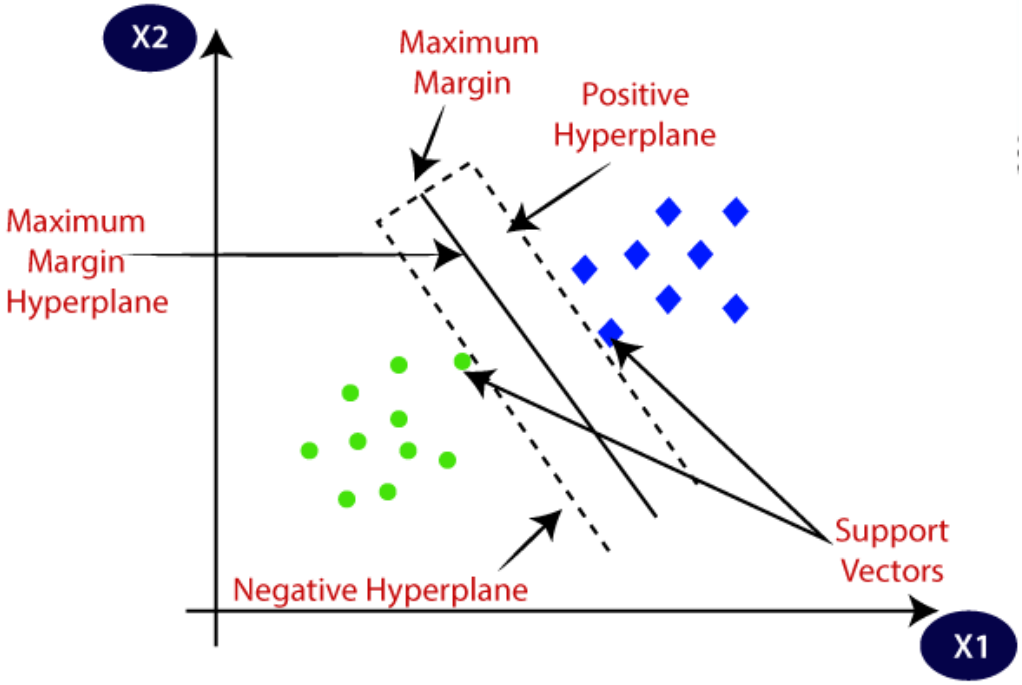

Figure 2: Classification of two different categories using a decision boundary or hyperplane (source: [18])

3.2 K-Nearest Neighbor (KNN)

KNN is a supervised learning algorithm capable of performing both classification and regression tasks. Assume we are given a dataset where X is a matrix of features from an observation and Y is a class label. k-nearest neighbors then, is a method of classification that estimates the conditional distribution of Y given X and classifies an observation to the class with the highest probability [19]. Given a positive integer k, k-nearest neighbors look at the k observations closest to a test observation x_o and estimates the conditional probability that it belongs to class j_i using equation 1;

$$Pr(Y = j|X = x_0) = \frac{1}{k}\sum_{i\in N_0} I(yi = j) \quad (1)$$

where $N\mathbf{0}$ is the set of k-nearest observations and I(yi=j) is an indicator variable that evaluates to 1 if a given observation (x_i, y_i) in N_0 is a member of class j, and 0 if otherwise. After estimating these probabilities, k-nearest neighbors assign the observation x_0 to the class which the previous probability is the greatest. Figure 3 illustrate how the algorithm works:

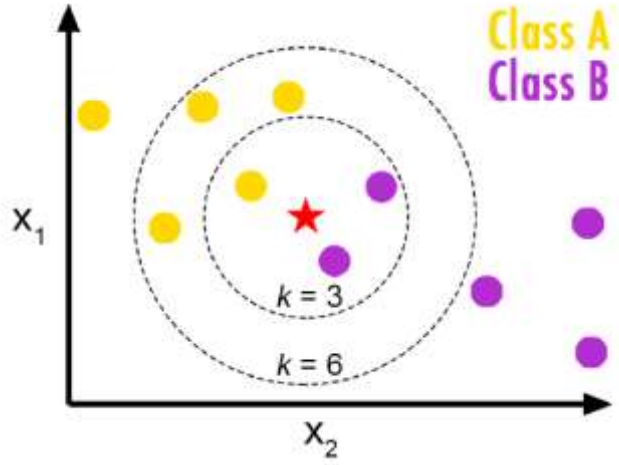

Figure 3: K Nearest Neigbhor Algorithm's Operation (source: [20])

- If we choose K=3, then we have 2 observations in Class B and one observation in Class A. So, we classify the red star to Class B.
- If we choose K=6, then we have 2 observations in Class B but four observations in Class A. So, we classify the red star to Class A.

To obtain k nearest points, the first step is calculating the distance between the input data point and other points in our training data. Suppose x is a point with coordinates $(x_1, x_2, \ldots, x_3)$and y is a point with coordinates $(y_1, y_2, \ldots, y_3)$, then the distance between these two points is:

$$d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \quad (2)$$

As most statistical learning model does, if K is small, then, we use a smaller region of data to learn. This may cause over-fitting. This may cause under-fitting. And if K is large, then we use a larger region of data to learn.

3.3 Decision Tree (DT)

A decision tree algorithm is a type of machine learning technique that employs a decision tree structure to generate predictions. This method operates on a tree-like framework of decisions and their potential outcomes. The algorithm functions by recursively dividing the dataset into smaller subsets, selecting the most relevant feature at each node of the tree. In this particular study, the ID3 algorithm was utilized to construct the decision tree. The ID3 algorithm creates decision trees through a top-down greedy search strategy, exploring the possible branches without any backtracking. As indicated by its name, a greedy algorithm consistently opts for the choice that appears to be the most advantageous at any given moment [18]. It follows the rule that a branch with an entropy of zero is a leaf node and a branch with an entropy more than zero needs further splitting. The steps are as follow;

(i) It begins with the original set S as the root node.

(ii) On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates Entropy (H) and Information gain (IG) of this attribute.

(iii) It then selects the attribute which has the smallest Entropy or Largest Information gain.

(iv) The set S is then split by the selected attribute to produce a subset of the data.

(v) The algorithm continues to recur on each subset, considering only attributes never selected before.

3.4 Naïve Bayes (NB)

The Naive Bayes algorithm is a supervised learning technique that utilizes Bayes' theorem, operating under the "naive" premise that all features are conditionally independent of one another when the class variable is known. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x_1 through x_n in equation 3:

$$p(y|x_1,\dots,x_n) = \frac{p(y)p(x_1,\dots,x_n|y)}{p(x_1,\dots,x_n)} \quad (3)$$

Using the naive conditional independence assumption in equation 4, that;

$$p(x_1 \mid y, x_1,\dots,x_{i-1},x_{i+1},\dots,x_n) = p(x_1 \mid y) \quad (4)$$

for all i, this relationship is simplified to equation 5;

$$p(y|x_1,\dots,x_n) = \frac{p(y)\ \prod_{i=1}^{n} p(x_i|y)}{p(x_1,\dots,x_n)} \quad (5)$$

Since $p(y|x_1,\dots,x_n)$ is constant given the input, the study uses the following classification rule in equation 6 and 7:

$$p(y|x_1,\dots,x_n) \propto p(y)\ \prod_{i=1}^{n} p(x_i|y) \quad (6)$$

$$y = argmax\, p(y)_y \prod_{i=1}^{n} p(x_i|y) \quad (7)$$

and Maximum A Posteriori (MAP) estimation to estimate $p(y)$ and $p(x_i|y)$; the former is then the relative frequency of class y in the training set [21].

4. RESULTS AND DISCUSSION

The experimental findings are detailed in terms of various metrics, including False Positive Rate (FPR), accuracy (ACC), recall, positive predictive value (PPV), sensitivity (SPC), negative predictive value (NVP), F-1 score, and recognition time for malaria and jaundice diseases. These metrics were evaluated using Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), and Naïve Bayes (NB) algorithms. The SVM model yielded the following results for predicting malaria and jaundice: a false positive rate of 15%, 15.4%, and 14.2%; an accuracy of 86.4%, 87.5%, and 87.7%; a recall of 90%, 88.3%, and 89.8%; a positive predictive value of 86.3%, 85.5%, and 86.7%; a specificity of 85%, 84.6%, and 85.8%; a negative predictive value of 89.7%, 88%, and 89.5%; and an F1 score of 87.9%, 86.8%, and 88.1%. The recognition times recorded were 77.37 seconds for jaundice, 85.28 seconds for malaria, and 81.39 seconds for uninfected cases.

The KNN model utilized for predicting malaria and jaundice demonstrated a false positive rate of 21.5%, 20.3%, and 19.4%, with corresponding accuracy rates of 81.6%, 82%, and 83.1%. The recall values were recorded at 84.7%, 84.4%, and 85.6%, while the positive predictive values stood at 80.4%, 81%, and 81.9%. Specificity was noted at 78.5%, 79.7%, and 80.6%, and the negative predictive values were 83.9%, 83.7%, and 84.9%. The F1 scores were 82.2%, 82.5%, and 83.5%, with processing times of 64.76 seconds, 72.67 seconds, and 66.9 seconds for jaundice, malaria, and uninfected cases, respectively. In contrast, the DT model for predicting malaria and jaundice exhibited a false positive rate of 35.4%, 30.7%, and 30.6%. Its accuracy was measured at 67.7%, 71.6%, and 71.9%, with recall values of 70.8%, 74%, and 74%. The positive predictive values were 67%, 70.9%, and 71.2%, while specificity was recorded at 64.6%, 69.3%, and 69.4%. The negative predictive values were 68.9%, 72.7%, and 73.1%, and the F1 scores were 68.7%,

72.3%, and 72.6%, with processing times of 49 seconds, 56.4 seconds, and 52.98 seconds for jaundice, malaria, and uninfected cases, respectively. The results produced by the Naive Bayes (NB) model for the prediction of malaria and jaundice are as follows: a false positive rate of 29.9%, 26.6%, and 26.6%; an accuracy of 73.3%, 75.8%, and 75.8%; a recall of 76.4%, 78.1%, and 78.1%; a positive predictive value of 72.3%, 74.9%, and 74.9%; a specificity of 70.1%, 73.4%, and 73.4%; a negative predictive value of 74.9%, 77.1%, and 77.1%; and an F1 score of 74.1%, 76.4%, and 76.4%. These results were achieved in 46.2 seconds, 50.1 seconds, and 50.1 seconds for jaundice, malaria, and uninfected cases, respectively.

The results indicate that the SVM technique achieved reductions of 4.89%, 6.53%, and 5.22% in false negative rates (FNR) for the malaria, jaundice, and uninfected datasets, respectively, when compared to the KNN technique. Additionally, the SVM technique demonstrated improvements in accuracy of 4.43%, 5.90%, and 4.73% for the same datasets over the KNN technique. Furthermore, recall rates increased by 3.95%, 5.28%, and 4.22% for malaria, jaundice, and uninfected datasets, respectively, when utilizing the SVM technique in comparison to KNN. The positive predictive value (PPV) also saw enhancements of 4.51%, 5.94%, and 4.19% for the respective datasets with the SVM technique over KNN. Moreover, the SVM technique resulted in increases of 4.89%, 6.58%, and 5.22% in specificity (SPC) for malaria, jaundice, and uninfected datasets, respectively, compared to KNN. The negative predictive value (NPV) improved by 4.29%, 5.82%, and 5.22% for the same datasets when using the SVM technique over KNN, which is the next best in the evaluated metrics. In terms of timing, the NB technique recorded reductions of 6.36 minutes, 2.80 minutes, and 5.93 minutes for malaria, jaundice, and uninfected datasets, respectively, compared to the DT technique, which is the next in timing. A comparative analysis of the four developed models is illustrated in Figures 4, 5, 6, 7, 8, and 9

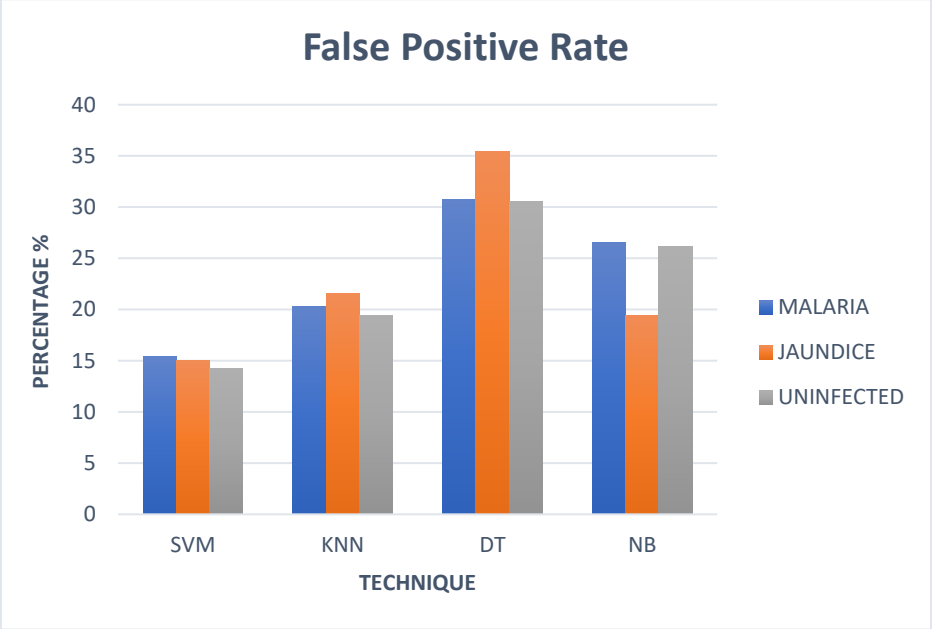

Figure 4: Prediction of Malaria and Jaundice Diseases based on False Positive Rate

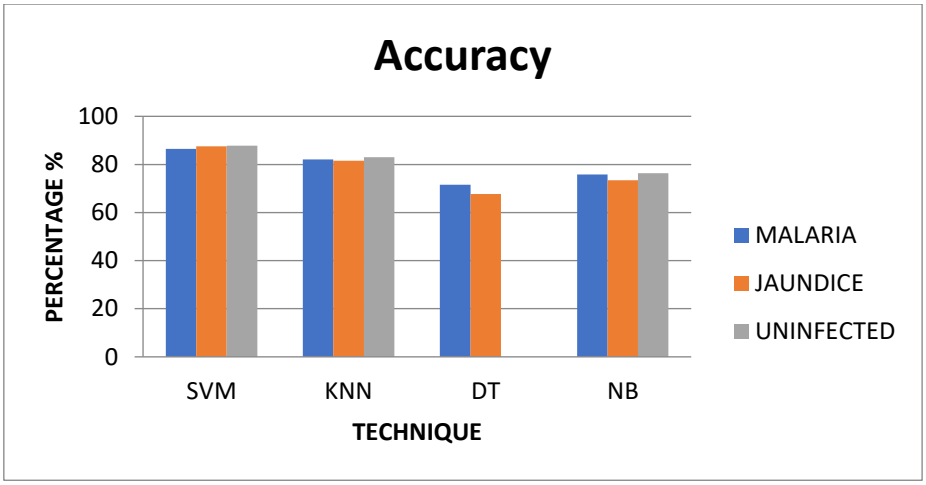

Figure 5: Prediction of Malaria and Jaundice Diseases based on Accuracy

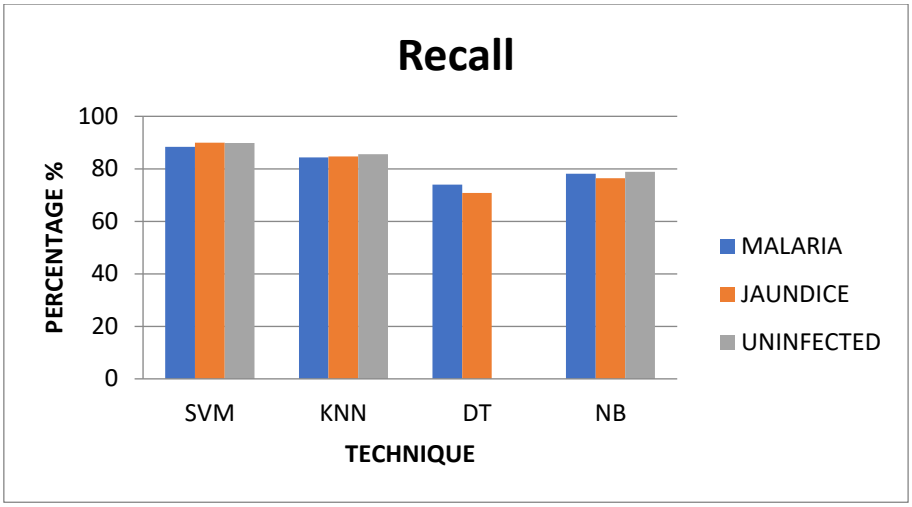

Figure 6: Prediction of Malaria and Jaundice Diseases based on Recall

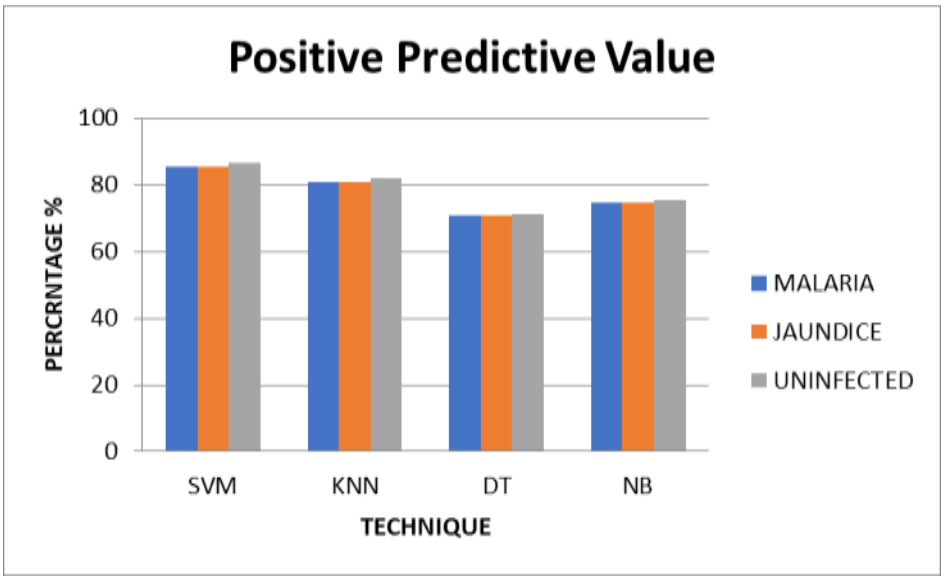

Figure 7: Prediction of Malaria and Jaundice Diseases based on Positive Predictive Value

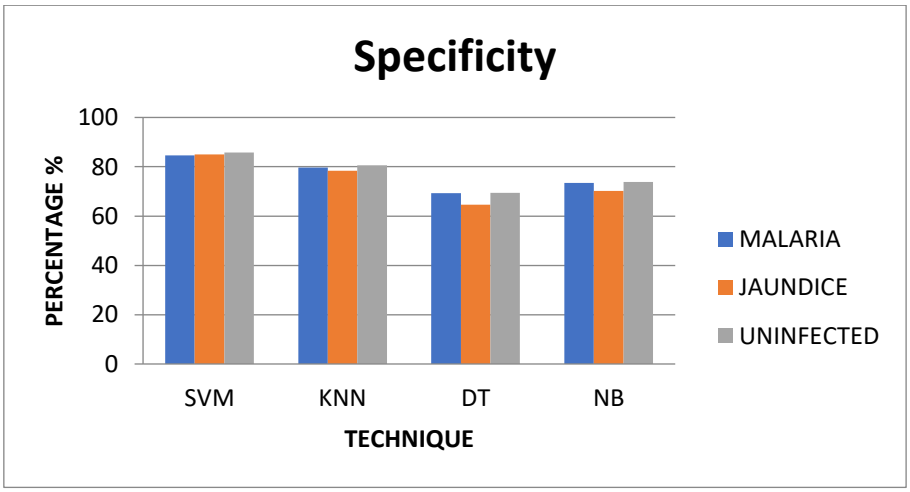

Figure 8: Prediction of Malaria and Jaundice Diseases based on Specificity

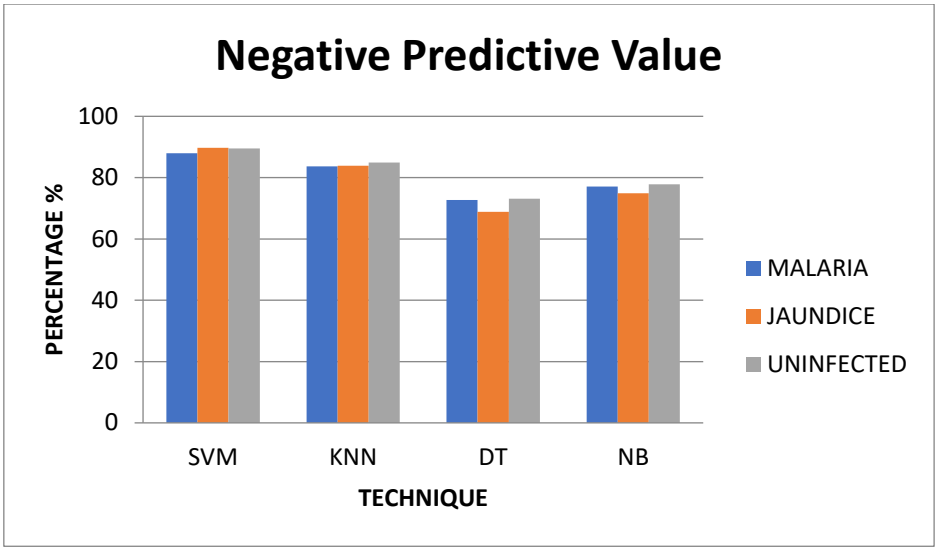

Figure 9: Prediction of Malaria and Jaundice Diseases based on Negative Predictive Value

5. CONCLUSION

This study assessed the performance of SVM, KNN, DT, and NB models in the detection and classification of diseases. A total of one thousand and forty images, categorized into malaria, jaundice, and uninfected datasets, were utilized for the evaluation of the methodologies. These images underwent training and testing with the models at various learning rate values. Throughout all evaluations, the SVM technique demonstrated superior recognition accuracy, reduced false positive rate, and enhanced negative prediction value, among other metrics. The findings indicate that, in terms of accuracy, false positive rate, sensitivity, and specificity, SVM surpassed the other methods examined in this research, while NB excelled in terms of processing time. This supports the conclusion that the SVM technique is more effective in predicting patients with malaria and jaundice compared to the other three algorithms analyzed.

REFERENCES

1. Edicha J. A., Hassan S. M. and Ocholi J. (2014). "An Analysis of Long Lasting insecticidal Treated Nets (LLIN) in Chanchaga LGA, Niger State, Nigeria". Confluence Journal of Environmental Studies, 9, 8-23. 2014.
2. Bell D. R., Jorgensen P., Christophel E. M., Palmer K. L. (2005). Malaria risk: estimation of the malaria burden. Nature. 2005;437:E3–E4.
3. Malaria Facts. CDC website. [Accessed October 10, 2008]. Available at: http://www.cdc.gov/malaria/facts.htm.
4. Chee Y. Y., Patrick H. Y. Chung, Rosanna M. S. Wong, Kenneth K. Y. Wong, (2018). Jaundice in infants and children: causes, diagnosis, and management. Hong Kong Medical Journal 2018. | Volume 24 Number 3 | June 2018 | www.hkmj.org.
5. Wan H. W. I. and Fadzilah S., (2006). Artificial Intelligence in Medical Application: an Exploration. School of Information Technology, University Utara Malaysia, 06010.
6. Mosabbir B. and Islam M. S. (2023). A New Ensemble Learning Approach to Detect Malaria from Microscopic Red Blood Cell Images, Sensors International, 4., https:doi.org/10.1016/j.sintl.2022.100209.
7. Dennis K. M., Victor W. L. and Mark O. (2024). A Machine Learning-Based Prediction of Malaria Occurrence in Kenya. American Journal of Theoretical and Applied Statistics 13(4) 65-72.
8. Dhrgam A. K., Noor N. T. Samara S. A. (2024). Malaria Diseases Prediction Based on Convolutional Neural Networks, Journal of Applied Engineering and Technological Science, 5(2), 1165-1181.
9. Ahmed Y. A., Saleem L. M., Ali A. and Javaan C. (2024). Real-Time Jaundice Detection in Neonatics Based on Machine Learning Models, BioMedInformatics, 4(1) 623-637.
10. Oguntimilehin A., Adetunmbi A. O., and Abiola O. B. (2015). "A Review of Predictive Models on Diagnosis and Treatment of Malaria Fever" IJCSMC, Vol. 4, Issue. 5, May 2015, pg.1087 – 1093
11. Samuel O. W., Omisore M. O., Ojokoh B. A., (2013). A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever. Expert Syst. Appl. 40 (10), 4164–4171. Elsevier Ltd. United Kingdom.
12. Fatumo S. A., Adetiba E. and Onaolapo J. O. (2013) "Implementation of xpertmaltyph: an expert system for medical diagnosis of the complications of malaria and typhoid". IOSR Journal of Computer Engineering (IOSRJCE), ISSN: 2278-8727Volume 8, Issue 5.2013.

13. Djam X. Y., Wajiga G. M., Kimbi Y. H. and Blamah N. V. (2011). A fuzzy expert system for the management of malaria. Int. J. Pure Appl. Sci. Technol. 5 (2), 84–108. Academic Publications Ltd., Bulgaria.

14. Aminu, E. F., Ogbonnia, E. O. and Shehu, I. S., (2016). A predictive symptoms-based system using support vector machines to enhanced classification accuracy of malaria and jaundice. Int. J. Mathematical Sci. Comput. (IJMSC) 2 (4), 54–66. Modern Education and Computer Science Press, Hong Kong.

15. Boruah I. and Kakoty S., (2019). Analytical study of data mining applications in malaria prediction and diagnosis. Int. J. Comput. Sci. Mobile Comput. 8 (3), 275–284.

16. Femi E. A., Joseph B. A., Roseline O. O., Sakinat O. F. and Adebola O. A. (2020). A decision support system for multi-target disease diagnosis: A bioinformatics approach. *Heliyon* (2020)

17. Ousman K. Jimoh O. A. and Pear M. H.(2024). Predicting Malaria Outbreak in the Gambia Using Machine Learning Techniques, doi:10.137/journal.pon.0299386, 19(5): 234-240

18. Madugu J. O., Olajide B. O., Okpor J. and Andrew I. W. (2023). Smart User Consumption Analysis Using Support Vector Machine and Multilayer Perceptron, Asian Research Journal of Current Science 5(1): 163-170.

19. Olajide B. O., James O., Chukwudi J. I., Temitope O. O., and Adeogun A. E. (2024). A Child Dietary Prescription Model using Artificial Neural Network and Fuzzy Logic, International Journal of research Publication and Reviews, 5(9): 750-758.

20. Okpor J., Olajide B. O., Tsokwa I. D. and Ehizuenlen E. P. (2024). An ECG-Based Model for Heart Arrythmia Classification Using Deep Learning, FUW Trends in Science and Technology Journal, 9(2): 218-224.

21. Ram K. S., Yas S., Yuvraj V., Himanshu K. and Shivdeepsm S. (2024). Diseases Prediction Using ML. International Journal of research Publication and Reviews 5(5):5368-5371.