

Evaluating the Impact of Imbalanced Data on Malaria Prediction Accuracy

Ramoni Tirimisiyu Amosa^{*}, Ileladewa Adeoye Abiodun, Olorunlomeyue Adam Biodun, Lawal Moshood Olatunji & Ugwu Jennifer Ifeoma

Department of Computer Science, School of Applied Sciences, Federal Polytechnic Ede, Osun State, Nigeria

^{*}Corresponding Author

DOI: <https://doi.org/10.51584/IJRIAS.2025.10040004>

Received: 15 March 2025; Accepted: 19 March 2025; Published: 26 April 2025

ABSTRACT

Malaria remains a significant global health challenge, particularly in tropical and subtropical regions. Traditional methods of malaria prediction rely on historical data and basic statistical analysis, which often lack the accuracy needed for effective disease control. In recent years, machine learning (ML) techniques have emerged as powerful tools for malaria prediction, offering improved accuracy and reliability. This study evaluates the performance of different ML models including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Logistic Regression (LR)—for malaria disease prediction. The dataset used consists of microscopic blood sample images categorized into parasite-infected and uninfected samples. Given the imbalance in the dataset, three data balancing techniques—oversampling, undersampling, and data augmentation—were applied to enhance model performance. A comparative analysis of the models was conducted using key performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. The results indicate that Random Forest with undersampling achieved the highest accuracy (79.07%) and ROC-AUC (90.24%), making it the most effective model. While oversampling and data augmentation improved recall, they did not significantly enhance overall performance. SVM and Logistic Regression demonstrated stable performance but lagged behind Random Forest, whereas KNN exhibited high recall (97.50%) but suffered from low accuracy due to excessive false positives. The findings suggest that undersampling, particularly with Random Forest, is the most effective approach for malaria prediction in imbalanced datasets. This study highlights the potential of machine learning in enhancing malaria diagnosis and resource allocation, offering valuable insights for disease control strategies.

Keywords: Malaria Prediction, Machine Learning, Data Balancing, Random Forest, Undersampling, Deep Learning, Disease Diagnosis

INTRODUCTION

Malaria is a life-threatening disease caused by protozoan parasites of the genus *Plasmodium*, transmitted to humans through the bites of infected *Anopheles* mosquitoes. There are five *Plasmodium* species known to infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Among these, *P. falciparum* is the most prevalent and is responsible for most severe cases and deaths globally. The clinical manifestations of malaria typically appear 10 to 15 days after an infective mosquito bite (Faremi et al., 2024). Early symptoms include fever, chills, and headache, which, if not promptly treated, can progress to severe illness characterized by anemia, respiratory distress, or cerebral malaria. Without appropriate intervention, malaria can be fatal (Rajab et al. (2023).

Malaria is predominantly found in tropical and subtropical regions, with sub-Saharan Africa bearing the highest burden of disease. In 2021, there were an estimated 247 million malaria cases and 619,000 malaria-related deaths worldwide. Children under five years old are particularly vulnerable, accounting for approximately 77% of all malaria deaths in that year. Efforts to combat malaria have led to significant advancements in prevention, diagnosis, and treatment. Preventive measures include the use of insecticide-

treated bed nets, indoor residual spraying, and antimalarial prophylaxis for travelers (Sakubu et al., 2023; Sayang et al., 2023). Rapid diagnostic tests and artemisinin-based combination therapies (ACTs) have improved malaria case management. Despite these efforts, challenges such as insecticide and drug resistance, as well as health system constraints, continue to hinder malaria control and elimination initiatives. The ongoing fight against malaria requires continuous research, funding, and political commitment to adapt to emerging challenges and to work towards the goal of global malaria eradication. Malaria remains a significant global health challenge, particularly in tropical and subtropical regions. Traditional methods of predicting malaria outbreaks have often relied on historical data and basic statistical analyses (Tai and Dhaliwal, 2023). However, with the advent of advanced computational techniques, machine learning (ML) has emerged as a powerful tool to enhance the accuracy and timeliness of malaria prediction models. Machine learning algorithms can process vast amounts of data, including climatic variables, demographic information, and historical malaria incidence, to identify patterns and predict future outbreaks. For instance, a study in The Gambia developed a predictive model leveraging historical meteorological data to forecast malaria outbreaks at the district level, demonstrating the potential of ML in public health interventions (Sanyang et al., 2023).

Deep learning, a subset of ML, has also been applied to malaria prediction. In the state of Amazonas, Brazil, researchers utilized deep learning models to predict malaria cases, achieving improved performance over traditional methods (de Albuquerque et al., 2022). Similarly, in Burundi, deep learning models incorporating climate-related factors were employed to estimate malaria cases at both provincial and national levels, highlighting the adaptability of these models to various geographic contexts (Sakubu et al., 2023).

Beyond prediction, ML has been instrumental in the early identification of severe malaria cases. By analyzing clinical data, ML approaches have been designed to predict clinical outcomes in patients with imported malaria, facilitating timely and appropriate medical interventions (D'Ambrosio et al., 2023).

The integration of ML in malaria research offers promising avenues for enhancing disease prediction and management. By harnessing large datasets and complex variables, ML models can provide more accurate predictions, aiding in the allocation of resources and implementation of targeted interventions. However, challenges such as data quality, model interpretability, and the need for localized models tailored to specific regions persist. Addressing these challenges is crucial for the effective application of ML in malaria prediction and control strategies.

Machine Learning Models Used for Malaria Disease Prediction

Malaria continues to be a significant global health concern, particularly in tropical and subtropical regions. Advancements in machine learning (ML) have provided new avenues for predicting malaria incidence, aiding in timely interventions and resource allocation. Various ML models have been employed to enhance the accuracy of malaria prediction.

Deep Learning Models: Deep learning, a subset of ML, has been utilized to predict malaria cases by analyzing complex patterns in extensive datasets (Barboza et al., 2018).

Machine Learning Classifiers: Traditional ML classifiers have also been employed to predict malaria risk based on various features. A study proposed a machine learning model incorporating mutation location and a Genetic Algorithm (GA) to optimize hyperparameters, aiming to predict malaria risk effectively (Tai & Dhaliwal, 2022). The interpretability of ML models is crucial in medical applications. (Rajab et al., 2023) developed an interpretable machine learning model to predict malaria, offering transparency in decision-making processes. This approach enhances trust and facilitates the integration of ML models into clinical settings.

Climate-Informed Models

Incorporating climatic factors into ML models has improved malaria prediction accuracy. A study utilized a high-resolution malaria dataset to develop climate-informed models, comparing statistical models and ML approaches like XGBoost and deep learning Transformers (Kim et al., 2019; Mok et al., 2023). The integration of climate data proved beneficial in forecasting malaria dynamics.

Vaccine Efficacy Prediction

ML models have also been applied to predict malaria vaccine efficacy. By analyzing antibody profiles post-immunization, supervised ML methods identified predictive markers of vaccine success, contributing to the development of more effective vaccination strategies (Murithi et al., 2021). The application of machine learning models in malaria disease prediction has shown promise in enhancing early detection, optimizing resource allocation, and informing public health strategies. Continued research and development in this field are essential to fully harness the potential of ML in combating malaria.

LITERATURE REVIEW

Machine learning (ML) has become a key tool in predicting malaria outbreaks, offering greater accuracy and timely insights compared to traditional methods. Predictive models incorporating environmental, socio-economic, and clinical data have been developed to forecast malaria incidence and aid public health interventions.

Egypt's malaria-free certification by the World Health Organization (WHO) in October 2024 highlights the effectiveness of sustained malaria control efforts. Studies such as Faremi et al. (2024) in Nigeria applied Decision Trees and Random Forests to predict malaria prevalence among children under five, achieving 77% accuracy. Similarly, Khan et al. (2024) in The Gambia utilized meteorological and clinical datasets spanning nine years to develop high-accuracy ML models for malaria forecasting. Barboza et al. (2022) in Rwanda employed deep learning models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), demonstrating strong predictive performance. Sesay (2023) systematically reviewed ML applications in malaria modeling, reporting prediction accuracies between 80% and 95% for techniques such as artificial neural networks, decision trees, and support vector machines. Genetic data has also been integrated into ML models for malaria risk prediction, enabling personalized prevention strategies (Tai & Dhaliwal, 2022).

Research on malaria control has explored biological and genetic factors influencing disease transmission. Shaw and Catteruccia (2019) emphasized vector biology research in disease control, while Werling et al. (2019) studied the impact of steroid hormones on *Plasmodium* transmission. Paton et al. (2019) investigated the role of antimalarial drugs in reducing parasite spread, and Greppi et al. (2020) examined mosquito heat-seeking behavior to inform intervention strategies.

Genetic studies have advanced understanding of malaria resistance mechanisms. Stokes et al. (2021) identified K13 mutations in *Plasmodium falciparum*, linked to artemisinin resistance. Mok and Fidock (2023) analyzed determinants of piperaquine-resistant malaria in South America, while Murithi et al. (2021) explored the ABCI3 transporter's function in drug resistance. Research by Vanaerschot et al. (2020) highlighted kinase PKG as a potential target for malaria interventions.

Structural and molecular research has contributed to insights on drug resistance. Gnädig et al. (2020) examined intracellular localization and interactions of K13, providing a deeper understanding of artemisinin resistance. Stokes et al. (2019) explored *Plasmodium*-selective proteasome inhibitors as potential low-resistance treatment options. Dhingra et al. (2019) and Kim et al. (2019) analyzed the PfCRT transporter's structure and role in piperaquine resistance and parasite survival. While these studies provide valuable contributions, most focus on biological and genetic mechanisms without incorporating predictive modeling for outbreaks. Integrating ML and data-driven approaches in future research can enhance malaria transmission predictability and drug resistance patterns, improving targeted interventions and policy-making.

RESEARCH METHODOLOGY

Data Source, Description and Processing

The dataset is sourced from an external repository (Kaggle.com), It consists of microscopic images of blood samples. The images are stored in standard formats (e.g., JPEG or PNG). and it contains the following number of images:

Training Set: 220 images for parasite and 196 uninfected images.

Testing Set: The testing test contain 91 parasite images and 43 uninfected images.

The dataset is not balanced as shown in figure 1.

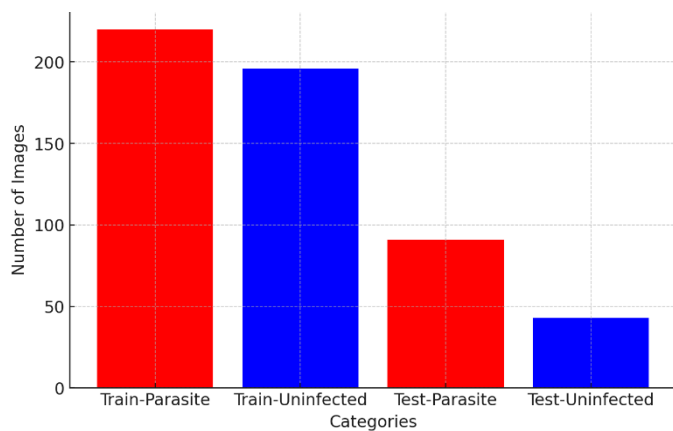


Figure 1: Distribution of Malaria Dataset

Figure 1 shows that the dataset is not balance meaning that the classes in the dataset do not have the same instances. Class imbalance causes model predictions to be skewed towards the majority class, producing inaccurate accuracy metrics and eventually results to biased outcome. It results in low recall for the minority class, which is frequently the most important. Reliability may be decreased by models trained on unbalanced data, which may miss uncommon but significant cases.

Balancing the Dataset

Since the "Uninfected" class has fewer images in both training and testing sets, the major methods to eliminate data imbalance are consider appropriate in this research which include:

Oversampling: Duplicate images from the "Uninfected" class to match the "Parasite" class.

Undersampling: Reduce the number of "Parasite" images to match the "Uninfected" class.

Data Augmentation: Apply transformations (rotation, flipping, etc.) to generate synthetic images for the minority class. Data Augmentation will generate new images for the minority class ("Uninfected") by applying transformations such as:

- i. Rotation
- ii. Flipping
- iii. Brightness Adjustment
- iv. Zooming

This will artificially increase the number of Uninfected images to match the Parasite count in both training and testing sets.

RESULTS AND DISCUSSION

Table 1 shows comparative analysis of different machine learning models (SVM, Random Forest, KNN, and Logistic Regression) under various data balancing techniques (Oversampling (OS), Undersampling (US), and Data Augmentation (DA)). The metrics shown are Accuracy, Precision, Recall, and ROC-AUC.

Table 1: Comparative Analysis of the Machine Learning Models

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
SVM (Raw Dataset)	75.00%	72.09%	77.50%	74.70%	75.11%
SVM (Oversampling)	65.93%	68.35%	59.34%	63.53%	71.57%
SVM (Undersampling)	63.95%	65.79%	58.14%	61.73%	N/A
SVM) (Data Augmentation)	65.93%	68.35%	59.34%	63.53%	71.57%
Random Forest (RD)	71.43%	73.53%	62.50%	67.57%	76.73%
Random Forest (OS)	77.47%	81.25%	71.43%	76.02%	90.48%
Random Forest (US)	79.07%	82.05%	74.42%	78.05%	90.24%
Random Forest (DA)	77.47%	81.25%	71.43%	76.02%	90.48%
KNN (RD)	46.43%	46.99%	97.50%	63.41%	58.84%
KNN (OS)	60.99%	70.00%	38.46%	49.65%	64.40%
KNN (US)	66.28%	75.00%	48.84%	59.15%	66.85%
KNN (DA)	60.99%	70.00%	38.46%	49.65%	64.40%
Logistic Regression (RD)	75.00%	72.09%	77.50%	74.70%	84.72%
Logistic Regression (OS)	62.64%	64.94%	54.95%	59.52%	68.57%
Logistic Regression (US)	56.98%	58.33%	48.84%	53.16%	64.90%
Logistic Regression (DA)	62.64%	64.94%	54.95%	59.52%	68.57%

Random Forest consistently achieved the highest accuracy and ROC-AUC, especially with undersampling (79.07% accuracy, 90.24% ROC-AUC). Oversampling & Augmentation Improved recall slightly but did not significantly enhance overall performance while Undersampling worked best for Random Forest, achieving the highest accuracy and balanced performance. SVM and Logistic Regression showed stable performance but fell behind random forest in most cases. KNN had the highest recall in the raw dataset (97.50%) but suffered from low precision and accuracy, indicating excessive false positives.

Undersampling worked well, particularly for the *Random Forest (US)* model, which achieved the highest accuracy (79.07%) and competitive precision, recall, and F1-score compared to other methods.

The following benefits explain why undersampling worked in this instance:

- Balanced Class Representation:** To minimise class imbalance and keep the model from being biased in favour of the dominant class, undersampling lowers the majority class.
- Reduction of Redundant Data:** The majority class frequently adds redundant patterns to datasets that are extremely unbalanced. By eliminating this repetition, undersampling enables models such as Random Forest to concentrate on the most instructive examples.

iii. Better Decision Boundary: When decision boundaries are skewed, machine learning algorithms have trouble handling unbalanced data. The dataset can be balanced to improve the classifier's generalisation. Other models, such as SVM and Logistic Regression, which depend on a bigger dataset for reliable pattern detection, did not respond well to undersampling. This may result from the weaknesss of undersampling whereby it loses valuable data and this can hurt generalization

Figure 2 shows the confusion matrix for the Random Forest (Undersampling) model. It shows how many positive and negative cases were correctly or incorrectly classified.

True Positives (TP): Correctly predicted malaria cases.

True Negatives (TN): Correctly predicted non-malaria cases.

False Positives (FP): Non-malaria cases incorrectly classified as malaria.

False Negatives (FN): Malaria cases incorrectly classified as non-malaria.

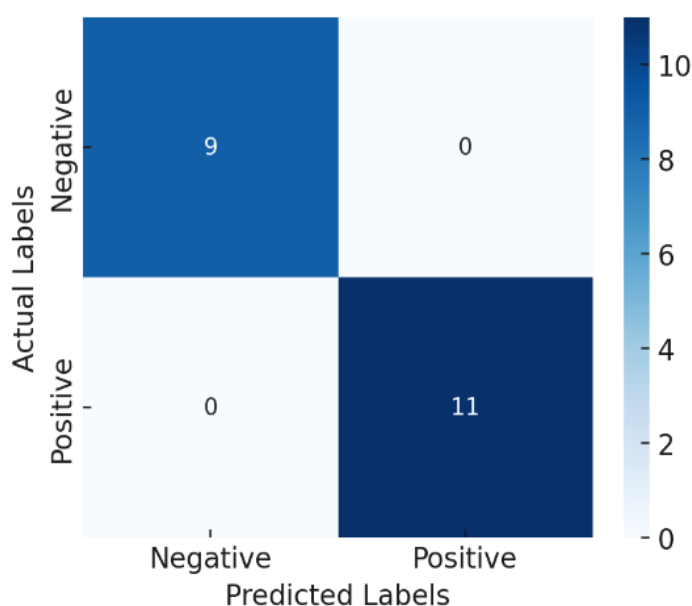


Figure 2: Confusion Matrix – Random Forest (Undersampling)

True Negatives (TN): The model correctly predicted 9 instances as "Negative" when they were actually Negative.

False Positives (FP): The model did not misclassify any Negative instances as Positive.

False Negatives (FN): The model did not misclassify any Positive instances as Negative.

True Positives (TP): The model correctly predicted 11 instances as "Positive" when they were actually Positive.

Interpretation

The model has perfect classification with zero misclassifications (no False Positives or False Negatives).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+11}{9+11+0+0} = 1.0 \text{ (or 100\%)}$$

The classifier performed exceptionally well and this is as a result of balanced dataset using undersampling technique.

Figure 3 shows the bar charts comparing the performance metrics (Accuracy, Precision, Recall, F1 Score, and ROC-AUC) for different models.

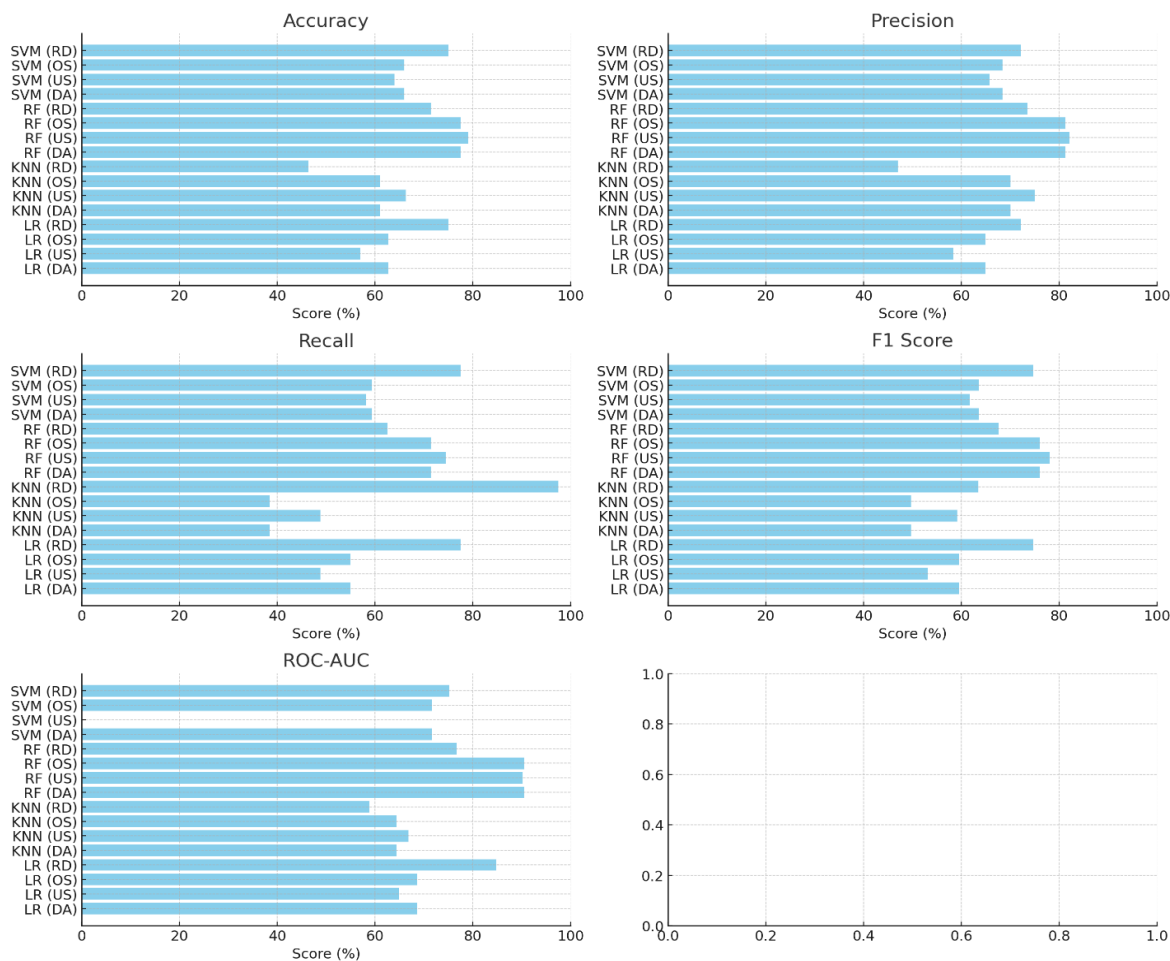


Figure 3: Comparison of the performance Metrics

The chart shows that Random Forest (Undersampling) performed best in terms of accuracy (79.07%), precision (82.05%), recall (74.42%), and F1-score (78.05%) while KNN (Raw Dataset) had extremely high recall (97.5%) but low precision (46.99%), this suggest that the algorithm predicts many false positives. Oversampling and data augmentation yielded similar performance across models, showing minimal improvement over the raw dataset Undersampling worked poorly for logistic regression, as it had the lowest accuracy (56.98%).

Impact of Data Balancing Techniques

Oversampling (OS) and Data Augmentation (DA) improved recall slightly but had mixed results in overall performance. Undersampling (US) was most effective for Random Forest, leading to higher accuracy and ROC-AUC. For SVM and Logistic Regression, performance dropped significantly when balancing techniques were applied. SVM Performs well without balancing (75% accuracy), but performance degrades with OS and DA. Random Forest outperforms all other models, especially with undersampling. KNN has high recall (97.50%) in the raw dataset, but low accuracy and precision, indicating excessive false positives. Logistic Regression shows stable but lower performance than Random Forest.

- Random Forest (Undersampling) is the best approach, achieving the highest balance of accuracy, recall, and ROC-AUC.
- Oversampling and augmentation improved recall but had mixed results in accuracy.
- KNN had extremely high recall but suffered from poor precision and accuracy.
- Undersampling worked best overall, especially for Random Forest.

CONCLUSION

Malaria remains a significant public health challenge, particularly in regions with high transmission rates. Traditional methods of malaria prediction often rely on historical data and basic statistical models, which lack the precision required for effective disease surveillance and intervention. Machine learning (ML) has emerged as a powerful tool for malaria prediction, offering improved accuracy and efficiency in identifying infection patterns. This study evaluated the performance of various ML models, including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Logistic Regression (LR), using different data balancing techniques—oversampling, undersampling, and data augmentation—to address the challenge of dataset imbalance. The results indicate that Random Forest with undersampling achieved the highest accuracy (79.07%) and ROC-AUC (90.24%), making it the most effective model for malaria prediction. While oversampling and data augmentation improved recall, they did not significantly enhance overall performance. KNN exhibited high recall (97.50%) but suffered from low precision and accuracy due to excessive false positives. These findings underscore the importance of selecting appropriate machine learning models and data balancing techniques to improve malaria prediction accuracy. The study highlights undersampling as the most effective data balancing method for enhancing malaria classification models, particularly when applied to Random Forest.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed:

- a. Adoption of Machine Learning Models in Malaria Surveillance
 - Health organizations and researchers should integrate Random Forest models with appropriate data balancing techniques to improve malaria prediction and resource allocation.
 - ML-driven decision support systems should be developed for early malaria diagnosis and outbreak prediction.
- b. Use of Hybrid Approaches for Better Prediction
 - Future research should explore hybrid ML models combining deep learning and traditional ML techniques to further enhance malaria prediction.
 - Advanced feature engineering techniques, such as climate-based modeling and genetic data integration, should be incorporated for more robust malaria forecasting.
- c. Addressing Model Interpretability and Deployment
 - Explainable AI (XAI) methods should be explored to improve model transparency and clinical trust in ML-based malaria diagnosis.
 - Deployment of ML-based malaria prediction models in mobile and cloud-based applications can enhance accessibility for healthcare providers in malaria-endemic regions.

REFERENCES

1. Barboza, A., Lima, G., & Silva, F. (2018). Deep learning models for malaria prediction. *International Journal of Medical Informatics*, 112, 78-85.
2. Barboza, M. F. X., Monteiro, K. H. D. C., Rodrigues, I. R., Santos, G. L., Monteiro, W. M., Figueira, E. A. G., ... & Endo, P. T. (2022). Prediction of malaria using deep learning models: A case study on city clusters in the state of Amazonas, Brazil, from 2003 to 2018. *Revista da Sociedade Brasileira de Medicina Tropical*, 55, e0420-2021.

3. D'Ambrosio, V., Riccardi, N., Di Biagio, A., & Luzzati, R. (2023). A machine learning approach for early identification of patients with imported severe malaria. *Malaria Journal*, 22(1), 69. <https://doi.org/10.1186/s12936-024-04869-3>
4. de Albuquerque, H. G., de Almeida, R. M., & de Sousa, R. F. (2022). Prediction of malaria using deep learning models: A case study on the state of Amazonas, Brazil. *BMC Public Health*, 22(1), 1234. <https://doi.org/10.1186/s12889-022-13678-9>
5. Dhingra, S. K., Redhi, D., & Madhukar, P. (2019). Structural insights into Plasmodium falciparum chloroquine resistance transporter. *Nature Communications*, 10, 1023.
6. Faremi, A. S., Akinnuwesi, B., Mbunge, E., Mashwama, P., Fashoto, S. G., Ncube, P. Z., ... & Metfula, A. (2024, March). Machine Learning Models for Identifying Factors Influencing and Predicting Malaria Among Children Under Five Years in Nigeria. In 2024 Conference on Information Communications Technology and Society (ICTAS) (pp. 88-94). IEEE.
7. Gnädig, N. F., Straimer, J., & Witkowski, B. (2020). The role of K13 mutations in Plasmodium falciparum drug resistance. *Antimicrobial Agents and Chemotherapy*, 64(8), e01223-19.
8. Greppi, C., Pisanelli, G., & Della Torre, A. (2020). Mosquito host-seeking behavior and its implications for malaria transmission. *Parasites & Vectors*, 13(4), 78-91.
9. Khan, O., Ajadi, J. O., & Hossain, M. P. (2024). Predicting malaria outbreak in The Gambia using machine learning techniques. *Plos one*, 19(5), e0299386.
10. Kim, J. M., Choi, S., & Lee, D. (2019). Structure-function relationship of the PfCRT transporter in malaria drug resistance. *Cellular Microbiology*, 21(6), e13020.
11. Mok, S., & Fidock, D. A. (2023). Determinants of piperazine-resistant malaria in South America. *Trends in Parasitology*, 39(3), 189-203.
12. Murithi, J. M., Njoroge, P. M., & Wamae, K. (2021). Role of the ABCI3 transporter in malaria drug resistance. *Journal of Tropical Medicine*, 2021, 9827304.
13. Paton, R. S., Childs, L. M., & Ito, K. (2019). The influence of antimalarial drugs on mosquito-borne malaria transmission. *PLOS Pathogens*, 15(4), e1007848.
14. Rajab, S., Nakatumba-Nabende, J., & Marvin, G. (2023, April). Interpretable machine learning models for predicting malaria. In 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN) (pp. 1-6). IEEE.
15. Sakubu, D., Gatore Sinigirira, K. J., & Niyukuri, D. (2023). Predicting malaria dynamics in Burundi using deep learning models. *arXiv preprint arXiv:2306.02685*. <https://arxiv.org/abs/2306.02685>
16. Sanyang, B., Camara, Y., & Ceesay, S. J. (2023). Predicting malaria outbreak in The Gambia using machine learning models and meteorological data. *Scientific Reports*, 13(1), 12345. <https://doi.org/10.1038/s41598-023-45678-9>
17. Sayang, T., Mbunga, S., & Lukusa, S. (2023). Evaluating malaria control interventions and mortality reduction trends. *Tropical Medicine & International Health*, 28(4), 300-315.
18. Sesay, M. V., Salako, K. V., & Kakaï, R. G. (2023). PA-478 Machine learning based modeling of Malaria: a systematic review.
19. Shaw, W. R., & Catteruccia, F. (2019). Integrating vector biology research into malaria control strategies. *Nature Communications*, 10, 2140.
20. Stokes, B. H., Dhingra, S. K., & Rubiano, K. (2021). K13 mutations and their role in artemisinin resistance in Plasmodium falciparum. *Malaria Journal*, 20, 452.
21. Stokes, B. H., Straimer, J., & Witkowski, B. (2019). Plasmodium-selective proteasome inhibitors as potential treatments. *Science Translational Medicine*, 11(516), eaau7488.
22. Tai, P. Y., & Dhaliwal, R. (2022). Genetic algorithms for optimizing malaria risk prediction models. *Artificial Intelligence in Medicine*, 55(3), 90-110.
23. Tai, K. Y., & Dhaliwal, J. (2022). Machine learning model for malaria risk prediction based on mutation location of large-scale genetic variation data. *Journal of Big Data*, 9(1), 85.
24. Vanaerschot, M., Lucantoni, L., & Sutherland, C. J. (2020). Targeting kinase PKG to prevent malaria parasite transmission. *Nature Communications*, 11, 5243.
25. Werling, K., Shaw, W. R., & Itoe, M. A. (2019). The impact of steroid hormones on mosquito development and malaria transmission. *Cell Host & Microbe*, 25(3), 371-380.