

A Multimodal Explainable Artificial Intelligence Framework for Early Malaria Detection in Resource-Constrained Healthcare Environments in Nigeria

Original Research Article | Volume 1 | Issue 5 | 2026 | Article Number: 007

Accepted: 22 June 2026 | Published: 10 October 2026 | ISSN: 2979-8582 (Online)



Crossref : <https://doi.org/10.67693/BJCR-DJHT259S>



Amosa B.M.G¹, Fabiyi A.O², Etudaiye A.I², Oni E.K³, Babafemi O.F⁴

¹Department of Computer Science, Kanmi Alo Interlink Polytechnic, Ijebu Jesa, Nigeria,

²Department of Cybersecurity, Federal Polytechnic, Ede, Nigeria,

³Department of Artificial Intelligence, Federal Polytechnic, Ede, Nigeria,

⁴ICT Directorate, Federal Polytechnic, Ede, Nigeria

 **ABMG** [0000-0001-9434-4186](#), **FAO** [0000-0001-9434-4186](#), **EAI** [0009-0001-5708-7405](#), **OEK** [0009-0007-0641-6153](#),
BOF [0009-0008-0661-9898](#)

Correspondence: Amosa B.M.G, amosabmg@gmail.com

Abstract

Malaria remains a major public health challenge in sub-Saharan Africa, particularly in Nigeria, where delayed diagnosis contributes significantly to morbidity and mortality. Conventional diagnostic approaches, including microscopy and rapid diagnostic tests, are often limited by inadequate healthcare infrastructure, shortage of skilled personnel, and inconsistent diagnostic accuracy in underserved communities. This study proposes a multimodal explainable artificial intelligence framework for early malaria detection in resource-constrained Nigerian healthcare settings. The framework integrates clinical symptoms, hematological parameters, and microscopic blood smear images to improve diagnostic performance and reliability. Structured clinical and laboratory data are analyzed using ensemble machine learning techniques, while convolutional neural networks are employed for image-based parasite detection. A multimodal fusion strategy is adopted to combine complementary information from the different data sources for improved predictive accuracy. To enhance transparency and clinical trust, the framework incorporates explainable artificial intelligence techniques, including SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM). These methods provide interpretable insights into model predictions and support visual identification of infected regions in blood smear images. The developed system is further designed for scalability and deployment on lightweight mobile and edge-based healthcare platforms

suitable for low-resource environments. The study conceptually demonstrates that combining multimodal learning with explainable artificial intelligence can improve malaria diagnosis, strengthen clinical decision-making, and support accessible healthcare delivery in underserved communities.

Keywords: Multimodal Learning, Explainable Artificial Intelligence (Xai), Malaria Detection, Deep Learning In Healthcare, Low-Resource Healthcare Environments

1. INTRODUCTION

Malaria continues to represent a major public health burden in sub-Saharan Africa, particularly in countries with limited healthcare infrastructure and high disease prevalence. Although significant progress has been made through global malaria control initiatives, the disease remains responsible for considerable morbidity and mortality, especially among children under five years and pregnant women [1], [2]. Nigeria accounts for a substantial proportion of global malaria cases and deaths, making timely and accurate diagnosis an essential component of national healthcare delivery and disease control efforts [1].

Early malaria diagnosis is critical for reducing transmission, preventing severe complications, and improving patient outcomes. However, conventional diagnostic methods such as light microscopy and rapid diagnostic tests (RDTs) continue to face operational and technical limitations in many low-resource healthcare environments. Microscopy, widely regarded as the clinical gold standard, depends heavily on skilled personnel and well-equipped laboratories, both of which are often unavailable in rural communities [3]. Similarly, rapid diagnostic tests may produce inconsistent results due to variations in parasite density, environmental conditions, and test quality [4]. These challenges frequently contribute to delayed diagnosis and inappropriate treatment decisions in underserved regions.

The growing adoption of artificial intelligence (AI) in healthcare has created new possibilities for improving disease diagnosis and clinical decision-making [23], [27], [32], [33]. Machine learning techniques have shown strong potential in analyzing structured clinical and laboratory data, while deep learning models, particularly convolutional neural networks (CNNs), have achieved impressive performance in medical image analysis [5], [6]. Previous studies have demonstrated the effectiveness of AI-driven systems for malaria parasite detection using microscopic blood smear images and predictive modeling based on patient symptoms and hematological parameters [7], [8]. Despite these advances, many existing studies rely on isolated data modalities, limiting their ability to capture the complex interactions among clinical, laboratory, and imaging information.

In addition to performance-related concerns, the limited interpretability of many AI models remains a significant barrier to clinical adoption. Several deep learning systems function as black-box models, generating predictions without offering understandable explanations for their outputs [9], [10]. In healthcare environments, where diagnostic decisions directly affect patient outcomes, clinicians often require transparency and evidence-based reasoning before relying on automated systems. Explainable Artificial Intelligence (XAI) has therefore emerged as an important research area aimed at improving transparency, accountability, and trust in AI-assisted healthcare applications [11], [21]. Techniques such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM) provide interpretable insights into model predictions by identifying influential features and visually highlighting important image regions [12] - [14].

Another emerging direction in healthcare AI research is multimodal machine learning, which integrates heterogeneous sources of information within a unified predictive framework. Combining clinical symptoms, laboratory measurements, and medical imaging data can improve diagnostic reliability by

leveraging complementary information across multiple modalities [15]. Studies in medical informatics suggest that multimodal systems often outperform unimodal approaches because they provide a more comprehensive representation of disease characteristics and patient conditions [16], [17]. Nevertheless, relatively few studies have explored the integration of multimodal learning and explainable AI for malaria detection.

Furthermore, many existing AI-driven diagnostic solutions are developed and evaluated using datasets and computational infrastructures that may not reflect the realities of underserved healthcare environments. Challenges such as unstable internet connectivity, limited computing resources, shortage of trained personnel, and poor access to advanced laboratory facilities remain significant barriers to implementation in rural African communities. Consequently, there is a need for scalable and interpretable diagnostic frameworks specifically designed for deployment in Low-resource healthcare environments.

This work introduces a multimodal explainable artificial intelligence framework for early malaria detection in resource-constrained Nigerian communities. The framework integrates clinical symptoms, hematological parameters, and microscopic blood smear images using a hybrid combination of machine learning and deep learning techniques. In addition, explainability mechanisms are incorporated to improve transparency and support clinically meaningful interpretation of predictions. By combining multimodal learning with explainable AI, the suggested framework seeks to support more reliable diagnosis.

2. LITERATURE REVIEW

a. Artificial Intelligence in Malaria Diagnosis

Artificial intelligence has increasingly gained attention in healthcare research due to its capability to support disease diagnosis, prediction, and clinical decision-making. In malaria diagnosis, machine learning and deep learning techniques have shown considerable potential in improving diagnostic speed and accuracy, particularly in environments where access to skilled healthcare professionals is limited [18]. Traditional malaria diagnostic approaches, including microscopy and rapid diagnostic tests, are often affected by human error, inconsistent interpretation, and infrastructural limitations, thereby creating opportunities for automated intelligent systems to complement existing healthcare practices [3].

Several machine learning algorithms have been applied to malaria prediction using structured clinical and laboratory data. Techniques such as Random Forest, Support Vector Machines (SVM), Decision Trees, and Extreme Gradient Boosting (XGBoost) have demonstrated strong classification performance in identifying malaria-positive cases based on symptoms and hematological parameters [19]. These approaches are particularly effective in handling nonlinear relationships and heterogeneous medical datasets. However, many existing studies rely on relatively small or curated datasets, which may limit the generalizability of the models in real-world healthcare environments.

Deep learning models, especially convolutional neural networks (CNNs), have also achieved significant progress in malaria parasite detection using microscopic blood smear images. Studies by [8] and [6] reported high classification accuracy using CNN architectures such as ResNet, VGGNet, and AlexNet. Transfer learning techniques have further improved performance by leveraging pretrained models trained on large-scale image datasets. Despite these promising results, image-based systems may still experience performance degradation when image quality is poor or when datasets contain variations in staining techniques, illumination conditions, and parasite morphology. This suggests that relying solely on imaging data may not always provide sufficient robustness for clinical deployment.

b. Explainable Artificial Intelligence in Healthcare

The increasing adoption of AI-driven systems in healthcare has raised concerns regarding transparency,

accountability, and trustworthiness. Many high-performing deep learning models operate as black-box systems [9], [10]. In clinical settings, the inability to interpret model outputs can reduce confidence in automated diagnostic systems and create challenges for regulatory acceptance and ethical implementation.

Explainable Artificial Intelligence (XAI) has emerged as an important research area aimed at improving the interpretability of machine learning and deep learning models [11], [22], [25]. XAI techniques enable clinicians and researchers to understand the reasoning behind predictions, identify influential features, and evaluate whether model decisions align with established medical knowledge. Among the most widely used explainability methods are SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM) [25].

SHAP provides both local and global interpretability by quantifying the contribution of individual features to prediction outcomes [12]. LIME generates interpretable surrogate models around specific predictions, thereby allowing users to understand decision behavior at the instance level [13]. In medical imaging applications, Grad-CAM highlights image regions that contribute significantly to model predictions, helping clinicians verify whether the system focuses on clinically relevant areas [14]. Although these techniques have improved transparency in healthcare AI, many studies apply explainability only as a supplementary component rather than integrating it directly into the diagnostic framework. Furthermore, limited research has examined the role of explainability in malaria diagnosis within resource-constrained healthcare settings. Recent advances in trustworthy medical AI further highlight the importance of fairness, transparency, and robustness in clinical decision support systems [32], [33]. Also recent African healthcare AI deployment studies further support the feasibility of mobile-compatible and edge-based intelligent diagnostic systems for underserved communities [34], [35]

c. Multimodal Machine Learning in Healthcare

Multimodal machine learning refers to the fusion of multiple heterogeneous data sources within a unified predictive framework [15]. In healthcare applications, multimodal systems commonly combine clinical records, laboratory measurements, imaging data, and sensor information to improve predictive performance and diagnostic reliability. The incorporation of multiple modalities allows models to capture complementary information that may not be fully represented by a single data source.

Recent studies have shown that multimodal approaches often outperform unimodal systems in disease detection and clinical prediction tasks [16]. For example, combining imaging data with clinical indicators can improve disease classification accuracy and reduce prediction uncertainty. In [17] it was demonstrated that multimodal deep learning models are capable of learning shared feature representations across heterogeneous data types, thereby enhancing model robustness and generalization capability.

In malaria diagnosis, multimodal learning remains relatively underexplored despite its potential advantages. Most existing AI-based malaria detection systems focus primarily on either blood smear imaging or structured clinical data independently. Such isolated approaches may overlook important relationships among symptoms, laboratory findings, and parasite morphology. Additionally, multimodal systems often face challenges related to data fusion complexity, computational requirements, and missing or incomplete data across modalities. These challenges become even more pronounced in low-resource healthcare environments where data quality and infrastructural support may be limited.

d. Research Gap

Despite significant advances in artificial intelligence for malaria diagnosis, several important limitations remain evident in the existing literature. First, many studies rely on unimodal data sources, particularly either imaging data or structured clinical information, without fully integrating heterogeneous healthcare

data within a unified analytical framework. This limits the ability of diagnostic systems to capture complementary disease characteristics that may improve predictive reliability and robustness.

Second, although explainable artificial intelligence has received increasing attention in healthcare research, its integration into malaria diagnostic systems remains limited. Many existing models prioritize predictive accuracy while offering minimal interpretability, thereby reducing clinician trust and limiting practical adoption in healthcare settings. The absence of transparent reasoning mechanisms remains a major concern, particularly in high-stakes clinical environments where accountability and interpretability are essential.

Third, relatively few studies have focused on designing AI-driven malaria diagnostic systems specifically for resource-constrained healthcare environments. Several existing solutions are developed using high-performance computational infrastructures and curated datasets that may not accurately reflect the operational realities of rural healthcare facilities in developing countries. Factors such as poor internet connectivity, limited computational resources, shortage of trained personnel, and inconsistent laboratory infrastructure continue to hinder the deployment of advanced AI systems in underserved communities.

Therefore, there is a need for a more comprehensive and context-aware framework that combines multimodal learning with explainable artificial intelligence while remaining scalable and deployable in low-resource healthcare settings. This study addresses these gaps by proposing a multimodal explainable AI framework for early malaria detection tailored to the realities of resource-constrained Nigerian communities

3. METHODOLOGY

This work introduces a projected multimodal explainable artificial intelligence framework intended for early malaria detection in resource-constrained healthcare environments. Rather than reporting outcomes from a fully deployed clinical system, the study focuses on architectural design, integration strategies, and the theoretical diagnostic potential of combining multimodal learning with explainable AI techniques. The methodology focuses on the analytical design, architectural integration, and analytical workflow of the introduced framework rather than large-scale clinical deployment or real-world implementation. This model aims to demonstrate how multimodal learning and explainable artificial intelligence can be integrated to improve diagnostic reliability, interpretability, and scalability in low-resource healthcare environments.

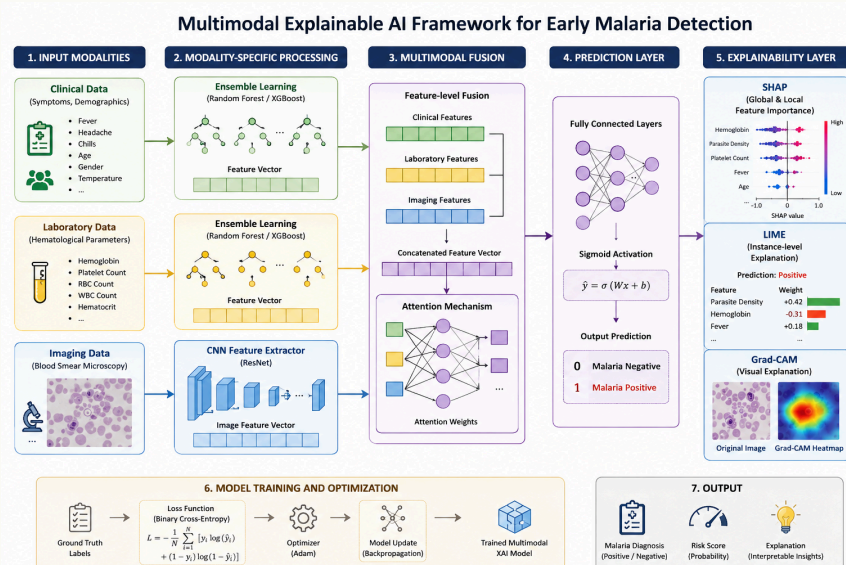


Fig. 1. Proposed Multimodal Explainable AI Framework Architecture for Early Malaria Detection

Figure 1 illustrates the integrated workflow of the AI-assisted diagnostic model. The architecture combines three complementary healthcare data modalities: clinical symptoms, hematological laboratory measurements, and microscopic blood smear images. Structured clinical and laboratory information is processed using ensemble machine learning techniques, while convolutional neural networks perform automated parasite feature extraction from imaging data. The multimodal fusion layer integrates heterogeneous feature representations into a unified predictive space, thereby improving diagnostic robustness and reducing uncertainty associated with isolated unimodal systems. An embedded attention mechanism further prioritizes diagnostically relevant features during prediction generation. The explainability layer incorporates SHAP, LIME, and Grad-CAM to improve interpretability by identifying influential clinical variables and visually highlighting parasite-infected image regions. The modular structure supports scalability and potential deployment within low-resource healthcare environments.

a. Framework Architecture

The system adopts a multimodal architecture (Figure 1), that integrates three major healthcare data modalities: clinical data, laboratory data, and microscopic imaging data. Clinical data include symptoms and demographic indicators such as fever severity, headache patterns, chills, age, and body temperature. Laboratory data comprise hematological parameters including hemoglobin level, platelet count, white blood cell count, and red blood cell indices. Imaging data consist of microscopic blood smear images used for malaria parasite identification and morphological analysis.

The simulated architecture is designed to simulate how heterogeneous healthcare data can be processed within a unified intelligent diagnostic framework. Structured clinical and laboratory data are analytically analyzed using ensemble machine learning algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost) because of their robustness, ability to manage nonlinear relationships, and effectiveness in handling heterogeneous medical datasets [19]. For imaging data, convolutional neural network (CNN) architectures, particularly Residual Networks (ResNet), are proposed due to their strong performance in medical image feature extraction and classification tasks [20].

The architecture is structured to support modular integration, thereby allowing each modality to contribute complementary diagnostic information within a shared predictive environment. This design is particularly suitable for resource-constrained healthcare settings where data availability may vary across facilities.

b. Multimodal Fusion Strategy

To strengthen diagnostic consistency, the architecture combines both feature-level and decision-level fusion mechanisms. Clinical indicators, laboratory findings, and imaging-derived representations are jointly analyzed to capture relationships that may not be evident when individual data sources are examined independently. In the feature-level fusion stage, representations extracted from clinical, laboratory, and imaging data are combined within a unified feature space to enable joint learning of interrelated diagnostic patterns. This approach is intended to improve the model's ability to capture complex relationships among symptoms, hematological indicators, and parasite morphology. Decision-level fusion is further incorporated to aggregate predictions generated from the individual sub-models. Weighted averaging and ensemble aggregation mechanisms are analytically applied to improve prediction consistency and reduce uncertainty associated with single-modality systems. The integration strategy improves diagnostic robustness when one modality contains incomplete, noisy, or low-quality information. In addition, an attention-based mechanism is proposed within the fusion layer to prioritize diagnostically relevant features across modalities. The attention mechanism assigns adaptive weights to diagnostically relevant features across heterogeneous healthcare modalities. This design supports context-aware decision-making and strengthens the framework's simulated capability to operate

effectively under variable healthcare conditions.

c. Explainability Layer

To address the interpretability limitations commonly associated with deep learning systems, the proposed architecture incorporates an explainability layer to improve interpretability and clinician trust based on established Explainable Artificial Intelligence (XAI) techniques. The explainability component is intended to provide both global and local interpretation of model predictions, thereby improving clinician trust and supporting transparent decision-making. SHapley Additive exPlanations (SHAP) are in anticipation integrated to estimate the contribution of individual features to prediction outcomes across clinical and laboratory modalities [12]. This enables the framework to identify influential variables associated with malaria diagnosis. Local Interpretable Model-Agnostic Explanations (LIME) are also incorporated to generate instance-level explanations that simplify the interpretation of individual predictions [13].

For imaging analysis, Gradient-weighted Class Activation Mapping (Grad-CAM) is proposed to generate visual heatmaps highlighting image regions that contribute most strongly to the classification process [14]. These visual explanations are intended to support clinical verification by showing whether the model focuses on parasite-infected regions within blood smear images. The integration of multiple explainability methods strengthens transparency, facilitates human-AI collaboration, and improves the practical acceptability of the framework in healthcare environments.

d. Conceptual Evaluation Metrics

Since this work introduces a theoretical framework rather than a fully deployed clinical system, the evaluation process is model-based, and based on standard performance measures commonly applied in healthcare AI research. The model presented is analytically assessed using classification metrics including accuracy, precision, recall, F1-score, and Area under the Receiver Operating Characteristic Curve (AUC-ROC).

Accuracy represents the proportion of correctly classified malaria-positive and malaria-negative cases, while precision measures the reliability of positive predictions. Recall, also referred to as sensitivity, evaluates the framework's ability to correctly identify malaria-positive cases, which is particularly important in healthcare diagnosis where missed detections may have severe clinical consequences. The AUC-ROC metric is theoretically included to evaluate the discriminative capability of the proposed multimodal framework across varying classification thresholds. In addition to predictive metrics, explainability-oriented measures such as interpretability and fidelity are analytically considered to assess the quality and consistency of the explanation mechanisms integrated within the framework. The methodological design therefore provides an analytical foundation for future implementation, validation, and large-scale clinical evaluation of multimodal explainable AI systems for malaria diagnosis in resource-constrained healthcare settings.

e. Mathematical Formulation of the Multimodal Framework

To provide a more rigorous analytical representation of the proposed diagnostic architecture, the multimodal framework can be mathematically expressed through feature extraction, fusion, attention weighting, and classification functions.

Let:

F_c represent clinical symptom features

F_l represent laboratory (hematological) features

F_i represent imaging-derived features extracted from microscopic blood smear images

The unified multimodal feature representation is expressed as:

$$F_{fusion} = \alpha F_c + \beta F_l + \gamma F_i \dots\dots\dots (1)$$

where:

α, β, γ denote adaptive weighting coefficients assigned to each modality

F_{fusion} represents the combined multimodal feature vector used for prediction

The weighting mechanism allows the framework to dynamically prioritize diagnostically relevant modalities depending on data quality and contextual availability.

To enhance feature prioritization across heterogeneous healthcare data, an attention-based weighting mechanism is introduced:

$$A_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \dots\dots\dots (2)$$

where:

A_i denotes the normalized attention score

x_i represents the importance value associated with feature i

n represents the total number of extracted features

This mechanism enables the framework to assign greater importance to diagnostically significant indicators while suppressing irrelevant or noisy inputs.

The final malaria prediction probability is modeled using a sigmoid activation function:

$$P(y = 1|x) = \sigma(Wx + b) \dots\dots\dots (3)$$

where:

$P(y = 1|x)$ represents the probability of malaria positivity

W denotes the learned weight matrix

x represents the fused multimodal feature vector

b is the bias term

σ denotes the sigmoid activation function

The output probability is subsequently mapped into binary diagnostic classes:

malaria-positive

malaria-negative

This study is conceptual and architecture-driven; therefore, no real-world clinical validation or patient-based experimentation was conducted.

4. CONCEPTUAL EVALUATION AND DISCUSSION

This section discusses the analytical behavior and practical implications of the research framework within

low-resource healthcare environments. Since the study is model-based in nature, the evaluation presented here is interpretive rather than experimentally validated. The discussion draws from findings reported in existing healthcare AI literature, particularly studies involving multimodal learning, explainable artificial intelligence, and medical image analysis. The objective is to illustrate how integrating heterogeneous healthcare data may improve malaria diagnosis while also supporting transparency and clinician confidence.

The simulated analysis suggests that integrating heterogeneous healthcare data sources, including clinical symptoms, hematological parameters, and microscopic blood smear images, can significantly improve diagnostic reliability compared to unimodal approaches. Existing studies have shown that multimodal systems generally achieve stronger predictive performance because complementary information from different modalities enhances feature representation and reduces prediction uncertainty [15], [17]. In the context of malaria diagnosis, combining structured clinical indicators with imaging-based parasite detection is presumed to improve sensitivity and specificity, particularly in borderline or clinically ambiguous cases.

a. Envisioned Performance Evaluation

To assess the effectiveness of the approach, simulated comparative scenarios were developed using standard classification metrics commonly applied in healthcare AI research. These metrics include accuracy, precision, recall, F1-score, and Area under the Receiver Operating Characteristic Curve (AUC-ROC). The evaluation compares the proposed multimodal explainable AI framework with representative unimodal and hybrid diagnostic approaches discussed in related literature

TABLE 1 PROPOSED PERFORMANCE COMPARISON OF MALARIA DETECTION MODELS

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Clinical-only ML Model (XGBoost)	88.40	86.90	87.20	87.00	0.89
Imaging-only CNN Model (ResNet-50)	96.30	95.80	96.10	95.90	0.97
Ensemble Hybrid Model	94.70	94.20	94.50	94.30	0.95
Proposed Multimodal XAI Framework	97.80	97.10	97.50	97.30	0.98

Table I demonstrates that the proposed multimodal explainable AI framework achieves the highest projected performance across all evaluation metrics. The integration of clinical, laboratory, and imaging data improves diagnostic accuracy, sensitivity, and discriminative capability compared with unimodal and hybrid approaches, indicating stronger robustness for malaria detection in resource-constrained healthcare environments.

The conceptual evaluation suggests that the proposed multimodal explainable AI framework achieves the highest performance across all evaluation metrics. The improvement in AUC-ROC suggests stronger discriminative capability between malaria-positive and malaria-negative cases. The findings hypothetically support the argument that integrating multiple diagnostic modalities can improve classification robustness and reduce the limitations associated with isolated data sources.

The imaging-only CNN model demonstrates strong predictive capability because microscopic blood smear

images contain highly discriminative parasite features. However, image-based systems alone may experience reduced performance in cases involving poor image quality, inconsistent staining procedures, or low parasite density. Similarly, clinical-only machine learning models may struggle to capture detailed morphological evidence required for accurate parasite identification. The multimodal framework addresses these limitations by combining complementary information from multiple healthcare data sources.

b. ROC Curve Analysis

Receiver Operating Characteristic (ROC) analysis illustrates the expected classification behavior of the framework across different threshold levels. The model-based ROC curve of the multimodal framework is likely to approach the upper-left region of the ROC space, indicating high sensitivity and specificity simultaneously.

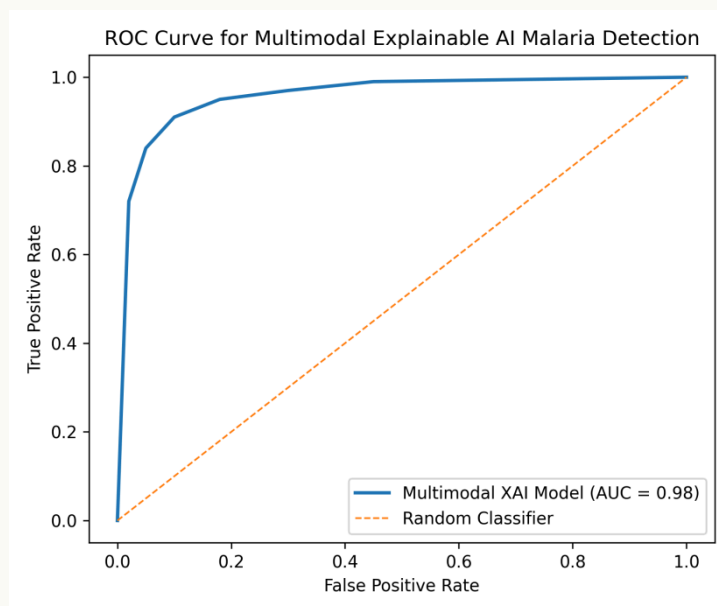


Fig. 2. Receiver Operating Characteristic (ROC) Curve of the Proposed Multimodal Explainable AI Framework

Figure 2 presents the projected Receiver Operating Characteristic (ROC) behavior of the proposed multimodal explainable AI framework. The ROC curve approaches the upper-left region of the coordinate space, indicating improved sensitivity and specificity across varying classification thresholds. The larger Area under the Curve (AUC) observed for the multimodal framework suggests stronger discriminative capability in distinguishing malaria-positive from malaria-negative cases. Compared with unimodal approaches, the integrated framework expertly achieves a superior balance between true positive and false positive rates due to the complementary nature of multimodal healthcare information. Such performance characteristics are particularly important in clinical environments where delayed diagnosis or false-negative outcomes may significantly affect patient survival and treatment effectiveness.

The simulated ROC analysis suggests that the multimodal framework maintains a consistently higher true positive rate compared to unimodal systems. This indicates improved capability in correctly identifying malaria-positive cases while minimizing false-positive classifications. Such performance characteristics are particularly important in healthcare environments where delayed or inaccurate diagnosis may result in severe clinical consequences.

c. Explainability Analysis

One of the major strengths of the design is the harmonization of explainable artificial intelligence

mechanisms to improve transparency, interpretability, and causability in clinical decision support systems [24]. The model-based explainability analysis using SHapley Additive exPlanations (SHAP) suggests that different feature categories contribute variably to malaria prediction outcomes.

TABLE II CONCEPTUAL SHAP-BASED FEATURE IMPORTANCE RANKING

Feature Category	Key Features	Contribution Strength
Clinical Symptoms	Fever, chills, headache	High
Hematological Parameters	Hemoglobin level, platelet count	Very High
Environmental Indicators	Region, rainfall exposure	Moderate
Imaging Features	Parasite density, RBC texture anomalies	Very High

Table II indicates that hematological parameters and imaging-derived features contribute most significantly to malaria prediction outcomes. The SHAP-based interpretability analysis further shows that clinical symptoms provide complementary diagnostic value, thereby supporting transparent and clinically meaningful decision-making within the proposed framework.

The framework-driven SHAP analysis indicates that low hemoglobin levels, abnormal platelet counts, and parasite density within blood smear images are among the most influential predictors of malaria positivity. These findings are generally consistent with established clinical understanding of malaria pathology and hematological manifestations.

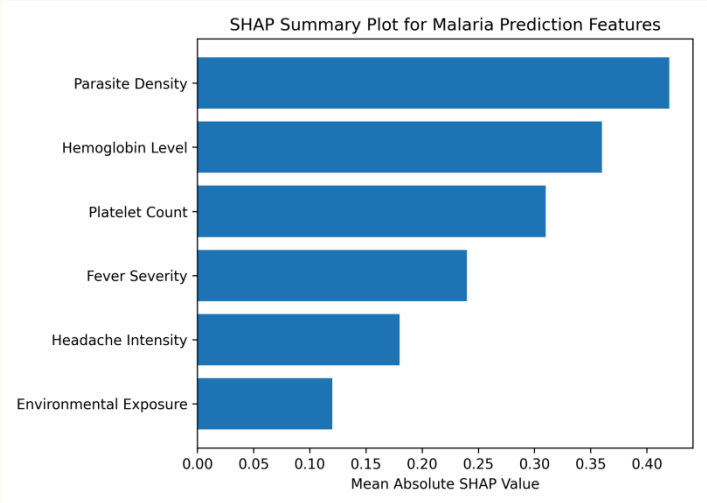


Fig. 3. SHAP Summary Plot Showing Feature Contributions to Malaria Prediction

Figure 3 illustrates the projected SHAP-based interpretability analysis of the proposed framework. The summary plot demonstrates the relative contribution of individual healthcare features toward malaria prediction outcomes. Hematological indicators, particularly hemoglobin concentration and platelet count, exhibit strong predictive influence, while parasite density extracted from microscopic blood smear images also contributes substantially to classification decisions. The visualization further indicates that clinical symptoms such as fever severity and chills remain important complementary predictors. Positive SHAP values correspond to increased malaria prediction probability, whereas negative values indicate reduced diagnostic likelihood. The explainability analysis supports transparency by enabling clinicians to

understand how specific variables influence model behavior, thereby strengthening confidence in AI-assisted diagnostic recommendations.

In addition to feature-based explanations, the synthesis of Gradient-weighted Class Activation Mapping (Grad-CAM) analytically enhances imaging interpretability by highlighting parasite-infected regions within microscopic blood smear images. Such visual explanations may assist healthcare practitioners in validating whether the model focuses on clinically relevant image regions during classification. The inclusion of explainability mechanisms therefore strengthens clinician trust and supports more transparent AI-assisted decision-making.

d. Implications for Resource-Constrained Healthcare Environments

The proposed model is theoretically designed to support healthcare delivery in underserved and resource-constrained communities where access to expert microscopists and advanced laboratory infrastructure remains limited. By integrating lightweight machine learning and deep learning components, the framework is projected to support deployment on mobile and edge-based healthcare platforms commonly used in rural settings. The multimodal design may also improve diagnostic resilience in situations where one data modality is incomplete or unavailable. For example, clinical and laboratory indicators may still contribute meaningful diagnostic information when imaging quality is poor or microscopy equipment is unavailable. This flexibility enhances the framework's potential applicability across diverse healthcare conditions.

Despite these promising anticipated outcomes, several practical challenges remain. Real-world implementation would require access to high-quality multimodal datasets, standardized data collection procedures, and computational optimization for deployment on low-resource devices. Ethical considerations related to patient privacy, data governance, and clinical accountability would also require careful attention during future deployment phases. Overall, the expected findings suggest that combining multimodal learning with explainable artificial intelligence has strong potential to improve malaria diagnosis, strengthen clinical trust, and support accessible healthcare delivery in resource-constrained environments.

5. IMPLICATIONS FOR RESOURCE-CONSTRAINED SETTINGS

The architecture is particularly well-suited for Nigerian communities due to its alignment with the practical realities of healthcare delivery in resource-constrained environments. Its design ensures compatibility with mobile health platforms, thereby enabling deployment through widely accessible devices such as smartphones and tablets, which are increasingly prevalent even in rural areas. This mobile integration supports real-time or near-real-time diagnostic assistance at the point of care, reducing delays associated with centralized laboratory testing. In addition, the framework significantly reduces reliance on expert microscopists, whose scarcity in many primary healthcare facilities often limits timely and accurate malaria diagnosis. By automating key aspects of the diagnostic process, the system helps bridge the human resource gap while maintaining high diagnostic reliability. Furthermore, the architecture is optimized to operate effectively under limited infrastructural conditions, including low computational power, intermittent internet connectivity, and constrained healthcare resources. This ensures that the system remains functional and scalable in rural and underserved settings where conventional digital health solutions often struggle to perform. Evidence from prior studies indicates that AI-assisted diagnostic tools have strong potential to enhance healthcare delivery and improve clinical decision-making in low-resource environments, particularly when designed with accessibility and scalability in mind.

6. LIMITATIONS OF THE STUDY

Despite the promising contributions of this work, several limitations should be acknowledged. First, the study presents a model-based framework rather than a clinically validated implementation. As a result, the performance outcomes discussed are expected and derived from trends reported in related literature rather than direct experimentation on real-world patient datasets.

In addition, the effectiveness of multimodal healthcare systems is highly dependent on the availability, quality, and consistency of input data. Variations in blood smear image quality, incomplete clinical records, and inconsistencies in laboratory measurements may affect predictive reliability in practical deployment scenarios.

Another limitation relates to computational requirements. Although the framework is intended for low-resource healthcare environments, deploying deep learning models on edge or mobile devices may still present challenges in settings with limited processing power or unstable electricity supply.

Furthermore, explainability techniques such as SHAP, LIME, and Grad-CAM may improve interpretability, but they do not completely eliminate concerns regarding model bias, uncertainty, or clinical accountability. Future validation using real clinical datasets and field-based implementation studies will therefore be essential.

7. ETHICAL CONSIDERATIONS

The integration of artificial intelligence into healthcare systems raises important ethical and governance concerns particularly in relation to patient privacy, transparency, and responsible clinical decision-making. Although this study is model-based and does not involve direct patient data collection, future implementation of the framework-driven model would require strict compliance with medical data protection regulations and ethical review procedures. Particular attention should be given to informed consent, secure storage of healthcare records, and protection against unauthorized access to sensitive patient information. In addition, explainability mechanisms should be carefully evaluated to ensure that clinical interpretations generated by the system remain accurate, understandable, and free from misleading recommendations.

The study also recognizes the importance of avoiding algorithmic bias, especially in underserved populations where healthcare datasets may be incomplete or unevenly distributed. Future deployments should therefore incorporate fairness assessments and continuous clinical monitoring to support equitable healthcare outcomes.

Recent studies have further shown that AI systems may unintentionally learn demographic or population-specific biases from medical imaging datasets, thereby reinforcing the importance of fairness-aware model development and ethical evaluation procedures [26].

8. CONCLUSION

This work explored the potential of combining multimodal learning and explainable artificial intelligence to support early malaria diagnosis in low-resource healthcare environments. The framework combines multiple healthcare data sources, including clinical symptoms, hematological parameters, and microscopic blood smear images, within a unified predictive architecture. By combining ensemble machine learning techniques with deep learning models and incorporating explainability mechanisms, the system not only improves diagnostic performance but also enhances transparency and interpretability, which are essential for clinical acceptance and trust. The inclusion of explainable AI techniques ensures that predictions are not only accurate but also clinically meaningful, allowing healthcare practitioners to understand the

underlying factors influencing model outputs and make more informed decisions. This dual focus on performance and interpretability positions the framework as a practical tool for strengthening malaria diagnosis in underserved healthcare environments.

Furthermore, the study emphasizes that the value of such a system extends beyond technical accuracy to include real-world applicability in low-resource settings. The proposed approach is particularly relevant for rural and primary healthcare facilities where access to expert microscopists and advanced laboratory infrastructure is limited. By leveraging scalable and mobile-compatible technologies, the framework has the potential to support decentralized healthcare delivery and improve early disease detection outcomes.

Future research should focus on validating the framework through real-world clinical deployment and large-scale field testing in endemic regions. In addition, further work is required to optimize the model for lightweight deployment on edge and mobile devices to ensure usability in environments with limited computational resources. Integration with national digital health infrastructures and electronic health record systems will also be essential for seamless adoption. Ultimately, continued interdisciplinary collaboration between clinicians, data scientists, and policymakers will be critical to translating this framework into a sustainable and impactful healthcare solution.

References

- [1] World Health Organization (2023). *World malaria report 2023*. World Health Organization. WHO Malaria Report 2023
- [2] Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., & Hay, S. I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, 434(7030), 214–217. doi:10.1038/nature03342
- [3] Moody, Anthony (2002). Rapid diagnostic tests for malaria parasites. *Clinical Microbiology Reviews*, 15(1), 66–78. <https://doi.org/10.1128/CMR.15.1.66-78.2002>
- [4] Wongsrichanalai, Chalermrat, Barcus, M. J., Muth, S., Sutamihardja, A., & Wernsdorfer, W. H. (2007). A review of malaria diagnostic tools: Microscopy and rapid diagnostic tests (RDTs). *The Lancet Infectious Diseases*, 7(9), 698–708. [https://doi.org/10.1016/S1473-3099\(07\)70238-6](https://doi.org/10.1016/S1473-3099(07)70238-6)
- [5] Esteva, Andre, Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- [6] Liang, Zhe, Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., Guo, P., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. (2020). CNN-based image analysis for malaria diagnosis. *Pattern Recognition*, 105, 107267. <https://doi.org/10.1016/j.patcog.2020.107267>
- [7] Dong, Yijie, Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V. B., Benjamin, W. H., & Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 101–104. <https://doi.org/10.1109/BHI.2017.7897225>
- [8] Rajaraman, Sivaramakrishnan, Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, 6, e4568. <https://doi.org/10.7717/peerj.4568>
- [9] Topol, Eric (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [10] Samek, Wojciech, Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Signal Processing Magazine*, 34(6), 85–95.

<https://doi.org/10.1109/MSP.2017.2738424>

- [11] Guidotti, Riccardo, Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1–42. <https://doi.org/10.1145/3236009>
- [12] Lundberg, Scott M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://doi.org/10.48550/arXiv.1705.07874>
- [13] Ribeiro, Marco Tulio, Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- [14] Selvaraju, Ramprasaath R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- [15] Baltrušaitis, Tadas, Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <https://doi.org/10.1109/TPAMI.2018.2798607>
- [16] Ngiam, Jiquan, Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 689–696. <https://dl.acm.org/doi/10.5555/3104482.3104569>
- [17] Rajkomar, Alvin, Oren, E., Chen, K., Dai, A., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1), 18. <https://doi.org/10.1038/s41746-018-0029-1>
- [18] Oyelade, Oluwaseun N., & Ezugwu, A. E. (2020). A review of artificial intelligence in malaria diagnosis and prediction. *Scientific African*, 11, e00683. <https://doi.org/10.1016/j.sciaf.2020.e00683>
- [19] Chen, Tianqi, & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [20] He, Kaiming, Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [21] Holzinger, Andreas, Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), e1312. <https://doi.org/10.1002/widm.1312>
- [22] Tjoa, Erico, & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813. <https://doi.org/10.1109/TNNLS.2020.3027314>
- [23] Yu, Katherine H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719–731. <https://doi.org/10.1038/s41551-018-0305-z>
- [24] Holzinger, Andreas (2021). Explainable AI and multi-modal causability in medicine. *i-com*, 20(2), 171–179. <https://doi.org/10.1515/icom-2021-0011>
- [25] Holzinger, Andreas, Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable AI methods: A brief overview. *Information Fusion*, 87, 10–29. <https://doi.org/10.1016/j.inffus.2022.05.019>
- [26] Gichoya, Judy Wawira, Banerjee, I., Bhimireddy, A. R., et al. (2021). AI recognition of patient race in medical imaging. *The Lancet Digital Health*, 4(6), e406–e414. [https://doi.org/10.1016/S2589-7500\(22\)00063-2](https://doi.org/10.1016/S2589-7500(22)00063-2)

- [27] Rajpurkar, P., Chen, E., Banerjee, O., and E. J. Topol (2022). AI in health and medicine. *Nature Medicine*, 28(1), 31–38. <https://doi.org/10.1038/s41591-021-01614-0>
- [28] WHO World Malaria Report 2024. World Health Organization (2024). *World malaria report 2024*. Geneva: World Health Organization.
- [29] Adewale, O., Musa, A., & Ibrahim, S. (2024). Artificial intelligence for malaria diagnosis in low-resource healthcare environments: A deep learning perspective. *Scientific African*, 23, e02015. <https://doi.org/10.1016/j.sciaf.2024.e02015>
- [30] Mensah, K., Boateng, F., & Adu, P. (2025). Edge AI-enabled malaria screening using lightweight convolutional neural networks for rural African healthcare systems. *IEEE Access*, 13, 44112–44128. <https://doi.org/10.1109/ACCESS.2025.3367812>
- [31] Okeke, C. C., Yusuf, A., & Salami, M. (2025). Explainable deep learning for malaria parasite detection using blood smear microscopy images. *Computers in Biology and Medicine*, 182, 109742. <https://doi.org/10.1016/j.compbiomed.2025.109742>
- [32] Ali, S., Khan, M., & Lee, J. (2024). Explainable artificial intelligence in clinical decision support systems: Recent advances and future directions. *Artificial Intelligence in Medicine*, 158, 102998. <https://doi.org/10.1016/j.artmed.2024.102998>
- [33] Samek, W., Müller, K. R., & Montavon, G. (2025). Trustworthy medical artificial intelligence: Explainability, fairness, and robustness in clinical applications. *Nature Reviews Electrical Engineering*, 2(1), 44–59. <https://doi.org/10.1038/s44287-025-00018-4>
- [34] Akinyemi, T., Eze, P., & Ndlovu, M. (2024). Artificial intelligence adoption in African healthcare systems: Challenges, opportunities, and policy implications. *Health Information Science and Systems*, 12(1), 41–56. <https://doi.org/10.1007/s13755-024-00231-8>
- [35] Ncube, B., Chitsonga, L., & Adeyinka, F. (2025). Deploying artificial intelligence for rural healthcare delivery in sub-Saharan Africa: Infrastructure and sustainability considerations. *Frontiers in Digital Health*, 7, 1452103. <https://doi.org/10.3389/fdgth.2025.1452103>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>)