IKR Publishers

IKR Journal of Multidisciplinary Studies (IKRJMS)
Journal homepage: https://ikrpublishers.com/ikrjms/
Volume-2, Issue-3 (May-June) 2026

OPEN ACCESS
ISSN: 3107-3999 (Online)

Bioinformatics Based Intelligent Framework for Real Time Malaria Severity Classification Using Ensemble Learning Technique

Gabriel Akibi Inyang[1] & Solomon, Fidelis Uma[2*]
Department of Computer Science, University of Cross River State, Nigeria

DOI:*10.5281/zenodo.20314706*

ARTICLE INFO

Article history:
Received : 07-05-2026
Accepted : 14-05-2026
Available online : 20-05-2026

Citation: Inyang, G. A., & Uma, S. F. (2026). Bioinformatics Based Intelligent Framework for Real Time Malaria Severity Classification Using Ensemble Learning Technique. *IKR Journal of Multidisciplinary Studies (IKRJMS), 2*(3), 1-15.

ABSTRACT **Original Research Article**

This study focuses on the development of a machine learning–based approach for classifying the severity of malaria, with particular attention to cerebral malaria. The dataset was compiled from healthcare records and included both clinical symptoms such as seizures, altered mental state, headache, vomiting, and focal neurological deficits and indicators reflecting household burden, including financial strain and caregiver stress. The data were carefully prepared and divided into training and testing sets to ensure a reliable evaluation of the model's performance. A Random Forest classifier was employed to distinguish between cerebral and non-cerebral malaria cases. By leveraging multiple decision trees, the model was trained to recognize patterns within the data and accurately predict the severity of the condition based on the observed features. The training process involved optimizing the model to improve its predictive capability across different symptom combinations and contextual factors. The results demonstrated strong performance, with the model achieving an accuracy of 0.97. Additional evaluation metrics further supported its effectiveness, with macro-average precision, recall, and F1 scores of 0.94, 0.91, and 0.92 respectively. Analysis of the dataset showed that 87.2% of cases were classified as non-cerebral malaria, while 12.8% were identified as cerebral malaria, reflecting known patterns in malaria-endemic regions. Importantly, symptoms such as seizures, altered mental status, and focal neurological deficits were found to be key indicators of severe malaria. Beyond clinical implications, the study also highlights the broader impact of the disease, particularly the financial and emotional burden placed on affected households. Overall, the findings suggest that machine learning techniques, especially Random Forest models, can serve as valuable tools in supporting more accurate diagnosis and improved management of malaria.

Keywords: Malaria, Severity, Burden, Machine Learning, Health.

**Corresponding author: Solomon, Fidelis Uma*
Department of Computer Science, University of Cross River State, Nigeria

Introduction

Malaria remains a major global public health concern, particularly in tropical and developing regions where it contributes significantly to illness and death. Beyond its direct medical effects, malaria also imposes considerable psychological, emotional, and socio-economic burdens on individuals and households. The concept of psychological malaria burden captures the stress, anxiety, and financial hardship experienced by patients and their caregivers, especially in severe cases such as cerebral malaria. Caregivers are often subjected to emotional strain and mental exhaustion as they manage the patient's condition, particularly when outcomes are uncertain. At the same time, the cost of treatment and the potential loss of income further intensify this burden, creating a cycle of hardship that affects the overall well-being of families and communities (Edet et al., 2025; Edet et al., 2024a). Similar multidimensional impacts have been observed in other health-related and data-driven studies, where both clinical and socio-economic factors influence outcomes (Ekong et al., 2022; Ekong et al., 2023).

Cerebral malaria, a severe neurological complication primarily caused by ***Plasmodium falciparum***, is among the most life-threatening forms of the disease. It is characterized by symptoms such as impaired consciousness, seizures, altered mental state, and focal neurological deficits, often requiring urgent and intensive medical care. Due to its severity, early detection and accurate classification are essential for improving survival rates and reducing long-term complications. Effective classification systems enable healthcare professionals to quickly identify high-risk patients and initiate appropriate interventions. In recent years, intelligent classification models have been successfully applied in various domains to improve decision-making and diagnostic accuracy, demonstrating their potential in addressing complex medical challenges (Edet & Ansa, 2023; Ebong et al., 2024).

Despite the well-documented clinical manifestations of malaria, its broader socio-economic and psychological effects remain relatively underexplored. Understanding these indirect impacts is crucial for developing comprehensive healthcare strategies that address not only the disease itself but also its consequences on patients and caregivers. Data-driven approaches have shown promise in capturing complex relationships between multiple factors, providing deeper insights into both health outcomes and associated burdens (Edet et al., 2024b; Ekong et al., 2024a). These approaches enable the integration of clinical indicators with contextual variables, offering a more complete understanding of disease severity and its implications.

Furthermore, advances in machine learning have enabled the development of models capable of analyzing large and complex datasets to support classification and prediction tasks. Such models have been applied across various fields, including healthcare, cybersecurity, and risk assessment, where they have demonstrated strong performance in identifying patterns and supporting informed decision-making (Ekong et al., 2024b; Edet et al., 2024c; Udoetor et al., 2024). The adaptability of these techniques makes them particularly suitable for handling the diverse and multidimensional nature of malaria-related data.

In addition, integrated frameworks that combine analytical techniques with intelligent systems have been shown to enhance the accuracy and efficiency of classification processes. These frameworks are capable of identifying critical features, detecting patterns, and providing actionable insights that can guide intervention strategies. Research in related domains further supports the effectiveness of such approaches in improving system performance and addressing complex challenges (Uwah & Edet, 2024; Edet et al., 2024d; Edet et al., 2024e). Applying similar methodologies to malaria classification can therefore contribute to more reliable and timely diagnosis.

Addressing cerebral malaria requires not only an understanding of its clinical presentation but also a recognition of its wider psychological and socio-economic impact. By integrating data-driven techniques with clinical knowledge, it becomes possible to improve detection, enhance patient care, and reduce the burden on affected households. This holistic perspective is essential for developing effective healthcare solutions that respond to both the medical and human dimensions of the disease (Edet et al., 2025; Ekong et al., 2024c).

This research is centered on the development of a real-time classification system designed to understand and predict the psychological and socio-economic impacts of cerebral malaria on affected households. The burden associated with cerebral malaria goes far beyond the individual patient, extending to caregivers, families, and even the wider community. Caregivers, in particular, often experience intense emotional strain, anxiety, and mental fatigue due to the prolonged and demanding nature of care, as well as the uncertainty surrounding patient outcomes. In many situations, these responsibilities limit their ability to maintain regular employment, resulting in reduced income and increased financial pressure. Consequently, the overall impact of cerebral malaria on households is multi-dimensional, affecting health, emotional stability, and economic well-being simultaneously (Edet et al., 2025; Edet et al., 2024a).

The socio-economic consequences of cerebral malaria can create a reinforcing cycle of hardship. As financial strain increases, access to adequate healthcare may become more limited, which in turn can worsen the severity of the condition and delay timely medical intervention. Despite the significance of these effects, existing research has largely focused on the clinical dimensions of malaria, with relatively little attention given to its psychological and economic implications. This gap highlights the need for a more comprehensive and integrated approach that combines clinical indicators with socio-economic variables to better understand the full scope of the disease’s impact. Data-driven frameworks in other domains have demonstrated the value of integrating multiple data sources to improve classification and decision-making processes (Edet et al., 2024b; Anthony Edet et al., 2025).

To address this gap, the present study proposes a real-time classification system that leverages advanced machine learning techniques, particularly the Random Forest algorithm, to assess malaria severity while incorporating psychological and socio-economic factors. By integrating clinical symptoms with variables such as caregiver burden, financial stress, and household income loss, the system is able to generate more comprehensive insights into disease severity. Similar intelligent classification systems have been successfully applied in various contexts, including healthcare analytics and risk assessment, where they enhance predictive accuracy and support informed decision-making (Edet et al., 2024c; Ekong et al., 2025). This approach enables healthcare professionals to move beyond purely clinical assessments and consider the broader context in which the disease occurs.

The model developed in this study is trained on a dataset that includes demographic information, clinical features, and socio-economic indicators. Through this integrated dataset, the system is capable of classifying cases into "cerebral malaria" and "non-cerebral malaria" categories with a high level of accuracy. Machine learning models have increasingly demonstrated their effectiveness in handling complex, multi-variable datasets, making them particularly suitable for applications in disease prediction and classification. Their ability to identify hidden patterns and relationships among variables allows for more accurate and timely predictions, which are essential for early intervention and improved patient outcomes (Edet et al., 2024d; Ekong et al., 2025).

Beyond its clinical utility, the proposed system also provides valuable insights into the relationship between psychological stressors and disease severity. Factors such as caregiver burden and financial instability are not only consequences of the disease but may also influence health outcomes by affecting access to care and adherence to treatment. Understanding these relationships is critical for effective resource allocation and policy formulation. Studies in related fields have shown that integrating economic and behavioral factors into analytical models can significantly improve the relevance and applicability of predictive systems (Edet et al., 2025; Edet et al., 2024a).

The implications of this research extend beyond the healthcare setting to inform public health planning and policy development. By quantifying the psychological and socio-economic burden of cerebral malaria, the system offers a more holistic perspective on the disease, enabling stakeholders to design targeted interventions that address both medical and non-medical challenges. This work contributes to the development of a more comprehensive framework for managing cerebral malaria one that not only focuses on diagnosis and treatment but also considers the broader social and economic realities faced by affected households (Anthony Edet et al., 2025; Edet et al., 2024b)

Literature Foundation

Yisa et al., 2023, proposed a work on ***Development of a Real-Time Classification System of Psychological Malaria Burden on Households***. This study critically examines the shortcomings of existing malaria information systems, particularly their limited capacity to effectively collect, integrate, and analyze data for meaningful decision-making. Many of these systems are fragmented in nature, operate in isolation from policy and intervention frameworks, and often fail to capture sufficient household-level data required for comprehensive analysis. As a result, they are unable to provide timely and actionable feedback that could support effective malaria control strategies. To address these limitations, the study introduces a real-time classification system aimed at improving malaria surveillance while also providing a more detailed understanding of the burden experienced at the household level. The system utilizes data related to the total cost of malaria burden incurred by households and applies a multi-classifier approach based on the K-Nearest Neighbors (KNN) algorithm. Through this approach, malaria burden is categorized into four distinct levels: no burden, low burden, high burden, and severe burden. Following classification, the outcomes are stored within a database management system, ensuring efficient tracking, monitoring, and planning for intervention strategies. The system achieved an accuracy rate of 89%, demonstrating its effectiveness and reliability in classifying malaria burden and supporting data-driven public health responses.

Yoon et al., 2021, proposed a work on ***An automated microscopic malaria parasite detection system using digital image analysis***. This research presents a significant advancement in malaria diagnostics by incorporating mathematical morphology techniques into the detection and classification of malaria parasites in stained blood smear images. Traditional diagnostic approaches, which rely heavily on manual visual inspection, are often associated with challenges such as subjectivity, inconsistency, and time inefficiency. To overcome these limitations, the study introduces computational image analysis methods that enhance both the speed and accuracy of diagnosis. Mathematical morphology, which focuses on the analysis of geometric structures within images, provides powerful tools for preprocessing and segmentation tasks. These techniques are particularly effective in biomedical image analysis, where precision is critical for identifying malaria parasites within complex and noisy microscopic images. By applying morphological operators, the system enhances image quality, enabling clearer visualization and more accurate segmentation of parasite structures. The approach also supports the classification of parasites across different stages of infection, thereby improving diagnostic depth. Overall, the study demonstrates how morphology-based techniques can automate and optimize malaria detection processes, reducing reliance on manual expertise while enabling scalable diagnostic solutions, particularly in regions with limited healthcare resources.

Jdey et al., 2024, proposed a work on ***Deep learning and machine learning for Malaria detection: overview, challenges and future directions***. This study underscores the critical role of evidence-based decision-making in public health, particularly in the context of malaria detection and control. It highlights how machine learning algorithms contribute to the efficient collection, processing, and analysis of large volumes of health-related data, ultimately supporting informed decision-making. The research places particular emphasis on image analysis, a key component of malaria surveillance systems, which has gained significant attention within the domains of computer vision and machine learning. By reviewing various machine learning and image processing techniques, the study explores how these approaches can be utilized to detect and predict malarial infections with improved accuracy. Deep learning methods, in particular, are identified as highly promising due to their ability to automatically extract complex features from medical images, thereby enhancing diagnostic performance and supporting clinicians in decision-making. However, the study also acknowledges several challenges associated with deep learning applications, including the need for large and well-annotated datasets, high computational requirements, difficulties in achieving real-time processing, and the limited interpretability of complex models. The work concludes by outlining future research directions aimed at addressing these challenges, with the goal of improving the scalability, efficiency, and transparency of machine learning systems for malaria detection.

Grignaffini et al., 2024, proposed a work on ***Computer-Aided Diagnosis Systems for Automatic Malaria Parasite Detection and Classification: A Systematic Review***. This study provides a comprehensive analysis of the role of artificial intelligence in improving malaria diagnosis, a critical requirement given the life-threatening nature of the disease and its widespread impact. While traditional light microscopy remains the standard method for malaria detection, it is often hindered by long processing times and the need for highly skilled professionals. To address these limitations, the study explores the development of computer-aided diagnosis systems that utilize machine learning and deep learning techniques to automate the detection and classification of malaria parasites in blood smear images. Adopting the PRISMA 2020 framework, the authors conducted an extensive review of literature from databases such as PubMed, Scopus, and arXiv, identifying 606 initial studies and narrowing them down to 135 relevant works for in-depth analysis. The findings reveal significant progress in the application of AI-driven techniques, demonstrating their effectiveness in improving diagnostic accuracy and efficiency. Additionally, the study highlights the emergence of mobile and web-based applications designed to support malaria diagnosis in resource-limited settings, where access to expert analysis may be constrained. These innovations offer scalable and accessible solutions, reducing dependence on specialized expertise while enhancing diagnostic capabilities. Overall, the review emphasizes the transformative potential of AI in malaria diagnostics and its role in improving healthcare delivery in endemic regions.

Silka et al., 2023, proposed a work on ***Malaria detection using advanced deep learning architecture***. This study introduces a novel approach to malaria diagnosis through the application of an advanced convolutional neural network (CNN) architecture. Recognizing malaria as a life-threatening disease that requires timely and accurate detection, the research focuses on improving diagnostic performance, particularly in developing regions where healthcare resources may be limited. The proposed CNN model is trained on a large dataset of blood smear images and is designed to automatically detect malaria parasites with a high degree of precision. The results of the study are particularly notable, with the model achieving an accuracy of 99.68%, significantly outperforming many existing diagnostic techniques. In addition to accuracy, the model demonstrates strong sensitivity and specificity, effectively distinguishing between infected and uninfected samples. The study also provides a detailed evaluation of the model’s performance across different malaria subtypes, highlighting its robustness and adaptability in handling diverse cases. By offering faster processing times and improved diagnostic reliability, the CNN-based approach presents a promising solution for large-scale deployment in malaria-endemic regions. The findings underscore the growing importance of deep learning technologies in infectious disease diagnostics and their potential to enhance healthcare systems by providing efficient, scalable, and accurate tools for disease detection

Methodology

The development of the classification system for cerebral and non-cerebral malaria severity follows a structured and systematic procedure designed to ensure accuracy, reliability, and real-time applicability. The methodology is outlined in sequential stages as follows:

Data Collection and Feature Selection

The dataset used in this study comprises essential medical and psychological indicators, including Seizures, Headache, Vomiting, Altered Mental Status, Focal Neurological Deficits, Financial Stress, and Caregiver Burden. These data were obtained from a combination of clinical health records, structured household surveys, and previously documented malaria-related studies. The selection of these features was guided by their established significance in determining malaria severity, particularly in distinguishing between cerebral and non-cerebral cases. Each variable was carefully chosen to ensure that both clinical symptoms and socio-economic factors contributing to disease impact were adequately represented in the dataset.

Data Preprocessing

The preprocessing stage involved a thorough cleaning of the dataset to eliminate inconsistencies, duplicate entries, and missing values that could affect model performance. Numerical features were normalized to maintain consistency in scale, while categorical variables were appropriately encoded to enable compatibility with the machine learning model. Variables such as caregiver burden and financial stress were further scaled to ensure they contributed proportionately during classification. To address potential class imbalance between cerebral and non-cerebral malaria cases, oversampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) were applied. This process ensures that the model remains unbiased and does not disproportionately favor the majority class during prediction.

Model Development and Training

The classification framework is built using the Random Forest algorithm, selected for its efficiency in handling complex, high-dimensional datasets and its ability to model non-linear relationships effectively. The dataset was divided into training and testing subsets using an 80:20 ratio, allowing for proper evaluation of the model's predictive capability. During training, cross-validation techniques were applied to enhance model reliability and reduce the risk of overfitting. The Random Forest model operates by generating multiple decision trees and combining their outputs to produce a final classification, thereby improving overall accuracy. This approach enables the system to effectively categorize malaria severity into cerebral and non-cerebral classes based on the input features.

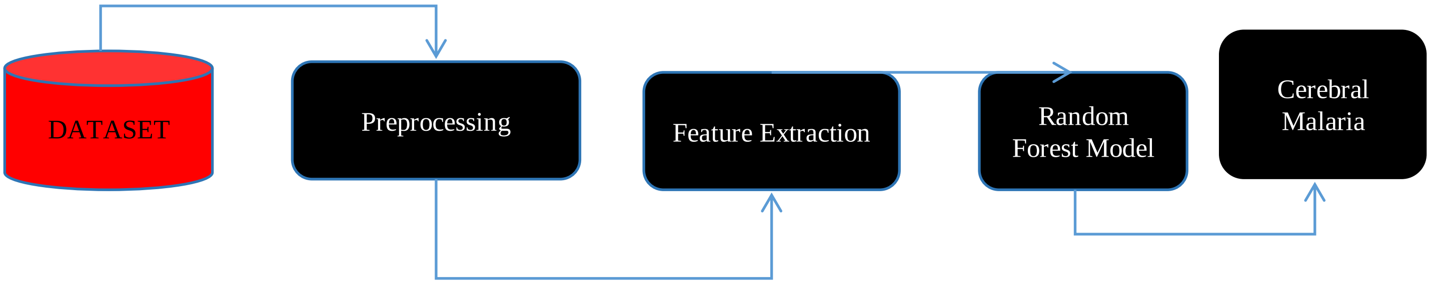

Figure 1: Black Diagram of the Proposed System

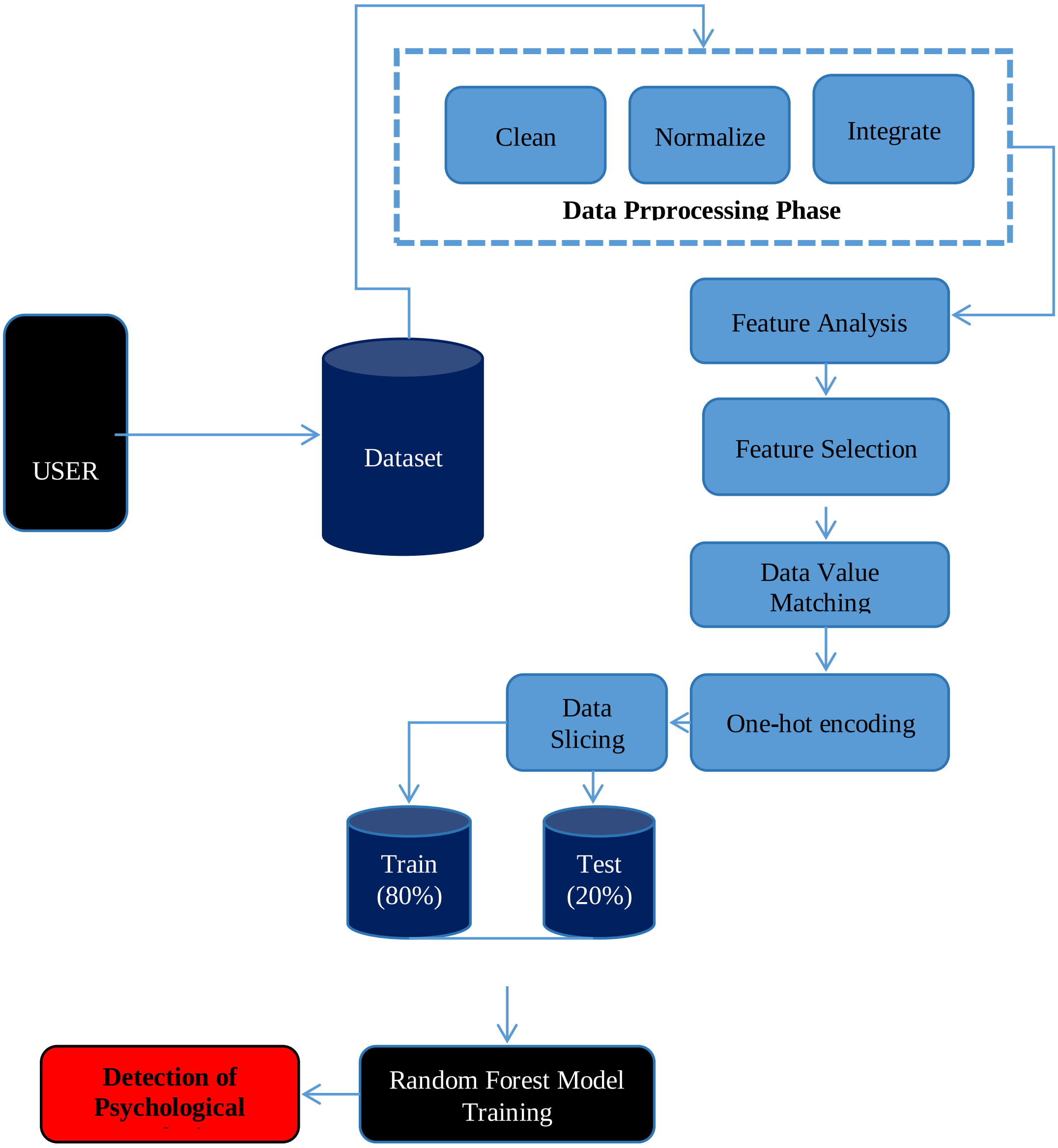

Figure 2: Conceptual Architecture of the Proposed System

Algorithm

Bootstrap Sampling

Let D represent the original dataset with size N: D = {X1, X2, ..., XN}

A bootstrap sample B is created from D by randomly selecting N data points **with replacement**, meaning some values may appear more than once:

B = {X1, X2, ..., XN} (with possible repetition)

Building Decision Trees

For each decision tree, nodes are split based on selected features:

a. A random subset of features F is chosen at each node: F is a subset of {F1, F2, ..., Fm}, where m is the total number of features

b. At each node, the best split is determined using a criterion such as Gini impurity:

Gini(t) = 1 − sum of (p_i squared) for all classes i in node t

Here, p_i represents the proportion of class i in node t

Growing the Trees

The splitting process continues recursively for each node until a stopping condition is reached. This could be a maximum tree depth or a minimum number of samples required at a leaf node.

Prediction

For a new input X_new, each tree in the forest makes its own prediction:

Prediction(t_i) = Class(X_new)

The final prediction is determined using majority voting across all trees:

Prediction(final) = mode of {Prediction(t1), Prediction(t2), ..., Prediction(tN)}

Here, N represents the total number of trees in the forest

Model Evaluation

The accuracy of the model on a test dataset T is calculated as: Accuracy = (1/N) × sum of I(Prediction(Ti) = Actual(Ti)) for all test instances Ti

Where:

I is an indicator function that equals 1 if the prediction is correct, and 0 if it is incorrect

N is the total number of test samples

Results and Discussion

In this section, the findings and overall development outcomes of the study are presented in detail. The dataset utilized for this research was obtained from a malaria burden and severity classification project, which is centered on examining both the clinical and psychological impacts of malaria, with particular emphasis on cerebral malaria. The data was gathered from a combination of structured surveys and medical records, capturing a comprehensive set of features that reflect not only the clinical manifestations of malaria but also the associated effects on households. The primary objective of the dataset is to support the classification of malaria cases into two distinct categories: Cerebral Malaria and Non-Cerebral Malaria. This classification plays a critical role in understanding the severity of the disease and its broader psychological implications on affected individuals and their families, thereby enabling more precise and targeted healthcare interventions.

The dataset is composed of several important clinical features, including Seizures, Altered Mental Status, Headache, Vomiting, and Focal Neurological Deficits. These variables are directly linked to the identification of malaria severity. For example, symptoms such as seizures, altered mental status, and focal neurological deficits are widely recognized as strong indicators of cerebral malaria, making them essential for accurately distinguishing between severe and non-severe cases. In addition, other symptoms like fever, chills, fatigue, and muscle pain are commonly observed across all malaria cases and provide valuable insight into the general clinical presentation of the disease, irrespective of its severity level.

Beyond clinical indicators, the dataset also incorporates household-level and socio-economic variables that reflect the broader burden of malaria. These include factors such as Financial Stress, Caregiver Burden, and Household Income Loss, all of which help to capture the psychological and economic strain experienced by affected families. Furthermore, variables like Access to Healthcare and Family Support System offer additional context regarding external influences on patient outcomes, as limited healthcare access and weak support systems can significantly worsen disease severity and recovery. By integrating both medical and socio-economic dimensions, the dataset provides a comprehensive perspective on malaria's impact, supporting effective classification of disease severity and contributing to more informed and holistic healthcare decision-making.

Table 1: Dataset Feature Information

Feature	**Data type**	**Data type Size**	**Feature Importance Ranking**	**Sample Type**	**Conversion**
Age	Integer	4 bytes	Medium	Continuous	No conversion needed
Gender	String	Varies	Low	Categorical (Male/Female)	Label Encoding (Male=1, Female=0)
Seizures	String	Varies	High	Categorical (Yes/No)	Binary (Yes=1, No=0)
Altered_Mental_Status	String	Varies	High	Categorical (Yes/No)	Binary (Yes=1, No=0)
Headache	String	Varies	Medium	Categorical (Yes/No)	Binary (Yes=1, No=0)
Vomiting	String	Varies	Medium	Categorical (Yes/No)	Binary (Yes=1, No=0)
Focal_Neuro_Defs	String	Varies	High	Categorical (Yes/No)	Binary (Yes=1, No=0)
Fever	String	Varies	Medium	Categorical (Yes/No)	Binary (Yes=1, No=0)
Chills	String	Varies	Low	Categorical (Yes/No)	Binary (Yes=1, No=0)
Fatigue	String	Varies	Medium	Categorical (Yes/No)	Binary (Yes=1, No=0)
Muscle_Pain	String	Varies	Low	Categorical (Yes/No)	Binary (Yes=1, No=0)
Financial_Stress	String	Varies	High	Categorical (High/Low)	Binary (High=1, Low=0)
Caregiver_Burden	String	Varies	High	Categorical (High/Low)	Binary (High=1, Low=0)
Household_Income_Loss	String	Varies	Medium	Categorical (Yes/No)	Binary (Yes=1, No=0)
Access_to_Healthcare	String	Varies	High	Categorical (Good/Poor)	Binary (Good=1, Poor=0)
Family_Support_System	String	Varies	High	Categorical (Good/Poor)	Binary (Good=1, Poor=0)
Malaria_Severity_Class	String	Varies	High	Categorical (Cerebral Malaria/Non-Cerebral Malaria)	Binary (Cerebral Malaria=1, Non-Cerebral Malaria=0)

```
# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)

# Print the evaluation results
print(f'Accuracy: {accuracy}')
print('Classification Report:')
print(classification_rep)
```

```
<ipython-input-13-8c2e768f224b>:20: FutureWarning: Downcasting behavior
  df_binary = df.replace(binary_conversion)
Accuracy: 0.9655
Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.99      0.98      1734
           1       0.91      0.82      0.86       266

    accuracy                           0.97      2000
   macro avg       0.94      0.91      0.92      2000
weighted avg       0.96      0.97      0.96      2000
```

Figure 3: Random Forest Classification Reports

In this research, the main objective is to classify malaria severity into two categories Cerebral Malaria and Non-Cerebral Malaria while also examining the burden of this condition on households and caregivers. This is achieved using a dataset that contains a combination of malaria-related symptoms, household burden indicators, and healthcare access variables. The initial step in the implementation involved setting up a connection between Google Drive and the Jupyter environment to enable seamless access to the dataset stored online. Once the connection was successfully established, the dataset was imported into the working environment for further processing and analysis.

After loading the data, the features were carefully processed and transformed into binary representations to make them suitable for machine learning analysis. These processed features were then used to train a Random Forest classification model. The Random Forest algorithm was selected because of its strength in handling complex and non-linear relationships within data, as well as its stability and effectiveness in classification tasks involving multiple features.

Following training, the model was evaluated using standard performance metrics. The results showed a high level of performance, with an accuracy of 97%, indicating that the model is highly reliable in correctly classifying malaria severity. In addition, the macro average precision, recall, and F1-score were recorded as 0.94, 0.91, and 0.92 respectively. These values suggest that the model does not only perform well overall but also maintains balanced performance across both classes, effectively distinguishing between Cerebral Malaria and Non-Cerebral Malaria without bias toward any particular category.

These results are consistent with the goal of the study, which is to develop a dependable classification system for malaria severity that can assist healthcare professionals in early identification of cerebral malaria cases and support better decision-making in resource allocation and patient management.

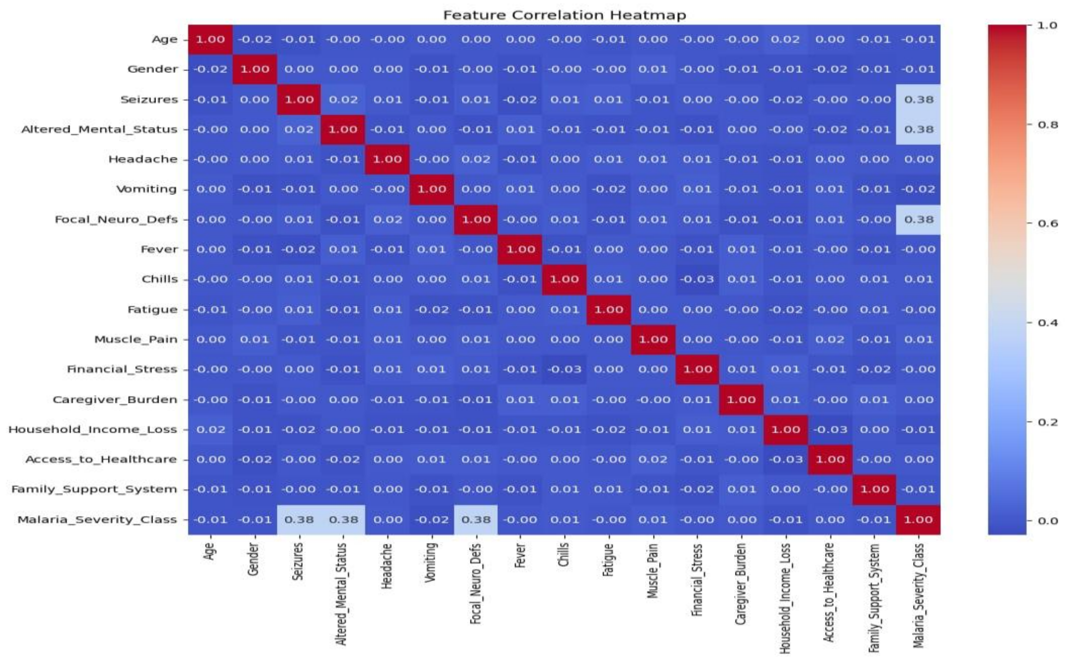

Figure 4: Heatmap of Feature Distribution Showing Relationships

A Heatmap of Feature Distribution Showing Relationships in this research provides a visual representation of how different variables in the dataset relate to one another, particularly in relation to malaria severity classification. It illustrates the correlation between clinical symptoms such as seizures, altered mental status, fever, vomiting, and focal neurological deficits, alongside socio-economic factors like caregiver burden, financial stress, and household income loss. Through this visualization, the study is able to uncover hidden patterns and dependencies that exist within the dataset.

For this research, the heatmap is particularly useful in showing how specific symptoms and household-related factors collectively influence the classification of cases into Cerebral Malaria and Non-Cerebral Malaria. It makes it easier to observe which features tend to move together and which ones have stronger relationships with disease severity outcomes. For instance, stronger correlations between neurological symptoms and severe malaria classification provide important insight into the most critical indicators of cerebral malaria. A properly constructed heatmap also helps to highlight the most influential features within the dataset, thereby supporting feature selection during model development. By identifying variables that are more strongly associated with malaria severity, it becomes easier to focus the learning process on the most meaningful inputs. This, in turn, contributes to improving the performance and reliability of the classification model by reducing noise and emphasizing relevant predictors. The heatmap serves as an essential analytical tool in this study, not only for understanding feature relationships but also for guiding the training process toward more accurate and meaningful predictions of malaria severity.

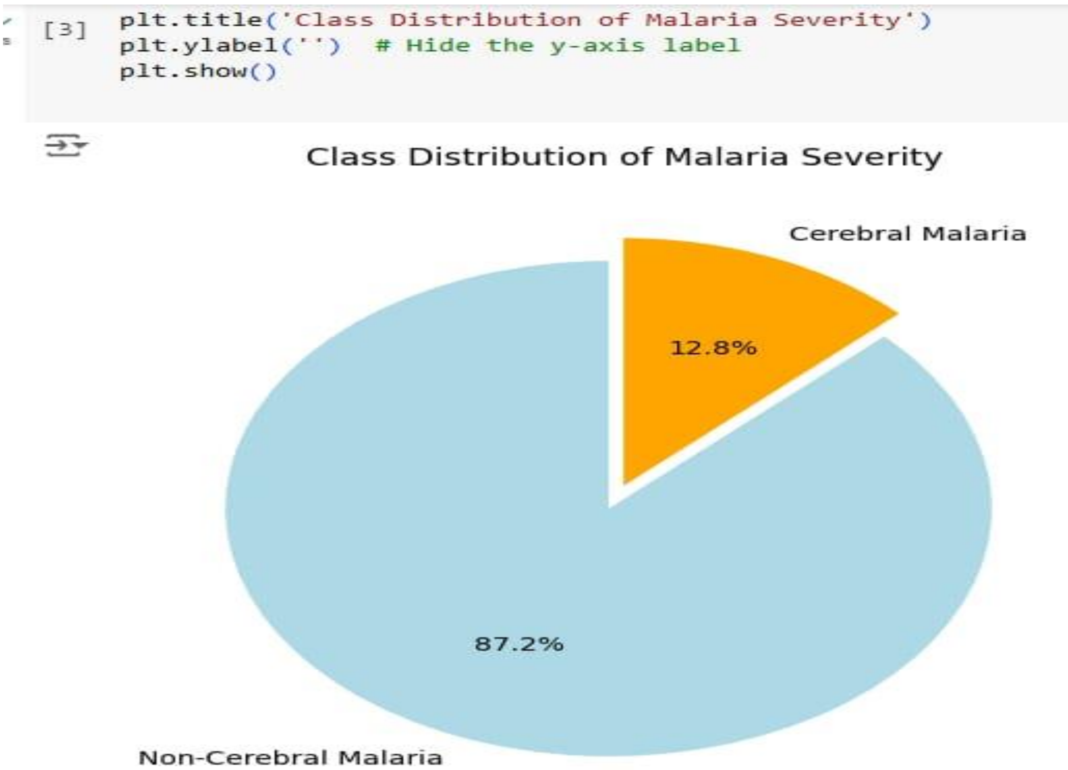

Figure 5: Distribution of Malaria Severity Classes

The pie chart illustrating the class distribution of the malaria dataset shows that 12.8% of the cases are classified as Cerebral Malaria, while 87.2% are categorized as Non-Cerebral Malaria. This distribution closely reflects real-world patterns, where Non-Cerebral Malaria occurs far more frequently than Cerebral Malaria. The higher proportion of Non-Cerebral Malaria cases is expected, as cerebral malaria is a severe but relatively uncommon complication of malaria, typically affecting high-risk groups such as young children and individuals with weakened immune systems. This imbalance in the dataset aligns with observations in malaria-endemic regions, where most reported cases fall under the non-severe category and are generally manageable with standard treatment approaches. Recognizing this distribution is important in understanding the structure of the dataset used in this study, as it directly influences how the classification model learns from the available data. From a modelling perspective, the observed class imbalance highlights the importance of ensuring that the Random Forest classifier is able to correctly differentiate between both categories without developing a bias toward the majority class. If not properly handled, such imbalance could lead the model to favor predictions of Non-Cerebral Malaria due to its higher frequency in the dataset. At the same time, this distribution provides meaningful insight for healthcare planning, as it emphasizes the need for strong focus on early detection and management of Non-Cerebral Malaria cases, while still ensuring that adequate resources and specialized attention are directed toward the smaller but more severe group of Cerebral Malaria patients.

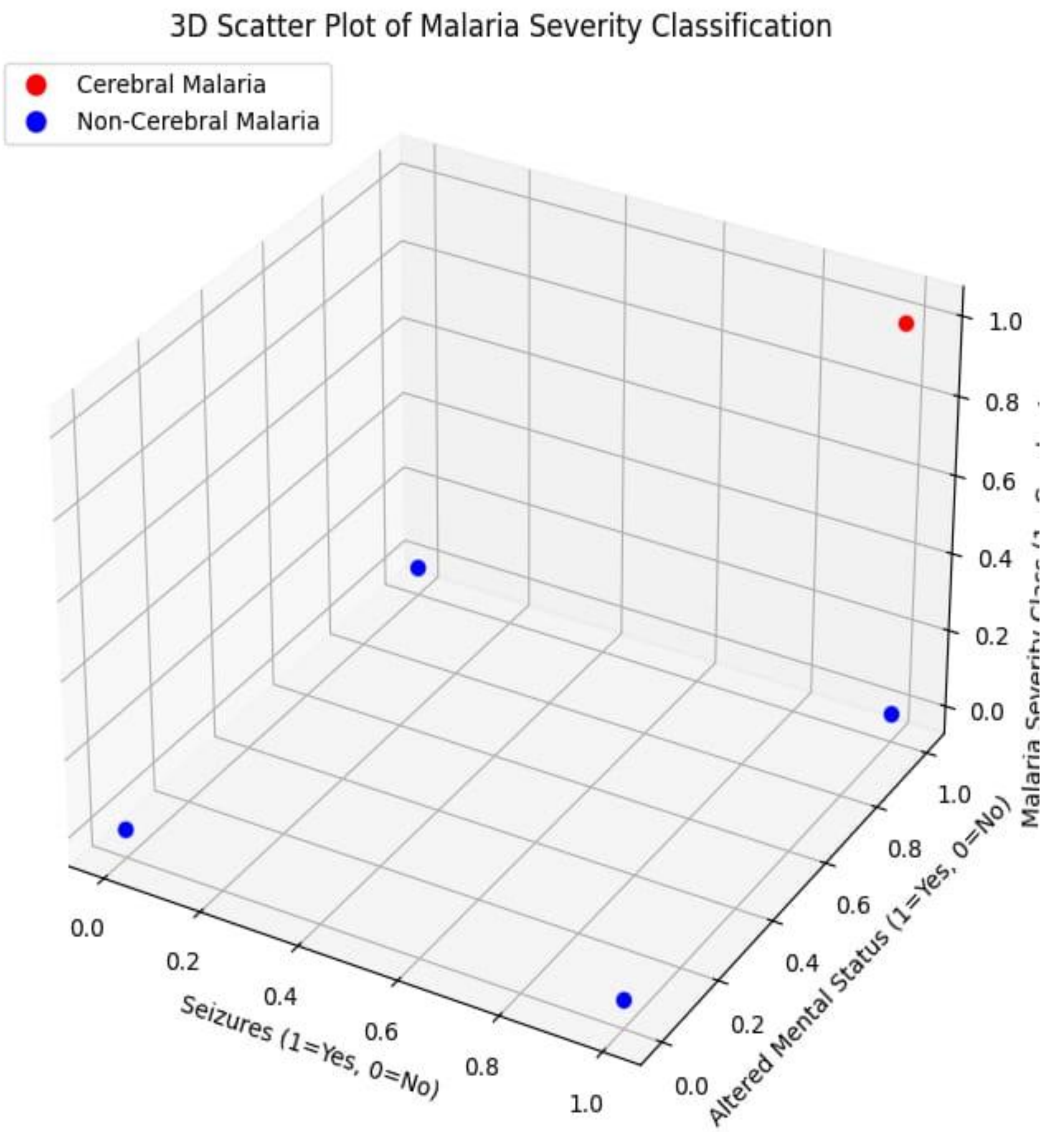

Figure 6: Seizures, Altered Mental Status and Malaria Severity Class

Seizures and Altered Mental Status are highly important features in the classification of malaria severity, especially in identifying cerebral malaria cases. Both symptoms are closely associated with the most severe neurological form of malaria, which affects the brain and central nervous system. The occurrence of seizures is often an early and critical indicator of cerebral involvement, reflecting abnormal electrical activity in the brain and significant neurological disruption. In the same way, Altered Mental Status manifesting as confusion, disorientation, reduced awareness, or even loss of consciousness further signals serious impairment of brain function. When these two symptoms appear together, they strongly increase the likelihood of a case being classified as cerebral malaria. Their combined presence provides a clear distinction between severe and non-severe cases, since non-cerebral malaria typically does not present with such pronounced neurological symptoms. This makes them essential variables in the classification process used by the model. In the context of the study, Seizures and Altered Mental Status contribute significantly to the decision-making capability of the classification system. They help the model to correctly identify high-risk cases that require urgent medical attention, thereby improving the overall reliability and clinical relevance of the predictions. Because of their strong association with disease severity, these features are given considerable importance during model training and evaluation. The inclusion of these symptoms not only improves classification accuracy but also enhances understanding of how malaria affects the nervous system. This supports better clinical decision-making, more timely intervention, and improved allocation of healthcare resources for patients presenting with severe malaria conditions.

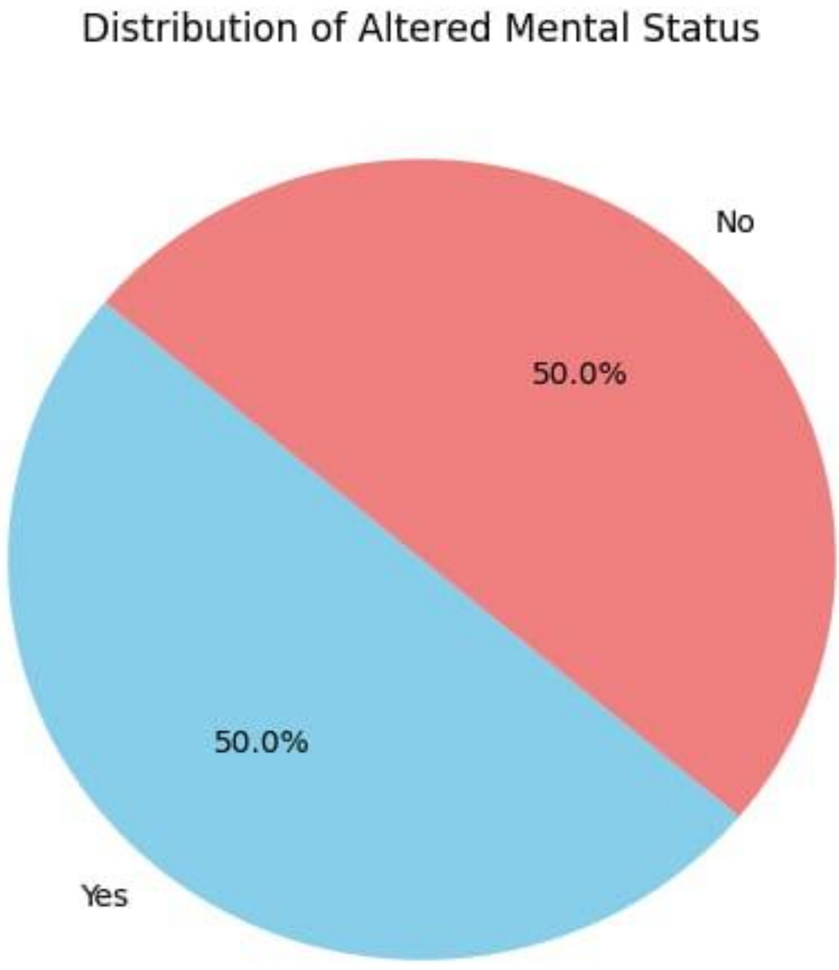

Figure 7: Altered Mental Status

Altered Mental Status (AMS) describes a notable change in an individual's cognitive functioning, particularly affecting awareness, orientation, attention, and overall level of consciousness. It may present in different forms, including confusion, disorientation, difficulty concentrating, memory impairment, and in more severe situations, partial or complete loss of consciousness. In cases of cerebral malaria, AMS is considered a key clinical sign because it strongly suggests involvement of the brain and central nervous system, which defines the severe form of the disease. The presence of AMS in a malaria patient is often an indication that the infection has progressed beyond typical symptoms and may have affected neurological functions. This makes it a critical marker for identifying cerebral malaria, as opposed to non-cerebral malaria, which does not usually present with significant cognitive or neurological disturbances. Its detection therefore becomes essential in separating severe cases from mild or uncomplicated ones. Within the classification process, Altered Mental Status serves as a highly influential feature in improving diagnostic accuracy. It enables the model to recognize patients who are at higher risk and who require urgent and intensive medical attention. By capturing this symptom, the classifier is better positioned to distinguish cerebral malaria cases from non-cerebral cases with greater reliability. In practical terms, incorporating AMS into the model helps ensure that high-risk patients are identified early, allowing for timely medical intervention. This is important in preventing severe complications such as permanent neurological damage or death. As a result, AMS plays a crucial role in supporting accurate classification and improving clinical decision-making in the management of malaria severity.

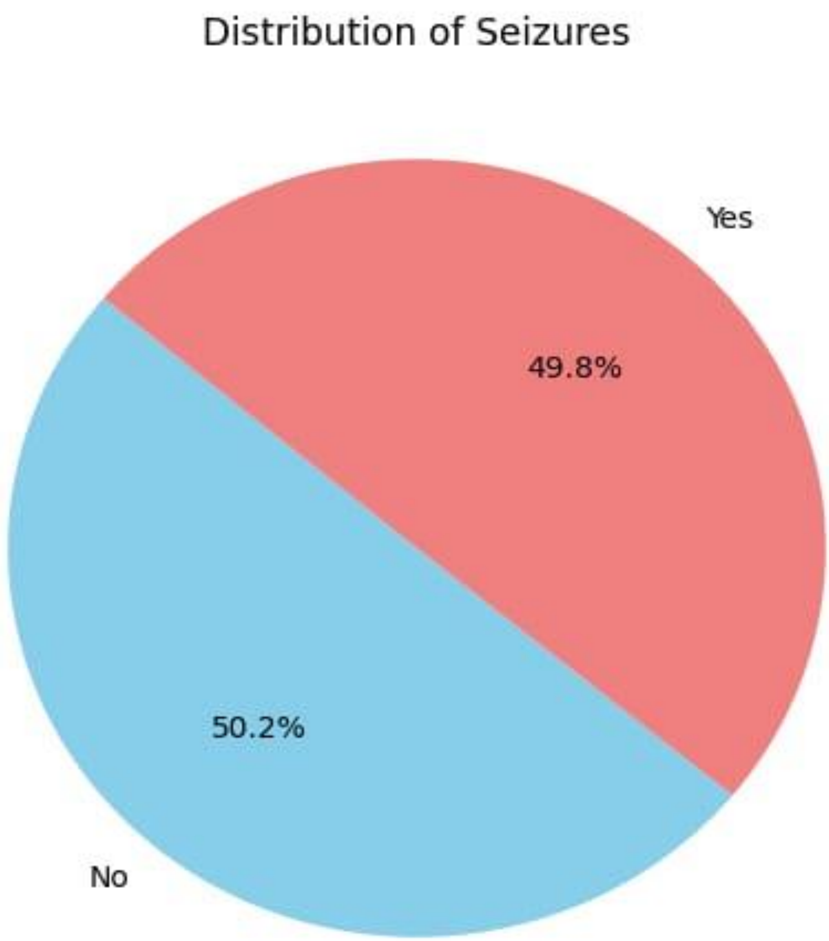

Figure 8: Seizures

Seizures are a prominent neurological symptom associated with cerebral malaria and are a strong indicator of severe brain involvement. In the dataset used for this study, approximately 49.8% of the population experienced seizures. A seizure is defined as a sudden and abnormal electrical disturbance in the brain that can result in uncontrolled body movements, loss of consciousness, and convulsions. In malaria patients, seizures typically occur when the parasite invades the central nervous system, leading to inflammation, disruption of normal neural activity, and impaired brain function. When seizures are observed in malaria cases, they often signify a dangerous progression of the disease, commonly linked to cerebral malaria, which represents the most severe and potentially fatal form of the infection if not treated promptly. Because of this strong association, seizures are regarded as a key clinical marker for identifying cerebral malaria and distinguishing it from less severe, non-cerebral cases. In the classification framework, the presence of seizures plays a significant role in determining malaria severity. It strongly influences the model's ability to correctly identify cases where the infection has progressed to a critical stage. The occurrence of seizures increases the likelihood that a patient is classified under cerebral malaria, making it an essential feature in the decision-making process of the model. By incorporating seizures as a core predictor, the system ensures that high-risk patients are quickly identified and appropriately categorized. When combined with other neurological indicators such as Altered Mental Status, seizures support more accurate classification and enable timely medical intervention. This early detection is vital in reducing the risk of long-term neurological complications and preventing mortality associated with severe malaria cases.

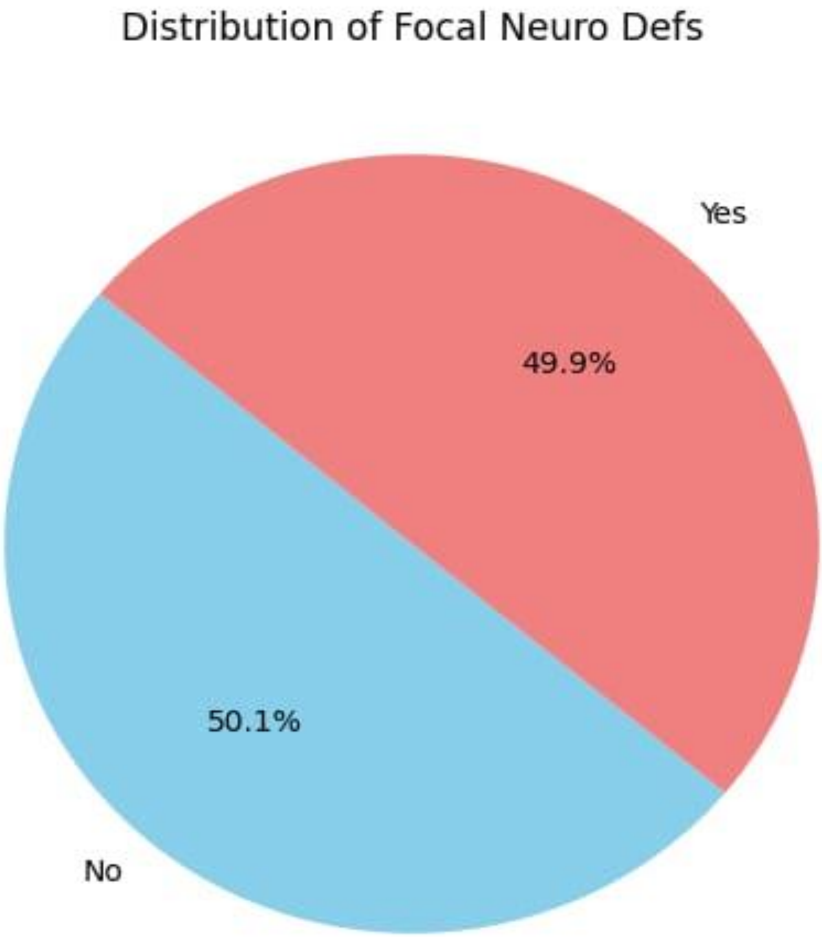

Figure 9: Focal Neuro Defs

Focal neurological deficits (Focal Neuro Defs) refer to impairments that affect specific, localized regions of brain function, depending on the particular area of the brain that has been affected. In the dataset used for this study, about 49.9% of the population presented with focal neurological deficits. These deficits may appear as weakness or paralysis in certain parts of the body, speech difficulties, visual disturbances, or altered sensory perception. In the context of cerebral malaria, focal neurological deficits occur as a result of the brain's inflammatory response to malaria infection, which leads to swelling, reduced blood flow, and disruption of normal neural activity in specific brain regions. These abnormalities are important clinical signs because they indicate direct neurological damage, which is a defining characteristic of severe malaria. If not promptly managed, such damage can result in long-term or permanent neurological complications. Focal neurological deficits are therefore key indicators used in distinguishing cerebral malaria from non-cerebral malaria. Their presence typically suggests that the infection has progressed to a more severe stage, requiring urgent and intensive medical intervention. In the classification process, these deficits significantly contribute to the model's ability to correctly identify severe malaria cases. By incorporating focal neurological deficits as a core feature, the system improves its capacity to detect cerebral malaria with higher accuracy. This allows for better prioritization of critical cases and ensures that patients receive timely and appropriate care. The inclusion of this feature strengthens the model's ability to assess disease severity and supports healthcare providers in reducing the risk of long-term neurological damage.

Conclusion

This research successfully tackled the challenge of classifying malaria severity by applying machine learning techniques, with particular emphasis on the Random Forest algorithm to differentiate between cerebral and non-cerebral malaria cases. The model was trained using a well-structured dataset that combined key clinical symptoms, socio-economic indicators, and demographic variables, enabling a more comprehensive and realistic prediction of malaria severity. Important clinical features such as seizures, altered mental status, headache, vomiting, and focal neurological deficits were used to guide the learning process. Based on these inputs, the model achieved an accuracy of 97%, demonstrating strong performance and reliability in distinguishing between the two classes. This high level of accuracy reflects the model's potential to support clinical decision-making and improve early detection of cerebral malaria, which is often associated with severe complications and higher mortality rates. Beyond clinical classification, the study also highlights the broader socio-economic implications of malaria, particularly in relation to household well-being and caregiver responsibilities. Cerebral malaria, in particular, places a heavy burden on families, contributing to financial stress, loss of household income, and increased caregiving demands. These challenges often extend the impact of the disease beyond the patient, affecting the emotional and economic stability of entire households. Caregivers frequently experience significant physical exhaustion, emotional strain, and financial pressure, which can reduce their ability to provide consistent and effective care. The findings suggest that malaria severity cannot be fully understood in isolation from its social and economic context. Instead, it reflects a combination of medical and household-level factors that interact to worsen outcomes. This creates a reinforcing cycle where health challenges and economic hardship intensify each other, particularly in vulnerable communities. The results of this study support the need for an integrated approach that combines accurate clinical classification with consideration of socio-economic conditions. Such an approach can improve healthcare delivery, support better allocation of medical resources, and contribute to reducing the overall burden of malaria on affected individuals and their families.

References

1. Ebong, O., Edet, A., Uwah, A., & Udoetor, N. (2024). Comprehensive Impact Assessment of Intrusion Detection and Mitigation Strategies Using Support Vector Machine Classification. *Research Journal of Pure Science and Technology*, *7*(2), 50-69.
2. Edet, A. E., & Ansa, G. O. (2023). Machine learning enabled system for intelligent classification of host-based intrusion severity. *Global Journal of Engineering and Technology Advances*, *16*(03), 041-050.
3. Edet, A., Chukwuemeke, O., Godwin, H., & Eyo, E. Intelligent Framework for Classification of Influenza Severity for Respiratory Disease Management. Research Journal of Pure Science and Technology, 8, (6), 29-46.
4. Edet, A., Ekong, B., & Attih, I. (2024). Machine Learning Enabled System for Health Impact Assessment of Soft Drink Consumption Using Ensemble Learning Technique. *International Journal Of Computer Science And Mathematical Theory*, *10*(1), 79-101.
5. Edet, A., Godwin, H., Chukwuemeke, O., & Hanson, Y. (2025). ECONOMIC AND COST-BENEFIT ANALYSIS OF DIABETES TREATMENT: CORRELATING TREATMENT COST WITH CLINICAL OUTCOMES USING MULTI-REGRESSION APPROACH. International Journal of General Scientific Research and Application, 1, (1), 1-19.
6. Edet, A., Inyang, S., Umoren, I., & Etuk, U. E. (2024). Machine Learning Approach for Classification of Cyber Threats Actors in Web Region. *Journal of Technology and Informatics (JoTI)*, *6*(1), 70-77.
7. Edet, A., Johnson, N., Godwin, H., Mregbe, J., and Etim, H. (2025). analytical framework for investigating privacy preserving techniques in iots ecosystem. International Journal of General Scientific Research and Application (IJGSRA), Vol. 1, No. 4
8. Edet, A., Silas, A., Ekaetor, E., Ebong, O., Isaac, E., & Udoetor, N. (2024). Data-Driven Framework for Classification and Management of Start-Up Risk For High Investment Returns. *Advanced Journal of Science, Technology and Engineering*, *4*(2), 81-102.
9. Edet, A., Udonna, U., Attih, I., & Uwah, A. (2024). Security Framework for Detection of Denial of Service (DoS) Attack on Virtual Private Networks for Efficient Data Transmission. *Research Journal of Pure Science and Technology*, *7*(1), 71-81.
10. Ekong, A., Ekong B., and Edet A. (2022). Supervised machine learning model for effective classification of patients with covid-19 symptoms based on bayesian belief network. *Researchers Journal of Science and Technology*, *2*(1), 27-33.
11. Ekong, A., James, G., Ekpe, G., Edet, A., & Dominic, E. A. (2024). Model For The Classification Of Bladder State Based On Bayesian Network. *International Journal of Engineering and Artificial Intelligence*, *5*(2), 33-47.
12. Ekong, A., James, G., Godwin Ekpe, I. J., Ekong, O., Ekong, B., Edet, A., & Okon, E. (2025). Support Vector Machine-Based Model for the Classification of Effective Antibiotic Combination for Pediatrics: A Comparative Analysis of Pre-COVID-19 and COVID-19 Era. *European Journal of Science, Innovation and Technology*, *5*(3), 190-210.
13. Ekong, B., Edet, A., Udonna, U., Uwah, A., & Udoetor, N. (2024). Machine Learning Model for Adverse Drug Reaction Detection Based on Naive Bayes and XGBoost Algorithm. *British Journal of Computer, Networking and Information Technology*, *7*(2), 97-114.
14. Ekong, B., Ekong, O., Silas, A., Edet, A. E., & William, B. (2023). Machine Learning Approach for Classification of Sickle Cell Anemia in Teenagers Based on Bayesian Network. *Journal of Information Systems and Informatics*, *5*(4), 1793-1808.
15. Grignaffini, F., Simeoni, P., Alisi, A., & Frezza, F. (2024). Computer-Aided Diagnosis Systems for Automatic Malaria Parasite Detection and Classification: A Systematic Review. *Electronics*, *13*(16), 3174.
16. Jdey, I., Hcini, G., & Ltifi, H. (2024). Deep learning and machine learning for Malaria detection: overview,

challenges and future directions. *International Journal of Information Technology & Decision Making*, *23*(05), 1745-1776.

17. Siłka, W., Wieczorek, M., Siłka, J., & Woźniak, M. (2023). Malaria detection using advanced deep learning architecture. *Sensors*, *23*(3), 1501.
18. Udoetor, N., Ansa, G., Ekong, A., & Edet, A. (2024). Intelligent System for Detection of Copyright-Protected Data for Enhanced Data Security. *Technology*, *7*(4), 58-80.
19. Uwah, A., & Edet, A. (2024). Customized Web Application for Addressing Language Model Misalignment through Reinforcement Learning from Human Feedback. *World Journal of Innovation And Modern Technology*, *8*(1), 62-71.
20. Yoon, J., Jang, W. S., Nam, J., Mihn, D. C., & Lim, C. S. (2021). An automated microscopic malaria parasite detection system using digital image analysis. *Diagnostics*, *11*(3), 527.