Utilizing Explainable AI and Biosensors for Clinical Diagnosis of Infectious Vector-Borne Diseases

Thavavel Vaiyapuri
College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Saudi Arabia
t.thangam@psau.edu.sa (corresponding author)

Received: 17 September 2024 | Revised: 13 October 2024 | Accepted: 16 October March 2024

 DOI: https://doi.org/10.48084/etasr.9026

ABSTRACT

Infectious Diseases (ID) are a significant global threat due to their epidemic nature and substantial impact on mortality rates. COVID-19 has proven this assertion by wreaking havoc on human wellness and healthcare resources. This has underscored the need for early ID diagnosis to restrict the spread and protect human lives. Recently, Artificial Intelligence (AI)-assisted biosensors have shown great potential to assist physicians in making decisions to minimize mortality rates. However, their adoption in clinical practice is still in its infancy, primarily due to the challenges faced by physicians to interpret decisions derived from these black-box systems. The objective of this study is to earn the trust of physicians to promote their acceptance and widespread adoption in healthcare. Against this backdrop, this research is a pioneering effort to investigate not only the diagnostic accuracy of several Machine Learning (ML) algorithms for ID but more specifically how to leverage the benefits of Shapley values to provide valuable insights regarding the contribution of clinical features for early ID diagnosis. This analysis examines four ML algorithms that stem from different theories, such as Random Forest Classifier (RFC), Gradient Boosting Classifier (GBC), Support Vector Classifier (SVC), and Multilayer Perceptron (MLP). The visual analysis results presented for local and global interpretation facilitate the observation of the marginal impact of each clinical feature on a patient-by-patient basis. Therefore, the results of this study are expected to aid practitioners in better evaluating the diagnostic decisions of the ML models developed and boost the use of AI-assisted biosensors for ID diagnoses.

Keywords-biosensors; machine learning; model agnostic methods; early pandemic diagnosis; SHAP framework; global and local explanation

I. INTRODUCTION

The emergence and reemergence of ID have historically posed a significant threat to human society due to their endemic nature and high degree of infection [1]. These phenomena provoke socioeconomic challenges that have widespread effects on the world. This issue is an escalating concern as it puts a substantial number of communities in a vulnerable position. The COVID-19 pandemic has confirmed the validity of this assertion, not only by spreading rapidly around the world with high fatality rates but also by profoundly disrupting the global economy [2]. In addition, several recent studies have highlighted that the impact of climate change can potentially affect and accelerate the spread of ID [3]. The World Health Organization (WHO) has acknowledged the critical importance of implementing more robust measures to prevent and control ID. The significance of research financing has been underscored in finding better solutions to combat vectors and the diseases they spread [4]. Keeping in mind the stark reality of these factors, it is imperative for researchers and policymakers to remain attentive and reassess the current approaches for predicting and managing new ID and revisit the global frameworks for controlling future pandemic diseases. The existing diagnostic modalities for ID depend predominantly on clinical laboratories, such as polymerase chain reaction, microscopy, culture, and enzyme-linked immunosorbent assay [5]. Most of these technologies are inaccessible to the majority of the global population due to their high cost, high complexity, centralization, and reliance on qualified technicians. Among these detection modalities, biosensors are at the forefront as a powerful alternative modality for timely ID diagnosis due to their rapid response, high sensitivity and specificity, lower cost, and ability for on-site detection [6]. Additionally, these devices have the potential to serve as point-of-care instruments within medical facilities or domestic settings due to their inherent simplicity and cost-effectiveness.

All biosensors, inevitably, exhibit inherent signal noise. Some biosensors have limited lifespans and unstable performance because they rely heavily on antibodies or aptamers as bioreceptors. The commercialization of most modern biosensors is hindered by their limited accuracy and dependability [7]. Therefore, researchers are actively seeking new ways to make biosensors more effective. Fortunately, the

AI revolution over the past decade has narrowed this gap [8]. Integration of AI with biosensors has resulted in the emergence of a novel class known as AI biosensors, instigating a transformative revolution. ML is a significant component of AI and has been utilized as a potent instrument for efficiently analyzing extensive datasets acquired from biosensors and subsequently extracting vital information. ML can offer innovative approaches to address the obstacles encountered by biosensors. Additionally, ML can facilitate the transformation of conventional biosensors into intelligent ones that can autonomously predict the species or concentration of analytes using a decision-making system [8]. The primary focus of this study is the analysis of sensing data using ML techniques. Several studies have explored various ML techniques for diagnosing diseases such as tuberculosis, malaria, and most recently, COVID-19. These models have shown promising results in terms of sensitivity and specificity, contributing to more precise diagnoses and the potential to reduce diagnostic delays [6, 9]. Research has also explored the application of deep learning for image-based diagnostics, particularly in interpreting medical imaging data such as chest X-rays and CT scans for ID detection [10, 11]. Despite these advances, one of the persistent challenges is the "black-box" nature of many ML models, which limits their interpretability and, consequently, their acceptance in clinical practice. Although prior studies have successfully implemented ML for ID diagnosis, there is a critical research gap regarding the application of explainable AI (XAI) techniques. To date, no studies have comprehensively investigated how XAI methods, such as Shapley values, can provide transparent and interpretable insights into the contributions of clinical features for ID diagnosis.

In recent years, XAI has gained traction in various healthcare fields, addressing the critical need for transparency in AI-driven decision-making [2, 12, 13]. Several successful XAI applications have been reported. For instance, in oncology, XAI methods have been employed to explain the predictive factors influencing cancer recurrence, helping physicians tailor personalized treatment plans [14]. Similarly, in cardiovascular disease, XAI models have provided transparency on the contribution of risk factors, such as cholesterol levels and blood pressure, to enhance trust and facilitate clinical decision-making [15]. In diabetic retinopathy detection, XAI has been used to visualize the areas of retinal images most relevant to disease classification, improving the interpretability of complex deep-learning models [16]. Impressed by the success of these studies in enhancing model transparency and clinical trust, this study aims to fill the literature gap by investigating the application of explainable ML techniques for the diagnosis of ID, an area where such approaches have yet to be comprehensively explored.

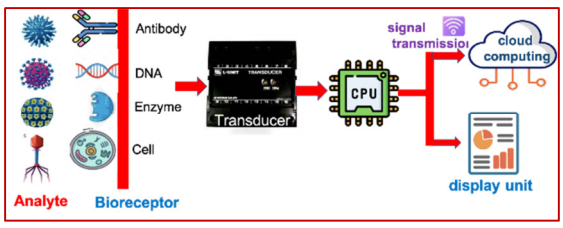


Fig. 1. A biosensor device schematic flow control within its modules.

II. MATERIALS AND METHODS

A. Biosensors and AI

The rapid growth of the global population and the rise in chronic lifestyle diseases have created an urgent need for timely therapeutic interventions. Experts highlight the critical role of biosensors in clinical management, noting their prompt responsiveness and specificity, which enable early diagnosis and treatment [5]. Advances in biomarker identification and technologies such as nanotechnology and microfluids have further expanded the potential of biosensors to improve healthcare. Since their introduction half a century ago, biosensors have revolutionized the treatment of various diseases. A biosensor is an analytical device designed to detect and measure biological changes, converting them into electrical signals for analysis [17]. It consists of several key components that work together to achieve accurate detection and diagnosis, as shown in Figure 1. The analyte is the target substance, such as glucose, urea, or drugs, whose concentration is measured. The bioreceptor is a molecule that specifically recognizes the analyte, including antibodies, enzymes, or DNA, which generate a measurable signal upon interaction. The transducer converts the biorecognition event into an electrical or optical signal, which is then processed by the processor. This unit amplifies and digitizes the signal for easier interpretation. Data are often transmitted through wireless communication systems, such as Bluetooth or 5G, to external devices for further analysis. Finally, the display unit presents the processed results, allowing real-time monitoring. Together, these components enable biosensors to provide precise, non-invasive, and timely detection of various biomarkers for disease diagnosis and monitoring [17]. Biosensors and AI are revolutionizing disease diagnosis by offering non-invasive, real-time monitoring of biomarkers and advanced data analysis. Biosensors track biomarkers with precision, while AI processes the data to identify disease patterns. Their integration enhances timely disease detection, improving treatment effectiveness and potentially saving lives.

B. Explainable AI (XAI)

Recent advances in AI and ML hold great promise for enhancing healthcare by improving clinical decision-making, reducing errors, and expanding access [18]. However, the adoption of ML in clinical practice remains limited due to the opaque nature of "black-box" models, which are difficult to understand and interpret. This lack of transparency, where models are not inherently understandable, creates challenges for both healthcare providers and developers [12, 19]. To overcome these issues, XAI has emerged as a key solution, aiming to enhance the interpretability of AI models by providing explanations that are understandable to humans, allowing users to grasp how decisions are made in these complex systems.

XAI focuses on improving explainability, the ability to assess a model's internal workings by human standards. By transforming black-box models into more transparent "glass-box" systems, XAI highlights the importance of specific features and uncovers the relationships between variables, making the decision-making process clearer [2, 15, 20]. Post-hoc explanation techniques offer instance-specific insights

without needing full knowledge of the model's internal mechanics. This combination of transparency, interpretability, and explainability is crucial to building trust in AI-driven healthcare, ensuring that both physicians and patients can rely on the accuracy and clarity of diagnoses and predictions generated by these systems.

C. Shapley Additive Explanation (SHAP)

Several techniques have been devised to assess the significance of variables within a model to enhance interpretability [21]. These techniques encompass correlation coefficients, hypothesis testing, and variance-based approaches. The SHAP approach has garnered significant attention among many methodologies due to its recent incorporation of game theory principles [16, 22].

The SHAP framework presents a comprehensive approach to diagnosis interpretation. The process involves calculating the contribution of each feature to the prediction to explain the prediction of a given instance X. Figure 2 provides a comprehensive overview of the process for utilizing SHAP to interpret the predictions generated by any given model. The SHAP method can be utilized in both global and local contexts. In the global context, it may provide explanations of the average impact of features on outputs. On the other hand, in the local context, it can estimate the contribution made by each feature toward specific predictions. The application of Shapley values provides two fundamental advantages. First, Shapley values can be computed for a wide range of models, extending beyond simple linear models. Additionally, each record possesses its own unique set of Shapley values [21].

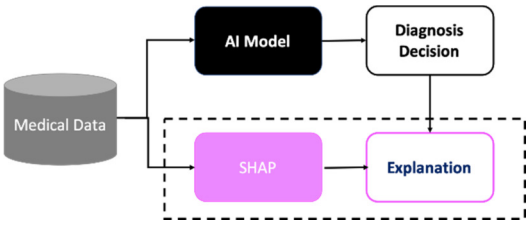


Fig. 2. SHAP integration for model explainability.

Given a predictive model f that uses a subset S of features from the total N input features to predict an output, Shapley values can be used to determine the contribution of each feature Θ_0 to the model's output as follows [22]:

$$\Theta_i = \sum_{S \subseteq N(i)} \frac{|S|!(N-|S|-1)!}{N!} [f(\cup S\{i\} - f(S)] \quad (1)$$

The above-computed values Θ_i can then be used to provide an explanation for an instance X using the additive feature attribution as follows:

$$g(S) = \Theta_0 + \sum_{i=1}^{N} S_i \, \Theta_i \quad (2)$$

Here g is the linear explanation model defined to replace the original model f and Θ_0 represents no inputs scenario. The SHAP method was utilized in this investigation to provide an interpretation of the model that exhibited the highest performance. Three distinct SHAP plot variants, namely summary, Force, and Waterfall plots, were employed to visually represent the impact and significance of each input parameter on the associated output. These visualizations have the potential to yield valuable insights regarding the behavior of the model, facilitating an improved understanding of the correlations between the independent input clinical features and the predicted diagnosis results.

```
Algorithm 1. Interpretability analysis of
XAI models using SHAP
Input:
  Biosensor dataset (BD): X(features),
  y (target labels)
Output:
  Tuned hyperparameters for RFC, GBC, SVC,
  and MLP, SHAP, summary and waterfall
  plots for interpretability comparison
Procedure CompareXAI {BD, ML (RFC, GBC,
  SVC, MLP)}
1: Load the biosensor dataset and split
   into (train, test)
2: Initialize hyperparameters for ML
   models as in Table I
3: for each model M in {RFC, GBC, SVC,
   MLP} do
     Apply BO and max CV accuracy for M
     Train M with hyperparameters
     optimized using BO
   end
4: Compute SHAP values on the test set
     Use shap.TreeExplainer for RFC, GBC
     Use shap.KernelExplainer for SVC, MLP
5: For interpretability analysis
     Generate summary plots for global
     explanations
     Generate waterfall plots for global
     explanations
6: Compare models on accuracy and
   interpretability
End procedure
```

D. Proposed Conceptual Framework

Figure 3 shows the conceptual framework of the developed diagnosis model, highlighting the complete process involved in developing the prediction model. In general, the data collected and preprocessed by the biosensor are stored in nearby servers or the cloud. Then, the candidate ML model is trained with the preprocessed data using Algorithm 1. Next, to better assist clinicians in making informed diagnostic decisions, the model is designed to integrate both global and local explanations. The developed model utilizes Shapley values to boost both model explainability and feature selection, which have been recognized to enhance the acceptability of medical diagnoses. The presence of irrelevant and redundant features within medical datasets has posed major barriers to current AI-based models, resulting in compromised diagnostic accuracy. The primary rationale behind this is that the inclusion of irrelevant and redundant features significantly increases the likelihood of overfitting and amplifies the complexity of the computational

process. Consequently, this model incorporates feature selection into the primary prediction phase to remove features that are deemed unnecessary or irrelevant. Once the model has been refined using the selected features, the diagnosis decision with a model-provided explanation can be shared with medical professionals for further verification.

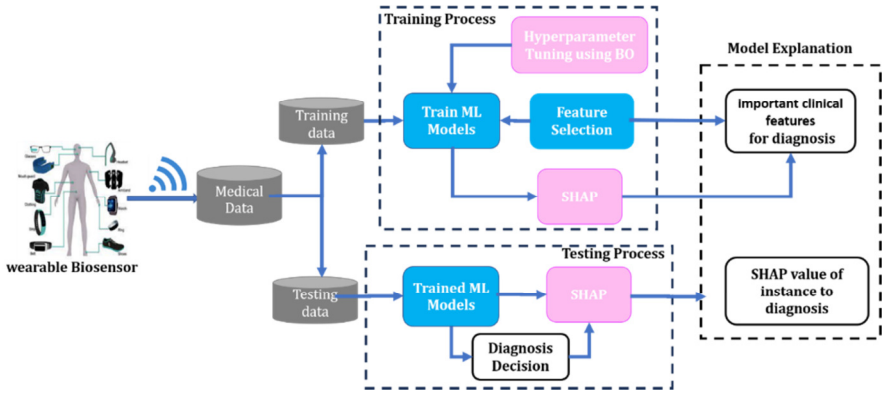


Fig. 3. Evaluation framework adopted for Interpretability comparison of ML models for ID diagnosis.

E. Model Selection

As there is no single optimal ML classifier method that applies to all applications, it is advisable to evaluate a variety of classifiers based on their performance, complexity, characteristics, and previous literature. In this context, four distinct ML algorithms, each rooted in different theoretical frameworks, were examined for their potential to improve diagnostic and interpretative capabilities in ID. These include Random Forest (RFC) from ensemble learning, Gradient Boosting (GBC), Support Vector machine derived from statistical learning theory (SVC), and neural networks (MLP). In conjunction with this, the impact of feature selection on diagnostic accuracy is examined, utilizing the advantages of their interpretative capabilities [9, 23, 24]. Moreover, the selection of these four ML algorithms is based on their demonstrated efficacy in medical diagnosis, as evidenced in the earlier literature [23, 25, 26].

TABLE I. BO OPTIMIZED HYPERPARAMETERS OF ML MODELS

ML model	Hyperparameter	Search Range	Optimal value
RFC	Max-depth	{3, 6, 9}	5
	n-estimators	{25, 50, 100}	87
GBC	Max-depth	{5, 15, 20}	15
	n-estimators	{25, 50, 100}	50
SVC	Kernel	{'rgf', 'linear'}	'rbf'
	C	{5, 10, 15, 20}	11
	Gamma	{5, 10, 15, 20}	10
MLP	Activation function	{'logistic', 'tanh','relu'}	'relu'
	Learning rate	{'constant', 'invscaling', 'adaptive'}	'invscaling'

F. Bayesian Optimization-based Hyperparameter Tuning

Hyperparameters are values that can significantly influence the learning process. When appropriately adjusted, these parameters can enhance both the accuracy and the generalizability of the model [25]. Hyperparameter optimization plays a crucial role in ML algorithms, as it seeks to refine hyperparameters to ensure optimal performance throughout the training process. To ensure equitable comparison, the hyperparameters of all the ML algorithms under investigation were meticulously optimized using Bayesian Optimization (BO). This method is unique in that it achieves the highest scores in the shortest number of iterations, making it an intelligent choice for complex models. In contrast to other hyperparameter tuning methods, BO allows one to choose a broad range for the search space as opposed to a limited number of hyperparameters. The method optimizes the search space at each iteration using the findings of previous iterations. Table I shows the hyperparameters tuned using BO for the ML models developed.

III. EXPERIMENTAL SETUP

A. Dataset Description

The dataset used in this study is a recent, publicly available dataset, making it an ideal choice for investigating and comparing ML models for interpretability in ID diagnosis [27]. It offers a valuable foundation for the application of XAI models in healthcare, particularly for diagnosing vector-borne ID such as dengue. Collected from 77 patients with dengue between 2009 and 2010, the dataset includes detailed information such as age, gender, length of hospital stays, clinical symptoms, laboratory test results, and patient diagnoses. Diagnoses are classified into three severity levels (mild, intermediate, and severe), allowing a comprehensive evaluation of model performance across different stages of the disease. This dataset, obtained from the Department of Microbiology, Universitas Indonesia, offers a practical basis to compare the interpretability of ML models in ID diagnosis.

B. Model Development and Implementation Details

The selected models were implemented using Python, leveraging a range of libraries, including scikit-learn for model building, skopt for hyperparameter tuning, and shap for interpretability assessment. Hyperparameter optimization was performed using the BayesSearchCV method from the skopt library, ensuring that each model was fine-tuned for optimal

performance [28]. To allow a fair comparison of model interpretability, SHAP values were employed to explain the contribution of individual features to model predictions at both global and local levels. For tree-based models (RFC, GBC), the TreeExplainer from SHAP was utilized, while the KernelExplainer was applied for non-tree models (SVC, MLP). SHAP summary plots were used to illustrate global feature importance, and SHAP waterfall plots provided detailed local explanations for specific instances.

IV. RESULTS AND DISCUSSION

A. Model Performance Analysis

To evaluate the performance of the ML models for ID diagnosis, the dataset was partitioned into training and testing in a 75:25 ratio. For a fair comparison, the same training set was adopted to train all the ML models using k-fold cross-validation. After training, all models were validated on the unseen test set. The investigation focuses on evaluating the effectiveness of the models for diagnosing ID across three aspects, as follows.

1) Quantitative Analysis

Table II presents the performance metrics derived from the confusion matrix for the four ML models, both with and without hyperparameter optimization. These results are also shown in Figure 4. The results clearly show that the optimized models outperformed the non-optimized ones across all performance measures. This improvement can be attributed to the role of hyperparameter optimization in fine-tuning key parameters such as learning rate, number of estimators, and maximum depth, which directly influence the model's ability to learn from the data. In non-optimized models, default settings may not capture the complexity of the dataset, leading to suboptimal performance.

The slight performance advantage observed in ensemble methods (RFC and GBC) compared to MLP and SVC can be justified by the inherent ability of ensemble techniques to reduce overfitting and improve robustness by combining multiple decision trees or learners. This allows them to better capture complex relationships in the data, making them more effective. Despite this, MLP and SVC also performed well, but the results indicate that ensemble learning approaches may be more suitable for this type of biosensor data.

TABLE II. QUANTITATIVE PERFORMANCE MEASURE FOR THE DEVELOPED ID DIAGNOSTIC MODELS

Selected XAI models	Diagnostic performance without hyperparameter optimization				Diagnostic performance with hyperparameter optimization			
	Accuracy	Sensitivity	Specificity	F1-Score	Accuracy	Sensitivity	Specificity	F1-Score
RFC	0.736	0.872	0.609	0.711	0.847	0.992	0.726	0.841
GBC	0.689	0.826	0.624	0.694	0.806	1.0	0.673	0.809
SVC	0.544	0.655	0.489	0.551	0.753	0.987	0.560	0.756
MLP	0.413	0.445	0.406	0.436	0.659	0.626	0.514	0.723

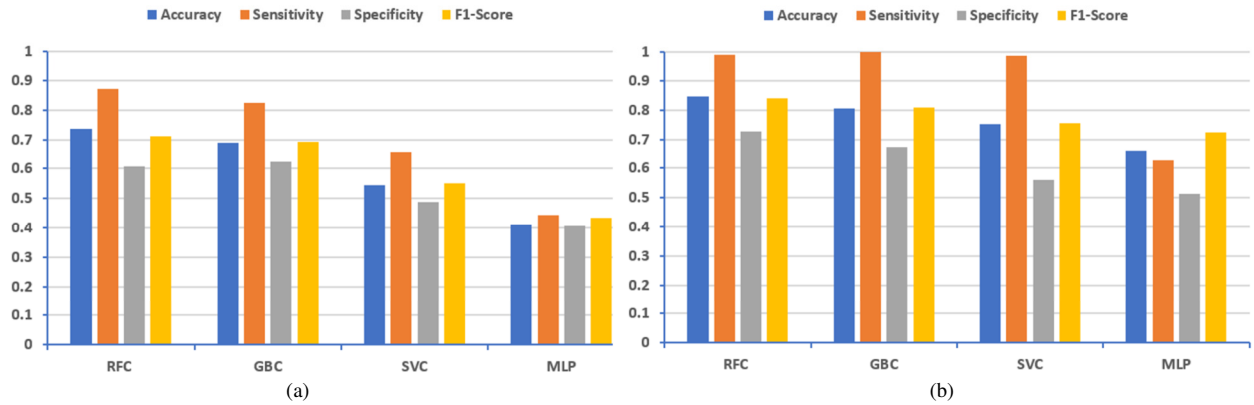
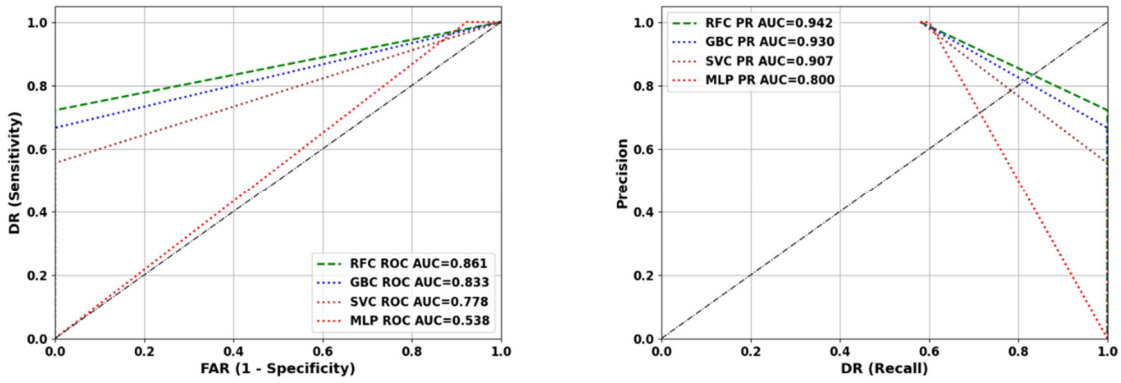


Fig. 4. Performance analysis graphs for the developed ID diagnostic models: (a) without hyperparameter optimization, (b) with hyperparameter optimization.

2) Sensitivity-Specificity Curve (ROC curve)

The ROC curve is crucial to understanding the trade-offs between false positives and false negatives, offering a comprehensive view of a model's performance across various decision boundaries. This is particularly important in the medical domain, where accurate classification decisions are essential. To evaluate the four ML models developed in this study, their performance was compared using ROC curves, as shown in Figure 5(a). The curves confirm the superior performance of the ensemble models, which outperformed the other models in terms of classification accuracy. Additionally, the Area Under the ROC curve (AUC), displayed in the figure's legend, quantifies the generalization ability of each model. However, the overall results suggest that the diagnostic performance of all models was affected by the limited size of the training dataset. Despite this, in real-world applications, the models are expected to perform better when provided with larger datasets, as training sample constraints will no longer be a limitation. In particular, the ROC results are consistent with the performance measures presented in Table II, further reinforcing the conclusion that the ensemble models are more effective in this context.



(a) (b)

Fig. 5. ROC (a) and Precision-Recall (b) curve analysis for the ID diagnostic models developed.

3) Precision-Recall Curve (PR Curve)

The PR curve is particularly valuable when evaluating ML models on imbalanced datasets, where the positive class is the primary focus. Unlike the ROC curve, the PR curve offers a more insightful perspective in such cases, as it emphasizes a model's ability to balance precision (minimizing false positives) and recall (maximizing true positives). This makes the PR curve especially useful in medical diagnoses, where datasets are often skewed, with a limited number of infected patient samples compared to a larger number of normal cases. Together, the PR and ROC curves provide a more comprehensive evaluation of model performance.

Figure 5(b) provides a comparison of the four ML models using the PR curve, offering deeper insights into how well the models handle the imbalanced dataset. The curve shows that, despite the dataset's imbalance, all models maintain consistent levels of precision and recall. This consistency indicates that the models are adept at identifying positive cases without generating an excessive number of false positives. The ensemble models, in particular, stand out as the superior choice, demonstrating better PR trade-offs compared to the MLP and SVC models. Their higher precision and recall values suggest that ensemble models, such as RFC and GBC, are more robust in capturing the complex patterns within the data, making them well-suited for building effective ID diagnosis models. This superior performance is likely due to ensemble models' ability to aggregate multiple learners, enhancing their discriminative power even in challenging scenarios such as imbalanced medical datasets.

B. Model Interpretability Analysis

It is essential to determine the primary clinical characteristics that influence diagnostic decision-making and establish their associations with patient infection to compare the diagnostic efficacy of the ML models developed. By utilizing the SHAP framework, it is possible to effectively attain the aforementioned objectives. These models offer interpretability, allowing for a clear understanding of the relationship between input features and output objectives.

1) Shapley Global Explanation

The evaluation of global interpretation in the ML models involves a qualitative analysis of the SHAP summary plot, which effectively illustrates the relative significance assigned by the model to different clinical features throughout the process of diagnosing ID. Shapley values explain the extent to which different clinical features of patients contribute, either positively or negatively, to the diagnosis of ID. Figure 6 visually represents this information. Upon careful examination of the results, it is evident that the primary two clinical features remain consistent across the two ensemble models, affirming their substantial contribution to ID diagnosis. In addition, Figure 6(a) reveals that the top three clinical features, namely hemoglobin, hematocrit, and platelet count, have the most variable Shapley values. Here, the low values of these clinical features correspond to patients with low red blood cells and are awarded positive Shapley values, which have a greater impact on the diagnosis probability of ID.

In accordance with [19, 27], the ensemble ML models developed have identified these three clinical features as the most essential on average for ID diagnosis. As a result, the RFC model is more likely to consider patient infection as values of these features decrease. Therefore, the use of SHAP facilitates medical professionals in examining the role of features within the developed model, as well as in the decision-making process.

2) Shapley Local Explanation

As stated previously, the use of local explanations can provide valuable support to healthcare providers in their decision-making processes about potentially contaminated patients. However, they play a crucial role even in the context of research. An illustrative example is the comparison between the explanation of a patient correctly identified as infected and a patient falsely identified as infected, thus facilitating the comprehension of factors that could potentially lead to confusion within the models. This study aims to showcase the application of the Shapley force and waterfall plots as a visual tool to interpret local explanations of Shapley values in the context of COVID-19 diagnoses. Column II in Figure 7 depicts the rationales that resulted in an accurate positive diagnosis for a given patient, whereas column I illustrates the justifications for a correct negative diagnosis. The SHAP force plots visually represent the features in a left-to-right manner, where the left side represents positive contributions highlighted in red, and the right side represents negative contributions highlighted in

blue. Similarly, waterfall plots depict features in a vertical layout, with positive contributions displayed at the top and negative contributions displayed at the bottom.

This study's findings align with [19, 27], demonstrating that decreased levels of white blood cells significantly influence the favorable prognosis of this particular patient, whereas, for non-infection, the increased levels of white blood cells and hemoglobin are significantly considered. The individual force and waterfall plots provide insights into the features that the models deem significant in predicting outcomes for specific patients. However, while it is possible to visually perceive the significant features that the models deem vital for certain patients, it is crucial to note that these models are not specifically designed for individual patients. Consequently, only the average risk for patients with comparable risk factors will be reflected in the diagnosis.

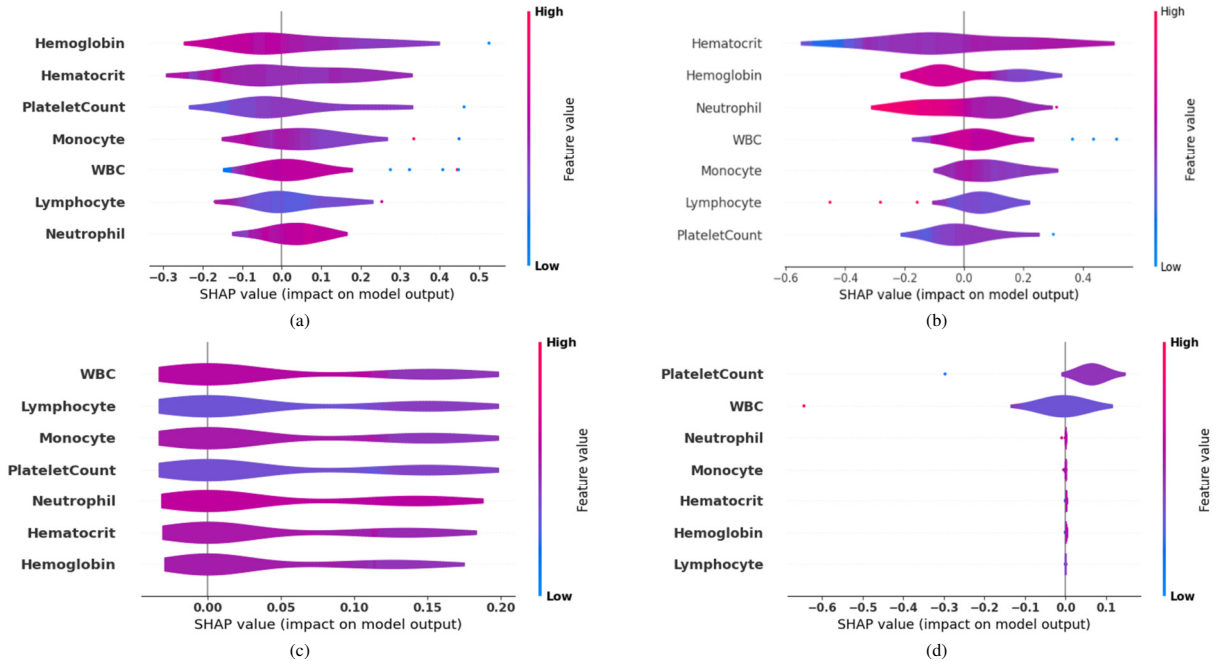


Fig. 6. Shap summary plot for the developed ID diagnostic models: (a) RFC, (b) GBC, (c) SVC, (d) MLP.

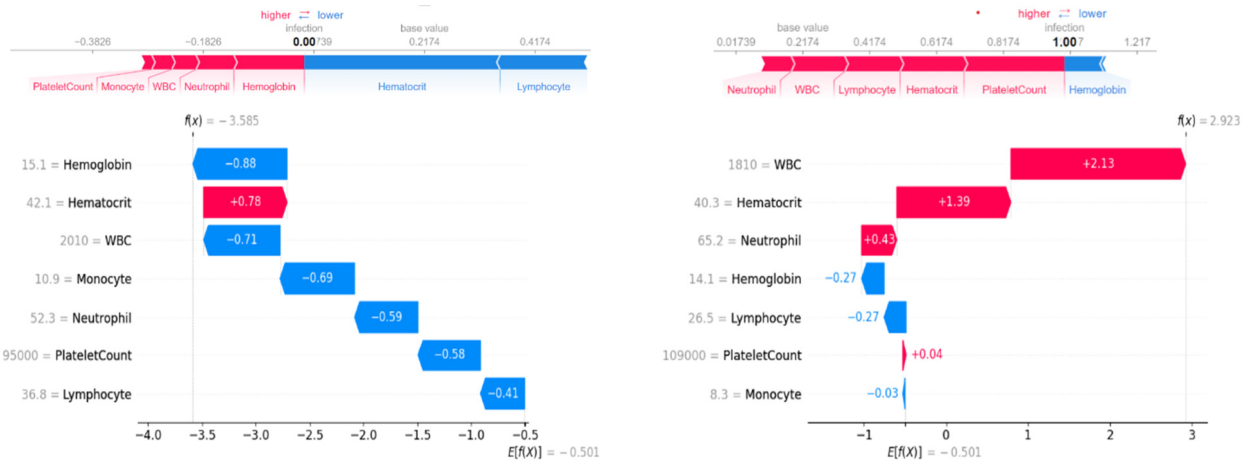


Fig. 7. SHAP force and corresponding waterfall plots for uninfected (column I) and infected patients (column II) of the developed GBC model.

Upon summarizing the findings, it becomes evident that a more effective model regarding AUC does not always imply a more precise model in terms of medical theory, and vice versa. A model may exhibit proficiency in one particular aspect while also demonstrating deficiencies in other areas. An example is the MLP, which had the lowest AUC. However, the MLP model aligns more closely with medical theory in terms of White Blood Corpuscles (WBC) and platelet count, since deviations from normal levels indicate problems with the immune system.

C. Comparison with Related Works

Most of the existing literature comparing the interpretability of ML models for vector-borne ID lacks a comprehensive and in-depth analysis. For example, the study in [29] compares the accuracy and interpretability of tree-based ensemble models for COVID-19 prognosis without hyperparameter tuning, leading to suboptimal model performance. In contrast, the study in [13] provides a more thorough analysis by evaluating both the accuracy and interpretability of tree-based ensembles with hyperparameter tuning on a COVID-19 dataset, successfully identifying key biomarkers critical for prognosis. Similarly, in [19], clinical, laboratory, and radiomics features are integrated to perform a detailed comparison between the SVC and RFC models in terms of performance and interpretability. The study in [12] expands the scope by exploring a variety of ML models, excluding SVC and RFC, to assess their performance and interpretability for COVID-19 prognosis.

Additionally, two recent studies [2, 30] focus on comparing the diagnostic accuracy and interpretability of various ML models but rely exclusively on global explanations, without providing insights into individual predictions and localized feature impacts. From this comparative analysis, it becomes evident that a significant portion of existing research tends to rely on basic interpretability techniques, such as SHAP and LIME, without fully exploring their broader implications for clinical decision-making and real-world applications. Moreover, most recent research has focused on identifying biomarkers for COVID-19, driven by the global urgency of the pandemic, while other critical IDs, such as dengue and malaria, that continue to affect millions worldwide have been overlooked. Notably, none of the existing studies have thoroughly examined the interpretability of different ML model categories on small datasets.

In summary, the combination of experimental findings and comparative analysis with related studies underscores the unique contributions of this research. It highlights both the strengths and limitations of the ML models developed for the early diagnosis of vector-borne ID, demonstrating their performance in terms of accuracy and interpretability. By focusing on both aspects, this study not only enhances diagnostic capabilities and supports physicians in making more informed and precise clinical decisions but also provides valuable guidance for CAD developers. This establishes a clear pathway to select the most appropriate ML models to design efficient CAD tools, tailored to the specific needs of vector-borne ID diagnosis.

V. KEY FINDINGS AND CONTRIBUTIONS

The key contributions of this study are summarized below, emphasizing the novel advancements it brings to the field of ID diagnosis by utilizing XAI techniques to interpret data from AI-assisted biosensors. These contributions address significant gaps in the current literature and provide valuable insights into improving the interpretability of AI-driven models, with the potential to enhance clinical decision-making and foster greater trust in AI tools among healthcare professionals.

- This study sets a new precedent by applying XAI techniques, specifically Shapley values, on small datasets to enhance transparency in AI-assisted ID diagnosis, particularly for the interpretation of biosensor data.
- This research pioneers the assessment of the interpretability of four distinct ML algorithms, RFC, GBC, SVC, and MLP, for the early detection of ID, providing insights through both local and global explanations.
- Analysis of both local and global explanations provides valuable insights into how specific clinical features influence the diagnosis both overall and on a patient-by-patient basis, thus improving the interpretability of the models.
- These findings are expected to assist healthcare professionals in evaluating AI-driven diagnostic tools more effectively and promote the adoption of AI-assisted biosensors in the early diagnosis of IDs.

VI. CONCLUSION

This study investigated the precision and interpretive capacity of ML models in improving the transparency and reliability of predictions made for ID diagnosis. The findings indicate that there is variation in accuracy among the developed ML models. This suggests that organizations should make thoughtful decisions while selecting the most suitable model for each specific problem. However, this work substantiates that ML models, specifically RFC and GBC, are robust and interpretable when confronted with complicated clinical data derived from real-world scenarios. Furthermore, the potential of these models to enhance diagnostic performance is evident even when trained using a small dataset.

The SHAP charts presented enhance the diagnostic models, making them more understandable by revealing their inner operations. Specifically, SHAP force and waterfall plots offer in-depth insights into clinical features that affect ID diagnosis, allowing physicians to quickly comprehend these factors. These results highlight the potential of integrating the SHAP framework to better understand the role of ML in ID diagnosis. Consequently, by establishing the diagnosis and interpretation through the developed models, healthcare providers can promptly identify patients at risk and administer suitable treatment in a promising manner. The results of this study make a valuable contribution to the continuous advancement of ML in the context of data-driven ID diagnosis. Additionally, these findings offer a clear direction for future research in this field. Moving forward, substantial prospects remain for further research using these methods. The ongoing advancement of ML techniques offers an opportunity to enhance the interpretability of predictions in healthcare by integrating supplementary or alternative explainable techniques.

ACKNOWLEDGMENT

The author extends her appreciation to Prince Sattam bin Abdulaziz University for funding this research work through project number PSAU/2024/01/29695.

REFERENCES

[1] R. E. Baker *et al.*, "Infectious disease in an era of global change," *Nature Reviews Microbiology*, vol. 20, no. 4, pp. 193–205, Apr. 2022, https://doi.org/10.1038/s41579-021-00639-z.

[2] S. Solayman, S. A. Aumi, C. S. Mery, M. Mubassir, and R. Khan, "Automatic COVID-19 prediction using explainable machine learning techniques," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 36–46, Jun. 2023, https://doi.org/10.1016/j.ijcce.2023.01.003.

[3] P. Van de Vuurst and L. E. Escobar, "Climate change and infectious disease: a review of evidence and research trends," *Infectious Diseases of Poverty*, vol. 12, no. 1, May 2023, Art. no. 51, https://doi.org/10.1186/s40249-023-01102-2.

[4] B. Chala and F. Hamde, "Emerging and Re-emerging Vector-Borne Infectious Diseases and the Challenges for Control: A Review," *Frontiers in Public Health*, vol. 9, Oct. 2021, https://doi.org/10.3389/fpubh.2021.715759.

[5] M. L. Sin, K. E. Mach, P. K. Wong, and J. C. Liao, "Advances and challenges in biosensor-based diagnosis of infectious diseases," *Expert Review of Molecular Diagnostics*, vol. 14, no. 2, pp. 225–244, Mar. 2014, https://doi.org/10.1586/14737159.2014.888313.

[6] S. A. A. Biabani and N. A. Tayyib, "A Review on the Use of Machine Learning Against the Covid-19 Pandemic," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8039–8044, Feb. 2022, https://doi.org/10.48084/etasr.4628.

[7] S. Jain *et al.*, "Internet of medical things (IoMT)-integrated biosensors for point-of-care testing of infectious diseases," *Biosensors and Bioelectronics*, vol. 179, May 2021, Art. no. 113074, https://doi.org/10.1016/j.bios.2021.113074.

[8] X. Jin, C. Liu, T. Xu, L. Su, and X. Zhang, "Artificial intelligence biosensors: Challenges and prospects," *Biosensors and Bioelectronics*, vol. 165, Oct. 2020, Art. no. 112412, https://doi.org/10.1016/j.bios.2020.112412.

[9] A. A. Theodosiou and R. C. Read, "Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician," *Journal of Infection*, vol. 87, no. 4, pp. 287–294, Oct. 2023, https://doi.org/10.1016/j.jinf.2023.07.006.

[10] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[11] K. R. Bhatele *et al.*, "COVID-19 Detection: A Systematic Review of Machine and Deep Learning-Based Approaches Utilizing Chest X-Rays and CT Scans," *Cognitive Computation*, vol. 16, no. 4, pp. 1889–1926, Jul. 2024, https://doi.org/10.1007/s12559-022-10076-6.

[12] A. Nambiar, H. S, and S. S, "Model-agnostic explainable artificial intelligence tools for severity prediction and symptom analysis on Indian COVID-19 data," *Frontiers in Artificial Intelligence*, vol. 6, Dec. 2023, https://doi.org/10.3389/frai.2023.1272506.

[13] K. Chadaga *et al.*, "Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers," *Scientific Reports*, vol. 14, no. 1, Jan. 2024, Art. no. 1783, https://doi.org/10.1038/s41598-024-52428-2.

[14] L. S. Wyatt, L. M. van Karnenbeek, M. Wijkhuizen, F. Geldof, and B. Dashtbozorg, "Explainable Artificial Intelligence (XAI) for Oncological Ultrasound Image Analysis: A Systematic Review," *Applied Sciences*, vol. 14, no. 18, Jan. 2024, Art. no. 8108, https://doi.org/10.3390/app14188108.

[15] A. Salih *et al.*, "Explainable Artificial Intelligence and Cardiac Imaging: Toward More Interpretable Models," *Circulation: Cardiovascular Imaging*, vol. 16, no. 4, Apr. 2023, Art. no. e014519, https://doi.org/10.1161/CIRCIMAGING.122.014519.

[16] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "Explainable AI for Retinoblastoma Diagnosis: Interpreting Deep Learning Models with LIME and SHAP," *Diagnostics*, vol. 13, no. 11, Jan. 2023, Art. no. 1932, https://doi.org/10.3390/diagnostics13111932.

[17] R. Karunakaran and M. Keskin, "Chapter 11 - Biosensors: components, mechanisms, and applications," in *Analytical Techniques in Biosciences*, C. Egbuna, K. C. Patrick-Iwuanyanwu, M. A. Shah, J. C. Ifemeje, and A. Rasul, Eds. Academic Press, 2022, pp. 179–190.

[18] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of Explainable AI Techniques in Healthcare," *Sensors*, vol. 23, no. 2, Jan. 2023, Art. no. 634, https://doi.org/10.3390/s23020634.

[19] F. Prinzi, C. Militello, N. Scichilone, S. Gaglio, and S. Vitabile, "Explainable Machine-Learning Models for COVID-19 Prognosis Prediction Using Clinical, Laboratory and Radiomic Features," *IEEE Access*, vol. 11, pp. 121492–121510, 2023, https://doi.org/10.1109/ACCESS.2023.3327808.

[20] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, "The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies," *Journal of Biomedical Informatics*, vol. 113, Jan. 2021, Art. no. 103655, https://doi.org/10.1016/j.jbi.2020.103655.

[21] O. O. Bifarin, "Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification," *PLOS ONE*, vol. 18, no. 5, 2023, Art. no. e0284315, https://doi.org/10.1371/journal.pone.0284315.

[22] Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost," *Computers, Environment and Urban Systems*, vol. 96, Sep. 2022, Art. no. 101845, https://doi.org/10.1016/j.compenvurbsys.2022.101845.

[23] N. Peiffer-Smadja *et al.*, "Machine learning for clinical decision support in infectious diseases: a narrative review of current applications," *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 584–595, May 2020, https://doi.org/10.1016/j.cmi.2019.09.009.

[24] V. Thavavel and M. Karthiyayini, "Hybrid Feature Selection Framework for Identification of Alzheimer's Biomarkers," *Indian Journal of Science and Technology*, vol. 11, no. 22, pp. 1–10, Jun. 2018, https://doi.org/10.17485/ijst/2018/v11i22/123310.

[25] A. Binbusayyis, H. Alaskar, T. Vaiyapuri, and M. Dinesh, "An investigation and comparison of machine learning approaches for intrusion detection in IoMT network," *The Journal of Supercomputing*, vol. 78, no. 15, pp. 17403–17422, Oct. 2022, https://doi.org/10.1007/s11227-022-04568-3.

[26] S. R. da S. Neto *et al.*, "Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review," *PLOS Neglected Tropical Diseases*, vol. 16, no. 1, 2022, Art. no. e0010061, https://doi.org/10.1371/journal.pntd.0010061.

[27] P. Silitonga, A. Bustamam, H. Muradi, W. Mangunwardoyo, and B. E. Dewi, "Comparison of Dengue Predictive Models Developed Using Artificial Neural Network and Discriminant Analysis with Small Dataset," *Applied Sciences*, vol. 11, no. 3, Jan. 2021, Art. no. 943, https://doi.org/10.3390/app11030943.

[28] T. Vaiyapuri, "Deep Learning Enabled Autoencoder Architecture for Collaborative Filtering Recommendation in IoT Environment," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 487–503, 2021, https://doi.org/10.32604/cmc.2021.015998.

[29] K. Chadaga *et al.*, "An explainable multi-class decision support framework to predict COVID-19 prognosis utilizing biomarkers," *Cogent Engineering*, vol. 10, no. 2, Dec. 2023, Art. no. 2272361, https://doi.org/10.1080/23311916.2023.2272361.

[30] Y. Fan, M. Liu, and G. Sun, "An interpretable machine learning framework for diagnosis and prognosis of COVID-19," *PLOS ONE*, vol. 18, no. 9, 2023, Art. no. e0291961, https://doi.org/10.1371/journal.pone.0291961.

AUTHORS PROFILE

Thavavel Vaiyapuri (IEEE member) is currently an Associate Professor in the Department of Computer Sciences at the College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University. She is a fellow of HEA, UK, and a DELL EMC2 certified associate in cloud computing and Big data analytics. She has authored/edited books with renowned publishers such as CRC Press and holds patents to her credit in the field of cybersecurity. With nearly 25 years of research and teaching experience, she has more than 100 research publications in impacted journals and international conferences. Her research interests include data science, security, computer vision, and high-performance computing.