Available online at
https://mijuitm.com.my

e-ISSN: 2735-0703

Mathematical Sciences and Informatics Journal 7(1) 2026, 81 – 92.

Mathematical Sciences and Informatics Journal

An Explainable AI Framework for Predicting Respiratory Tract Infection using Symptoms and Multidimensional Risk Factors in Resource-scarce Settings

Moses Ekpenyong[1*,2,3], Kingsley Attai[4,5], Daniel Asuquo[6], Okure Obot[1], Ekerette Attai[5], Kitoye Ebire Okonny[7], Faith-Valentine Uzoka[8] Christie Akwaowo[9] and Faith-Michael Uzoka[10]

[1]*Department of Computer Science, Faculty of Computing, University of Uyo, Nigeria*
[2]*STEM Centre, University of Uyo, Nigeria*
[3]*Centre for Research, University of Uyo, Nigeria*
[4]*Department of Mathematics and Computer Science, Ritman University, Ikot Ekpene, Nigeria*
[5]*Novena Computers and Technologies Limited, Uyo, Nigeria*
[6]*Department of Information Systems, Faculty of Computing, University of Uyo, Uyo, Nigeria*
[7]*ICT Centre, Ignatius Ajuru University of Education, Port Harcourt, Nigeria*
[8]*Texas Southern University, 3100 Cleburne St, Houston, TX, United States*
[9]*Institute of Health Research and Development, University of Uyo Teaching Hospital, Uyo, Nigeria*
[10]*Department of Mathematics and Computing, Mount Royal University, Calgary, AB T3E 6K6, Canada*

ARTICLE INFO

Article history:
Received 2 December 2025
Revised 20 February 2026
Accepted 11 March 2026
Online first
Published 30 April 2026

Keywords:
Digital health
Explainable AI
LIME
Machine learning
Respiratory disease
SMOTE

DOI:
10.24191/mij.v7i1.11575

ABSTRACT

Respiratory Tract Infections (RTIs) pose significant health challenges in resource-scarce settings, where access to diagnostic tools is limited. This study presents an explainable artificial intelligence (XAI) framework for predicting RTIs using patient symptoms and multidimensional risk factors. Leveraging a dataset of 4,870 patient records provided by the New Frontiers in Research Fund (NFRF), we addressed class imbalance, with 1,756 RTI-positive and 3,112 RTI-negative cases, using the Synthetic Minority Over-sampling Technique (SMOTE). This resulted in a balanced dataset of 6,258 instances (3,129 positive and 3,129 negative). We implemented and evaluated three traditional machine learning models: Random Forest, XGBoost, and Decision Tree. To enhance model transparency and support clinical trust, we integrated Local Interpretable Model-Agnostic Explanations (LIME) for feature-level interpretability. Among the models, XGBoost achieved the highest AUC-ROC of 0.8876, followed closely by Random Forest (0.8823), while Decision Tree performed lowest (0.7922). LIME explanations consistently identified Dizziness (DIZ) and Shock (SHK) as the most influential predictors of RTI, reflecting symptoms associated with severe infections. Fever was also highlighted, underscoring alignment with clinical understanding. This framework not only achieves high predictive performance but also promotes interpretability, making it suitable for real-world deployment in healthcare systems with limited resources.

[1*] Corresponding author. *E-mail address*: mosesekpenyong@uniuyo.edu.ng
https://doi.org/10.24191/mij.v7i1.11575

1. INTRODUCTION

In resource-scarce regions where diagnostic infrastructure is often limited, Respiratory Tract Infections (RTIs) remain a significant health concern. RTIs contribute to considerable morbidity and mortality, particularly among vulnerable groups such as children, the elderly, and individuals with weakened immune systems (Georgakopoulou et al., 2024). The primary causes of RTIs are bacteria and viruses, with recurrent infections presenting substantial challenges for healthcare providers (Schaad et al., 2016). Chronic illnesses and other risk factors can compromise immune responses, increasing susceptibility to various viral RTIs (Fragkou et al., 2022).

Early and accurate diagnosis of infectious diseases is essential for timely treatment and effective containment (Olatunji et al., 2024). Therefore, there is a need for alternative diagnostic methods that are not only scalable and precise but also sufficiently interpretable to adapt to challenging clinical settings. Advances in artificial intelligence (AI) and machine learning (ML) have shown promising results in disease prediction and clinical diagnosis. ML algorithms have demonstrated superior performance in forecasting patient outcomes, resource allocation, and disease diagnosis compared to traditional methods (Shrivastava et al., 2024). Research has demonstrated that the XGBoost (Asuquo et al., 2024; Yang & Guan, 2022; Sankar et al., 2022), random forest (Attai et al., 2024; Attai et al., 2025a, 2025b), and decision tree (Arora et al., 2024; Tanaka et al., 2025; Yu et al., 2024) algorithms are useful in the diagnosis, prediction, and treatment of medical conditions and have demonstrated effectiveness in handling high-dimensional and complex datasets. Previous studies have demonstrated the effectiveness of machine learning (ML) models in predicting Respiratory Tract Infections (RTIs). For instance, a Long Short-Term Memory (LSTM) model was applied for lower RTI prediction (Bukenya et al., 2025), while a deep learning approach was used for pulmonary disease prediction (Habtamu et al., 2025). Similarly, other studies have implemented various ML algorithms for predicting acute Respiratory Tract Infections (Yehuala et al., 2024; Kassaw et al., 2024). These studies collectively provide a foundation for understanding how ML techniques can support early detection and clinical decision-making in RTI diagnosis.

Although traditional machine learning models have achieved remarkable accuracy in disease prediction, they are often criticized for functioning as “black boxes,” offering little or no insight into how predictions are made. This lack of interpretability limits their adoption in clinical practice, where healthcare professionals must understand and trust the reasoning behind automated decisions. Explainable Artificial Intelligence (XAI) has emerged as a vital research direction to overcome this limitation by making model outputs more transparent and clinically interpretable. Recent studies have successfully applied XAI techniques to enhance the explainability of AI systems in healthcare, for instance, in medical imaging diagnostics (Abraham et al., 2025; Ukwuoma et al., 2025), cardiovascular disease prediction (Kiran et al., 2025; Yan et al., 2025), and infectious disease modeling (Amannah et al., 2025; Attai et al., 2025a, 2025b), demonstrating that interpretability can coexist with high predictive performance (Hassan et al., 2025). However, despite these advances, the integration of XAI into Respiratory Tract Infection (RTI) diagnosis remains limited, particularly in frameworks that combine symptoms and risk factors for model interpretation. This gap underscores the need for interpretable ML approaches tailored to RTI prediction in resource-limited settings. To address this, the present study employs the Local Interpretable Model-Agnostic Explanations (LIME) technique to identify and prioritize the most influential features contributing to RTI predictions, thereby improving transparency, clinical validation, and trust in AI-assisted medical decision-making.

In this study, we integrated LIME with traditional machine learning techniques, including Random Forest, XGBoost, and Decision Tree, to create an interpretable model for predicting RTIs based on symptoms and multidimensional risk factors, such as behavioral and environmental indicators. XGBoost, Random Forest, and Decision Tree algorithms were selected for predicting RTIs due to their proven performance, interpretability, and suitability for complex clinical datasets. Decision Trees provide transparent and easily interpretable structures, making them ideal for medical decision support (Balcan et

https://doi.org/10.24191/mij.v7i1.11575

al., 2024; Catalano et al., 2025). Random Forest enhances predictive accuracy by combining multiple trees and reducing overfitting, which is beneficial in heterogeneous healthcare data (Attai et al., 2024; Attai et al., 2025a, 2025b). XGBoost, a gradient-boosted ensemble method, efficiently captures nonlinear relationships among clinical features and has demonstrated high accuracy in disease prediction tasks (Jiang et al., 2025; Zhou et al., 2025). When integrated with LIME, these models provide both robust performance and explainability, enabling clinicians to understand model-driven predictions with confidence. This work focuses on healthcare settings with limited resources, where diagnostic support tools need to be accurate, easily deployable, inexpensive, and explainable.

2. METHODOLOGY

This study focused on developing predictive models for respiratory tract infection (RTI) using classical machine learning algorithms, integrated with an Explainable Artificial Intelligence (XAI) approach to enhance model interpretability. The proposed explainable framework for predicting Respiratory Tract Infections integrated both clinical symptoms and patient risk factors to enhance diagnostic accuracy and interpretability. The process began with input data collection, comprising relevant symptoms and risk factors associated with Respiratory Tract Infections as shown in Fig. 1. These data underwent preprocessing to handle missing values, normalize feature scales, and encode categorical variables. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, ensuring equitable representation of disease and non-disease cases. The balanced dataset was then used to train a machine learning model, whose parameters were optimized through GridSearchCV for improved performance and generalization. To enhance transparency, the framework incorporated Explainable Artificial Intelligence (XAI) using the Local Interpretable Model-Agnostic Explanations (LIME) technique, which identified and visualized the most influential features contributing to each prediction. Finally, the interpretable outputs were presented as part of a decision-support system to assist healthcare workers in making informed, data-driven, and transparent clinical decisions regarding Respiratory Tract Infection diagnosis.

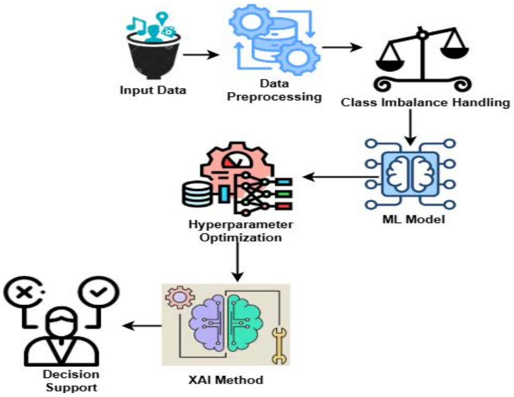

Fig. 1. Proposed Framework for Explainable Respiratory Tract Infection Prediction

2.1 Dataset and Preprocessing

The New Frontiers in Research Fund (NFRF) project provided 4,870 patient records used in this study (University of Uyo Teaching Hospital & Mount Royal University, 2024). Following the removal of entries with missing symptom or risk factor information, 4,868 complete records were retained for analysis. Table 1 presents the demographic distribution of male and female patients. During the preprocessing stage, all columns unrelated to the study objective, such as irrelevant symptoms, risk factors, and non-target diseases, were excluded. The dataset was then refined to include only the relevant features necessary for modeling

https://doi.org/10.24191/mij.v7i1.11575

Respiratory Tract Infections (RTIs). The final set of selected features used in the machine learning pipeline is summarized in Table 2. The dataset demonstrated a class imbalance, comprising, 1,756 RTI-positive cases and 3,112 RTI-negative cases. To mitigate this imbalance and enhance the fairness and robustness of model training, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. This approach generated synthetic examples for the minority class, resulting in a balanced dataset of 3,129 RTI-positive and 3,129 RTI-negative instances.

Table 1. Statistics of patients in the dataset

Patient Age (Years)	< 5	5 -12	13 – 19	20 – 64	≥ 65	Total
Female	419	323	213	1605	135	2695
Male	534	346	150	1012	133	2175
Total	953	669	363	2617	268	4870

Table 2. Symptoms and Risk Factors Used in the Study

SN	Symptoms/Risk Factors	SN	Symptoms/Risk Factors
1	Cough_initial dry (CGHDRY)	14	Fatigue (FTG)
2	Dry cough (DRYCGH)	15	Foul breath (FOLBRT)
3	Catarrh (CTRH)	16	Dizziness (DIZ)
4	Chest pain (CHSPN)	17	Shock (SHK)
5	Chest indraw (CHSIND)	18	Skin rash (SKNRSH)
6	Difficulty breathing (DIFBRT)	19	Indoor Air Pollution (EXPIDARPOL)
7	Wheezing (WHZ)	20	Smoking Exposure (SMSCHNSM)
8	Red eyes (REDEYE)	21	Overcrowding (OVCRW)
9	Red eyes, face, tongue (REDEYEFCTNG)	22	Poor Environmental Condition (PECON)
10	Sore throat (SRTRT)	23	Contact with an Infected Person (DRCOIFPS)
11	Mouth ulcer (MUTUCR)	24	Allergy (ALG)
12	Generalized rashes (GENRSH)	25	Underlying Chronic Illness (UNCHRIL)
13	Fever (FVR)	26	Poor Personal Hygiene (PPHYG)

2.2 Model Development and Performance Evaluation

Three supervised machine learning algorithms, Random Forest, XGBoost, and Decision Tree, were utilised to develop predictive models for RTI. These algorithms were chosen for their strong track records in clinical prediction tasks, ease of interpretability, and resilience against overfitting. GridSearchCV was employed for hyperparameter tuning to determine the best-performing settings for each classifier. In the case of XGBoost, the tuning process tested combinations of 100 and 200 estimators with maximum tree depths of 3 and 5, aiming to minimize overfitting while preserving accuracy. The Random Forest algorithm used a similar procedure, with the n_estimators parameter tested at 100 and 200, and max_depth evaluated at both unrestricted (None) and a controlled limit of 10. The Decision Tree model was optimized by assessing tree depths of None, 10, and 20 to find the most effective complexity level. This careful calibration of parameters ensured that each model was tailored for optimal learning and stable predictions in the context of the study. To assess model performance, a suite of classification metrics was employed: Accuracy, Precision, Recall, F1 Score, Logarithmic Loss (Log Loss), and area under the receiver operating characteristic curve (AUC-ROC). This multi-metric approach ensured a balanced evaluation, especially critical in healthcare applications where data imbalance and the cost of misclassification must be carefully managed.

2.3 Model Explainability with LIME and Experimental Environment

To improve the interpretability of the model predictions, the Local Interpretable Model-Agnostic Explanations (LIME) technique was adopted. LIME provides local explanations by identifying which features most significantly influenced individual prediction outcomes, thereby making the models' decisions more transparent to end users, particularly healthcare professionals. All model training, evaluation, and explainability analyses were conducted in Google Colaboratory, a cloud-based interactive

https://doi.org/10.24191/mij.v7i1.11575

development environment. The experiments were implemented using Python, with key libraries including scikit-learn for model development, imblearn for handling class imbalance with SMOTE, and LIME for generating explanation outputs.

3. RESULTS AND DISCUSSION

The XGBoost model showed a strong balance between correctly identifying patients with Respiratory Tract Infections (RTI) and those without. This was reflected in the confusion matrix in Fig. 2, where both the True Positive and True Negative values are high, while the False Positive and False Negative values are comparatively low. XGBoost achieved the highest accuracy (81.72%), indicating that it correctly classified most cases. Its precision (82.03%) and recall (81.72%) are well-balanced, meaning the model is both reliable in identifying true positives and minimizing false alarms. The F1 score (81.83%) further confirmed this balance between sensitivity and specificity. Additionally, XGBoost had the lowest log loss (0.4335), suggesting its probability estimates are both accurate and confident. The highest AUC-ROC (0.8876) indicated that XGBoost was effective at distinguishing between patients with and without RTI.

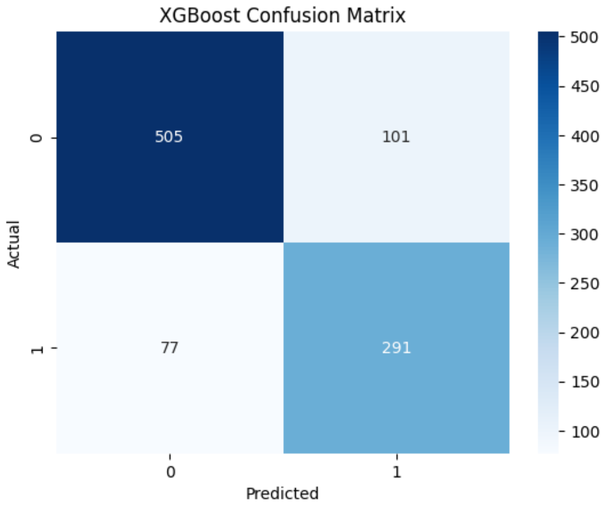

Fig. 2. XGBoost Confusion Matrix

The Random Forest model also performed well, with a slightly higher precision (82.24%) than XGBoost, indicating fewer False Positives, which is desirable in a medical setting because it helps avoid misdiagnosis and overtreatment as shown in Fig. 3. The confusion matrix was like XGBoost, but with a minor increase in FN and a reduction in FP. The Random Forest model's recall and F1 score were nearly identical to XGBoost, making it a very dependable alternative. However, its higher log loss (0.6174) compared to XGBoost suggests it may be less confident in its probability outputs. Nonetheless, the AUC-ROC of 0.8823 showed that it still had strong classification capability.

https://doi.org/10.24191/mij.v7i1.11575

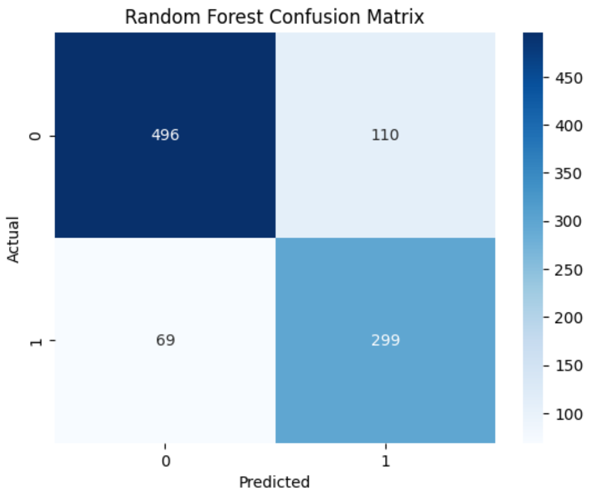

Fig. 3. Random Forest Confusion Matrix

Although the Decision Tree model was simple and interpretable, it underperformed compared to other models, as seen from its lower precision (78.75%) and recall (78.85%). As shown in its confusion matrix (Fig. 4), the FP value is relatively higher than in the other models, which resulted in a lower number of TN. The Decision Tree model had the lowest accuracy (78.85%), precision (78.75%), recall (78.85%), and F1 score (78.79%), suggesting that it was less reliable overall. More concerning was its extremely high log loss (2.7169), indicating that its predicted probabilities were far from actual outcomes, which was problematic in medical applications where confidence in predictions matters. The AUC-ROC (0.7922) also lagged, showing a weaker ability to distinguish between RTI and non-RTI cases. Given these weaknesses, the Decision Tree was not recommended for deployment in a clinical application.

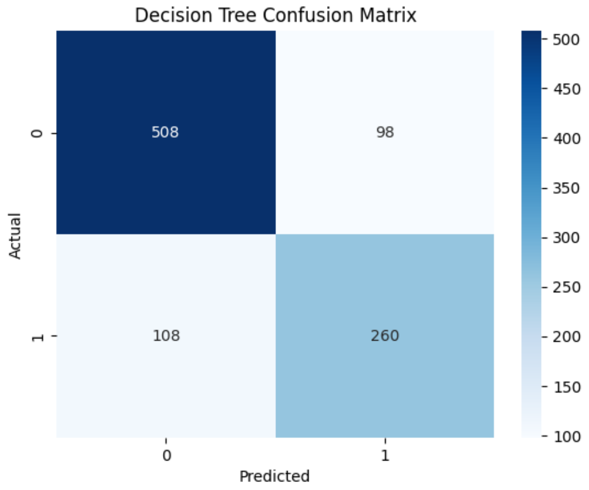

Fig. 4. Decision Tree Confusion Matrix

Table 3. Prediction model performance

	Accuracy	Precision	Recall	F1 Score	Log Loss	AUC-ROC
Random Forest	0.8162	0.8224	0.8162	0.8179	0.6174	0.8823
XGBoost	0.8172	0.8203	0.8172	0.8183	0.4335	0.8876
Decision Tree	0.7885	0.7875	0.7885	0.7879	2.7169	0.7922

https://doi.org/10.24191/mij.v7i1.11575

LIME was used to interpret the model’s individual prediction by identifying which symptoms and risk factors contributed most to the diagnostic outcome, and the results are illustrated in Fig. 5-7. The Y-axis presents the list of features that are important in determining the outcome of a specific instance. The X-axis shows the feature contribution weights, representing the degree and direction of influence each variable has on the model’s prediction. Values further from zero indicate a stronger influence, while values closer to zero have minimal impact. Features shown in red represent factors that decrease the likelihood of Respiratory Tract Infection in this specific prediction, whereas green bars indicate factors that increase the likelihood of infection.

For the XGBoost model, the LIME explanation in Fig. 5 showed that severe clinical symptoms, particularly shock (SHK) and dizziness (DIZ), were the strongest positive contributors to RTI prediction, consistent with their association with serious infections. Additional positive influences included mouth ulcer (MUTUCR), foul breath (FOLBRT), fatigue (FTG), skin rash (SKNRSH), generalized rash (GENRSH), and fever (FVR), highlighting the model’s sensitivity to inflammatory and systemic features. Symptoms such as cough_initial dry (CGHDRY), dry cough (DRYCGH), catarrh (CTRH), and difficulty breathing (DIFBRT) exhibited relatively small negative contributions, suggesting limited influence on this specific prediction. This minimal effect may have been attributed to the dataset’s composition, which originally included several febrile diseases and overlapping symptom profiles; after refining the dataset to focus exclusively on RTI, these features retained less discriminative power in this localized context. Environmental and hygiene-related indicators (PPHYG, OVCRW, PECON) also contributed to the explanation, underscoring the integration of social determinants with clinical symptoms.

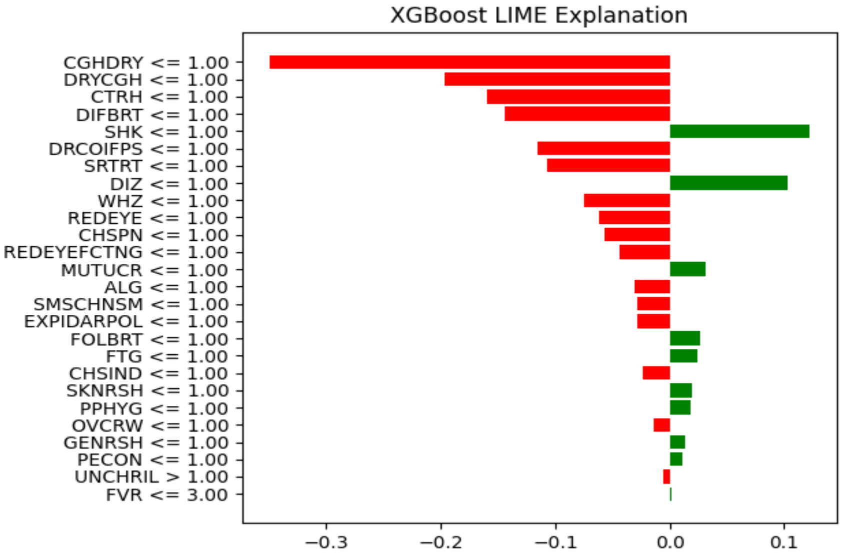

Fig. 5. XGBoost LIME diagram

For the Random Forest model, the LIME explanation in Fig. 6, showed that dizziness (DIZ) and shock (SHK) were the most influential positive contributors to RTI prediction, consistent with their role as markers of systemic illness and severe infection. Fever (FVR) also appeared as a key positive predictor, reinforcing the model’s alignment with clinical expectations for febrile respiratory illnesses. Additional features such as generalized rash (GENRSH), fatigue (FTG), skin rash (SKNRSH), and mouth ulcer (MUTUCR) further indicated the model’s sensitivity to systemic and dermatological symptoms that may have captured viral etiologies. At the same time, most cough-related features, including cough_initial dry (CGHDRY), dry cough (DRYCGH), catarrh (CTRH), and difficulty breathing (DIFBRT), exhibited strong

https://doi.org/10.24191/mij.v7i1.11575

negative contributions, suggesting limited discriminative value in this localized prediction. This reduced influence may have stemmed from the dataset's initial inclusion of multiple febrile illnesses with overlapping respiratory symptoms, which diminished the specificity of these features once the dataset was refined to focus exclusively on RTI. Finally, hygiene and environmental factors such as poor personal hygiene (PPHYG), overcrowding (OVCRW), and poor environmental condition (PECON) also contributed positively, reflecting the model's integration of socioeconomic determinants of infection risk. Overall, the LIME interpretation illustrated that Random Forest emphasized both severe systemic features and contextual risk factors, though its reliance on overlapping respiratory symptoms was reduced by dataset constraints.

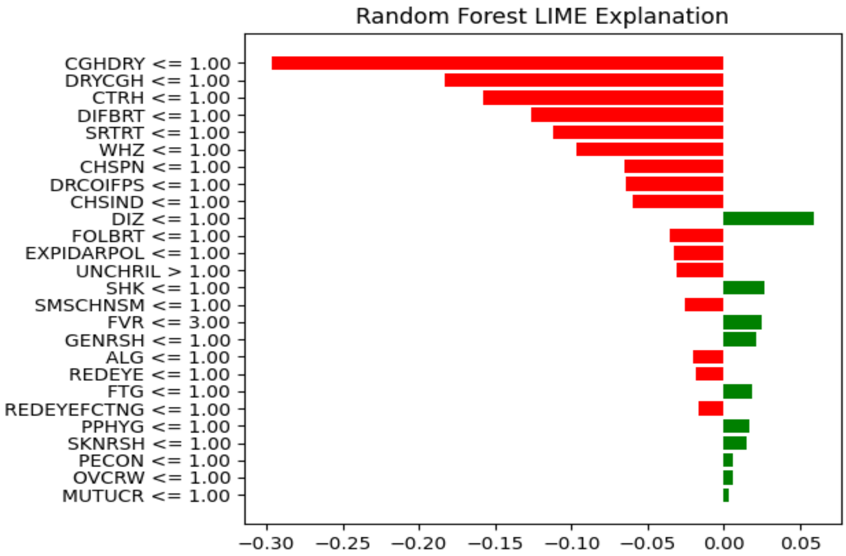

Fig. 6. Random Forest LIME diagram

For the Decision Tree model, Fig. 7 highlighted dizziness (DIZ), shock (SHK), and fatigue (FTG) as the strongest positive contributors to RTI prediction, underscoring the model's prioritization of systemic illness indicators. Dermatological features such as skin rash (SKNRSH) and generalized rash (GENRSH), along with chest indraw (CHSIND) and red eyes (REDEYE), also appeared as influential, suggesting that the model relied on more visible and clinically observable signs of infection. However, most cough-related features, including cough_initial dry (CGHDRY), catarrh (CTRH), dry cough (DRYCGH), and difficulty breathing (DIFBRT), dominated as negative contributors, indicating reduced discriminative power for respiratory symptoms in this localized prediction. This limitation likely reflected the dataset's initial inclusion of multiple febrile illnesses with overlapping respiratory manifestations, which diluted the predictive weight of these features when the dataset was narrowed to RTI-specific cases. Unlike Random Forest and XGBoost, the Decision Tree gave little weight to contextual risk factors such as overcrowding and poor hygiene, suggesting an underrepresentation of socioeconomic determinants. Overall, while the model's reliance on fewer, more observable features enhances interpretability, its narrower feature scope and reduced contextual sensitivity may have explained its comparatively lower predictive performance.

https://doi.org/10.24191/mij.v7i1.11575

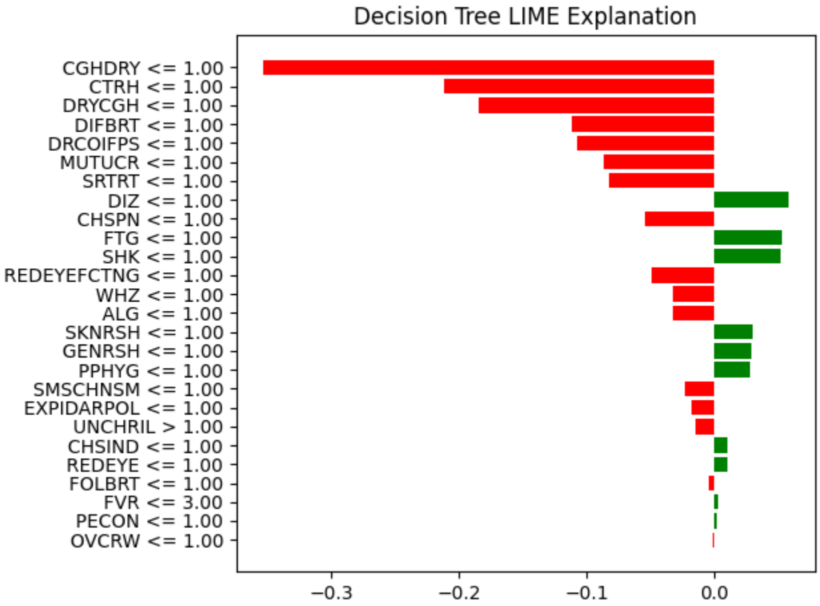

Fig. 7. Decision Tree LIME diagram

4. CONCLUSION

This study demonstrated the feasibility and effectiveness of integrating Explainable AI with traditional machine learning techniques to predict RTIs in resource-scarce environments. By using SMOTE to address class imbalance and LIME to support model interpretability, the proposed framework ensures fairness in training and transparency in predictions. Among the models tested, XGBoost emerged as the most accurate and robust classifier, with LIME explanations emphasizing clinically relevant symptoms such as dizziness, shock, and fever, thereby validating the framework's diagnostic alignment with medical realities. Nonetheless, the study was limited by the composition of the dataset, which originally included multiple febrile diseases with overlapping symptom profiles, reducing the discriminative power of certain features such as cough and difficulty breathing. To address these limitations, future work should expand the dataset to incorporate comorbidities, a broader range of clinical features, and additional diseases, thereby enhancing the generalizability and clinical utility of the models. Further studies should also explore integration into mobile-based diagnostic tools to support frontline healthcare workers and strengthen early detection of respiratory illnesses in low-resource settings.

5. ACKNOWLEDGEMENTS/FUNDING

This study was funded by the New Frontiers in Research Fund from April 2020 to March 2024 under grant number NFRFE-2019-01365.

6. CONFLICT OF INTEREST STATEMENT

The authors declare there is no conflict of interest in the subject matter or materials discussed in this manuscript.

https://doi.org/10.24191/mij.v7i1.11575

7. AUTHORS' CONTRIBUTIONS

Author Contributions: Conceptualization, F.-M.U., M.E., D. A., O.O. and K.A.; methodology, K.A., M.E., D.A., and F.-M.U.; validation, F.-M.U., K.A., D. E., O.O. and M.E.; formal analysis, K. A., D.A., E.A., F.-V.U. and E.D.; data curation, K. A., K.O. and E.A.; writing—original draft preparation, K. A., E.A., S.B., E.D., F.-V.U., C.A., K.O., and F.-M.U.; writing—review and editing, M.E., D.A., O.O., K. A., S.B., K. O., and F.-M.U.; supervision, F.-M.U., M.E., O.O., and D.A. project administration, O.O., M.E. and F.-M.U.; funding acquisition, F.-M.U. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Abraham, L. A., Palanisamy, G., & Veerapu, G. (2025). Transparent brain tumor detection using DenseNet169 and LIME. Scientific Reports, 15(1), 28185. https://doi.org/10.1038/s41598-025-13233-7

Amannah, C., Attai, K. F., & Uzoka, F. M. (2025). A Data-Driven Intelligent Methodology for Developing Explainable Diagnostic Model for Febrile Diseases. Algorithms, 18(4), 190. https://doi.org/10.3390/a18040190

Arora, A., Tsigelny, I. F., & Kouznetsova, V. L. (2024). Laryngeal cancer diagnosis via miRNA-based decision tree model. European Archives of Oto-Rhino-Laryngology, 281(3), 1391-1399. https://doi.org/10.1007/s00405-023-08383-1

Asuquo, D., Attai, K., Obot, O., Ekpenyong, M., Akwaowo, C., Arnold, K., & Uzoka, F. M. (2024). Febrile disease modeling and diagnosis system for optimizing medical decisions in resource-scarce settings. Clinical eHealth, 7, 52-76. https://doi.org/10.1016/j.ceh.2024.05.001

Attai, K., Ekpenyong, M., Amannah, C., Asuquo, D., Ajuga, P., Obot, O., Johnson, E., John, A., Maduka, O., Akwaowo, C., & Uzoka, F.-M. (2024). Enhancing the interpretability of malaria and typhoid diagnosis with explainable AI and large language models. *Tropical Medicine and Infectious Disease, 9*(9), Article 216. https://doi.org/10.3390/tropicalmed9090216

Attai, K. F., Amannah, C., Ekpenyong, M., Baadel, S., Obot, O., Asuquo, D., Attai, E., Uzoka, F.-V., Dan, E., Akwaowo, C., & Uzoka, F.-M. (2025a). Predicting Predisposition to Tropical Diseases in Female Adults Using Risk Factors: An Explainable-Machine Learning Approach. Information, 16(7), 520. https://doi.org/10.3390/info16070520

Attai, K. F., Amannah, C., Ekpenyong, M. E., Asuquo, D. E., Akputu, O. K., Obot, O. U., Ajuga, P. C., Obi, J. C., Maduka, O., Akwaowo, C., & Uzoka, F.-M. (2025). Developing an explainable artificial intelligence system for the mobile-based diagnosis of febrile diseases using random forest, LIME, and GPT. *Healthcare Informatics Research, 31*(2), 125–135. https://doi.org/10.4258/hir.2025.31.2.125

Balcan, M.-F., & Sharma, D. (2024). Learning accurate and interpretable decision trees. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* (pp. 288–307). PMLR.

Bukenya, L., Eyobu, O. S., & Oyana, T. J. (2025). An optimized LSTM model for diagnosis prediction of lower respiratory tract infections using a minimalistic data set. *IEEE Access, 13*, 60341–60354. https://doi.org/10.1109/ACCESS.2025.3540359

https://doi.org/10.24191/mij.v7i1.11575

Catalano, G. A. P. I., Brownlee, A., Cairns, D., Ainslie, R., & McCall, J. (2025, July). Interpretable decision trees to predict solution fitness. In Proceedings of the Genetic and Evolutionary Computation Conference (pp. 1099-1107). https://doi.org/10.1145/3712256.3726387

Fragkou, P. C., Moschopoulos, C. D., Reiter, R., Berger, T., & Skevaki, C. (2022). Host immune responses and possible therapeutic targets for viral respiratory tract infections in susceptible populations: a narrative review. Clinical Microbiology and Infection, 28(10), 1328-1334. https://doi.org/10.1016/j.cmi.2022.03.010

Georgakopoulou, V. E., Papalexis, P., & Trakas, N. (2024). Nanotechnology-based approaches for targeted drug delivery for the treatment of respiratory tract infections. Journal of Biological Methods, 11(4), e99010032. https://doi.org/10.14440/jbm.2024.0065

Habtamu, H., Abebe, M., & Mohapatra, S. K. (2025). Mobile application-based pulmonary disease prediction using respiratory sound and deep learning. International Journal of Bioinformatics Research and Applications, 21(3), 219-233. https://doi.org/10.1504/IJBRA.2025.146348

Hassan, S. U., Abdulkadir, S. J., Zahid, M. S. M., & Al-Selwi, S. M. (2025). Local interpretable model-agnostic explanation approach for medical imaging analysis: A systematic literature review. Computers in Biology and Medicine, 185, 109569. https://doi.org/10.1016/j.compbiomed.2024.109569

Jiang, H., Ji, L., Zhu, L., Wang, H., & Mao, F. (2025). XGBoost model for predicting erectile dysfunction risk after radical prostatectomy: development and validation using machine learning. Discover Oncology, 16(1), 810. https://doi.org/10.1007/s12672-025-02685-y

Kassaw, A. K., Bekele, G., Kassaw, A. K., & Yimer, A. (2024). Prediction of acute respiratory infections using machine learning techniques in Amhara Region, Ethiopia. Scientific Reports, 14(1), 27968. https://doi.org/10.1038/s41598-024-76847-3

Kiran, I., Ali, S., Alhussein, M., Aslam, S., & Aurangzeb, K. (2025). An AI-enabled framework for transparency and interpretability in cardiovascular disease risk prediction. *Computers, Materials & Continua, 82*(3), 3747–3773. https://doi.org/10.32604/cmc.2025.058724

Olatunji, A. O., Olaboye, J. A., Maha, C. C., Kolawole, T. O., & Abdul, S. (2024). Revolutionizing infectious disease management in low-resource settings: The impact of rapid diagnostic technologies and portable devices. International Journal of Applied Research in Social Sciences, 6(7), 1417-1432. https://doi.org/10.51594/ijarss.v6i7.1332

Sankar, S., Potti, A., Chandrika, G. N., & Ramasubbareddy, S. (2022). Thyroid disease prediction using XGBoost algorithms. J. Mob. Multimed, 18(3), 1-18. http://dx.doi.org/10.13052/jmm1550-4646.18322

Schaad, U. B., Esposito, S., & Razi, C. H. (2016). Diagnosis and management of recurrent respiratory tract infections in children: A practical guide. *Archives of Pediatric Infectious Diseases, 4*(1), Article e31039. https://doi.org/10.5812/PEDINFECT.31039

Shrivastava, A., Kotiyal, A., Habelalmateen, M. I., Rana, A., Devi, V. A., Rao, B. D., & Bansal, S. (2024). Leveraging XGBoost for predictive analytics in healthcare: Enhancing disease diagnosis. In *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)* (Vol. 7, pp. 1666–1672). IEEE. https://doi.org/10.1109/IC3I61595.2024.10829136

Tanaka, Y., Ando, H., Miyamoto, T., Yokokawa, Y., Ono, M., Asaka, R., Inuzuka, R., & Shiozawa, T. (2025). Usefulness of decision tree analysis of MRI features for diagnosis of placenta accreta spectrum in cases with placenta previa. *Japanese Journal of Radiology, 43*(3), 492–501. https://doi.org/10.1007/s11604-024-01684-3

Ukwuoma, C. C., Cai, D., Eziefuna, E. O., Oluwasanmi, A., Abdi, S. F., Muoka, G. W., Thomas, D., & Sarpong, K. (2025). Enhancing histopathological medical image classification for early cancer diagnosis using deep learning and explainable AI–LIME & SHAP. *Biomedical Signal Processing and Control, 100*, Article 107014. https://doi.org/10.1016/j.bspc.2024.107014

University of Uyo Teaching Hospital, & Mount Royal University. (2024). *NFRF project patient dataset with febrile diseases* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13756418.

Yehuala, T. Z., Fente, B. M., Wubante, S. M., & Derseh, N. M. (2024). Exploring machine learning algorithms to predict acute respiratory tract infection and identify its determinants among children under five in Sub-Saharan Africa. Frontiers in Pediatrics, 12, 1388820. https://doi.org/10.3389/fped.2024.1388820

Yan, Z., Zhang, Z., Pei, Y., & Yuan, C. (2025). Cardiovascular disease prediction model based on machine learning and LIME explainability analysis. In *2025 International Conference on Artificial Intelligence and Engineering Management (ICAIEM)* (pp. 127–130). IEEE. https://doi.org/10.1109/ICAIEM66060.2025.11139355

Yang, J., & Guan, J. (2022). A heart disease prediction model based on feature optimization and Smote-Xgboost algorithm. *Information, 13*(10), Article 475. https://doi.org/10.3390/info13100475

Yu, H., Zhao, J., & Sun, M. (2024). Classified diagnosis and treatment scheme of oral cosmetic restoration based on aesthetic analysis (part I): Basic concept, decision tree and clinical pathway. *Hua Xi Kou Qiang Yi Xue Za Zhi, 42*(1), 19–27. https://doi.org/10.7518/hxkq.2024.2023212

Zhou, L., Zhu, Q., Chen, Q., Wang, P., & Huang, H. (2025). Predicting hospital outpatient volume using XGBoost: a machine learning approach. Scientific Reports, 15(1), 17028. https://doi.org/10.1038/s41598-025-01265-y

https://doi.org/10.24191/mij.v7i1.11575