Journal of Information Systems and Informatics
Vol. 7, No. 2, June 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i2.1124 Published By DRPM-UBD

Predicting Respiratory Conditions Using Random Forest and XGBoost

Dhiyaussalam[1], Ahmad Yusuf[2], Isna Wardiah[3], Nitami Lestari Putri[4]

[1,2,3]Informatics Engineering, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia
[4]Smart City Information System, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia
Email: [1]salam@poliban.ac.id, [2]ahmadyusuf@poliban.ac.id, [3]isnawardiah@poliban.ac.id, [4]nitamiputri@poliban.ac.id

Abstract

This study examines the performance of Random Forest and XGBoost in predicting the diagnosis and severity of respiratory diseases using a simulated dataset of 2,000 patient records. The models were tested on two classification tasks: identifying disease types (e.g., pneumonia, influenza) and classifying severity levels (mild, moderate, severe). Both models achieved perfect accuracy in severity classification, with 1.0000 ± 0.0000 cross-validation scores, demonstrating strong stability under balanced class distributions. However, in the diagnosis task, Random Forest underperformed on minority classes, particularly pneumonia, with a recall of 0.18 and F1-score of 0.31. XGBoost, on the other hand, achieved superior results across all classes, including minority cases, with 0.9825 ± 0.0170 cross-validation accuracy and perfect test set performance. These findings highlight XGBoost's robustness in handling imbalanced and multiclass medical data, making it a promising candidate for clinical decision support. Future work should address class imbalance and explore explainability techniques to improve trust and transparency in real-world applications.

Keywords: Machine Learning, Random Forest, Respiratory Disease, Severity Classification, XGBoost

1. INTRODUCTION

Early and accurate diagnosis is a cornerstone of effective healthcare, particularly for respiratory illnesses such as pneumonia, influenza, bronchitis, and the common cold. Timely identification of these diseases leads to improved treatment outcomes, minimizes complications, and reduces healthcare costs. In pneumonia, for instance, early diagnosis supports targeted treatment and prevents severe outcomes like respiratory failure [1], [2]. Similarly, early detection of influenza enables the timely administration of antiviral drugs and helps mitigate transmission [3], [4]. For bronchitis and the common cold, accurate differentiation from other respiratory illnesses avoids unnecessary antibiotic use and contributes to better patient management [5].

Nevertheless, the diagnostic process remains complex due to nonspecific symptom presentation, patient variability, and limitations in conventional diagnostic tools. Many diseases, such as those caused by Mycoplasma pneumoniae, mimic other respiratory infections and present diagnostic ambiguity [6]. Moreover, in elderly populations, symptoms may be attenuated, leading to underdiagnosis or delayed care [7]. Even advanced imaging techniques suffer from sensitivity issues, as observed in various domains including oncology and dermatology [8]. These challenges are further compounded by the lack of universally accepted gold standards for several diseases.

To address these limitations, the integration of data-driven technologies, particularly machine learning, has gained traction in the medical field. Clinical data (such as patient demographics, medical history, and laboratory results) serve as valuable inputs for prediction models [9]. Symptom-based data also offer predictive potential, especially in chronic and mental health conditions [10].

A strong framework for enhancing diagnosis and classifying the severity of a disease is provided by combining machine learning algorithms with clinical, symptomatic, and sensor data. This context has been the subject of exceptional performance by tree-based machine learning methods, including Random Forest and XGBoost. These algorithms are commonly utilized in medicine due to their resilience, accuracy, and multiclass problem handling. Random Forest consolidates numerous decision trees to enhance generalizability and diminish variation, whereas XGBoost constructs trees in succession to rectify previous inaccuracies, adeptly managing imbalanced data [11], [12].

Both models have shown superior capabilities in managing noisy clinical data, identifying nonlinear interactions among features, and extracting meaningful patterns that are often difficult to detect using traditional statistical methods [13], [14]. Their ability to perform multiclass classification makes them particularly suitable for healthcare applications involving diverse diagnoses and severity levels. However, challenges remain, particularly regarding model interpretability and the practical integration of these tools into clinical workflows [15].

In order to better anticipate respiratory illnesses and their severity, numerous machine learning models have been investigated in earlier research. Albrecht et al. devised a multivariate time-series forecasting methodology to anticipate hospitalization rates for severe acute respiratory infections (SARI) utilizing models such as the Temporal Fusion Transformer (TFT) and DeepAR. Their research indicated that while multivariate models can enhance prediction accuracy, the dependability of forecasts fluctuates seasonally due to varying viral incidence patterns [16]. A comparison study was undertaken by Işık and Aydın (2023) to predict the severity of infection and symptoms based on gene expression data.

They utilized advanced feature selection techniques and machine learning models, specifically Random Forest and XGBoost. Their methodology surpassed benchmarks like DeepFlu, especially in forecasting symptom intensity across several respiratory viral datasets [17].

Other researchers have emphasized model interpretability and integration with clinical workflows. For example, the review by Ma and Shen focuses on interpretable machine learning models, particularly in the context of COVID-19, highlighting the importance of feature transparency and explainability for clinical decision-making. heir analysis shows that tree-based models, for example, Random Forest and XGBoost, are effective, making them suitable for clinical applications involving severity risk predictions [18]. Likewise, another study proposed leveraging image-derived environmental factors such as AQI to predict lung disease severity, combining domain knowledge with AI techniques to model external health determinants [19]. Finally, Kumar et al. underscored the advantages and trade-offs of different AI model types, noting that ensemble methods enhance accuracy but require computational resources. This balances model selection between interpretability, robustness, and resource efficiency [20].

Considering all of these factors, the purpose of this research is to construct and evaluate tree-based machine learning models, specifically Random Forest and XGBoost, for predicting disease diagnosis and severity classification. This study utilizes a simulated dataset of 2,000 patient records, including demographic profiles, symptom information, and sensor-based vital measurements. The goal is to assess the predictive performance of these ensemble methods and explore their potential to enhance diagnostic accuracy within digital healthcare systems.

Despite the increasing use of machine learning (ML) in medical contexts, there remains a lack of focused research comparing ensemble models' effectiveness in jointly predicting disease type and severity level from structured clinical data. This study seeks to fill that gap by addressing two key research questions: (1) How accurately can Random Forest and XGBoost predict respiratory disease diagnosis and severity based on simulated clinical features? Moreover, (2) Which algorithm demonstrates superior generalization ability and interpretability in multiclass classification scenarios? To investigate these questions, we design and test two classification models using features such as demographic details, symptoms, and physiological indicators. The findings are expected to contribute to developing intelligent decision-support tools in healthcare.

2. METHODS

This study utilized a simulated medical dataset comprising 2,000 patient records, each containing demographic details, symptom profiles, sensor-based health

indicators, and corresponding disease diagnoses and severity levels. The dataset was designed to reflect common respiratory conditions, including pneumonia, influenza, bronchitis, the common cold, and healthy cases. Each patient entry includes structured attributes such as age, gender, three reported symptoms, and physiological sensor readings, including heart rate (beats per minute), body temperature (°C), blood pressure (systolic/diastolic mmHg), and oxygen saturation (%). A diagnostic (a multiclass categorization of disease kind) and a severity (a multiclass categorization of condition severity: mild, moderate, and severe) were utilized as target variables. Due to their demonstrated robustness and capacity to handle complicated, multicategorical healthcare data, Random Forest and XGBoost were used in this study to create prediction models for both goals. Figure 1 shows the overall flow of process in the approach.

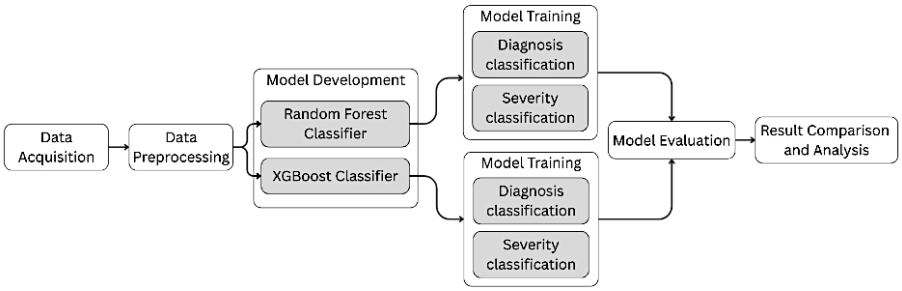

Figure 1. The overall process flow of the research.

2.1. Data Acquisitions

A publicly available dataset containing 2,000 simulated patient records was used in this study. The dataset was initially constructed to support research and experimentation in medical machine-learning applications and is openly accessible for academic use. Each record consists of a well-structured set of features, including demographic details (age and gender), three randomly assigned symptoms from a predefined list, and physiological sensor data such as heart rate (bpm), body temperature (°C), blood pressure (systolic/diastolic in mmHg), and oxygen saturation (%). In addition to these input features, each record includes two target labels: a rule-based diagnosis (i.e., pneumonia, influenza, bronchitis, cold, or healthy) and a severity classification (mild, moderate, or severe).

Although the dataset does not contain real patient data, it was designed to mimic realistic clinical patterns based on expert-informed rules, ensuring plausible associations between symptoms, sensor readings, and disease outcomes [21]. However, as a synthetic dataset, it may not fully capture the noise, irregularities, and edge cases commonly found in real-world clinical environments. This

limitation should be considered when interpreting the generalizability of the models developed in this study.

2.2. Data Preprocessing

The dataset underwent a series of preprocessing operations prior to model building to make it ready for machine learning analysis. The dataset was initially examined for absent or inconsistent values; however, due to its synthetic production, it displayed no missing elements. Subsequently, categorical variables like gender, symptoms, diagnosis, and severity were transformed into numerical representations. One-hot encoding was specifically employed for symptoms and gender to prevent the introduction of ordinal correlations, whilst label encoding was utilized for target variables (diagnosis and severity) to enable multiclass classification [22].

One-hot encoding was applied to categorical symptom variables to ensure that the machine learning models could interpret them without imposing ordinal assumptions [23]. Label encoding was used for the target variables (diagnosis and severity) as they represent discrete class labels [24]. These encoding techniques are widely adopted for tree-based models, which do not require feature scaling and are robust to categorical splits when encoded appropriately [25].

Numerical values were employed to characterize sensor measurements, including heart rate, temperature, blood pressure (systolic and diastolic), and oxygen saturation. The variables were normalized using Min-Max scaling. It is imperative to allocate 80% of the dataset for training and the remaining 20% for testing to ensure a fair and impartial evaluation of the models across all classes.

2.3. Model Development

This study employed two tree-based machine learning models: Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to develop classifiers capable of predicting illness categories and their severity levels. These algorithms were selected for their efficacy in modeling nonlinear patterns, handling multiclass classification, and mitigating overfitting.

The Random Forest approach combines the forecasts of numerous decisions trees and determines the final result via a majority vote. A number of critical hyperparameters were fine-tuned to improve the model's performance. These included the overall tree count, the maximum depth per tree, and the min sample needed at each terminal node [26].

At the same time, XGBoost builds its decision trees in a sequential fashion, with the goal of improving prediction accuracy with each building. Critical hyperparameters, including the learning rate, the number of estimators, the tree depth, the subsample proportion, and the regularization factors (L1 and L2), were meticulously modified during the process of model optimization in order to improve performance [27].

The two models were independently trained and assessed for the following prediction tasks: (1) diagnostic classification and (2) severity classification. To optimize the hyperparameters, we used grid search and then used k-fold cross-validation with a k value of 5 to make sure it works for everyone and avoid overfitting [28].

2.4. Model Evaluation

In order to measure how well the built models performed, we used standard classification measures to examine the RF and XGBoost classifiers. The metrics employed include accuracy, precision, recall, and F1-score, detailed as shown in Equation 1, 2, 3, and 4, calculated for each class and averaged using macro and weighted approaches [29].

$$\text{accuracy} = \frac{(\text{True Positive}+\text{True Negative})}{(\text{True Positive}+\text{True Negative}+\text{False Positive}+\text{False Negative})} \quad (1)$$

$$\text{precision} = \frac{(\text{True Positive})}{(\text{True Positive}+\text{False Positive})} \quad (2)$$

$$\text{recall or sensitivity} = \frac{(\text{True Positive})}{(\text{True Positive}+\text{False Negative})} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision}+\text{recall})} \quad (4)$$

Additionally, confusion matrices were generated to visualize the distribution of true versus predicted labels across all classes. This helped in identifying specific classes where the models performed well or struggled, and provided insights into common misclassification patterns. Model evaluation was conducted separately for each task:

1) Diagnosis prediction: a 5-class classification task (e.g., Pneumonia, Flu, Bronchitis, Cold, Healthy)
2) Severity prediction: a 3-class classification task (Mild, Moderate, Severe)

To ensure the reliability and generalizability of model performance, k-fold cross-validation (with k=5) was applied to the training data. This approach helps minimize overfitting by validating the model on multiple subsets of the data [30].

Each model was trained and validated across five distinct folds, with average performance metrics recorded to obtain a more stable and generalized evaluation. In addition, a final evaluation was performed on a held-out 20% test set that was never seen during training or cross-validation. This strict separation ensures that the reported results reflect each model's true generalization ability.

Grid search hyperparameter tuning was conducted during cross-validation to further optimize model performance. Key parameters such as the number of trees, maximum depth, and minimum samples per leaf were tuned for Random Forest. For XGBoost, tuning included the learning rate, number of estimators, maximum depth, subsample ratio, and regularization terms (L1/L2). These optimization steps contributed to enhancing the accuracy and robustness of both models [31].

3. RESULTS AND DISCUSSION

3.1. Prediction for Diagnosis Classification

The Random Forest and XGBoost models were evaluated for their ability to predict respiratory disease diagnoses based on clinical and sensor data. XGBoost achieved perfect classification performance on the test set, with accuracy, precision, recall, and F1-score all reaching 1.00 across all diagnostic categories, including Bronchitis, Cold, Flu, Healthy, and Pneumonia (see Table 2). In contrast, Random Forest exhibited slightly lower performance, particularly in identifying Pneumonia cases, where it achieved a recall of 0.18 and an F1-score of 0.31. This led to a macro-average F1-score of 0.84 and a test set accuracy of 0.97 for Random Forest, as shown in Table 1.

Tabel 1. Classification report for diagnosis prediction using Random Forest.

Class	Precision	Recall	F1-Score
Healthy	1.0	1.0	1.0
Cold	1.0	1.0	1.0
Bronchitis	0.97	1.0	0.98
Flu	0.89	0.98	0.93
Pneumonia	1.0	0.18	0.31

These discrepancies highlight that, although Random Forest performed well in detecting standard classes like Cold and Healthy, it struggled with minority classes, leading to misclassifications that could have clinical consequences in real-world settings.

The confusion matrices for diagnosis prediction (Figure 2) reflect this contrast. XGBoost showed perfect class separation, while Random Forest demonstrated misclassification specifically for Pneumonia, evidenced by off-diagonal values.

Cross-validation results also support these findings. Random Forest achieved a mean cross-validation accuracy of 0.8225 ± 0.0339, indicating some variability and reduced generalizability. In comparison, XGBoost maintained a high mean cross-validation accuracy of 0.9825 ± 0.0170, confirming its consistency and robustness across data splits. The best-performing hyperparameters for each model, obtained through grid search, were:

1) Random Forest: 'n_estimators': 200, 'max_depth': None, 'min_samples_split': 2
2) XGBoost: 'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100

These results confirm that while both models can handle multiclass classification for respiratory diagnoses, XGBoost outperforms Random Forest in predictive performance and stability, especially when dealing with underrepresented or imbalanced classes.

Tabel 2. Classification report for diagnosis prediction using XGBoost.

Class	Precision	Recall	F1-Score
Healthy	1.0	1.0	1.0
Cold	1.0	1.0	1.0
Bronchitis	1.0	1.0	1.0
Flu	1.0	1.0	1.0
Pneumonia	1.0	1.0	1.0

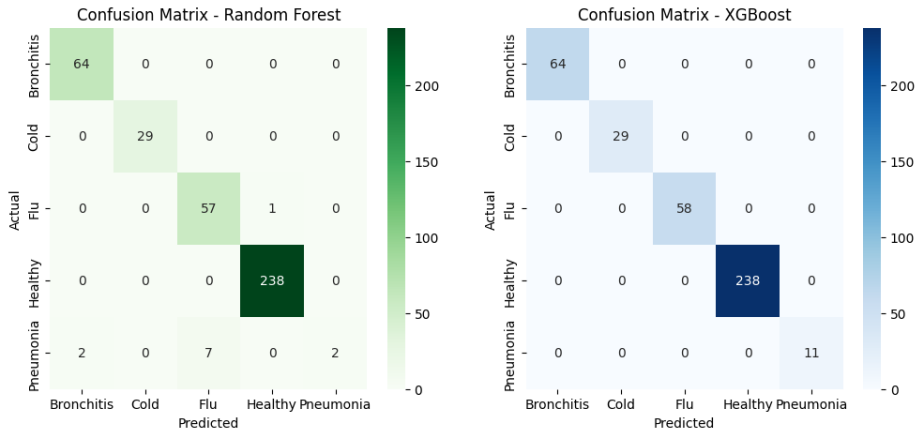

Figure 2. Confusion matrices for diagnosis prediction using Random Forest (left) and XGBoost (right).

3.2. Prediction Results for Severity Classification

In the severity classification task, both Random Forest and XGBoost models achieved perfect classification results on the test set. Each class (Mild, Moderate,

and Severe) was predicted with 100% precision, recall, and F1-score, resulting in an overall accuracy of 1.00 for both models (see Table 3 and Table 4). This level of performance demonstrates the models' strong capability to distinguish between varying levels of clinical severity, especially given the availability of diagnosis as a supporting feature in the input data. The balanced class distribution and the discriminative power of features likely contributed to this outcome.

Tabel 3. Classification Report for Severity Prediction (Random Forest).

Class	Precision	Recall	F1-Score
Mild	1.0	1.0	1.0
Moderate	1.0	1.0	1.0
Severe	1.0	1.0	1.0

Beyond test set results, 5-fold cross-validation was conducted to assess generalization consistency. Both models yielded a mean cross-validation accuracy of 1.0000 ± 0.0000, indicating exceptional stability across folds and the absence of performance variance. These findings reinforce the reliability of the models in multiclass classification settings where misclassification between severity levels can have critical clinical implications. The best-performing hyperparameters for each model, found through grid search, were as follows:

1) Random Forest: n_estimators=10, max_depth=None, min_samples_split=2
2) XGBoost: n_estimators=10, max_depth=3, learning_rate=0.1

Such configurations balanced depth and complexity, allowing the models to achieve perfect accuracy without overfitting.

Tabel 4. Classification Report for Severity Prediction (XGBoost).

Class	Precision	Recall	F1-Score
Mild	1.0	1.0	1.0
Moderate	1.0	1.0	1.0
Severe	1.0	1.0	1.0

Figure 3 presents the confusion matrices for severity classification using Random Forest (left) and XGBoost (right). Both models achieved perfect classification for all three severity categories (Mild, Moderate, and Severe) with no misclassified instances. Every sample in each class was accurately predicted, resulting in a clear diagonal structure in both matrices. The Random Forest model and the XGBoost model correctly classified all 1,330 Mild, 292 Moderate, and 378 severe cases. This ideal result confirms the models' exceptional ability to differentiate severity levels, particularly when combined with diagnosis information as input. The lack of off-diagonal values highlights the robustness and reliability of both models in high-stakes, multiclass medical prediction tasks.

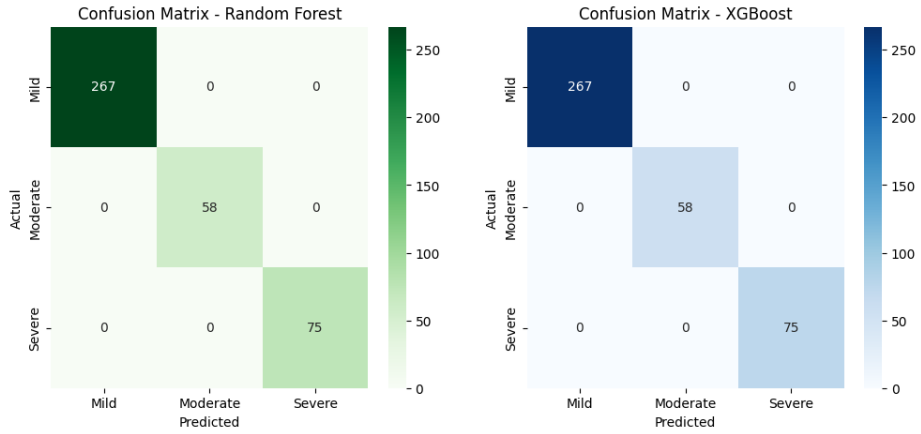

Figure 3. Confusion matrices for severity classification using Random Forest (left) and XGBoost (right).

3.3. Discussion

The findings of this study confirm the effectiveness of ensemble machine learning models (Random Forest and XGBoost) in classifying disease diagnoses and severity levels based on structured clinical data. XGBoost consistently demonstrated superior performance across both prediction tasks, achieving perfect classification (100% accuracy, precision, recall, and F1-score) on the test sets and high cross-validation accuracy, indicating strong generalization capabilities.

In the diagnosis classification task, XGBoost achieved perfect performance, successfully identifying all five diagnostic categories without error. Conversely, Random Forest showed signs of model instability and misclassification, particularly in identifying Pneumonia. Despite achieving high precision (1.00) for Pneumonia, the recall dropped significantly to 0.18, with an F1-score of just 0.31. This indicates that the model frequently failed to detect actual Pneumonia cases, likely due to the underrepresentation of this class in the dataset—only 44 out of 2,000 total instances were labeled as Pneumonia. Such class imbalance can lead tree-based models like Random Forest to bias predictions toward more dominant classes, especially when cost-sensitive or resampling techniques are not applied.

The confusion matrix for Random Forest further illustrates this issue, showing clear misclassifications primarily for the Pneumonia category. This weakness in minority class prediction could have significant clinical implications if deployed without further model refinement. In contrast, XGBoost effectively handled these challenges, maintaining perfect recall and F1-score across all diagnostic classes, including the rare Pneumonia cases. Its gradient boosting framework and regularization mechanisms likely contributed to better learning from minority

classes and reducing overfitting, as reflected in its high cross-validation accuracy of 0.9825 ± 0.0170, compared to Random Forest's 0.8225 ± 0.0339.

In the severity classification task, both Random Forest and XGBoost achieved perfect results on the test set, correctly classifying all samples in the Mild, Moderate, and Severe categories. Both models also demonstrated exceptional training stability, achieving a cross-validation accuracy of 1.0000 ± 0.0000, indicating perfect consistency across folds. These results suggest that when the class distribution is relatively balanced, as in the severity dataset, both Random Forest and XGBoost can deliver reliable and accurate multiclass classification. Figure 4, which presents cross-validation accuracy across both prediction tasks, confirms these trends. XGBoost exhibited greater stability and generalization in diagnosis, while Random Forest matched its performance only in the severity prediction, where class distributions were more even.

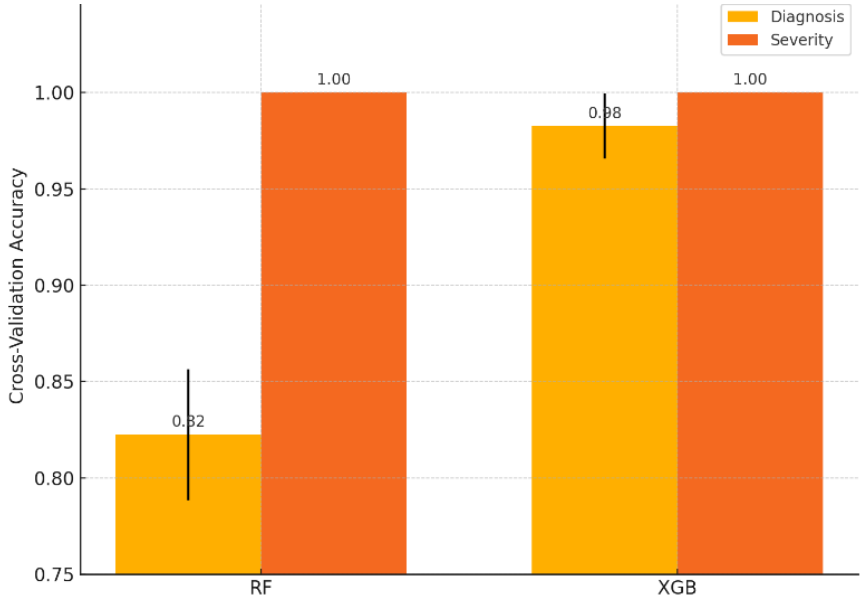

Figure 4. 5-fold cross-validation accuracy of Random Forest and XGBoost for diagnosis and severity prediction.

In summary, while both models showed high potential, XGBoost is more robust and reliable, particularly in scenarios involving class imbalance and rare event prediction. The relatively poor recall and F1-score of Random Forest for Pneumonia suggest the need for future research to address data imbalance through methods such as oversampling, class-weight adjustment, or synthetic data generation (e.g., SMOTE). Additionally, interpretability techniques such as SHAP

or LIME are recommended for future implementation to support transparent clinical decision-making.

4. CONCLUSION

This study demonstrated the effectiveness of Random Forest and XGBoost in predicting both the diagnosis and severity of respiratory diseases using a simulated clinical dataset. While both models achieved perfect accuracy on the test set for the severity classification task, their performance in the diagnosis task varied significantly. XGBoost consistently outperformed Random Forest, achieving perfect classification results and higher cross-validation accuracy (0.9825 ± 0.0170) across all diagnostic categories, including minority classes such as Pneumonia. In contrast, Random Forest exhibited reduced sensitivity, particularly in identifying Pneumonia, where it achieved a recall of only 0.18 and an F1-score of 0.31. This limitation is likely due to the class imbalance in the dataset, where Pneumonia cases were severely underrepresented. Such imbalance can bias model predictions toward dominant classes, highlighting the need for further strategies to enhance model fairness and recall in real-world deployment scenarios.

Despite these differences, both models performed equally well in the severity classification task, with 1.0000 ± 0.0000 cross-validation accuracy and flawless test set results. These findings suggest that tree-based ensemble models are highly reliable when balanced class distributions and that XGBoost suits imbalanced or complex multiclass problems. Future work should focus on validating these results using real-world clinical data and addressing class imbalance through oversampling, synthetic data generation, or cost-sensitive learning. Integrating explainable AI techniques such as SHAP or LIME can further enhance model interpretability and support transparent clinical decision-making.

REFERENCES

[1] H. Zhu, J. Dong, X. Xie, and L. Wang, "Comparison between the molecular diagnostic test and chest X-ray combined with multi-slice spiral CT in the diagnosis of lobar pneumonia," *Cell. Mol. Biol.*, vol. 67, no. 3, pp. 129–132, 2021, doi: 10.14715/cmb/2021.67.3.18.

[2] P. Zatovkaňuková and J. Slíva, "The potential dangers of whooping cough: a case of rib fracture and pneumothorax," *BMC Infect. Dis.*, vol. 24, no. 1, pp. 0–5, 2024, doi: 10.1186/s12879-024-10192-8.

[3] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, 2020, doi: 10.1016/S0140-6736(20)30211-7.

[4] J. Czubak, K. Stolarczyk, A. Orzel, M. Frączek, and T. Zatoński, "Comparison of the clinical differences between COVID-19, SARS, influenza, and the common cold: A systematic literature review.," *Adv. Clin. Exp. Med. Off. organ Wroclaw Med. Univ.*, vol. 30, no. 1, pp. 109–114, Jan. 2021, doi: 10.17219/acem/129573.

[5] J. Qu, C. Yang, F. Bao, S. Chen, L. Gu, and B. Cao, "Epidemiological characterization of respiratory tract infections caused by Mycoplasma pneumoniae during epidemic and post-epidemic periods in North China, from 2011 to 2016," *BMC Infect. Dis.*, vol. 18, no. 1, pp. 1–8, 2018, doi: 10.1186/s12879-018-3250-2.

[6] M. Oberoi, R. Kulkarni, and T. Oliver, "An Unusual Case of Myocarditis, Left Ventricular Thrombus, and Embolic Stroke Caused by Mycoplasma pneumoniae," *Cureus*, vol. 13, no. 3, pp. 0–4, 2021, doi: 10.7759/cureus.14170.

[7] T. A. Rowe *et al.*, "Reliability of nonlocalizing signs and symptoms as indicators of the presence of infection in nursing-home residents," *Infect. Control Hosp. Epidemiol.*, vol. 43, no. 4, pp. 417–426, 2022, doi: 10.1017/ice.2020.1282.

[8] L. Han, "Prediction of hepatocellular carcinoma and Edmondson-Steiner grade using an integrated workow of multiple machine learning algorithms," 2023, [Online]. Available: https://doi.org/10.21203/rs.3.rs-2905568/v1

[9] P. Jabbari, N. Taraghikhah, F. Jabbari, S. Ebrahimi, and N. Rezaei, "Body Mass Index as a Predictor of Symptom Duration in COVID-19 Outpatients," *Disaster Med. Public Health Prep.*, vol. 17, no. 6, 2023, doi: 10.1017/dmp.2022.185.

[10] M. Esposito *et al.*, "Depressive symptoms and insecure attachment predict disability and quality of life in psoriasis independently from disease severity," *Arch. Dermatol. Res.*, vol. 313, no. 6, pp. 431–437, 2021, doi: 10.1007/s00403-020-02116-8.

[11] D. Meng, J. Xu, and J. Zhao, "Analysis and prediction of hand, foot and mouth disease incidence in China using Random Forest and XGBoost," *PLoS One*, vol. 16, no. 12 December, pp. 1–16, 2021, doi: 10.1371/journal.pone.0261629.

[12] P. Yang and B. Yang, "Development and validation of predictive models for diabetic retinopathy using machine learning," *PLoS One*, vol. 20, no. 2 February, pp. 1–13, 2025, doi: 10.1371/journal.pone.0318226.

[13] Y. Han and S. Wang, "Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study," *Front. Public Heal.*, vol. 11, 2023, doi: 10.3389/fpubh.2023.1271595.

[14] N. Acharya, P. Kar, M. Ally, and J. Soar, "Predicting Co-Occurring Mental Health and Substance Use Disorders in Women: An Automated Machine Learning Approach," *Appl. Sci.*, vol. 14, no. 4, 2024, doi: 10.3390/app14041630.

[15] Y. Xiao, Y. Chen, R. Huang, F. Jiang, J. Zhou, and T. Yang, "Interpretable machine learning in predicting drug-induced liver injury among tuberculosis patients: model development and validation study," *BMC Med. Res. Methodol.*, vol. 24, no. 1, pp. 1–17, 2024, doi: 10.1186/s12874-024-02214-5.

[16] S. Albrecht *et al.*, "Forecasting severe respiratory disease hospitalizations using machine learning algorithms," *BMC Med. Inform. Decis. Mak.*, vol. 0, 2024, doi: 10.1186/s12911-024-02702-0.

[17] Y. Emre and Z. Ayd, "Comparative analysis of machine learning approaches for predicting respiratory virus infection and symptom severity," pp. 1–26, 2023, doi: 10.7717/peerj.15552.

[18] M. Shen, Jinzhi; Ke, "Review of Interpretable Machine Learning Models for Disease Prognosis".

[19] A. Mahajan, C. Kulkarni, and S. Mate, "Predicting Lung Disease Severity Via Image-Based Aqi Analysis Using Deep Learning Techniques," pp. 1–11.

[20] P. Yadav, V. Rastogi, A. Yadav, and P. Parashar, "Artificial Intelligence : A promising tool in diagnosis of respiratory diseases," *Intell. Pharm.*, vol. 2, no. 6, pp. 784–791, 2024, doi: 10.1016/j.ipha.2024.05.002.

[21] Warner, "Disease Diagnosis Dataset." [Online]. Available: https://www.kaggle.com/datasets/s3programmer/disease-diagnosis-dataset

[22] D. Ali, M. M. S. Missen, and M. Husnain, "Multiclass Event Classification from Text," *Sci. Program.*, vol. 2021, no. 1, 2021, doi: 10.1155/2021/6660651.

[23] A. Mansoori, M. Zeinalnezhad, and L. Nazarimanesh, "Optimization of Tree-Based Machine Learning Models to Predict the Length of Hospital Stay Using Genetic Algorithm," *J. Healthc. Eng.*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/9673395.

[24] F. Pargent, F. Pfisterer, J. Thomas, and B. Bischl, "Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features," *Comput. Stat.*, vol. 37, no. 5, pp. 2671–2692, 2022, doi: 10.1007/s00180-022-01207-6.

[25] S. Mumtaz and M. Giese, "Hierarchy-based semantic embeddings for single-valued & multi-valued categorical variables," *J. Intell. Inf. Syst.*, vol. 58, no. 3, pp. 613–640, 2022, doi: 10.1007/s10844-021-00693-2.

[26] Dhiyaussalam, A. Wibowo, F. A. Nugroho, E. A. Sarwoko, and I. M. A. Setiawan, "Classification of Headache Disorder Using Random Forest Algorithm," in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 2020, pp. 1–5. doi: 10.1109/ICICoS51170.2020.9299105.

[27] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[28] Dhiyaussalam and S. Uyun, "Optimization of Random Forest Hyperparameters with Genetic Algorithm in Classification of Lung Cancer," *6th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2023 - Proceeding*, pp. 82–88, 2023, doi: 10.1109/ISRITI60336.2023.10467686.

[29] A. Maulana *et al.*, "Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm," *Infolitika J. Data Sci.*, vol. 1, no. 1, pp. 1–7, 2023, doi: 10.60084/ijds.v1i1.72.

[30] S. Montaha, S. Azam, A. K. M. Rakibul Haque Rafid, S. Islam, P. Ghosh, and M. Jonkman, *A shallow deep learning approach to classify skin cancer using down-scaling method to minimize time and space complexity*, vol. 17, no. 8 August. 2022. doi: 10.1371/journal.pone.0269826.

[31] S. Baharvand and H. Ahmari, *Application of Machine Learning Approaches in Particle Tracking Model to Estimate Sediment Transport in Natural Streams*, vol. 38, no. 8. 2024. doi: 10.1007/s11269-024-03798-9.

[32] T. Inoue *et al.*, "XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury," *Neurotrauma Reports*, vol. 1, no. 1, pp. 8–16, 2020, doi: 10.1089/neur.2020.0009.