JCBI
Journal of Computing & Biomedical Informatics Volume 07 Issue 02
ISSN: 2710 - 1606 2024

Research Article
https://doi.org/10.56979/702/2024

Optimized XGBoost-Based Model for Accurate Detection and Classification of COVID-19 Pneumonia

Fazal Malik[1*], Muhammad Suliman[1], Muhammad Qasim Khan[1], Noor Rahman[2], Khairullah Khan[3], Mohammad Khan[1]

[1]Department of Computer Science, Iqra National University Peshawar, Khyber Pakhtunkhwa (KPK), Pakistan.
[2]Department of Computer science and Engineering, AL- Fayha College, 6480 Al Fayha, Al Jubayl 31961, Saudi Arabia.
[3]Department of Computer Science, University of Science & Technology, Bannu, Khyber Pakhtunkhwa (KPK), Pakistan.
[*]Corresponding Author: Fazal Malik. Email: fazal.malik@inu.edu.pk

Received: June 21, 2024 **Accepted:** August 18, 2024 **Published:** September 01, 2024

Abstract: Accurately diagnosing COVID-19 pneumonia is a critical global health challenge, particularly for vulnerable populations. Existing diagnostic methods often lack precision due to limited algorithm sophistication and insufficient dataset validation. This study addresses these issues by introducing a customized XGBoost algorithm for classifying COVID-19 pneumonia. The methodology follows a four-phase approach: (1) data acquisition from a comprehensive GitHub dataset, (2) data preprocessing with augmentation and normalization, (3) model training using XGBoost, and (4) evaluation against existing models. The model achieves an average accuracy of 87.35%, demonstrating superior performance in accuracy and diagnostic precision compared to current methods. The findings of this research provides a systematic framework for improving pneumonia classification and sets the stage for future AI-driven healthcare advancements in respiratory diseases.

Keywords: COVID-19 Pneumonia; XGBoost; Classification; Chest X-rays; Prediction.

1. Introduction

Pneumonia holds a position as a chronic problem in the context of world population health, being an inflammatory process in the lungs that occurs in patients of any age but is a greater threat to babies and elderly people. An analysis was made in concern with Computed Tomography (CT) scans that included 1658 COVID-19 patients and 1027 Community-acquired Pneumonia (CAP) patients. For classification, they employed a disease-size-aware random forest technique, popularly recognized as Unsupervised Size-Aware Forest (USAF from which they gained remarkable results [1]. The condition has a variety of symptoms and possible complications, and depending on the pathogens the condition severity is variable. Contrarily, viral infections in most cases cause relatively mild illness, though bacterial pneumonia of severe forms and even mortal with some strains, could pose serious threats to such vulnerable groups as newborns [2].

Early diagnosis is crucial for effective health management and achieving desired outcomes. Diagnostic methods are vital for pathogen identification in healthcare-associated infections, guiding antibiotic stewardship and infection control. CAP is a heterogeneous group caused by various microbes, including viruses and fungi [3].

Knowledge of pneumonia types is crucial for effective management. Pneumonia can be bacterial, viral, or fungal, with distinct diagnostic and treatment approaches. A classification system helps clinicians determine appropriate investigations, select antimicrobial therapies, and develop preventive measures. Additionally, walking pneumonia, with cold-like symptoms, and fungal pneumonia present unique challenges [4]. Pneumonia progresses in four stages: congestion (24 hours), red Hepatization (blood, neutrophils, and fibrin in alveoli), gray Hepatization (grayish-brown or yellowish lung tissue due to fibrin

and blood changes), and resolution (enzymes break down exudates, promoting lung repair). Proper diagnosis at each stage is essential for effective treatment [5].

There are different types of pneumonia, including bacterial, viral, walking, fungal, Hospital-acquired pneumonia (HAP) and CAP which possess their clinical effects and challenges. HAP is a severe hospital-acquired infection that often declines in an organization because of drug-resistant microorganisms and immunosuppressed individuals. [6]. Bacterial pneumonia and viral pneumonia constitute predominant forms of the disease, each presenting distinctive clinical features and management strategies.

Pneumonia includes bacterial, viral, walking, fungal, HAP, and CAP, each with distinct clinical challenges. HAP, often linked to drug-resistant microorganisms and immunosuppressed patients, is a severe hospital infection [6]. Bacterial and viral pneumonias are the most prevalent forms. COVID-19, caused by the novel coronavirus 2019-nCoV, poses significant global health risks, transmitting rapidly among people [7]. Key symptoms include fever, cough, and shortness of breath [8]. Deep transfer learning using models like Alexander Neural Network (AlexNet), Google Neural Network (GoogleNet), and Residual Neural Network (ResNet18) is being explored for pneumonia X-ray classification, with severe cases leading to organ failure and death [9]. Young people and men may be more susceptible to pneumonia, as patients often have higher breathing rates than healthy individuals [10]. X-ray imaging, revealing haziness and opacity in lung structures, is essential for diagnosing pneumonia [11]. Advancements in machine learning models and medical imaging have significantly improved the diagnosis and classification of COVID-19 and other pneumonia types, including CAP [12].

Additionally, a technique that uses ordinary non-virological data to forecast the rise in atypical pneumonia cases brought on by novel infections has been announced [13]. Deep learning is a well-established technique for classifying diseases; GoogLeNet and AlexNet are two examples of models that have been investigated for using deep transfer learning and Generative Adversarial Networks (GANs) to separate COVID-19 patients from normal cases [14]. As evidenced by the evaluation of deep learning models for identifying COVID-19, including ResNet, Inception, and GoogLeNet, robust frameworks are needed [15]. Additionally, building on earlier research that linked blood chemistry to the risk of COVID-19, this work evaluates how well XGBoost predicts age from routine blood tests [16].

The XGBoost algorithm combined with logistic regression (XGBoost + LR) evaluated data from 140 COVID-19 patients and 144 healthy controls, suggesting its potential for screening in hotels, care homes, and crowded areas [17]. A retrospective analysis of 678 CT scans from six centers (2018–2022) showed that XGBoost models accurately predicted entry and post-entry outcomes, emphasizing smoking as a mortality predictor [18]. XGBoost also demonstrated potential in identifying pathogens from clinical data in children hospitalized with respiratory infections (2010–2018) [19]. Recent studies highlight its use in automating COVID-19 detection from chest X-rays [20]. An automated approach for pneumonia classification from chest X-ray images using the XGBoost algorithm is introduced. Bayesian optimization refines feature representation and optimizes hyperparameters, improving classification accuracy [21].

Pneumonia, particularly during the COVID-19 pandemic, presents major public health challenges due to its diverse symptoms and the demand for precise diagnosis and classification. Current diagnostic methods for pneumonia suffer from limitations such as outdated algorithms, inadequate validation across datasets, low accuracy, and poor adaptability to evolving disease patterns.

To address these issues, this study introduces a customized XGBoost algorithm tailored for COVID-19 pneumonia classification. The research consists of four primary phases: (1) Dataset Acquisition—Utilizing the latest GitHub dataset with Python and Anaconda Jupyter Notebook for data collection; (2) Data Processing and Analysis—Employing histograms and scatter plots to preprocess and analyze data, optimizing it for machine learning (ML) algorithms; (3) Supervised Learning XGBoost Application with Data Augmentation—Applying the XGBoost algorithm enhanced by data augmentation techniques to boost model accuracy and robustness; and (4) Performance Evaluation—Assessing the developed model against existing methods to measure accuracy and effectiveness.

The key contributions of this research include: This study introduces a specialized XGBoost model for COVID-19 pneumonia classification, offering a significant advancement in the application of machine learning (ML) techniques to address the unique challenges posed by COVID-19 pneumonia. By employing histograms and scatter plots for data processing, this research ensures that the dataset is optimally prepared for ML algorithms, thus improving the quality of the analysis. The integration of data

augmentation techniques enhances the model's reliability and prediction accuracy, addressing challenges associated with limited data availability. The model is evaluated against existing methodologies, showcasing superior accuracy, which demonstrates its effectiveness in improving COVID-19 pneumonia classification. This research provides critical insights into predictive modeling, offering valuable contributions to public health strategies, including resource optimization during pandemics.

This methodology ensures accurate and reliable disease classification while offering predictive modeling insights for future disease trends, crucial for public health planning and resource optimization.

The subsequent sections comprise a comprehensive analysis of current state-of-the-art methods in the Literature Review, a proposed framework and methodology in the Methodology section, an in-depth discussion of experimental results in the Results and Discussion section, and a conclusion with recommendations for future research in the Conclusion section.

2. Literature Review

The literature review presents a comprehensive overview of studies in respiratory disease diagnosis, particularly focusing on COVID-19 and pneumonia.

An integrated approach combining the XGBoost algorithm and logistic regression (XGBoost + LR) was applied to analyze data from 140 COVID-19 patients and 144 healthy controls. This method shows significant potential for widespread use in environments such as hotels, nursing homes, and other crowded settings, enabling medical personnel to conduct efficient screenings. [22]. The study used various machine learning algorithms, like Naive Bayes, Linear Regression, Random Forest, XGBoost, Adaptive Boosting, K-nearest neighbor, Kernel Support Vector Machine (SVM), and Back Propagation Neural Networks. The evaluation was done with 5-fold cross-validation. Results show that an XGBoost-based clinical model can accurately identify high-risk patients at admission [23]. In a COVID-19 pathogenesis study, XGBoost was applied to predict disease severity using multi-omics data. The model, trained on 80% of the data and tested on an independent set, used 297 features selected using the hybrid approach. The results demonstrated strong discrimination between intensity levels, indicating strong performance under cross-validation [24].

In the December 2019 Wuhan coronavirus outbreak, XGBoost analyzed gene expression data from COVID-19 patients in the Gene Expression Omnibus (GEO) database. Identified potential diagnostic markers associated with viral transcription and COVID-19 pathways [25]. Hospitalized children with respiratory infections (2010-2018) had clinical data within 24 hours of admission used to build predictive models for six pathogens. XGBoost's potential in aiding clinicians to identify these pathogens was shown, optimizing diagnostic testing and possibly reducing medical costs [26]. This study responds to the urgent demand for precise COVID-19 screening tools by proposing two framework models. One integrates a Convolutional Neural Network (CNN) feature extractor with XGBoost for classification, while the other utilizes a CNN paired with a feedforward neural network. The COVID-19 CT-2A (CovidxCT-2A) dataset is employed for evaluation [27]. The recent COVID-19 outbreak calls into question the accurate diagnosis of pneumonia. The study presents the neutrosophic method, which classifies lung opacity images into correct (T), incorrect (F), and uncorrected (I) sets to reduce opacity. Preprocessing enhances images with primary alpha and beta enhancement operations. The enhanced images are fed into XGBoost for classification [28].

In COVID-19 prediction models, several machine-learning algorithms were explored, including logistic regression (LR), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost). In particular, XGBoost demonstrated effective COVID-19 prediction capabilities, further enhanced by interpretation tools such as Local Interpretable Model Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) [29]. Recent research and medical practice prioritize automatic COVID-19 detection with chest X-ray (CXR) images. Our Computer-Aided Diagnosis using Chest X-ray (CAD-CXR) system, incorporating Haar-like features and the XGBoost classifier, improves detection. We preprocess images, extract features, and classify them using various machine learning algorithms. Evaluation is conducted on publicly available datasets with three train-test splits [20]. A study implementing Random Forest classification in a COVID-19 Reverse Transcription Polymerase Chain Reaction (RT-PCR) test classification system alongside an auto-encoder algorithm was proposed, utilizing a dataset from a Brazilian hospital where the auto-encoder preprocesses features before classification with

Random Forest [30]. The pathology of H1N1, CAP, and sepsis was investigated, suggesting the potential of group mass disturbance for clinical outcomes in CAP and sepsis patients [31].

Volumetric Extraordinary Learning Machine (ELM) and k-nearest Neighbor (K-nn) AI were applied for breath sound analysis, demonstrating high accuracy through perceptual mode reduction and feature extraction [32]. Chest radiograph classification was enhanced using AlexNet (MAN) [33]. Dataset preparation processes were explored, proving effective in both engineering and real datasets [34]. The effectiveness of generative adversarial networks (GANs) was Demonstrated for pneumonia chest x-ray recognition, outperforming a related task in precision, recall, and F1 score [35]. An approach to diagnose COVID-19 with chest radiography was proposed, which shows promise despite the limitations of the data set [36]. Emphasis is placed on a methodological framework for the identification of COVID-19 [37]. The applicability of Artificial Intelligence (AI) in health status assessment has been explored [38]. Attention has focused on the diagnosis of pneumonia using X-ray imaging [39]. Effective methods have been proposed for the detection of tuberculosis and pneumonia [40, 41]. The role of machine learning in clinical areas, particularly in pneumonia identification, has also been discussed [42]. The potential of AI in clinical decision-making has been explored [43]. A meta-study analyzing explicit CT designs has been conducted [44]. A web application for pneumonia differentiation has been developed [45]. Additionally, a fully automated system for COVID-19 identification has been established [46]. Evolutionary Neural Network (ENN) and SVM expectation models were introduced in a previous study [47].

Another study focused on pollution-related biomarkers and clinical symptoms for pneumonia prediction [48]. Random Forest's superior performance in breast tumor detection using the Wisconsin Breast Cancer Database was demonstrated in prior research [49]. Additionally, the Random Forest algorithm, combined with Synthetic Minority Over-sampling Technique (SMOTE) preprocessing, was applied for fault identification in power transformers, effectively addressing complexity [50]. Random Forest was employed to classify chest X-ray images for detecting COVID-19, pneumonia, or normal cases, while XGBoost was applied to predict high-risk COVID-19 patients based on factors such as age, sex, diabetes, and hypertension. This model, implemented through a Python Flask web application, aids in patient prioritization and resource allocation [51, 52]. The global strain on healthcare systems during the COVID-19 pandemic has been mitigated by AI advancements, with models like XGBoost enhancing severity forecasting and supporting medical practitioners [53]. Additionally, COVID-19 detection through audio features such as cough and breathing, along with SHAP and data augmentation techniques, is explored [54].

3. Methodology

This study employs a thorough four-phase methodology for COVID-19 pneumonia classification utilizing XGBoost, as depicted in Figure 1 and outlined in Algorithm 1.

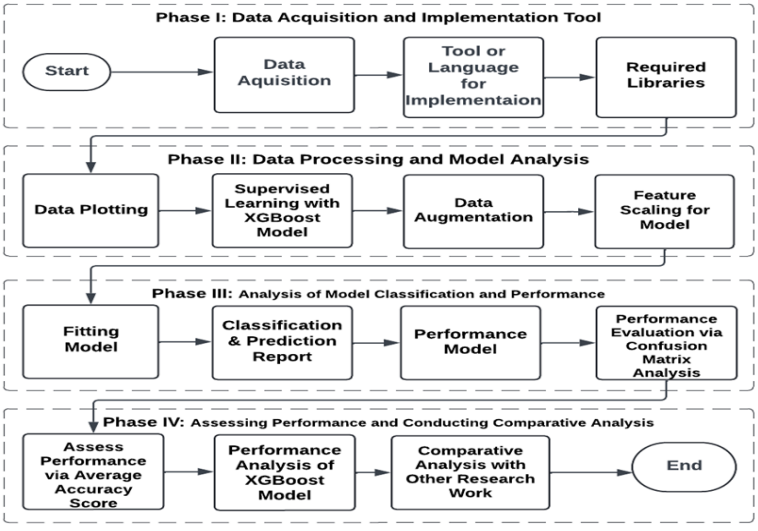

Figure 1. Block diagram of Proposed work

Algorithm 1: Research on Pneumonia Detection Evaluation
Input: Pneumonia X-ray images sourced from GitHub
Output: Evaluation of Model Performance
Step 1. Data Processing Phase
//Load the dataset containing pneumonia X-ray images, provided by WHO.
WHO_dataset ← LoadWHODataSet()
//Extract demographic information (e.g., age, gender) from the dataset
demog_info ← ExtractDemographicInfo (WHO_dataset)
//Preprocess the data to clean and format it for further analysis
preproc_data ← PreprocessData (WHO_dataset, demog_info)
Step 2. Data Visualization Phase
//Generate a plot to visualize the distribution of demographic information.
Demog_Distrib_Plot ← Visualize_DemographicDistribution(demog_info)
//Plot the prevalence of symptoms using the preprocessed data.
Symp_Preval_Plot ← PlotSymptomsPrevalence(preproc_data)
Step 3. Model Building for Feature Extraction
//Initialize XGBoost model that will be used for feature extraction
XGBoost_model ← InitializeXGBoostModel()
//Train the XGBoost model using the X-ray images and demographic data
train_models ← TrainModels(XGBoost_model, COVID19_XRay_images, demog_data)
//Optimize the trained models for better accuracy and performance
optim_models ← OptimizeModels(train_models)
Step 4. Performance Evaluation Phase
//Split the preprocessed data into training and testing sets for evaluation.
train_set, test_set ← SplitDataset(preproc_data)
//Evaluate the optimized models using a confusion matrix.
conf_matrix ← EvaluateModels(optim_models, test_set)
//Calculate performance metrics such as accuracy, sensitivity, and specificity
metrics ← CalculateMetrics(conf_matrix, accuracy, sensitivity, specificity)
Step 5. Comparative Analysis of Results
//Compare the accuracy of this study with previous research on pneumonia classification.
accu_compar ← CompareAccuracy(base_study, existing_research)
//Identify areas where the model shows substantial improvement over past methods
improv_areas ← IdentifyAreasOfSubstantialImprovement()
//Discuss the implications of the results and potential applications of the findings.
implic_and_applic ← DiscussImplicationsAndApplications()
Step 6. End

XGBoost, a robust gradient-boosting framework, is selected for its proficiency in handling structured data and delivering accurate predictions. The methodology includes three phases: (1) Data preprocessing, where raw data, including chest X-ray images, is cleaned and normalized; (2) Model training, where XGBoost is trained on preprocessed data, optimizing parameters for better performance; and (3) Evaluation, where the model is tested on a separate dataset, and metrics like precision, recall, and F1-score are computed.

3.1. Phase I: Data Acquisition and Implementation Tool

3.1.1. Data Collection and Preprocessing

For this study, we have acquired the latest dataset from GitHub, containing X-ray images of COVID-19 (pneumonia) cases. The dataset comprises over 800 chest X-ray images, encompassing both COVID-19 cases and other pneumonia cases [55]. Previous research studies have also utilized this dataset for classification purposes.

3.1.2. Tool Selection: Python and Anaconda Jupyter Notebook

Python, coupled with Anaconda Jupyter Notebook, serves as the primary tool for this investigation. Python's versatility in data mining, artificial intelligence (AI), and particularly in machine learning makes it an ideal choice. Anaconda Jupyter Notebook, an open-source web-based application, offers an intuitive

interface conducive to efficient coding and visualization. Its compatibility with Python enhances the ease of implementation, highlighting Python's widespread adoption in machine learning tasks.

3.1.3. Importing Required Libraries

Specialized libraries are essential for performing various functions within the study. Key among these is Keras, utilized for constructing neural network models in deep learning applications. Additionally, other libraries are imported for data preprocessing and visualization. Numpy, renowned for its capabilities in numerical and matrix analysis, aids in data manipulation. Meanwhile, Matplotlib and Seaborn serve as indispensable data visualization libraries, facilitating the presentation of data in visual formats, thereby enhancing comprehension and analysis.

3.2. Phase II: Data Processing and Model Analysis

3.2.1. Data Plotting

To gain insights into the COVID-19 pneumonia dataset, effective data plotting techniques are essential. Various visualization methods such as histograms, boxplots, scatter plots, and dimensionality reduction techniques like Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) are employed to visualize feature distribution, relationships, and intricate patterns within the data. Grids of example images are utilized to highlight characteristic features, facilitating a comprehensive understanding of the dataset [56].

Graphical representation of data is paramount in the analysis process. To achieve this, we utilized OpenCV, Matplotlib, and Seaborn libraries, which offer robust capabilities for data visualization. These libraries enable the creation of visually appealing and informative plots. As illustrated in Figure 2, an X-ray image is displayed for normal and abnormal classification, providing a visual representation of the dataset. Our dataset comprises X-ray images sourced from GitHub, forming the cornerstone of our research endeavors.

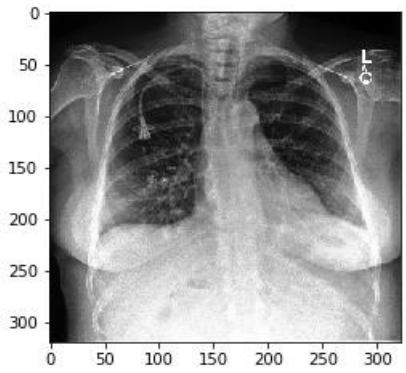

Figure 2. Chest X-ray Image Depicting Pneumonia Affliction

Following data visualization, the subsequent step involves building an XGBoost model for data analysis. XGBoost is a powerful machine learning algorithm known for its efficiency and performance in handling structured data. Unlike convolutional neural networks (CNNs), which are primarily used for image processing tasks, XGBoost excels in tabular data analysis. By leveraging its ensemble learning approach and gradient boosting framework, XGBoost enables the creation of predictive models capable of capturing complex relationships within the data [57].

3.2.2. Supervised Learning with XGBoost

Supervised learning with XGBoost begins with a well-defined training dataset comprising input features and corresponding output labels. In the context of applications like COVID-19 pneumonia classification using chest X-ray images, the input features typically represent pixel values extracted from the X-ray images. These pixel values serve as essential descriptors of the image's characteristics. On the other hand, the output labels indicate the classification of the image, specifically whether it represents COVID-19 pneumonia or not. This supervised learning approach enables the XGBoost algorithm to learn the underlying patterns and relationships between input features and output labels, facilitating accurate classification tasks in various domains, including medical imaging analysis [58].

- *XGBoost Classifier*

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm widely used in classification tasks, including COVID-19 pneumonia classification using chest X-ray images. XGBoost improves classification accuracy by iteratively training decision trees, with each tree correcting errors from previous ones. This boosting process refines predictions, resulting in high accuracy. XGBoost is well-suited for complex, high-dimensional datasets, making it ideal for medical imaging tasks like identifying patterns of COVID-19 pneumonia, such as opacities and infiltrates in the lungs. It handles missing data and outliers effectively, offers interpretable results, and is highly customizable through Hyperparameters tuning. Its versatility and robustness make XGBoost a valuable tool in COVID-19 diagnosis and research.

3.2.3. Data Augmentation

Data augmentation is a key technique in machine learning, particularly for classifying COVID-19 pneumonia using chest X-rays. Random Forest models, enhance the dataset by generating synthetic data through methods like rotation, flipping, scaling, cropping, brightness/contrast adjustment, and noise addition. This improves model generalization, mitigates overfitting, and enhances performance without needing extra real-world samples.

- *Data Augmentation Techniques in XGBoost*

Data augmentation enhances XGBoost model performance, especially in COVID-19 pneumonia classification using chest X-ray images. Unlike deep learning models like CNN, XGBoost benefits from diverse training data. Augmentation techniques—such as rotation, flipping, scaling, cropping, brightness/contrast adjustment, and noise addition—generate synthetic data, improving model generalization and accuracy on unseen images. This approach boosts performance, mitigates overfitting, and eliminates the need for additional real-world samples.

3.2.4. Feature Scaling for Optimal XGBoost Performance

Feature scaling is a critical preprocessing step to optimize the performance of XGBoost in COVID-19 pneumonia classification, especially when dealing with chest X-ray images. Unlike deep learning models that automatically handle feature scaling, XGBoost requires manual preprocessing to standardize the range of independent variables or features in the dataset. Feature scaling ensures uniformity in scale across all features, preventing certain features from dominating others. In the case of chest X-ray images, feature scaling involves scaling pixel intensity values to a smaller range, such as 0 to 1 or -1 to 1. Additionally, demographic or clinical features such as age, body temperature, and blood oxygen levels are scaled to bring them to a comparable range. This preprocessing aids the XGBoost algorithm in effectively learning from diverse features and improves its performance and robustness in accurately classifying COVID-19 pneumonia cases based on chest X-ray images.

3.3. Phase III: Analysis of Model Classification and Performance

3.3.1. Model Training Using the XGBoost Technique

In this phase, we apply the XGBoost technique for COVID-19 pneumonia classification. Data preparation begins with assembling a labeled chest X-ray dataset, distinguishing between COVID-19 and non-COVID-19 pneumonia. The dataset is split into training and testing subsets. Feature extraction identifies relevant characteristics from the X-ray images, serving as input for the XGBoost model. The model, using an ensemble of decision trees, predicts the output class, enhancing accuracy. Model evaluation is performed using metrics like accuracy, precision, recall, and F1-score, with parameter tuning for optimization.

3.3.2. XGBoost-Based Classification and Prediction Outcomes

This section presents a detailed analysis of the classification and prediction results obtained using the XGBoost model. Figure 3 illustrates the classification and prediction results derived from the proposed XGBoost model.

This visualization evaluates the XGBoost model's effectiveness in classifying COVID-19 pneumonia cases using accuracy, precision, recall, and F1-score. Real-time validation on chest X-ray images confirms its performance in clinical settings, ensuring reliability in real-world scenarios. Detailed interpretation of the model's decision-making process highlights key features contributing to accurate predictions. This comprehensive analysis demonstrates the efficacy of XGBoost in distinguishing between normal and abnormal pneumonia cases, providing accurate predictions based on image patterns and features.

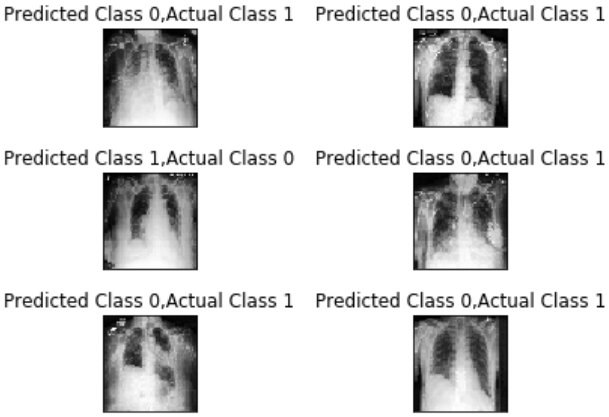

Figure 3. Classification and Prediction results derived from the proposed model.

3.3.3. Assessing Model Performance through Confusion Matrix Analysis

A confusion matrix is essential for evaluating a classification algorithm's performance. It offers insights into the algorithm's accuracy and the types of errors it generates, helping identify areas for improvement and overall effectiveness.

- *Performance Evaluation Using Confusion Matrix*

The confusion matrix is a key tool in machine learning for evaluating model performance. It compares actual and predicted labels, highlighting classification accuracy and error types like false positives and false negatives. This detailed analysis aids in refining models, improving diagnostic accuracy, and enhancing clinical interpretation.

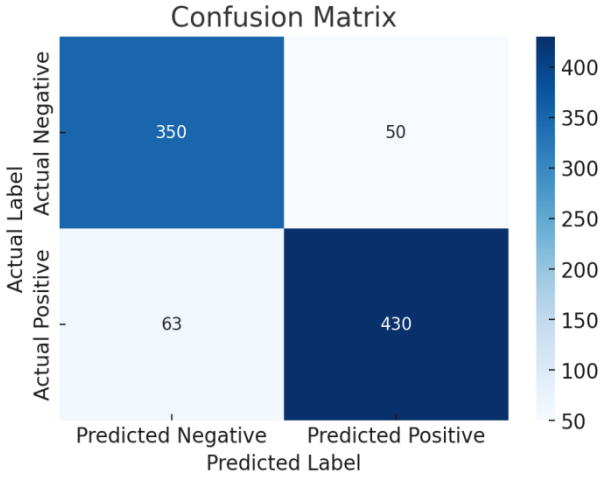

Figure 4. Confusion Matrix Generated by XGBoost

Figure 4 displays the confusion matrix produced by the XGBoost model, showing actual and predicted labels, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These values are essential for calculating performance metrics and assessing the model's predictive accuracy.

- *Confusion Matrix Components*

The Confusion Matrix, consisting of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), reflects classifier performance:

TP: Accurate positive predictions (430)

TN: Accurate negative predictions (350)

FP: Negative cases misclassified as positive (50)

FN: Positive cases misclassified as negative (63)

Using these values, the model's precision, recall, and accuracy are calculated. This analysis provides a comprehensive evaluation of the proposed XGBoost model's effectiveness in classification tasks, highlighting areas for potential optimization.

4. Results and Discussion

4.1. Phase IV: Assessing Performance and Conducting Comparative Analysis

This phase evaluates the accuracy of the proposed XGBoost model, comparing it with existing studies. Key performance metrics, including average accuracy, recovery rates, and F1 scores, are calculated to assess the model's precision and recall. These metrics provide valuable insights into the model's effectiveness and ability to correctly classify data instances.

4.1.1. Average Accuracy Score to Assess Model Performance

The average accuracy score is a key metric for evaluating the XGBoost model. It combines precision, recall, and F1 scores to assess the model's effectiveness. Precision measures the percentage of correctly classified relevant results, while recall reflects the proportion of relevant results identified. The F1 score, which balances precision and recall, offers a comprehensive performance measure. These metrics are derived from the confusion matrix, which summarizes the model's classification performance across different categories.

4.1.2. Evaluation Metrics for COVID-19 Pneumonia Classification using XGBoost

Evaluating the performance of the proposed XGBoost model for COVID-19 pneumonia classification is crucial for determining its diagnostic effectiveness. Key metrics, including accuracy, precision, recall, and F1-score, are essential for this assessment.

- *Performance Metrics*

Accuracy (AC): Represents the proportion of correctly classified cases, calculated as:

$$Accuracy(AC) = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Here, TP (True Positive) and TN (True Negative) denote correctly classified COVID-19 pneumonia cases and non-COVID-19 cases, respectively.

Precision (PR): Precision measures the proportion of true positive predictions for COVID-19 pneumonia. It is calculated using the formula:

$$Precision\ (PR) = \frac{TP}{TP+FP} \quad (2)$$

Recall (RE): Recall quantifies the proportion of true positive predictions within all actual COVID-19 cases. It is calculated using the formula:

$$Recall\ (RE) = \frac{TP}{TP+FN} \quad (3)$$

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both metrics. It is calculated using the formula:

$$F1\ Score = 2 \times \frac{PR \times RE}{PR+RE} \quad (4)$$

- *Model Evaluation Results*

Utilizing these formulas, our XGBoost model's performance metrics were derived from the confusion matrix. Precision (PR) achieved 89.58%, recall (RE) attained 87.22%, F1-Score manifested as 88.39%, and accuracy (AC) reached 87.35%.

- *Interpretation of Results*

These results underscore the effectiveness of the XGBoost model, with an accuracy of 87.35% indicating its ability to correctly classify cases in the test set. The F1-score of 88.39% showcases a balanced measure of precision and recall, indicating robust performance in identifying positive cases while minimizing false positives, as shown in Table 1.

Table 1. Performance Measures of the Model Showing Accuracy, Precision, Recall, and F1 Score

Algorithm	Precision	Recall	F1 Score	Accuracy
XGBoost Classifier	89.58%	87.22%	88.39%	87.35%

- *Discussion on Performance Measures*

The high precision of 89.58% implies a significant likelihood of accurate positive case predictions, crucial for mitigating unnecessary interventions or misdiagnoses. However, the recall of 87.22% highlights a potential oversight of true positive cases, suggesting the need for further refinements to enhance sensitivity and improve positive case detection.

The XGBoost model demonstrates promising performance in COVID-19 pneumonia classification, but continuous refinement and optimization are necessary for further enhancing its diagnostic accuracy and reliability.

4.1.3. Performance Analysis of the Proposed Model Utilizing the XGBoost Technique

The performance analysis of our proposed model, which utilizes the XGBoost technique, is a critical aspect of evaluating its effectiveness in classifying COVID-19 pneumonia cases. The model has demonstrated promising performance, achieving an impressive overall accuracy of 87.35%. This accuracy metric, as presented in Table 1, indicates the proportion of correctly classified cases out of the total cases evaluated.

- *Precision, Recall, and F1-Score Analysis*

The precision of 89.58% indicates that a vast majority (89.58%) of predicted positive cases are true positives. This high precision minimizes the risk of misdiagnoses and unnecessary interventions, ensuring that positive cases are accurately identified.

With a recall of 87.22%, there is a possibility that a small portion (12.78%) of true positive cases might be overlooked. While the majority of positive cases are detected, a small fraction may go undetected. Investigating potential reasons for this discrepancy, such as data characteristics or model architecture, could lead to improvements in sensitivity and ensure that all positive cases are identified.

The F1-score of 88.39% serves as a balanced measure of precision and recall, highlighting the model's effectiveness in accurately classifying both normal and abnormal cases. This balanced metric demonstrates the model's ability to maintain a high level of precision while also effectively capturing positive cases.

- *Clarification of Normal vs. Abnormal Classification*

It's important to clarify that our model categorizes each image as normal or abnormal, rather than explicitly distinguishing between "normal" and "pneumonia" cases. This distinction ensures that the model focuses on identifying any abnormalities present in the images, including those indicative of pneumonia, without making explicit differentiations within the "abnormal" category.

Figure 5 shows the distribution of Abnormal (Pneumonia) and Normal X-ray images in the dataset, with 493 classified as Abnormal and 400 as Normal out of 893 images. This data underpins the model's accuracy evaluation.

Classifying these images provides insights into the model's performance, particularly in balancing false positives and false negatives. A high false positive rate leads to unnecessary interventions, while a high false negative rate risks missed diagnoses, guiding further model refinement.

The XGBoost model significantly contributes to COVID-19 pneumonia classification, demonstrating high accuracy. While enhancing recall is needed, the overall performance suggests promising potential for improving diagnostic accuracy and patient care. Further research in AI-powered medical diagnosis remains promising.

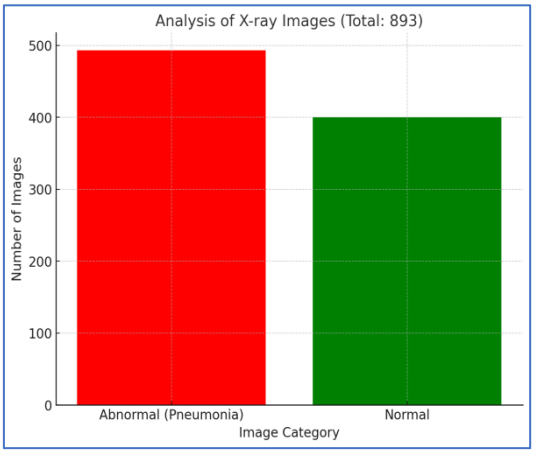

Figure 1. Distribution of Abnormal (Pneumonia) and Normal X-ray Images

4.1.4. Comparative Analysis with Existing Research

In our study on COVID-19 pneumonia recognition, we conduct a comprehensive comparative analysis with methodologies proposed by NourEldeen M. Khalifa et al. [9] and Ieracitano, Cosimo, et al. [59], as shown in Table 2.

Table 2. Comparative Analysis of Proposed and Existing Research Models for X-ray Image Classification

Authors	Data Set	Algorithm	Accuracy
Khalifa, N. M., et al. [9]	X-ray Images	Alexnet, Googlenet, and Restnet18	78.70%
Ieracitano, Cosimo, et al. [59]	X-ray Images	Fuzzy-CovNNet	81.00%
Our proposed research work	X-ray Images	XGBoost Classifier	87.35%

Khalifa et al. applied deep transfer learning techniques using AlexNet, GoogleNet, and ResNet18 for pneumonia detection in X-ray images, focusing on binary classification between normal and abnormal cases. Their approach achieved 78.70% accuracy, primarily targeting pneumonia diagnosis.

Similarly, Ieracitano et al. developed the CovNNet model, integrating fuzzy logic with deep learning to distinguish COVID-19 pneumonia from other interstitial pneumonia types, achieving 81% accuracy through image features and fuzzy edge data from various datasets.

Our proposed XGBoost model surpasses these methodologies, achieving 87.35% accuracy, significantly outperforming Khalifa et al. and Ieracitano et al., as presented in Table 2. Additionally, the XGBoost model reduced false positives, achieving an F1-score of 88.39%, demonstrating superior classification performance, particularly for COVID-19 pneumonia. While our model shows slightly lower recall than Ieracitano et al.'s, its specialized focus on distinguishing normal from abnormal pneumonia cases, especially in the context of COVID-19, enhances precision and overall performance.

This study advances pneumonia classification by achieving high accuracy and precision, filling a research gap for vulnerable populations. The XGBoost model, developed using Python and Jupyter Notebook, ensures data quality, optimizes feature extraction, and enhances performance, offering a reproducible framework for future investigations, particularly in COVID-19 pneumonia classification.

5. Conclusion and Future Directions

This research addresses the critical challenge of COVID-19 pneumonia classification by developing and applying an optimized XGBoost algorithm. The study's comprehensive four-phase methodology, covering data acquisition, processing, analysis, and model evaluation, effectively demonstrates the efficacy of XGBoost in distinguishing between normal and abnormal pneumonia cases. The model's precision and recall capabilities are underscored through a comprehensive analysis of key performance metrics, including average accuracy scores, recovery rates, and F1 scores. Achieving an accuracy rate of 87.35%, the model exhibits proficiency in accurately classifying relevant cases, surpassing the performance of existing methodologies. This success can be attributed to the careful consideration of technical intricacies, such as architecture design and training specifics, which enhance the model's reliability and reproducibility. The systematic framework presented in this study not only advances COVID-19 pneumonia classification but also sets the stage for future investigations, especially concerning vulnerable populations. Acting as a cornerstone, this research enhances diagnostic accuracy and patient care in the COVID-19 pneumonia classification. Ongoing research endeavors are directed towards refining and integrating deep learning algorithms into clinical practice, aiming to support informed public health interventions.

Future research should prioritize enhancing sensitivity, improving recall rates, and validating datasets on a larger scale to enhance model performance. Efforts should also focus on broadening the model's applicability across diverse demographics and healthcare settings by integrating real-time data for continuous optimization and adaptation.

References

1. Shi, Feng, Liming Xia, Fei Shan, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, He Sui, and Dinggang Shen. "Large-Scale Screening of COVID-19 from Community Acquired Pneumonia Using Infection Size-Aware Classification." 2020.
2. Wang, Kefan, and Yanping Shi. "Discussion on the Syndrome Differentiation Treatment of Pneumonia and Cough in Children." MEDS Chinese Medicine 5, no. 8 (2023): 121-126.
3. File Jr, Thomas M., and Julio A. Ramirez. "Community-Acquired Pneumonia." New England Journal of Medicine 389, no. 7 (2023): 632-641.
4. Yu, Xinxin, Shuai Zhang, Jingxu Xu, Yong Huang, Hao Luo, Chencui Huang, Pei Nie, et al. "Nomogram Using CT Radiomics Features for Differentiation of Pneumonia-Type Invasive Mucinous Adenocarcinoma and Pneumonia: Multicenter Development and External Validation Study." American Journal of Roentgenology 220, no. 2 (2023): 224-234.
5. Ma, Hai-Ran, Bi-Ying Deng, Jing Liu, Peng Jiang, Yan-Lei Xu, Xiu-Yun Song, Jie Li, et al. "Lung Ultrasound to Diagnose Infectious Pneumonia of Newborns: A Prospective Multicenter Study." Pediatric Pulmonology 58, no. 1 (2023): 122-129.
6. Alshammari, Mohammed Kanan, Mzoun Abdulaziz Alotaibi, Ahad Sanad AlOtaibi, Hanan Tareq Alosaime, Mona Awadh Aljuaid, Budur Mohammed Alshehri, et al. "Prevalence and Etiology of Community-and Hospital-Acquired Pneumonia in Saudi Arabia and Their Antimicrobial Susceptibility Patterns: A Systematic Review." Medicina 59, no. 4 (2023): 760.
7. Kakde, Aditya, Nitin Arora, Durgansh Sharma, and Subhashchander Sharma. "Multi-Spectral Classification and Recognition of Breast Cancer and Pneumonia." Polish Journal of Medical Physics and Engineering 26, no. 1 (2020). https://doi.org/10.2478/pjmpe-2020-0001.
8. Chen, Kuan-Fu, Tsai-Wei Feng, Chin-Chieh Wu, Ismaeel Yunusa, Su-Hsun Liu, Chun-Fu Yeh, and Shih-Tsung Han, et al. "Diagnostic Accuracy of Clinical Signs and Symptoms of COVID-19: A Systematic Review and Meta-Analysis to Investigate the Different Estimates in a Different Stage of the Pandemic Outbreak." Journal of Global Health 13 (2023).
9. Khalifa, M. NE, F. Smarandache, and M. Loey. "COVID-19 Chest X-Ray Images Diagnosis: A Neutrosophic and Deep Transfer Learning Approach." 2020.
10. Xu, Xiaowei, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, et al. "A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia." Engineering 6, no. 10 (2020): 1122-1129.
11. Moorthy, Chinnadurai Ganesa, and Ganesamoorthy Udhaya Sankar. "Analysis on Electromagnetic Waves of CT Scanners and MRI Scanners for Applications." World Scientific News 188 (2024): 1-14.
12. Zhou, Min, Dexiang Yang, Yong Chen, Yanping Xu, Jin-Fu Xu, Zhijun Jie, Weiwu Yao, et al. "Deep Learning for Differentiating Novel Coronavirus Pneumonia and Influenza Pneumonia." Annals of Translational Medicine 9, no. 2 (2021).
13. Jung, Sung-mok, Ryo Kinoshita, Robin N. Thompson, Natalie M. Linton, Yichi Yang, Andrei R. Akhmetzhanov, and Hiroshi Nishiura. "Epidemiological Identification of a Novel Pathogen in Real Time: Analysis of the Atypical Pneumonia Outbreak in Wuhan, China, 2019–2020." Journal of Clinical Medicine 9, no. 3 (2020): 637.
14. Loey, Mohamed, Florentin Smarandache, and Nour Eldeen M. Khalifa. "Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning." Symmetry 12, no. 4 (2020): 651.
15. Janizek, Joseph D., Gabriel Erion, Alex J. DeGrave, and Su-In Lee. "An Adversarial Approach for the Robust Classification of Pneumonia from Chest Radiographs." In Proceedings of the ACM Conference on Health, Inference, and Learning, 69-79. 2020.

16. Qomariyah, Nunung Nurul, Ardimas Andi Purwita, Maria Seraphina Astriani, Sri Dhuny Atas Asri, and Dimitar Kazakov. "An XGBoost Model for Age Prediction from COVID-19 Blood Test." In 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 446-452. IEEE, 2021.
17. Dong, Chunjiao, Yixian Qiao, Chunheng Shang, Xiwen Liao, Xiaoning Yuan, Qin Cheng, Yuxuan Li, et al. "Non-contact Screening System Based for COVID-19 on XGBoost and Logistic Regression." Computers in Biology and Medicine 141 (2022): 105003.
18. Sharifi-Kia, Ali, Azin Nahvijou, and Abbas Sheikhtaheri. "Machine Learning-based Mortality Prediction Models for Smoker COVID-19 Patients." BMC Medical Informatics and Decision Making 23, no. 1 (2023): 129.
19. Chang, Tu-Hsuan, Yun-Chung Liu, Siang-Rong Lin, Pei-Hsin Chiu, Chia-Ching Chou, Luan-Yin Chang, and Fei-Pei Lai. "Clinical Characteristics of Hospitalized Children with Community-acquired Pneumonia and Respiratory Infections: Using Machine Learning Approaches to Support Pathogen Prediction at Admission." Journal of Microbiology, Immunology and Infection 56, no. 4 (2023): 772-781.
20. Shaheed, Kashif, Qasiar Abbas, and Munish Kumar. "Automatic Diagnosis of CoV-19 in CXR Images Using Haar-like Feature and XgBoost Classifier." Multimedia Tools and Applications (2024): 1-23.
21. El-Ghandour, Mohammed, and Marwa Ismael Obayya. "Pneumonia Detection in Chest X-ray Images Using an Optimized Ensemble with XGBoost Classifier." Multimedia Tools and Applications (2024): 1-31.
22. Dong, Chunjiao, Yixian Qiao, Chunheng Shang, Xiwen Liao, Xiaoning Yuan, Qin Cheng, Yuxuan Li, et al. "Non-contact Screening System Based for COVID-19 on XGBoost and Logistic Regression." Computers in Biology and Medicine 141 (2022): 105003.
23. Liu, Qin, Baoguo Pang, Haijun Li, Bin Zhang, Yumei Liu, Lihua Lai, Wenjun Le, et al. "Machine Learning Models for Predicting Critical Illness Risk in Hospitalized Patients with COVID-19 Pneumonia." Journal of Thoracic Disease 13, no. 2 (2021): 1215.
24. Sun, Chaoyang, Yong Bai, Dongsheng Chen, Liang He, Jiacheng Zhu, Xiangning Ding, and Lihua Luo, et al. "Accurate Classification of COVID-19 Patients with Different Severity via Machine Learning." Clinical and Translational Medicine 11, no. 3 (2021).
25. Song, Xianbin, Jiangang Zhu, Xiaoli Tan, Wenlong Yu, Qianqian Wang, Dongfeng Shen, and Wenyu Chen. "XGBoost-based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers." Frontiers in Public Health 10 (2022): 926069.
26. Chang, Tu-Hsuan, Yun-Chung Liu, Siang-Rong Lin, Pei-Hsin Chiu, Chia-Ching Chou, Luan-Yin Chang, and Fei-Pei Lai. "Clinical Characteristics of Hospitalized Children with Community-Acquired Pneumonia and Respiratory Infections: Using Machine Learning Approaches to Support Pathogen Prediction at Admission." Journal of Microbiology, Immunology and Infection 56, no. 4 (2023): 772-781.
27. Dumakude, Aphelele, and Absalom E. Ezugwu. "Automated COVID-19 Detection with Convolutional Neural Networks." Scientific Reports 13, no. 1 (2023): 10607.
28. Jennifer, J. Sofia, and T. Sree Sharmila. "A Neutrosophic Set Approach on Chest X-rays for Automatic Lung Infection Detection." Information Technology and Control 52, no. 1 (2023): 37-52.
29. Yagin, Fatma Hilal, İpek Balikci Cicek, Abedalrhman Alkhateeb, Burak Yagin, Cemil Colak, Mohammad Azzeh, and Sami Akbulut. "Explainable Artificial Intelligence Model for Identifying COVID-19 Gene Biomarkers." Computers in Biology and Medicine 154 (2023): 106619.
30. Sihananto, Andreas Nugroho, Eristya Maya Safitri, Arif Widiasan Subagio, Muhammad Dafa Ardiansyah, and Aditya Primayudha. "Classification of COVID-19 RT-PCR Test Results Using Auto-encoder and Random Forest." Nusantara Science and Technology Proceedings (2023): 237-243.

31. de Jong, Tristan, Victor Guryev, and Yury M. Moshkin. "Discovery of Pharmaceutically-Targetable Pathways and Prediction of Survivorship for Pneumonia and Sepsis Patients from the Viewpoint of Ensemble Gene Noise." bioRxiv (2020): 2020-04.
32. Neili, Zakaria, Mohamed Fezari, and Abdelghani Redjati. "ELM and K-nn Machine Learning in Classification of Breath Sounds Signals." International Journal of Electrical and Computer Engineering 10, no. 4 (2020): 3528-3536.
33. Varshni, Dimpy, Kartik Thakral, Lucky Agarwal, Rahul Nijhawan, and Ankush Mittal. "Pneumonia Detection Using CNN-Based Feature Extraction." In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1-7. IEEE, 2019.
34. Zhao, Peng, Jia-Wei Shan, Yu-Jie Zhang, and Zhi-Hua Zhou. "Exploratory Machine Learning with Unknown Unknowns." Artificial Intelligence 327 (2024): 104059.
35. Khalifa, Nour Eldeen M., Mohamed Hamed N. Taha, Aboul Ella Hassanien, and Sally Elghamrawy. "Detection of Coronavirus (COVID-19) Associated Pneumonia Based on Generative Adversarial Networks and a Fine-Tuned Deep Transfer Learning Model Using Chest X-ray Dataset." In International Conference on Advanced Intelligent Systems and Informatics, 234-247. Cham: Springer International Publishing, 2022.
36. Hall, Lawrence O., Rahul Paul, Dmitry B. Goldgof, and Gregory M. Goldgof. "Finding COVID-19 from Chest X-rays Using Deep Learning on a Small Dataset." arXiv preprint arXiv:2004.02060 (2020).
37. Ilyas, Muhammad, Hina Rehman, and Amine Naït-Ali. "Detection of COVID-19 from Chest X-ray Images Using Artificial Intelligence: An Early Review." arXiv preprint arXiv:2004.05436 (2020).
38. Evans, Ethan D., Claire Duvallet, Nathaniel D. Chu, Michael K. Oberst, Michael A. Murphy, Isaac Rockafellow, David Sontag, and Eric J. Alm. "Predicting Human Health from Biofluid-Based Metabolomics Using Machine Learning." Scientific Reports 10, no. 1 (2020): 17635.
39. Nagamounika, R., C. N. S. V. Sri, A. Harshitha, K. L. Tejaswi, and P. R. S. M. Lakshmi. "Prediction of Pneumonia Disease by Using Deep Convolutional Neural Networks." Journal of Engineering Sciences Criterion 17 (2020): 18.
40. Liebenlito, Muhaza, Yanne Irene, and Abdul Hamid. "Classification of Tuberculosis and Pneumonia in Human Lungs Based on Chest X-ray Image Using Convolutional Neural Network." InPrime: Indonesian Journal of Pure and Applied Mathematics 2, no. 1 (2020): 24-32.
41. Yao, Shangjie, Yaowu Chen, Xiang Tian, Rongxin Jiang, and Shuhao Ma. "An Improved Algorithm for Detecting Pneumonia Based on YOLOv3." Applied Sciences 10, no. 5 (2020): 1818.
42. Chadha, Raman. "A Novel Approach for Detecting Pneumonia in Machine Learning." International Journal of Trend in Innovative Research (2019).
43. Chumbita, Mariana, Catia Cillóniz, Pedro Puerta-Alcalde, Estela Moreno-García, Gemma Sanjuan, Nicole Garcia-Pouton, Alex Soriano, Antoni Torres, and Carolina Garcia-Vidal. "Can Artificial Intelligence Improve the Management of Pneumonia?" Journal of Clinical Medicine 9, no. 1 (2020): 248.
44. Ebner, Lukas, Stergios Christodoulidis, Thomai Stathopoulou, Thomas Geiser, Odile Stalder, Andreas Limacher, Johannes T. Heverhagen, Stavroula G. Mougiakakou, and Andreas Christe. "Meta-Analysis of the Radiological and Clinical Features of Usual Interstitial Pneumonia (UIP) and Nonspecific Interstitial Pneumonia (NSIP)." PloS One 15, no. 1 (2020): e0226084.
45. Shuaib, Khan Maseeh, Solkar Ahmed Shahid, Ansari Almas Javed, and Mohammed Zaid. "Pneumonia Detection Through X-ray Using Deep Learning." IOSR Journal of Computer Engineering 22, no. 1 (2020): 08-11.
46. Li, Lin, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, et al. "Artificial Intelligence Distinguishes COVID-19 from Community-Acquired Pneumonia on Chest CT." Radiology (2020).
47. Liao, Yu-Hsuan, Zhong-Chuang Wang, Fu-Gui Zhang, Maysam F. Abbod, Chung-Hung Shih, and Jiann-Shing Shieh. "Machine Learning Methods Applied to Predict Ventilator-Associated Pneumonia with Pseudomonas aeruginosa Infection via Sensor Array of Electronic Nose in Intensive Care Unit." Sensors 19, no. 8 (2019): 1866.

48. Feng, Cong, Lili Wang, Xin Chen, Yongzhi Zhai, Feng Zhu, Hua Chen, Yingchan Wang, et al. "A Novel Triage Tool of Artificial Intelligence-Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics." MedRxiv (2020): 2020-03.
49. Minnoor, Manas, and Veeky Baths. "Diagnosis of Breast Cancer Using Random Forests." Procedia Computer Science 218 (2023): 429-437.
50. Prasojo, Rahman Azis, Muhammad Akmal A. Putra, Meyti Eka Apriyani, Anugrah Nur Rahmanto, Sherif S. M. Ghoneim, Karar Mahmoud, Matti Lehtonen, and Mohamed M. F. Darwish. "Precise Transformer Fault Diagnosis via Random Forest Model Enhanced by Synthetic Minority Over-Sampling Technique." Electric Power Systems Research 220 (2023): 109361.
51. Shaheed, Kashif, Piotr Szczuko, Qaisar Abbas, Ayyaz Hussain, and Mubarak Albathan. "Computer-Aided Diagnosis of COVID-19 from Chest X-ray Images Using Hybrid Features and Random Forest Classifier." In Healthcare, vol. 11, no. 6, p. 837. MDPI, 2023.
52. Yenurkar, Ganesh Keshaorao, Sandip Mal, Vincent O. Nyangaresi, Anshul Hedau, Prajwal Hatwar, Shreyas Rajurkar, and Juli Khobragade. "Multifactor Data Analysis to Forecast an Individual's Severity over the Novel COVID-19 Pandemic Using Extreme Gradient Boosting and Random Forest Classifier Algorithms." Engineering Reports 5, no. 12 (2023): e12678.
53. Khanna, Varada Vivek, Krishnaraj Chadaga, Niranjana Sampathila, Srikanth Prabhu, and Rajagopala Chadaga. "A Machine Learning and Explainable Artificial Intelligence Triage-Prediction System for COVID-19." Decision Analytics Journal 7 (2023): 100246.
54. Shamma, Aashitha L., Susmitha Vekkot, Deepa Gupta, Mohammed Zakariah, and Yousef Ajami Alotaibi. "Development of a Non-Invasive COVID-19 Detection Framework Using Explainable AI and Data Augmentation." Journal of Intelligent & Fuzzy Systems Preprint (2024): 1-14.
55. Bhosale, Yogesh H., and K. Sridhar Patnaik. "PulDi-COVID: Chronic Obstructive Pulmonary (Lung) Diseases with COVID-19 Classification Using Ensemble Deep Convolutional Neural Network from Chest X-Ray Images to Minimize Severity and Mortality Rates." Biomedical Signal Processing and Control 81 (2023): 104445.
56. Garcea, Fabio, Alessio Serra, Fabrizio Lamberti, and Lia Morra. "Data Augmentation for Medical Imaging: A Systematic Literature Review." Computers in Biology and Medicine 152 (2023): 106391.
57. Ni, Ansong, Srini Iyer, Dragomir Radev, Veselin Stoyanov, Wen-tau Yih, Sida Wang, and Xi Victoria Lin. "Lever: Learning to Verify Language-to-Code Generation with Execution." In International Conference on Machine Learning, 26106-26128. PMLR, 2023.
58. Tyagi, Amit Kumar, ed. Automated Secure Computing for Next-Generation Systems. John Wiley & Sons, 2023.
59. Ieracitano, Cosimo, Nadia Mammone, Mario Versaci, Giuseppe Varone, Abder-Rahman Ali, Antonio Armentano, Grazia Calabrese, et al. "A Fuzzy-Enhanced Deep Learning Approach for Early Detection of COVID-19 Pneumonia from Portable Chest X-Ray Images." Neurocomputing 481 (2022): 202-215.