Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

A Hybrid Approach for COVID-19 Pneumonia Detection and Classification using Machine and Deep Learning Techniques

Fazal Malik*
Department of Computer Science, Iqra National University Peshawar, Khyber Pakhtunkhwa, Pakistan. Corresponding Author Email: fazal.malik@inu.edu.pk

Muhammad Inam ul Haq
Department of Computer Science & Bioinformatics, Khushal Khan Khattak University, Karak, Pakistan. muhammad.inamulhaq@kkkuk.edu.pk

Afsheen Khalid
Center for Excellence in IT, Institute of Management Sciences, Peshawar, Khyber Pakhtunkhwa, Pakistan. afsheen.khalid@imsciences.edu.pk

Dilawar Khan
Computer Science & IT Department, University of Engineering and Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan. dilawarkhan@uetpeshawar.edu.pk

Muhammad Javed
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa, Pakistan. drjaved@ustb.edu.pk

Ashraf Ullah
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa, Pakistan. ashrafbth@gmail.com

Muhammad Qasim Khan
Department of Computer Science, Iqra National University Peshawar, Khyber Pakhtunkhwa, Pakistan. m.qasim@inu.edu.pk

Abstract
Pneumonia, including COVID-19 pneumonia, remains a significant global health challenge, with early and accurate diagnosis being essential for effective management. Traditional diagnostic methods often struggle with low accuracy and limited adaptability to evolving disease patterns. The COVID-19 pandemic highlighted the need for more robust, adaptive diagnostic approaches.This study develops an integrated machine learning (ML) framework for classifying COVID-19 pneumonia using Random Forest, AdaBoost, XGBoost, and Convolutional Neural Networks (CNNs). The research aims to optimize these algorithms, improve data preprocessing, and evaluate their performance compared to traditional methods.The proposed framework consists of four phases: (1) Dataset Acquisition: Utilization of a GitHub dataset, processed in Python and Anaconda Jupyter Notebook; (2) Data Processing and Analysis: Histograms and scatter plots for dataset preprocessing; (3) Model Application and Optimization: Random Forest and AdaBoost for ensemble learning, XGBoost with data augmentation for enhanced accuracy, and CNNs for extracting intricate patterns from X-ray images; and (4) Performance Evaluation: A comparative analysis of

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

the integrated model against traditional methods to assess improvements in accuracy, reliability, and adaptability. Results demonstrate that XGBoost achieves the highest performance, with an accuracy of 87.35%, precision of 89.58%, recall of 87.22%, and an F1-score of 88.39%. CNN performs well in precision (88.60%), while AdaBoost provides balanced precision and recall. The integrated ML framework significantly enhances pneumonia classification accuracy, offering a promising tool for COVID-19 diagnostics. These findings emphasize the importance of model selection to optimize both precision and recall, potentially improving clinical decision-making and patient outcomes.

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

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].

Pneumonia encompasses various types, including bacterial, viral, walking, fungal, Hospital-acquired pneumonia (HAP), and Community-acquired pneumonia (CAP), each with unique clinical challenges. HAP is a severe hospital infection often associated with drug-resistant microorganisms and immunosuppressed patients [6]. Bacterial and viral pneumonia 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

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

symptoms include fever, cough, and shortness of breath Young people and men may be more susceptible to pneumonia, as patients often have higher breathing rates than healthy individuals [8]. X-ray imaging, revealing haziness and opacity in lung structures, is essential for diagnosing pneumonia [9]. Advancements in machine learning models and medical imaging have significantly improved the diagnosis and classification of COVID-19 and other pneumonia types, including CAP [10].

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 [11]. 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 [12]. As evidenced by the evaluation of deep learning models for identifying COVID-19, including ResNet, Inception, and GoogLeNet, robust frameworks are needed [13]. 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 [14].

An ensemble support vector kernel random forest-based hybrid equilibrium Aquila optimization (ESVMKRF-HEAO) approach presented for accurate skin cancer prediction. Utilizing the HAM10000 dataset, the model demonstrates robust classification into five categories [15].

A cloud-random forest (C-RF) model for CHD risk assessment, combining cloud model and random forest techniques introduced. The C-RF utilizes a weighted attribute determination algorithm based on decision-making trial and evaluation laboratory and cloud model. Empirical analysis performed using dataset from Kaggle demonstrates the effectiveness of random forest [16].

A deep forest-based model proposed for early-stage COVID-19 diagnosis from CXR images, highlighting the effectiveness of Random Forest technique [17]. AdaBoost (Adaptive Boosting), developed in 1997, employs weak classifiers and an exponential loss function to enhance performance, with further research exploring its varied applications and capabilities [18]. AdaBoost excels in classifying output classes from input variables, providing rapid and superior mapping, a notable improvement over techniques like decision trees, regression, and nearest neighbors [19]. Utilizing AdaBoost, this research enhances pneumonia detection amidst the COVID-19 epidemic, addressing symptoms such as fever and cough, primarily spread through saliva droplets. Other techniques are reviewed for comprehensive analysis [20]. AdaBoost, an Artificial Intelligence (AI) ensemble technique, improves epidemic spread factor classification by selecting relevant attributes objectively. Its utilization in COVID-19 growth rate classification experiments showcases effective containment measure correlation with growth factors [21]. Introducing a COVID-19 Prediction Support System is utilizing AdaBoost, among other Machine Learning techniques, integrated with routine blood testing, demonstrating enhanced predictive accuracy over conventional methods [22].

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 [23]. The XGBoost algorithm combined with logistic

ISSN Online: 3007-3154
ISSN Print: 3007-3146

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 [24]. 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 [25]. XGBoost also demonstrated potential in identifying pathogens from clinical data in children hospitalized with respiratory infections (2010–2018) [26]. Recent studies highlight its use in automating COVID-19 detection from chest X-rays [27]. 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 [28].

Pneumonia, particularly during the COVID-19 pandemic, poses significant public health challenges due to its diverse symptoms and the need for precise diagnosis and classification. Current diagnostic methods often rely on outdated algorithms, exhibit low accuracy, and lack adaptability to evolving disease patterns, emphasizing the need for innovative approaches.

This study introduces an integrated machine learning framework combining Random Forest, AdaBoost, XGBoost, and Convolutional Neural Networks (CNNs) for COVID-19 pneumonia classification. The proposed research framework comprises four phases: (1) Dataset Acquisition: The latest GitHub dataset is utilized, with Python and Anaconda Jupyter Notebook facilitating data collection; (2) Data Processing and Analysis: Histograms and scatter plots preprocess and analyze the dataset to enhance machine learning (ML) readiness; (3) Model Application and Optimization: Random Forest and AdaBoost provide robust ensemble learning foundations, XGBoost incorporates data augmentation for improved accuracy, and CNNs extract intricate patterns from X-ray images to enhance classification precision; and (4) Performance Evaluation: Comparative analysis evaluates the integrated model against existing methods, highlighting improvements in accuracy, reliability, and adaptability.

This integrated approach significantly advances ML applications for pneumonia diagnosis by leveraging the strengths of multiple algorithms. Random Forest and AdaBoost address data variability and imbalances, while XGBoost ensures efficient hyperparameter optimization and CNNs contribute deep feature extraction from medical images. These techniques collectively overcome challenges such as limited data availability and variability in disease presentation. The study's contributions include a comprehensive ML-based framework tailored for COVID-19 pneumonia classification. The proposed methodology optimally processes datasets, integrates data augmentation to enhance reliability, and demonstrates superior classification accuracy compared to traditional methods. These advancements offer valuable insights for predictive modeling and resource optimization, aiding public health strategies during pandemics and beyond. By combining diverse ML techniques, this framework ensures accurate and reliable disease classification while offering critical insights into future disease trends, enabling effective public health planning.

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

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

results in the Results and Discussion section, and a conclusion with recommendations for future research in the Conclusion section.

Literature Review

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

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 [29]. Random Forest's superior performance in breast tumor detection using the Wisconsin Breast Cancer Database was demonstrated in prior research [30]. 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 [31]. 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 [32, 33]. 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 [34].

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. [35]. 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 [36]. 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 [37].

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 [38]. 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 [39]. 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

ISSN Online: 3007-3154
ISSN Print: 3007-3146

feedforward neural network. The COVID-19 CT-2A (CovidxCT-2A) dataset is employed for evaluation [40]. 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 [41].

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) [42]. 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 [27].

An AdaBoost algorithm is introduced in a study, improving model performance by effectively handling rough knowledge and enhancing sensitivity and regression capacity [43]. Utilizing AdaBoost in 2019 Novel Coronavirus (nCoV) analysis, patient history, country, age, and gender are, integrated to enhance death prediction accuracy by over sevenfold, showcasing the potential of machine learning in early outcome estimation. This highlights the value of integrating disease history into predictive models, promising improved patient treatment and healthcare system relief [44]. Using AdaBoost, this study enhances COVID-19 case identification via symptom analysis, aiding in self-assessment and patient triage, while leveraging Non-dominated Sorting Genetic Algorithm II (NSGA-II) for feature selection to boost accuracy [45]. AdaBoost's performance in COVID-19 diagnosis surpasses Random Forest post-feature selection, indicating the potential for heightened disease classification accuracy [46]. Using AdaBoost alongside other techniques like Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and k-nearest neighbors (KNN), this study on COVID-19 patient blood samples achieves top predictive performance, emphasizing age's significance and strong associations among Lactate Dehydrogenase (LD), C-reactive Protein (CRP), and leukocytes [47]. Introducing Vulture Based Adaboost-Feedforward Neural (VbAFN), a novel AdaBoost-based scheme for early COVID-19 severity prediction from chest X-ray images, incorporating preprocessing, feature extraction, and segmentation for accurate data input [48].

Utilizing AdaBoost with decision tree estimators, this study enhances COVID-19 patient severity assessment with a novel parameter tuning process, showcasing competitive accuracy through extensive experimentation on the University of California Irvine (UCI) and COVID-19 datasets [49]. Proposing a Hybrid Disease Detection Principle (HDDP), that combines AdaBoost with Convolutional Neural Network (CNN) for COVID-19 detection from Lung Computed Tomography (CT) scans, offering enhanced accuracy and efficiency in disease identification [50].

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

The pathology of H1N1, CAP, and sepsis was investigated, suggesting the potential of group mass disturbance for clinical outcomes in CAP and sepsis patients [51]. 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 [52]. Chest radiograph classification was enhanced using AlexNet (MAN) [53]. Dataset preparation processes were explored, proving effective in both engineering and real datasets [54]. 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 [55]. An approach to diagnose COVID-19 with chest radiography was proposed, which shows promise despite the limitations of the data set [56]. Emphasis is placed on a methodological framework for the identification of COVID-19 [57]. The applicability of Artificial Intelligence (AI) in health status assessment has been explored [58]. Attention has focused on the diagnosis of pneumonia using X-ray imaging [59]. Effective methods have been proposed for the detection of tuberculosis and pneumonia [60, 61]. The role of machine learning in clinical areas, particularly in pneumonia identification, has also been discussed [62]. The potential of AI in clinical decision-making has been explored [63]. A meta-study analyzing explicit CT designs has been conducted [64]. A web application for pneumonia differentiation has been developed [65]. Additionally, a fully automated system for COVID-19 identification has been established [66]. Evolutionary Neural Network (ENN) and SVM expectation models were introduced in a previous study [67].

Another study focused on pollution-related biomarkers and clinical symptoms for pneumonia prediction [68]. A Deep Convolutional Neural Organization (DCNN) model-based approach for classifying pneumonia images was presented [69]. A non-arranged convolutional neural architecture for pneumonia event detection was proposed, exhibiting better accuracy [70], Convolutional neural networks (CNNs) were used to identify pneumonia, demonstrating the accuracy of the VGG16 design [71]. A lightweight pneumonia detection model using CNNs with three kernel sizes proposed in a study in which the outputs are combined with a weighted ensemble, whose threshold can be adjusted according to the desired diagnostic sensitivity and specificity. An ensemble deep learning approach for classifying digital chest X-ray images using different kernel sizes is examined in this study [72]. Additionally, COVID-19 detection through audio features such as cough and breathing, along with SHAP and data augmentation techniques, is explored [73].

Methodology

This study employs an integrated four-phase framework for COVID-19 pneumonia classification by leveraging the strengths of Random Forest, AdaBoost, XGBoost, and CNN techniques as depicted in Figure 1 and outlined in Algorithm 1. The proposed approach ensures improved accuracy and robustness by combining ensemble learning models with deep learning methods for precise feature extraction and classification.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

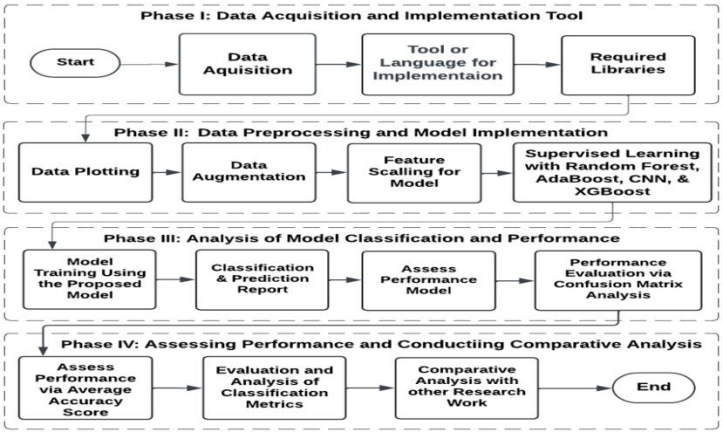

Figure 1. Block diagram of Proposed work

Figure 2.

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.

1.1. WHO_dataset ← LoadWHODataSet()

//Extract demographic information (e.g., age, gender) from the dataset

1.2. demog_info ← ExtractDemographicInfo (WHO_dataset)

//Preprocess the data to clean and format it for further analysis

1.3. preproc_data ← PreprocessData (WHO_dataset, demog_info)

Step 2. ***Data Visualization Phase***

//Generate a plot to visualize the distribution of demographic information.

2.1. Demog_Distrib_Plot ← Visualize_DemographicDistribution(demog_info)

//Plot the prevalence of symptoms using the preprocessed data.

2.2. Symp_Preval_Plot ← PlotSymptomsPrevalence(preproc_data)

Step 3. ***Model Building for Feature Extraction***

//Initialize models that will be used for feature extraction

3.1. RF_model ← InitializeRFModel()

3.2. AdaBoost_model ← InitializeAdaBoostModel()

3.3. CNN_model ← InitializeAdaCNNModel()

3.4. XGBoost_model ← InitializeXGBoostModel()

//Train the models using the X-ray images and demographic data

3.5. train_models ← TrainModels(RF_model, AdaBoost_model, CNN_model, XGBoost_model, COVID19_XRay_images, demog_data)

//Optimize the trained models for better accuracy and performance

3.6. optim_models ← OptimizeModels(train_models)

Step 4. ***Performance Evaluation Phase***

//Split the preprocessed data into training and testing sets for evaluation.

4.1. train_set, test_set ← SplitDataset(preproc_data)

//Evaluate the optimized models using a confusion matrix.

4.2. conf_matrix ← EvaluateModels(optim_models, test_set)

//Calculate performance metrics such as accuracy, sensitivity, and specificity

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

4.3. metrics ← CalculateMetrics(conf_matrix, precision, recall, F1_score,accuracy)
Step 5. Comparative Analysis of Results
//Compare the accuracy of this study with previous research on pneumonia classification.
5.1. accu_compar ← CompareAccuracy(base_study, existing_research)
//Identify areas where the model shows substantial improvement over past methods
5.2. improv_areas ← IdentifyAreasOfSubstantialImprovement()
//Discuss the implications of the results and potential applications of the findings.
5.3. implic_and_applic ← DiscussImplicationsAndApplications()
Step 6. End

The proposed research framework involves four stages: (1) **Data Acquisition**, using a GitHub chest X-ray dataset preprocessed in Python via Anaconda Jupyter Notebook for quality analysis; (2) **Data Processing and Feature Analysis**, employing scatter plots, histograms, and Random Forest for feature selection to enhance efficiency; (3) **Supervised Learning Implementation**, integrating CNNs with data augmentation for detailed feature extraction, alongside AdaBoost and XGBoost for robust pneumonia classification and improved accuracy; and (4) **Performance Evaluation**, comparing models using accuracy, precision, recall, and F1-score, demonstrating superior results in COVID-19 pneumonia classification. This framework provides a strong foundation for disease prediction and advances healthcare research by addressing critical challenges in pneumonia diagnosis and classification.

Phase I: Data Acquisition and Implementation Tool
Data Collection and Preprocessing
Data acquisition and preprocessing are fundamental. Raw data must be collected and cleaned (e.g., removing missing values, handling outliers) before any further steps.

For this study, we acquire a GitHub dataset containing 893 chest X-ray images, including COVID-19 and other pneumonia cases [74]. To enhance classification accuracy, we incorporate Random Forest, AdaBoost, XGBoost, and CNN techniques. Random Forest effectively handles data heterogeneity, while AdaBoost boosts weak classifiers to achieve higher accuracy. XGBoost, with its gradient boosting framework, optimizes model performance through hyperparameter tuning and feature selection. CNNs excel in extracting spatial features from X-ray images, making them highly effective for image-based disease classification. Previous studies also utilize this dataset, highlighting its relevance for pneumonia classification research.

Tool Selection for Implementation
Python, combined with Anaconda Jupyter Notebook, is utilized for this research due to its versatility in data mining, AI, and machine learning. Python supports implementing advanced algorithms such as Random Forest, AdaBoost, XGBoost,

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

and CNN, making it an optimal choice for diverse classification tasks. Anaconda Jupyter Notebook, an open-source web-based application, facilitates efficient coding, data visualization, and model evaluation, enhancing the overall workflow. Its seamless integration with Python ensures robust execution of machine learning techniques, further emphasizing its prominence in data-driven research.

Importing Required Libraries

Importing essential libraries (e.g., NumPy, Pandas, and scikit-learn) is crucial for data preprocessing, visualization, and model development. Specialized libraries play a critical role in the study, with key libraries like Keras being essential for constructing neural network models in deep learning applications. The proposed model integrates techniques such as Random Forest, AdaBoost, XGBoost, and Convolutional Neural Networks (CNN) to enhance classification accuracy. Additionally, libraries like Numpy facilitate numerical and matrix analysis, aiding in data manipulation. Matplotlib and Seaborn serve as vital tools for data visualization, presenting data in visual formats to enhance comprehension and analysis.

Phase II: Data Processing and Model Implementation

Data Plotting

Data visualization facilitates the understanding of patterns, distributions, and correlations, aiding in feature selection and model design. Effective data visualization techniques are crucial for gaining insights into the COVID-19 pneumonia dataset. Various methods, including histograms, boxplots, scatter plots, and dimensionality reduction techniques such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), are employed to visualize feature distribution, relationships, and patterns within the data. Example image grids are utilized to highlight characteristic features, enhancing understanding of the dataset [75].

Graphical representation plays a key role in data analysis. OpenCV, Matplotlib, and Seaborn libraries are leveraged to create visually appealing and informative plots. As shown in Figure 2, an X-ray image is presented for normal and abnormal classification, providing a visual representation of the dataset. The dataset consists of X-ray images sourced from GitHub, which forms the foundation of our research.

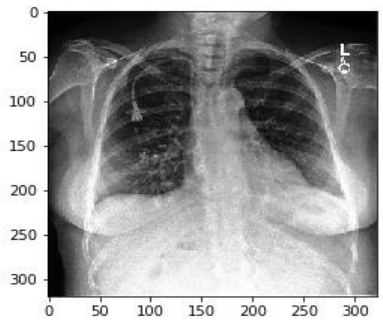

Figure 3. Chest X-ray Image Depicting Pneumonia Affliction

Following data visualization, the proposed model integrates Random Forest, AdaBoost, XGBoost, and CNN techniques to enhance classification performance. Random Forest contributes to capturing non-linear relationships and handling imbalanced datasets effectively, while AdaBoost emphasizes reducing bias

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

through iterative learning. XGBoost, renowned for its efficiency in tabular data analysis, applies a gradient-boosting framework to uncover complex patterns within the dataset. Complementing these, CNNs focus on extracting spatial features, particularly from image-based data. This combined approach leverages the strengths of these algorithms, ensuring robust analysis and accurate predictions across diverse data types [76].

Data Augmentation

Augmentation is applied to expand the dataset, particularly in image processing scenarios, to ensure diversity and robustness in the training data. Data augmentation is essential for enhancing the performance of machine learning models, particularly in the classification of COVID-19 pneumonia using chest X-ray images. By applying techniques such as rotation, flipping, scaling, cropping, brightness/contrast adjustment, and noise addition, synthetic data is generated, enabling models to generalize better without needing additional real-world samples. This approach not only mitigates overfitting but also improves model accuracy, especially for models like XGBoost, which benefits from diverse training data.

XGBoost optimizes these augmented features through its robust algorithm, refining the classification process. Integrating ensemble methods like Random Forest and AdaBoost further strengthens the model. Random Forest improves robustness by averaging multiple decision trees, reducing variance, while AdaBoost enhances performance by re-weighting misclassified instances. When combined with Convolutional Neural Networks (CNN), which extract complex spatial features from the augmented images, this hybrid approach offers a comprehensive solution for pneumonia classification, leveraging the strengths of both traditional machine learning and deep learning techniques to achieve superior accuracy and robustness.

Feature Scaling for Optimal Machine Learning and Deep Learning Techniques

Scaling ensures that features have uniform ranges, improving the performance and convergence of ML and DL algorithms.

Random Forest Feature Scaling

Random Forest is relatively robust to feature scaling due to its reliance on decision trees, which split based on feature thresholds rather than feature magnitudes. However, applying feature scaling may still slightly enhance performance, especially when Random Forest is used alongside other algorithms in ensemble methods. While optional, min-max scaling or standardization can be applied if the model is combined with others that require scaling. Features such as patient demographics (e.g., age, body temperature) should be standardized to ensure consistent processing.

AdaBoost Feature Scaling

Similar to Random Forest, AdaBoost's performance is not heavily dependent on feature scaling, as it leverages weak learners (usually decision trees). However, scaling may improve convergence when non-tree-based base learners are used. It is optional to apply min-max scaling or standardization to maintain uniformity in

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

ensemble settings. Additionally, scale demographic features and categorical variables (e.g., gender, medical history) to improve the consistency of base learners.

XGBoost Feature Scaling

XGBoost is highly sensitive to feature scaling, particularly when handling mixed data types (e.g., numeric and categorical). Uniform feature scaling enhances gradient calculations, optimizing model performance. Min-max scaling or standardization should be applied (e.g., scaling age, blood oxygen levels, and body temperature). Pixel intensity values should be normalized to a range of 0–1 or -1–1 for consistency. This step is essential for ensuring fair weight distribution across features during optimization and tree construction.

CNN Feature Scaling

CNNs inherently handle feature scaling through internal operations, such as batch normalization and activation functions. However, preprocessing image data to a standard scale is crucial for effective training. Normalize pixel intensity values to a range of 0–1 (by dividing by 255 for RGB images). Standardize images by subtracting the mean pixel value and dividing by the standard deviation. Scaling enhances convergence, reduces computational load, and stabilizes gradients during training.

Supervised Learning with Proposed Model

The proposed models (Revised Random Forest, AdaBoost, XGBoost, and CNN) are applied to the preprocessed and scaled data. The proposed model integrates Random Forest, AdaBoost, XGBoost, and CNN to enhance supervised learning for tasks such as COVID-19 pneumonia classification from chest X-ray images. Input features, representing pixel values extracted from X-rays, serve as descriptors of the image characteristics, while output labels indicate the classification, such as COVID-19 pneumonia or other categories.

XGBoost captures intricate feature interactions with high computational efficiency, while Random Forest and AdaBoost contribute to robust feature selection and reduce overfitting by leveraging ensemble learning. CNN excels in extracting spatial and hierarchical features from X-ray images, complementing the ensemble techniques. The combined model enables the identification of complex patterns and relationships between input features and output labels, ensuring accurate and reliable classification in medical imaging analysis[77].

Revised Random Forest Classifier

Random Forest is an ensemble learning technique that constructs multiple decision trees and aggregates their outputs for accurate predictions. Known for its ability to handle imbalanced and noisy datasets, it effectively identifies critical features from medical imaging, enhancing its application in diagnosing COVID-19 pneumonia. By mitigating overfitting and managing incomplete data, Random Forest excels in distinguishing pneumonia from other respiratory conditions, supporting precise classification even in complex datasets.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

AdaBoost Classifier

AdaBoost (Adaptive Boosting) enhances model performance by iteratively focusing on misclassified instances and assigning higher weights to them. This approach strengthens weak learners, enabling the detection of nuanced patterns in chest X-ray images, such as subtle pixel intensity variations indicative of pneumonia. Its adaptive nature ensures high precision in scenarios with limited labeled data and complements other classifiers in addressing the challenges of pneumonia classification. AdaBoost's ability to emphasize difficult cases makes it an effective choice for early and accurate diagnosis, contributing significantly to the proposed hybrid model.

XGBoost Classifier

XGBoost (eXtreme Gradient Boosting) is a gradient-boosting algorithm optimized for high-dimensional datasets, widely recognized for its efficiency and accuracy in classification tasks. By iteratively refining decision trees, it identifies subtle patterns such as opacities and infiltrates in chest X-rays of COVID-19 pneumonia cases. XGBoost effectively handles missing data, manages outliers, and allows for advanced hyperparameter tuning, ensuring reliable and interpretable outcomes. Its adaptability and robust performance make it a cornerstone for medical imaging and predictive modeling, particularly in COVID-19 research. Data augmentation techniques further enhance its classification accuracy and model robustness.

CNN (Convolutional Neural Networks) Classifier

CNNs are deep learning architectures specialized in image classification, particularly effective in extracting spatial features from medical images. In the context of COVID-19 pneumonia, CNNs identify key visual patterns, such as lung opacities and structural abnormalities, which are crucial for diagnosis. Models like ResNet and GoogLeNet have proven highly accurate in distinguishing COVID-19 pneumonia from other types. Their ability to learn hierarchical features allows for precise predictions, even with limited data.

In the proposed model, CNNs are integrated with tree-based algorithms to enhance feature extraction. This combination leverages deep learning for improved classification accuracy, complementing the strengths of other machine learning techniques.

Results and Discussion

The Results and Discussion section evaluates the proposed model integrating Random Forest, AdaBoost, XGBoost, and CNN techniques for classifying pneumonia X-ray images into Normal and Abnormal categories. The analysis focuses on confusion matrix components—True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN)—as well as performance metrics, including precision, recall, F1-score, and accuracy.

Phase III: Analysis of Model Classification and Performance

Model Training Using the Proposed Model

Training the models with the prepared dataset is a critical step to evaluate their performance. In this phase, we train a classification model for COVID-19 pneumonia using a combination of Random Forest, AdaBoost, XGBoost, and

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

Convolutional Neural Networks (CNN). The process begins with preparing a labeled chest X-ray dataset, distinguishing between COVID-19 and non-COVID-19 pneumonia. The dataset is divided into training and testing subsets. Feature extraction identifies critical characteristics from the X-ray images, which serve as input for the models.

For the ensemble models, we use Random Forest and AdaBoost, each contributing to the classification process through their respective strengths in decision tree learning and boosting techniques. The XGBoost model, known for its gradient boosting capabilities, further refines the classification by optimizing tree-based predictions. Simultaneously, the CNN architecture extracts spatial hierarchies from the images, enabling the model to learn complex features for improved performance.

Model evaluation involves using accuracy, precision, recall, and F1-score metrics, with parameter tuning across all techniques to optimize performance. By combining these advanced models, the system achieves enhanced prediction accuracy, making it robust for classifying COVID-19 pneumonia in chest X-ray images.

XGBoost-Based Classification and Prediction Outcomes

Individual model outcomes, especially for XGBoost, should be assessed to benchmark performance. 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.

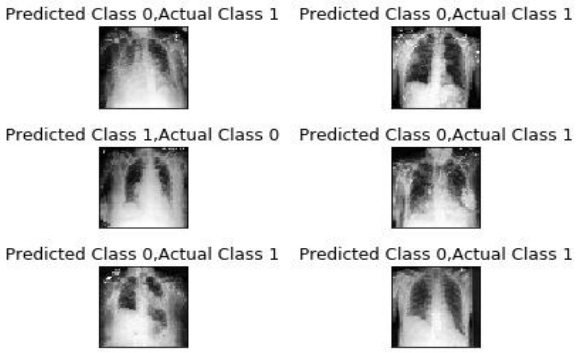

Figure 4. Classification And Prediction Results Derived From The Proposed Model

This visualization evaluates the effectiveness of the XGBoost model in classifying COVID-19 pneumonia cases based on 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. A detailed interpretation of the model's decision-making process highlights key features that contribute 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.

Assessing Model Performance through Confusion Matrix Analysis

Confusion matrix analysis provides a detailed evaluation of model accuracy, sensitivity, and specificity. A confusion matrix is crucial for assessing the performance of a classification algorithm. It provides valuable insights into the

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

algorithm's accuracy, highlighting the types of errors it produces and identifying areas for improvement. This helps in evaluating the model's 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.

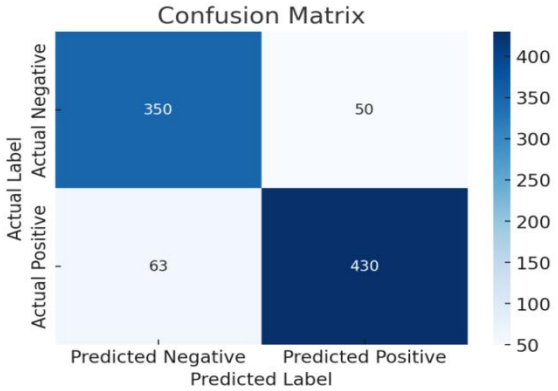

Figure 5. Confusion Matrix Generated by XGBoost

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), provides a detailed assessment of classifier performance:

- True Positive (TP): Correctly predicted positive cases (430).
- True Negative (TN): Correctly predicted negative cases (350).
- False Positive (FP): Negative cases incorrectly classified as positive (50).
- False Negative (FN): Positive cases incorrectly classified as negative (63).

These values are used to compute key performance metrics, including precision, recall, and accuracy, offering a thorough evaluation of the model's classification ability. This analysis highlights the effectiveness of the proposed XGBoost model, while also identifying areas for further optimization and improvement. The confusion matrices are generated in a similar manner for the other proposed techniques, including Random Forest, AdaBoost, and CNN.

Phase IV: Assessing Performance and Conducting Comparative Analysis

This phase evaluates the performance of the proposed XGBoost model by comparing it with existing studies. Key metrics, such as accuracy, recall, and F1-score, are calculated to assess the model's precision and ability to correctly classify instances. These performance indicators offer valuable insights into the model's effectiveness in achieving reliable and accurate classifications.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

Average Accuracy Score to Assess Model Performance

Aggregate metrics like average accuracy quantify model reliability across different datasets or configurations. 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.

Evaluation Metrics for COVID-19 Pneumonia Classification using XGBoost

XGBoost-specific metrics allow fine-grained insights into its strengths and weaknesses for pneumonia classification. 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:

$$AC = \frac{TP + TN}{TP + FN + TN + FN} \qquad (1)$$

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:

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

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

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

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:

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

Clarification of Normal vs. Abnormal Classification

Differentiating between normal and abnormal cases helps contextualize the model's predictions. 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.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

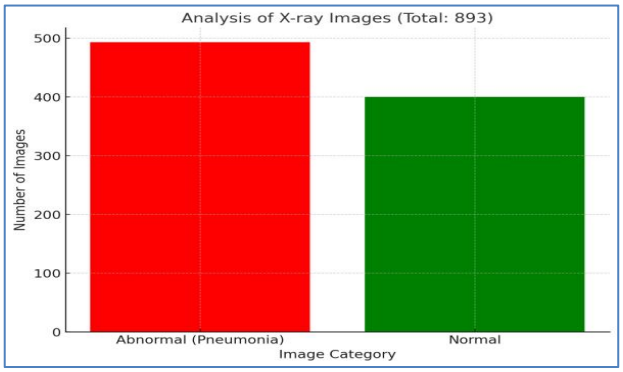

Figure 6. Distribution of Abnormal (Pneumonia) and Normal X-ray Images

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.

Analysis and Discussion of Performance Metrics for Proposed Models

This step synthesizes the results to highlight the contributions of the models, including insights from comparative studies. This section presents an in-depth analysis of the performance of the proposed model, which utilizes multiple techniques—Random Forest, AdaBoost, Convolutional Neural Network (CNN), and XGBoost—for classifying COVID-19 pneumonia cases. The results, presented in Table 1, indicate the performance of each model across key metrics: Precision, Recall, F1-Score, and Accuracy.

Table 1. Performance Analysis of Metrics for Proposed Model

Techniques	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
Random Forest	86.63	80.09	83.23	82.29
AdaBoost	87.31	83.90	85.57	84.49
CNN	88.60	83.74	86.10	85.70
XGBoost	89.58	87.22	88.39	87.35

The Random Forest model demonstrates a solid performance with precision of 86.63%, recall of 80.09%, F1-score of 83.23%, and accuracy of 82.29%. While Random Forest has decent overall performance, its recall score suggests a relatively higher number of false negatives compared to other models. This is significant, as it may fail to correctly classify a proportion of COVID-19 pneumonia cases, which can have serious consequences in a clinical setting. The precision is reasonably high, meaning that the model tends to avoid false positives. However, the recall improvement is a target area for further

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

optimization, potentially through fine-tuning hyperparameters or employing additional feature engineering.

The AdaBoost model achieves a precision of 87.31%, recall of 83.90%, F1-score of 85.57%, and accuracy of 84.49%. The results show an improvement in both precision and recall over Random Forest, with a more balanced trade-off between these metrics. AdaBoost's recall (83.90%) suggests a better ability to correctly classify positive cases, thereby reducing the false negatives. Furthermore, its F1-score of 85.57% reflects an overall strong balance between precision and recall, marking AdaBoost as a robust classifier for COVID-19 pneumonia detection.

The CNN model yields an accuracy of 85.70%, precision of 88.60%, recall of 83.74%, and F1-score of 86.10%. CNN, which utilizes deep learning to capture spatial hierarchies and patterns in X-ray images, outperforms both Random Forest and AdaBoost in terms of precision and F1-score. CNN's precision is the highest among the models, indicating a strong tendency to avoid false positives. However, its recall score slightly lags behind that of AdaBoost. Despite this, CNN demonstrates excellent potential in accurately identifying COVID-19 pneumonia, particularly in scenarios where high precision is paramount.

The XGBoost model outperforms all other models with the highest precision (89.58%), recall (87.22%), F1-score (88.39%), and accuracy (87.35%). XGBoost's superior performance across all metrics indicates its capability to handle complex data relationships effectively, likely due to its implementation of gradient boosting and tree-based learning. The high recall ensures that fewer true positive cases are missed, while the elevated precision highlights its accuracy in avoiding false positives. This makes XGBoost an ideal choice for classifying COVID-19 pneumonia, where both precision and recall are critical in the medical context.

Discussion of Comparative Results

From the analysis of performance metrics, it is evident that XGBoost outperforms the other models across all metrics, making it the most effective technique for COVID-19 pneumonia classification in this study. Its high precision and recall suggest that it is able to both accurately classify pneumonia cases and minimize missed cases, which is vital for medical decision-making.

While CNN demonstrates superior performance in terms of precision, it falls slightly behind in recall compared to XGBoost and AdaBoost. This can be attributed to the specific nature of CNNs, which, although excellent at detecting intricate image patterns, may not always balance false positives and false negatives optimally without proper tuning.

AdaBoost, with a balanced F1-score, showcases a good trade-off between precision and recall, performing significantly better than Random Forest. Its ability to improve classification accuracy through boosting weak learners is evident in the results.

The Random Forest model, while still useful, lags behind the other techniques in both precision and recall, suggesting that it is less capable of handling the complexity of classifying COVID-19 pneumonia cases, especially when compared to the advanced boosting methods (XGBoost and AdaBoost) and deep learning-based CNN.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

Substantial Insights and Contributions

The results of this study offer key insights into the effectiveness of different machine learning techniques for the classification of COVID-19 pneumonia. The XGBoost model's superior performance in both precision and recall positions it as the most reliable model for this task. This is especially critical in clinical applications where misclassifications can have serious consequences. AdaBoost, while slightly less accurate, offers a well-rounded performance with good balance between recall and precision, making it a strong alternative in scenarios where both false positives and false negatives need to be minimized.

The application of CNNs highlights the potential of deep learning for image-based classification tasks, where CNNs excel in capturing complex patterns in X-ray images. However, further improvements in recall could make CNN a more competitive choice for this task.

The proposed models—especially XGBoost—can significantly contribute to enhancing diagnostic accuracy and decision-making processes in the healthcare sector, particularly in detecting COVID-19 pneumonia from chest X-ray images. These findings suggest the importance of model selection in healthcare applications, as the trade-off between precision and recall must be carefully considered depending on the specific clinical needs.

The incorporation of diverse machine learning techniques in this study enriches the ongoing research into COVID-19 pneumonia detection and can pave the way for more robust, accurate, and adaptable models in future medical applications.

Comparative Analysis with Existing Research

Comparing results with prior studies provides context, validates the findings, and identifies advancements made by the proposed approach. This study introduces a comprehensive methodology for COVID-19 pneumonia classification, leveraging advanced ML and DL techniques, including Random Forest (RF), AdaBoost, CNN, and XGBoost. The proposed models were benchmarked against existing research by Khalifa et al. [14] and Ieracitano et al. [78], as summarized 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. [14]	X-ray Images	Alexnet, Googlenet, and Restnet18	78.70
Ieracitano, Cosimo, et al. [78]	X-ray Images	Fuzzy-CovNNet	81.00
Our proposed research work	X-ray Images	Random Forest	82.29
		AdaBoost	84.49
		CNN	85.70
		XGBoost	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

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

pneumonia types, achieving 81% accuracy through image features and fuzzy edge data from various datasets.

The proposed XGBoost classifier achieves the highest accuracy of 87.35%, surpassing the algorithms used in prior studies, including AlexNet, GoogLeNet, and ResNet18 (78.70%) as well as Fuzzy-CovNNet (81.00%). CNN, a deep learning (DL) technique, also demonstrates superior performance with an accuracy of 85.70%, further emphasizing the effectiveness of DL in image classification tasks. Ensemble methods such as AdaBoost (84.49%) and Random Forest (RF) (82.29%), while slightly less accurate than CNN and XGBoost, consistently outperform the approaches employed in earlier studies.

All approaches utilize chest X-ray datasets for classification, providing a comparable evaluation baseline. The integration of ensemble techniques and deep learning (DL) models in the proposed methodology significantly enhances the ability to extract meaningful patterns and adapt to complex datasets. In comparison, Khalifa et al. [14] relied on transfer learning models (AlexNet, GoogLeNet, and ResNet18), which, while effective, lack the benefits of custom model tuning and advanced feature engineering employed in this study, resulting in improved classification performance. Ieracitano et al. [78] implemented Fuzzy-CovNNet, which combines fuzzy logic with convolutional layers; however, it is surpassed by CNN and XGBoost in this study due to their superior capacity for feature extraction and effective handling of complex data representations.

Random Forest (RF) provides competitive accuracy, and its ensemble nature makes it suitable for initial experimentation and baseline comparisons. AdaBoost enhances model performance by focusing on misclassified instances, achieving a balance between computational efficiency and accuracy. CNN excels in capturing spatial features from X-ray images, resulting in robust classification performance. XGBoost delivers the highest accuracy due to its gradient-boosting framework, efficient handling of missing data, and capability to optimize performance through hyperparameter tuning.

The superior accuracy of the proposed models enhances reliability in distinguishing between normal and abnormal (COVID-19 pneumonia) cases, supporting clinicians in making timely and precise diagnoses. The implementation of ML and DL techniques ensures scalability and adaptability, facilitating their integration into automated diagnostic systems for real-time applications.

Achieving an accuracy of 87.35%, the XGBoost classifier establishes a new benchmark in COVID-19 pneumonia classification using chest X-rays. The comparative analysis highlights the significance of integrating traditional ensemble techniques with advanced deep learning (DL) models to improve diagnostic precision. This study effectively bridges the gap between existing methodologies and practical diagnostic applications, contributing to the development of robust and automated solutions for pandemic response and future healthcare challenges.

Conclusion

This study develops an integrated machine learning framework for the classification of COVID-19 pneumonia using Random Forest, AdaBoost, XGBoost, and CNNs. The research aims to optimize diagnostic accuracy by applying ensemble learning and deep learning methods to chest X-ray images. Results

ISSN Online: 3007-3154
ISSN Print: 3007-3146

indicate that XGBoost outperforms other models, achieving the highest accuracy of 87.35%, followed closely by CNN, AdaBoost, and Random Forest. The findings indicate that XGBoost outperforms other techniques, achieving the highest accuracy and reliability. The study introduces a optimized approach that integrates multiple machine learning algorithms, improving classification performance compared to traditional diagnostic methods. The results, particularly from XGBoost, demonstrate enhanced accuracy, precision, and recall in detecting pneumonia, especially COVID-19, from chest X-ray images. The integrated framework can significantly improve pneumonia diagnostic processes, offering a more reliable and adaptable solution for clinical settings. The study's results highlight the potential for AI-driven models to assist in early disease detection, optimizing resource allocation, and improving healthcare outcomes, especially during pandemics.

Future research should focus on enhancing recall, particularly for XGBoost, to reduce missed diagnoses. Expanding the dataset to include a wider variety of real-world chest X-ray images would also improve model generalizability. Additionally, further optimization of the hybrid model could lead to more precise real-time diagnostics and decision support in clinical environments.

Conflicts of Interest

The authors declare no conflicts of interest related to this research study.

References

1. Feng Shi†, Liming Xia†, Fei Shan†, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, & He Sui, Dinggang Shen. (2020). Large-Scale Screening of COVID-19 from Community Acquired Pneumonia using Infection Size-Aware Classification.
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, Yasmen Bejad AlOtaibi 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. Aditya Kakde, Nitin Arora, Durgansh Sharma, & Subhashchander Sharma. (2020). Multi-spectral classification and recognition of breast cancer and

ISSN Online: 3007-3154
ISSN Print: 3007-3146

pneumonia. Polish Journal of Medical Physics and Engineering, 26(1), doi:10.2478/pjmpe-2020-0001.
8. 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.
9. 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.
10. 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).
11. 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.
12. 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.
13. 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, pp. 69-79. 2020.
14. NE, M. Khalifa, F. Smarandache, and M. Loey. "COVID-19 Chest X-Ray Images Diagnosis: A Neutrosophic and Deep Transfer Learning Approach." (2020).
15. Kalpana, B., Reshmy, A. K., Pandi, S. S., & Dhanasekaran, S. (2023). OESV-KRF: Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases. Biomedical Signal Processing and Control, 85, 104779.
16. Wang, J., Rao, C., Goh, M., & Xiao, X. Risk assessment of coronary heart disease based on cloud-random forest. Artificial Intelligence Review, (2023), 56(1), 203-232.
17. Singh, D., Kumar, V., Kaur, M., & Kumari, R. Early diagnosis of COVID-19 patients using deep learning-based deep forest model. Journal of Experimental & Theoretical Artificial Intelligence, (2023), 35(3), 365-375.
18. Souza E.N.D., Matwin S. Improvements to adaboost dynamic, Advances in Artificial Intelligence. Lecture Notes in Computer Science. 2012:293–298.
19. Shhadat, Ihab, Amena Hayajneh, and Ziad A. Al-Sharif. "The use of machine learning techniques to advance the detection and classification of unknown malware." Procedia Computer Science 170 (2020): 917-922.
20. Lai, Chih-Cheng, Tzu-Ping Shih, Wen-Chien Ko, Hung-Jen Tang, and Po-Ren Hsueh. "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges." International journal of antimicrobial agents 55, no. 3 (2020): 105924.
21. Singer, Gonen, and Matan Marudi. "Ordinal decision-tree-based ensemble approaches: The case of controlling the daily local growth rate of the COVID-19 epidemic." Entropy 22, no. 8 (2020): 871.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

22. El Gannour, Oussama, Soufiane Hamida, Bouchaib Cherradi, and Abdelhadi Raihani. "Enhancing early detection of COVID-19 with machine learning and blood test results." Multimedia Tools and Applications (2024): 1-31.
23. 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), pp. 446-452. IEEE, 2021.
24. 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.
25. 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.
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. 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.
28. 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.
29. 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.
30. Minnoor, Manas, and Veeky Baths. "Diagnosis of breast cancer using random forests." Procedia Computer Science 218 (2023): 429-437.
31. Prasojo, Rahman Azis, Muhammad Akmal A. Putra, Meyti Eka Apriyani, Anugrah Nur Rahmanto, Sherif SM Ghoneim, Karar Mahmoud, Matti Lehtonen, and Mohamed MF Darwish. "Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique." Electric Power Systems Research 220 (2023): 109361.
32. 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.
33. 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 novel COVID-19 pandemic using extreme gradient boosting and random forest classifier algorithms." Engineering Reports 5, no. 12 (2023): e12678.
34. Khanna, Varada Vivek, Krishnaraj Chadaga, Niranjana Sampathila, Srikanth Prabhu, and Rajagopala Chadaga. "A machine learning and explainable

ISSN Online: 3007-3154
ISSN Print: 3007-3146

artificial intelligence triage-prediction system for COVID-19." Decision Analytics Journal 7 (2023): 100246.
35. Soui, M., Mansouri, N., Alhamad, R., Kessentini, M., & Ghedira, K. (2021). NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient's symptoms. Nonlinear dynamics, 106(2), 1453-1475.
36. 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.
37. Sun, Chaoyang, Yong Bai, Dongsheng Chen, Liang He, Jiacheng Zhu, Xiangning Ding, Lihua Luo et al. "Accurate classification of COVID-19 patients with different severity via machine learning." Clinical and translational medicine 11, no. 3 (2021).
38. 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.
39. 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.
40. Dumakude, Aphelele, and Absalom E. Ezugwu. "Automated COVID-19 detection with convolutional neural networks." Scientific Reports 13, no. 1 (2023): 10607.
41. 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.
42. 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.
43. Bai, Yun, Jingjing Xie, Dongqiang Wang, Wanjuan Zhang, and Chuan Li. "A manufacturing quality prediction model based on AdaBoost-LSTM with rough knowledge." Computers & Industrial Engineering 155 (2021): 107227.
44. Khan, Koffka, and Emilie Ramsahai. "Maintaining proper health records improves machine learning predictions for novel 2019-nCoV." BMC Medical Informatics and Decision Making 21, no. 1 (2021): 172.
45. Soui, Makram, Nesrine Mansouri, Raed Alhamad, Marouane Kessentini, and Khaled Ghedira. "NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient's symptoms." Nonlinear dynamics 106, no. 2 (2021): 1453-1475.
46. Aremu, T. B., Akinyemi Moruff OYELAKIN, M. B. Akanbi, H. F. Sulaimon, and K. M. Odeyale. "Machine Learning Techniques for Early Classification of COVID-19 Disease in Patients." Journal of Information Communication Technologies and Robotic Applications 13, no. 1 (2022): 8-14.
47. Mazloumi, Rahil, Seyed Reza Abazari, Farnaz Nafarieh, Amir Aghsami, and Fariborz Jolai. "Statistical analysis of blood characteristics of COVID-19

ISSN Online: 3007-3154
ISSN Print: 3007-3146

patients and their survival or death prediction using machine learning algorithms." Neural Computing and Applications 34, no. 17 (2022): 14729-14743.
48. Mary, S. Roselin, Vinit Kumar, KJ Prasanna Venkatesan, R. Satish Kumar, Naga Padmaja Jagini, and Amedapu Srinivas. "Vulture-based AdaBoost-feedforward neural frame work for COVID-19 prediction and severity analysis system." Interdisciplinary Sciences: Computational Life Sciences 14, no. 2 (2022): 582-595.
49. Sevinç, Ender. "An empowered AdaBoost algorithm implementation: A COVID-19 dataset study." Computers & Industrial Engineering 165 (2022): 107912.
50. Prabha, B., Sandeep Kaur, Jaspreet Singh, Praful Nandankar, Sanjiv Kumar Jain, and Harikumar Pallathadka. "Intelligent predictions of Covid disease based on lung CT images using machine learning strategy." Materials Today: Proceedings 80 (2023): 3744-3750.
51. 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 view point of ensemble gene noise." bioRxiv (2020): 2020-04.
52. Neili, Zakaria, Mohamed Fezari, and Abdelghani Redjati. "ELM and K-nn machine learning in classification of Breath sounds signals." Int J Electr Comput Eng 10, no. 4 (2020): 3528-3536.
53. 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), pp. 1-7. IEEE, 2019.
54. Zhao, Peng, Jia-Wei Shan, Yu-Jie Zhang, and Zhi-Hua Zhou. "Exploratory machine learning with unknown unknowns." Artificial Intelligence 327 (2024): 104059.
55. 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, pp. 234-247. Cham: Springer International Publishing, 2022.
56. 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).
57. 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).
58. 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.
59. 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." J Eng Sci Criterion 17 (2020): 18.
60. Liebenlito, Muhaza, Yanne Irene, and Abdul Hamid. "Classification of tuberculosis and pneumonia in human lung based on chest x-ray image using

ISSN Online: 3007-3154
ISSN Print: 3007-3146

convolutional neural network." InPrime: Indonesian Journal of Pure and Applied Mathematics 2, no. 1 (2020): 24-32.
61. 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.
62. Chadha, Raman. "A Novel Approach for Detecting Pneumonia in Machine Learning." International Journal of Trend in Innovative Research (IJTIIR) (2019).
63. 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.
64. 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.
65. 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 (IOSR-JCE) 22, no. 1 (2020): 08-11.
66. 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).
67. 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.
68. 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.
69. El Asnaoui, Khalid, Youness Chawki, and Ali Idri. "Automated Methods for Detection and Classification of Pneumonia Based on X-ray Images Using Deep Learning." In Artificial Intelligence and Blockchain for Future Cybersecurity Applications, 257-284. Cham: Springer International Publishing, 2021.
70. Verma, G., & Prakash, S. (2020). Pneumonia Classification using Deep Learning in Healthcare. International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-4, February 2020.
71. Hidayatullah, R. C., & Violina, S. (2020). Convolutional Neural Network Architecture and Data Augmentation for Pneumonia Classification from Chest X-Rays Images. International Journal of Innovative Science and Research Technology, ISSN No: 2456-2165, 2020.
72. Bhatt, Harsh, and Manan Shah. "A Convolutional Neural Network Ensemble Model for Pneumonia Detection Using Chest X-ray Images." Healthcare Analytics 3 (2023): 100176.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 7 (July) (2025)

73. 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 1." Journal of Intelligent & Fuzzy Systems Preprint (2024): 1-14.
74. 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.
75. 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.
76. 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, pp. 26106-26128. PMLR, 2023.
77. Tyagi, Amit Kumar, ed. Automated Secure Computing for Next-Generation Systems. John Wiley & Sons, 2023.
78. 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.