JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Optimizing Pneumonia Diagnosis during COVID-19: Utilizing Random Forest for Accurate Classification and Effective Public Health Interventions

Fazal Malik[1*], Muhammad Suliman[1], Irfan ullaha[1], Shehla Shaha[1], Muhammad Qasim Khan[1], and Abd Ur Rub[2]

[1]Department of Computer Science, Iqra National University Peshawar, Khyber Pakhtunkhwa (KPK), Pakistan.
[2]School of Electronics and Information, Northwestern Polytechnical University, Xi'an Shaanxi, China.
[*]Corresponding Author: Fazal Malik. Email: fazal.malik@inu.edu.pk

Received: December 11, 2023 **Accepted:** March 21, 2024 **Published:** June 01, 2024

Abstract: In the quest for precise diagnosis and classification of pneumonia, particularly intensified by the Coronavirus 2019 (COVID-19) pandemic, this research work presents an optimized Random Forest algorithm based mechanism specifically tailored for COVID-19 pneumonia classification. The research methodology encompasses four critical phases: data acquisition from a current COVID-19 X-ray image dataset on GitHub, data processing and analysis using histograms and scatter plots, application of supervised learning with Random Forest enhanced by data augmentation techniques, and performance evaluation through comparative analysis with existing methods. Our proposed model achieved an accuracy score of 82.29% on average, demonstrating significant precision and recall capabilities. Results indicate that the Random Forest model outperforms current methodologies, providing a robust framework for future pneumonia classification research. This study underscores the potential for improved diagnostic accuracy and patient care, highlighting the model's utility in supporting public health interventions and optimizing resource allocation in the context of COVID-19 pneumonia.

Keywords: COVID-19 Pneumonia; Random Forest; Classification; Chest x-rays; Prediction.

1. Introduction

Pneumonia stands as a formidable challenge in global public health, characterized by its inflammatory impact on the lungs, affecting individuals across various age groups but notably presenting higher risks among newborns and those aged over 65 [1]. The condition manifests through a spectrum of symptoms and potential complications, with severity often dictated by the underlying pathogens, while viral infections typically manifest milder symptoms, bacterial pneumonia, particularly with virulent strains, can lead to grave consequences, especially among immune-compromised populations such as newborns [2].

Accurate and a timely medical diagnosis is paramount for effective disease management and improved patient outcomes. Identifying specific pathogens not only guides treatment strategies but also informs broader public health interventions and facilitates the implementation of targeted antibiotic stewardship programs [3]. In this context, understanding the diverse types of pneumonia becomes crucial for clinical decision-making and preventive measures. Pneumonia emerges from various infectious sources, including bacteria, viruses, and fungi, each necessitating distinct diagnostic and therapeutic approaches for optimal patient care. A well-established classification system aids clinicians in categorizing pneumonia based on its source of infection, enabling tailored investigations, antimicrobial therapy selection, and preventive strategies [4].

The landscape of the pneumonia types is comprehensively delineated, encompassing Hospital-Acquired Pneumonia (HAP), bacterial pneumonia, community-acquired pneumonia, viral pneumonia, walking pneumonia, and fungal pneumonia. Each subtype carries unique clinical implications and challenges, demanding nuanced approaches to diagnosis, treatment, and prevention. Hospital-acquired pneumonia (HAP) represents a severe nosocomial infection, compounded by compromised immunity and

the proliferation of multidrug-resistant pathogens within healthcare settings [5]. Similarly, Community-Acquired Pneumonia (CAP) spans a diverse array of microbial origins, ranging from common viruses to fungal pathogens [3]. Bacterial pneumonia and viral pneumonia constitute predominant forms of the disease, each presenting distinctive clinical features and management strategies.

Furthermore, walking pneumonia, characterized by milder symptoms akin to a severe cold, and fungal pneumonia, posing significant risks to immune-compromised individuals, adds layers of complexity to the pneumonia landscape [4].

Pneumonia, a common respiratory illness, progresses through distinct stages delineated by physiological changes within the lungs. Pneumonia progression has four stages i.e. Congestion, Red Hepatization, Gray Hepatization, and Resolution. Initially, within 24 hours of infection, Congestion ensues, characterized by vascular congestion, alveolar fluid accumulation, and observable lung prominence, often accompanied by coughing and dyspnea. Subsequently, Red Hepatization emerges, typified by lung discoloration resembling the liver, attributed to the infiltration of blood, neutrophils, and fibrin into alveolar spaces. Following this phase, Gray Hepatization occurs, marked by a grayish-brown or yellowish hue in the lungs due to vascular changes and fibrin deposition. Finally, Resolution denotes the recuperative phase, during which enzymatic breakdown of exudates facilitates their absorption by macrophages or clearance by fibroblasts, culminating in lung restoration. Understanding these stages is pivotal for clinical diagnosis and effective management of pneumonia-related complications [6].

The occurrence of COVID-19, caused by the novel coronavirus 2019-nCoV, marked a significant global health crisis. COVID-19, an abbreviation for "Coronavirus Disease 2019," is a member of the coronavirus family, known for its ability to spread rapidly among humans [7]. Characterized by fever, cough, and shortness of breath [8], COVID-19 poses substantial risks, with severe cases leading to pneumonia, respiratory diseases, and, in extreme instances, organ failure and death [4, 9]. Men and young individuals may be particularly vulnerable, with pneumonia cases often exhibiting faster breath rates compared to healthy counterparts [10]. Amongst the diagnostic challenges, X-ray imaging serves as a crucial tool in identifying pneumonia, revealing distinctive patterns of haziness and opacity in lung structures [11]. As the world grapples with the multifaceted impact of COVID-19, understanding its symptoms and diagnostic methodologies remains paramount for effective disease management and prevention efforts.

In recent years, medical imaging techniques have played a pivotal role in the medical diagnosis and classification of respiratory diseases like COVID-19 and community-acquired pneumonia (CAP). Various research works have explored the application of advanced algorithms and machine learning models to improve disease detection and classification accuracy.

Feng Shi et al [1] conducted a comprehensive study involving Computed Tomography (CT) examinations on 1658 COVID-19 and 1027 CAP patients. They utilized a Disease-Size-Aware Random Forest technique (USAF) for classification, achieving notable performance.

In the field of medical imaging, Aditya Kakde et al [7] stressed the significance of employing diverse phantom representations to enhance disease detection accuracy. Xu, X., et al [10] employed a convolutional neural network to screen COVID-19 pneumonia, distinguishing between normal and abnormal cases. Similarly, Altaf Hussain et al [12] utilized MobileNet for organizing pneumonia chest radiographs.

Furthermore, Sung-mok Jung et al [13] presented a method for predicting the increase in atypical pneumonia cases caused by new pathogens, using routine non-virological information. Rodolfo M. Pereira et al [14] proposed a COVID-19 pneumonia classification system employing a convolutional neural network across different datasets.

Kalpana, B.,et al; [15] present a novel ensemble support vector kernel random forest-based hybrid equilibrium Aquila optimization (ESVMKRF-HEAO) approach for accurate skin cancer prediction. Utilizing the HAM10000 dataset, the model demonstrates robust classification into five categories.

Wang, J.,et al; [16] introduce a novel Cloud-Random Forest (C-RF) model for Coronary Heart Disease (CHD) risk assessment, combining cloud model and random forest techniques. The C-RF utilizes a weighted attribute determination algorithm founded on decision-making trials and evaluation laboratory and cloud model. Empirical analysis performed using a dataset from Kaggle demonstrates the effectiveness of random forest.

The effectiveness of deep learning based models in disease classification has been extensively explored. Mohamed Loey et al [17] investigated deep learning models like Googlenet and Alexnet,

distinguishing COVID-19 and normal cases, utilizing Generative Adversarial Networks (GANs) and deep learning transfer models.

Additionally, Nour Eldeen M. Khalifa et al proposed a deep transfer learning method for X-ray images, utilizing models like Alexnet, Googlenet, and Resnet18, classifying normal and abnormal pneumonia cases [9]. Muhammad Ilyas et al [18] reviewed various deep learning structures including ResNet, Inception, and Googlenet for COVID-19 detection, emphasizing the necessity for a robust identification framework.

Muhammad E. H. Chowdhury et al [19] improved COVID-19 pneumonia diagnosis using advanced X-ray images, image preprocessing techniques and deep learning methods. Tawsifur Rahman et al [20] made significant strides in identifying bacterial and viral pneumonia through computerized X-ray images. They employed pre-trained deep Convolutional Neural Networks (CNNs) such as AlexNet, ResNet18, D247, DenseNet201, and SqueezeNet, across distinct classifications. Their study holds promise for applications in rapid airport pneumonia screening. Min Zhou et al [21] developed early identification equipment for confirmed NCP cases and focal patients, demonstrating accurate results in chest examinations for COVID-19 and other types of pneumonia.

Lastly, Singh, D. et al; [22] proposed a deep forest-based model for the diagnosis of COVID-19 in early-stages from Chest X-ray (CXR) images, highlighting the effectiveness of the Random Forest technique.

COVID-19 has underscored the critical need for accurate and efficient diagnostic tools to manage and mitigate the spread of the disease. Pneumonia, a severe complication associated with COVID-19, requires prompt and precise detection to ensure timely intervention and treatment. Traditional diagnostic methods, while effective, can be time-consuming and may lack the precision needed for early detection. In response to these challenges, this study proposes the optimized Random Forest algorithm specifically tailored for classifying COVID-19 pneumonia.

This research makes several significant contributions to the field:

- Data Acquisition: Utilizing the latest GitHub dataset with Python and Anaconda Jupyter Notebook to gather relevant data, ensuring that the dataset is comprehensive and up-to-date.
- Data Processing and Analysis: Employing advanced preprocessing techniques, such as histograms and scatter plots, to enhance the suitability of the data for machine learning algorithms.
- Algorithm Development: Implementing a supervised learning Random Forest algorithm, augmented by data augmentation methods, to improve the robustness of the model and performance for classifying COVID-19 pneumonia.
- Performance Evaluation: Conducting a thorough comparison of key metrics derived from the developed model with those of existing studies to assess its accuracy and effectiveness.

These phases collectively ensure a robust and accurate disease classification while also exploring predictive modeling capabilities to anticipate future trends in disease patterns. The insights gained from this research are invaluable for informing public health interventions and optimizing resource allocation.

The preceding sections of the research paper will be examined into a comprehensive discussion of the current state-of-the-art methods (Section 2), introduce the proposed framework and technique (Section 3), present a detailed discussion on experimental results (Section 4), and conclude with future research directions (Section 5).

2. Literature Review

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

Sihananto [23] proposed a study that implements Random Forest classification in a COVID-19 Reverse Transcription Polymerase Chain Reaction (RT-PCR) test classification system alongside an Auto-encoder algorithm. Utilizing a dataset from a Brazilian hospital, the Auto-encoder preprocesses features, followed by a Random Forest for classification. Rehman and Naz [24] investigated the pathology of H1N1, CAP, and sepsis, suggesting the potential of group mass disturbance for clinical outcomes in CAP and sepsis patients. Z. Neili, M. Fezari et al [25] applied volumetric Extraordinary Learning Machine (ELM) and k-Nearest Neighbor (K-nn) AI for breath sound analysis, demonstrating high accuracy through perceptual mode reduction and feature extraction.

Authors [26] [27] proposed a novel deep learning system for pneumonia detection using mobile learning, with X-ray images from the Maternal and Child Medical Center dataset. In this study, Authors enhanced [27] enhanced chest radiograph classification using AlexNet Multi-Axis Network (MAN). Yu-Jie Zhang [28] explored dataset preparation processes, proving effective in engineering and real datasets. Khalid EL ASNAOUI et al [29] proposed a method for pneumonia image classification using the Deep Convolutional Neural Organization (DCNN) model. Nour Eldeen M. Khalifa [30] demonstrated the effectiveness of Generative Adversarial Networks (GAN) for pneumonia Chest X-ray Recognition, outperforming related work in accuracy, recall, and F1 score. Lawrence O. Hall, Rahul Paul et al [31] proposed an approach for diagnosing COVID-19 with chest X-rays, promising despite dataset limitations. Muhammad Ilyas [32] emphasized procedural frameworks for COVID-19 identification. Ethan D. Evans et al [33] explored AI's applicability in health status diagnosis. R. Nagamounika et al [34] focused on pneumonia diagnosis through X-ray images. Muhaza Liebenlito et al [35] and Shangjie Yao et al [36] proposed efficient methods for tuberculosis and pneumonia detection, respectively.

Raman Chadha et al [37] discussed machine learning's role in clinical areas, particularly pneumonia identification. Mariana Chumbita et al [38] explored AI's potential in clinical decision-making. Lukas Ebner et al [39] conducted a meta-study analyzing explicit tomography (CT) designs. Khan Maseeh Shuaib et al [40] developed a web application for pneumonia differentiation. A. Raghavendra Reddy et al [41] proposed a deep learning algorithm for precise chest X-ray predictions. Lin Li et al [42] established a fully automated system for COVID-19 identification. Yu-Hsuan Liao et al [43] introduced Ensemble Neural Network (ENN) and Support Vector Machine (SVM) expectation models.

Cong [44] focused on pollution-related biomarkers and clinical symptoms for pneumonia prediction. Kh Tohidul Islam et al [45] explored deep learning for pneumonia diagnosis. Garima Verma et al [46] proposed an unsupervised CNN model for pneumonia recognition. Joseph F. Pierre et al [47] utilized a pre-trained CNN model for diagnostic purposes. Deniz Yagmur Urey et al [48] developed a procedural framework for prompt pneumonia detection. Ieracitano, Cosimo, et al [49] introduced Covariant Neural Network (CovNNet) for rapid patient triage. Sun, Z., et al. [50] proposed an optimized Random Forest method, refining dot product calculations and cosine similarity assessments, leading to superior accuracy, G-means, and out-of-bag data scores validated through non-parametric tests.

Mustofa, F., et al. [51] noted suboptimal results with the PIMA Indians dataset, highlighting dataset quality. Using Random Forest (RF), this study compared PIMA Indians and Abelvikas datasets, demonstrating RF's superior performance on the latter due to its comprehensive glucose features. Minnoor, M. et al. [52] demonstrated Random Forest's superior performance in breast tumor detection using the Wisconsin Breast Cancer Database. Prasojo [53] applied the Random Forest algorithm with Synthetic Minority Over-sampling Technique (SMOTE) preprocessing for fault identification in power transformers, addressing complexity.

Shaheed, K. et al. [54] utilized Random Forest to classify images of chest X-rays for the detection of COVID-19, following feature extraction methods. Wang et al. [55] utilized Random Forest in predicting viral pneumonia onset and severity based on clinical factors or Personally Identifiable Information (PII), while [56] [70] employed Random Forest to identify factors associated with Pneumocystis pneumonia. [69] and [57] employed Random Forest in constructing an evidence-based diagnostic algorithm for adult asthma. In respiratory failure risk prediction, [58] emphasized Random Forest's stable performance within ensemble learning, facilitating early Intensive Care Unit (ICU) intervention for pneumonia cases.

Mittal e. [59] presented COVINet, integrating Residual Neural Network (ResNet-101) and K-Nearest Neighbors, demonstrating superior performance in automated COVID pneumonia detection compared to conventional CNN-based and classical Machine Learning (ML) methods. A study employs the Random Forest technique alongside a convolutional neural network for the classification of COVID-19 from lung CT images [60]. Similarly, the Random Forest technique is utilized for community-acquired pneumonia severity classification using features extracted from chest X-ray images, achieving notable performance when all features are combined [61] [68].

3. Methodology

This research adopts a comprehensive three-phase methodology for the classification of COVID-19 pneumonia using deep learning, illustrated in Figure 1 and Algorithm 01.

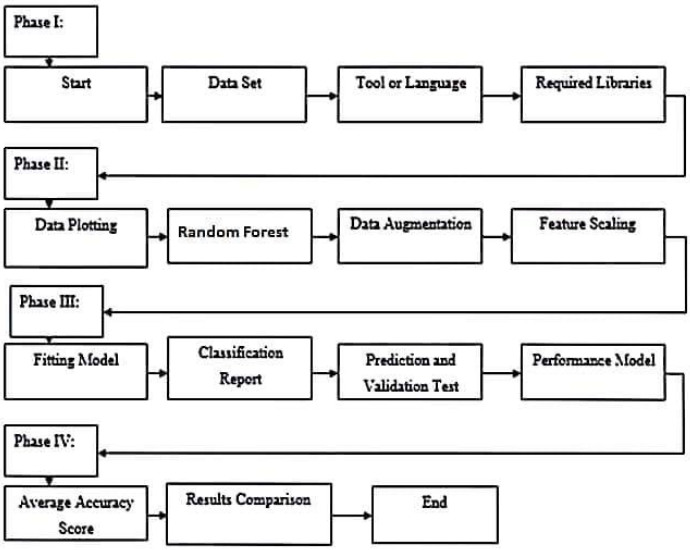

Figure 1. Block diagram of proposed work

Algorithm 01: Advanced Pneumonia Recognition Study

1. ***Data Processing Phase***
    1. *WHO_dataset ← LoadWHODataSet()*
    2. *demog_info ← ExtractDemographicInfo (WHO_dataset)*
    3. *preproc_data ← PreprocessData (WHO_dataset, demog_info)*
2. ***Data Visualization Phase***
    1. *Demog_Distrib_Plot ← Visualize_DemographicDistribution(demog_info)*
    2. *Symp_Preval_Plot ← PlotSymptomsPrevalence(preproc_data)*
3. ***Model Building for Feature Extraction***
    1. *Random_Forest_models ← InitializeRandomForesModels()*
    2. *train_models ← TrainModels(Random_Forest_models, COVID19_XRay_images, demog_data)*
    3. *optim_models ← OptimizeModels(train_models)*
4. ***Performance Evaluation Phase***
    1. *train_set, test_set ← SplitDataset(preproc_data)*
    2. *conf_matrix ← EvaluateModels(optim_models, test_set)*
    3. *metrics ← CalculateMetrics(conf_matrix, accuracy, sensitivity, specificity)*
5. ***Comparative Analysis of Results***
    1. *accu_compar ← CompareAccuracy(base_study, existing_research)*
    2. *improv_areas ← IdentifyAreasOfSubstantialImprovement()*
    3. *implic_and_applic ← DiscussImplicationsAndApplications()*
6. ***End***

COVID-19 has emphasized the urgent need for accurate and efficient diagnostic tools to manage and control the spread of the disease. Pneumonia, a serious complication associated with COVID-19, necessitates rapid and precise detection to ensure timely intervention and treatment. Traditional diagnostic methods, although effective, can be time-consuming and may lack the necessary precision for early detection. Addressing these challenges, this study proposes the development and implementation of an optimized Random Forest algorithm specifically designed for classifying COVID-19 pneumonia. This research offers several significant contributions to the field:

Data Acquisition: Utilizing the latest GitHub dataset with Python and Anaconda Jupyter Notebook to gather relevant data, ensuring the dataset is comprehensive and up-to-date.

Data Processing and Analysis: Employing advanced preprocessing techniques, such as histograms and scatter plots, to enhance the suitability of the data for machine learning algorithms.

Algorithm Development: Implementing a supervised learning Random Forest algorithm, enhanced by data augmentation methods, to make the model more robust and efficient in classifying COVID-19 pneumonia.

Performance Evaluation: Conducting a thorough comparison of key metrics derived from the developed model with those of existing studies to assess its accuracy and effectiveness.

These phases collectively ensure robust and accurate disease classification while also exploring predictive modeling capabilities to anticipate future trends in disease patterns. The insights gained from this research are invaluable for informing public health interventions and optimizing resource allocation.

3.1. Phase I: Data Acquisition and Implementation Tool

3.1.1. Data Set

In this research, we are using the latest data set from GitHub. The data set consists of over a 1000 COVID-19 and other pneumonia X-ray images [62]. In existing relevant research, the same data set has been used for classification tasks.

3.1.2. Tool and Programming Language

This research is utilizing Python and Anaconda Jupyter Notebook, using Python's versatility in data mining, AI (Artificial Intelligence), and deep learning. Jupyter Notebook, an open-source web-based application, provided a user-friendly interface for efficient coding and visualization, showcasing Python's widespread adoption in machine learning tasks.

3.1.3. Required Libraries

Here the specific libraries will be imported for performing special functions like keras which is used for implementing neural network models of deep learning. It also has other important libraries for data pre-processing and data visualization. Numpy is using for numerical and matrices analysis. Matplotlib and Seaborne is data visualization libraries to present data in visual form.

3.2. Phase II: Data Processing and Model Analysis

3.2.1. Data Plotting

Effectively understanding the COVID-19 pneumonia dataset involves crucial data plotting. Techniques like histograms, boxplots, scatter plots, and dimensionality reduction methods (such as Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE)), visualize feature distribution, relationships, and intricate patterns. Grids of example images highlight characteristic features, providing a comprehensive approach for nuanced dataset interpretation [63].

Presenting data graphically is crucial, and for that, we utilized Open Source Computer Vision Library (OpenCV), Matplotlib, and Seaborn libraries. Figure 2 displays an X-ray image used for normal and abnormal classification. Our dataset, consisting of X-ray images sourced from GitHub, serves as the foundation for our research. The subsequent step involves building a convolutional neural network model for image processing [64].

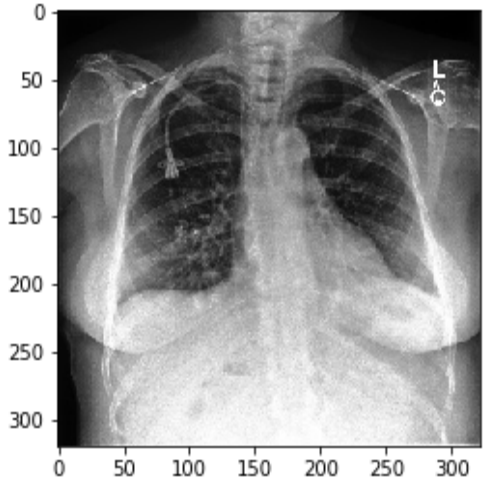

Figure 2. Example of a Chest X-ray Image Showing Pneumonia Affliction

3.2.2. Supervised Learning

The supervised learning process starts with a training dataset that contains input features and corresponding output labels. For example, for COVID-19 pneumonia classification using chest X-ray images, the input features could be pixel values of the X-ray images, and the output labels would indicate whether the image represents COVID-19 pneumonia or not [65].

3.2.3. Random Forest

1. Building Decision Trees

Random Forest is an ensemble learning method and internally consists of multiple decision trees. Each decision tree of Random Forest is built using a subset of the training data and a random selection of features. This helps reduce the overfitting problem and makes the model more general.

2. Learning Patterns

Decision trees in Random Forest learns to identify reoccurring patterns and the relationships between input features and output labels in the training data. For instance, in the case of COVID-19 pneumonia classification, decision trees may learn to recognize visual patterns in chest X-ray images that are indicative of the presence or absence of COVID-19 pneumonia.

3. Aggregating Predictions

After training of all the trees in decision trees, predictions are made by aggregating the outputs of individual trees. In the case of classification tasks, such as COVID-19 pneumonia classification, the final prediction is often determined by a majority voting scheme, where the most common prediction among all the trees is chosen as the final output.

4. Generalization to New Data

Trained Random Forest model can has the ability to make predictions on new or unseen data. For instance, it can classify previously unseen chest X-ray images into COVID-19 pneumonia and other pneumonia categories based on the patterns learned during training.

Supervised learning in the context of Random Forest involves training multiple decision trees on labeled data to learn patterns and feature relationships and output labels, and then aggregating the predictions of these trees to make accurate predictions on new data.

3.2.4. Data Augmentation

Data augmentation is a widely utilized technique in machine learning, especially in tasks like classifying COVID-19 pneumonia using chest X-ray images. In Random Forest models, data augmentation enriches the training dataset by generating synthetic data from existing samples. Techniques include rotating, flipping, scaling, cropping, adjusting brightness/contrast, and adding noise to chest X-ray images. These variations provide the model with a more diverse dataset, enhancing its capability to generalize and to make accurate predictions on unseen data (images). Data augmentation not only improves model performance and robustness but also mitigates overfitting by introducing variability into the training data without requiring additional real-world samples [67].

3.2.5. Feature Scaling

Feature scaling is essential for optimizing the performance of Random Forest in COVID-19 pneumonia classification. This preprocessing step standardizes the range of independent variables or features in the dataset, ensuring uniformity in scale across all features. For chest X-ray images, feature scaling involves scaling pixel intensity values to a smaller range, like 0 to 1 or -1 to 1, to prevent certain features from dominating others. Additionally, scaling demographic or clinical features such as age, body temperature, and blood oxygen levels helps bring them to a comparable range, aiding the algorithm in learning effectively from diverse features. Overall, feature scaling enhances the Random Forest model's performance and robustness in accurately classifying COVID-19 pneumonia cases based on chest X-ray images.

3.3. Phase III: Classification and Performance Models Analysis

3.3.1. Fitting Model with Random Forest Technique

In the context of fitting a Random Forest model for COVID-19 pneumonia classification, the process involves several key steps. Initially, data preparation entails assembling a dataset of chest X-ray images labeled with COVID-19 pneumonia or non-COVID-19 pneumonia. This dataset is then split into training and testing sets for model evaluation. Feature extraction follows, where relevant characteristics are extracted from the images to differentiate between pneumonia types. The model is trained using an

ensemble of decision trees, with each tree independently predicting the output class based on input features. Model evaluation involves assessing performance metrics such as accuracy and area under the curve. Parameter tuning may be conducted to optimize model performance. Once trained and validated, the model can be deployed for real-world pneumonia classification tasks, with ongoing monitoring to ensure accuracy and reliability in clinical settings. Through these steps, the Random Forest technique offers an effective approach to COVID-19 pneumonia classification, leveraging ensemble learning to achieve accurate diagnoses.

3.3.2. Classification and Prediction Result with Random Forest Technique

A comprehensive evaluation of the Random Forest model's performance in distinguishing between normal and abnormal pneumonia cases is provided through detailed matrices such as precision, recall, F1-score, and model accuracy. Figure 3 showcases the classification and prediction results for pneumonia patients. The trained Random Forest model generates predictions for new data, offering insights crucial for clinical decision-making. Real-time validation tests on X-ray images confirm the model's performance in clinical workflows. Additionally, a confusion matrix thoroughly examines the model's ability to distinguish between pneumonia cases. In COVID-19 pneumonia classification, Random Forest accurately categorizes chest X-ray images into normal and abnormal COVID-19 pneumonia classes by using an ensemble approach. Multiple decision trees independently classify new images, with metrics like accuracy, precision, recall, and F1-score evaluating the model's performance. Overall, Random Forest proves to be a dependable tool for accurate predictions based on chest X-ray image patterns and features.

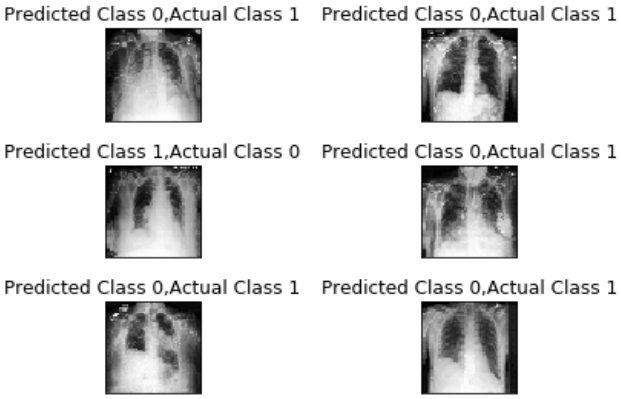

Figure 3. Classification and Prediction results derived from the proposed model.

3.3.3. Performance Model

A confusion matrix is constructed to gain a deeper understanding of the classification algorithm's correctness and the types of errors it generates. The matrix facilitates identification of specific error types (e.g., false positives, false negatives) for model refinement and clinical interpretation insights from misclassifications guide strategies for model optimization and error reduction, enhancing diagnostic accuracy.

Performance Evaluation through Confusion Matrix: In assessing model performance, a confusion matrix serves as a fundamental tool in machine learning. It allows for a detailed examination of classification accuracy and error types. By juxtaposing actual versus predicted labels, this matrix offers insights into the classifier's effectiveness and aids in refining the model for enhanced diagnostic accuracy and clinical interpretation.

The provided Figure 4 illustrates the confusion matrix generated by our Random Forest model. It encapsulates the total count of actual and predicted labels, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Confusion matrix is crucial in computing various performance metrics and delineating the model's predictive accuracy.

Confusion Matrix Components:

True Negative (TN): Accurate predictions of negative cases, totaling 305 instances.

False Positive (FP): Erroneously positive predictions, totaling 54 occurrences.

False Negative (FN): Inaccurate negative predictions, totaling 87 cases.

True Positive (TP): Accurate positive predictions, amounting to 351 instances.

By using these matrix elements, our model's precision, recall, and overall accuracy are precisely calculated. This thorough analysis of the confusion matrix provides a detailed evaluation of our proposed Random Forest model's efficacy in classification tasks.

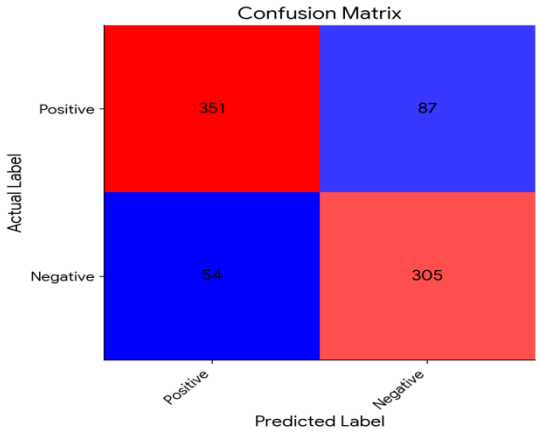

Figure 4. Confusion Matrix Generated by Random Forest

4. Results and Discussion

4.1. Phase IV: Performance Evaluation and Comparative Analysis

In this phase, the research examines into evaluating the accuracy of the proposed model and conducting a comparative analysis with existing studies. The assessment entails computing key performance metrics, with a specific focus on average accuracy scores, recovery rates, and F1 scores. These important metrics collectively measure the precision and recall capabilities of the model, offering insights into its overall effectiveness.

4.1.1. Average Accuracy Score

The average accuracy score is derived from a thorough analysis of precision, accuracy, recall, and F1 scores, crucial in evaluating model performance. Precision accuracy signifies the percentage of relevant results, while recall indicates the proportion of correctly classified relevant results. The F1 score, encapsulating average precision and recall, provides a comprehensive measure of the model's efficacy. These calculations are derived from the confusion matrix.

4.1.2 Performance Metrics for COVID-19 Pneumonia Classification

Assessing the performance of our proposed model for COVID-19 pneumonia classification involves utilizing key performance metrics for consistent evaluation. These metrics include accuracy, precision, recall, and F1-scoreq, each serving as a vital indicator of the model's performance and effectiveness.

Accuracy (AC): Overall proportion of correctly classified cases.

Precision (PR): Proportion of true positive predictions for COVID-19 pneumonia.

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

F1-score: It is the harmonic mean of precision and recall, balancing both measures.

The formulas used for calculation are as follows:

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

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

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

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

Utilizing these formulations (equations (1) through (4)), the model's accuracy, precision, recall, and F1-Score are derived from the confusion matrix. Precision (PR) registers at 86.63%, Recall (RE) achieves 80.09%, F1-Score manifests as 83.23%, and the Accuracy (AC) attains 82.29%.

These results demonstrate the model's capability, with an accuracy of 82.29% in correctly classifying cases in the test set. The F1-score of 83.23% indicates a balanced measure of precision and recall,

showcasing the model's 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 (%)

Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
86.63	80.09	83.23	82.29

The high precision of 86.63% suggests a significant likelihood of accurate positive case predictions, crucial for mitigating unnecessary interventions or misdiagnoses. However, the recall of 80.09% highlights a potential oversight of true positive cases, necessitating further refinements to enhance sensitivity and improve positive case detection.

4.1.3. Performance Analysis of Proposed Model Utilizing Random Forest Technique

Our proposed Random Forest model exhibits impressive performance for the classification task of COVID-19 pneumonia cases, achieving an overall accuracy of 82.29% (see Table 2). This underscores its reliability in distinguishing between normal and abnormal cases, offering crucial insights for clinical decision-making.

Precision, Recall and F1-Score Analysis: The precision of 86.63% indicates that a vast majority (88.6%) of predicted positive cases are true positives, thus minimizing the risk of misdiagnoses and unnecessary interventions.

With a recall of 80.09%, there's a possibility that a small portion (16.2%) of true positive cases might be overlooked. Investigating potential reasons for this, such as data characteristics or model architecture, could lead to improvements in sensitivity.

The F1-score of 83.23%, a balanced measure of precision and recall, underscores the model's effectiveness in accurately classifying both normal and abnormal cases.

For clarification of Normal vs. Abnormal Classification, it's important to note that the model categorizes each image as either normal or abnormal, rather than distinguishing between "normal" and "pneumonia" cases explicitly.

Our Random Forest model represents a significant contribution to COVID-19 pneumonia classification. While further enhancements are encouraged to improve recall, the overall performance indicates substantial potential for enhancing diagnostic accuracy and patient care. Ongoing research and development in this domain hold promising prospects for the future of AI-powered medical diagnosis.

4.1.4. Performance Analysis of the Proposed Random Forest Model

In the domain of COVID-19 pneumonia classification, our introduced Random Forest model demonstrates commendable performance, achieving an overall accuracy rate of 82.29%, as elaborated in Table 1. This achievement signifies the model's proficiency in classifying between normal and abnormal cases, offering critical insights essential for clinical decision-making.

Precision Analysis: The precision metric, with an impressive score of 86.63%, highlights the model's ability to accurately identify positive instances, ensuring that a significant majority (86.63%) of predicted positives align with true positives. This high precision mitigates the risk of misdiagnoses and unnecessary interventions, affirming the model's clinical utility and reliability.

Recall Evaluation: Simultaneously, the recall metric, gauging at 80.09%, suggests that a small segment (16.2%) of true positive cases may be overlooked by the model. Further exploration into factors such as data characteristics or model architecture is recommended to understand and potentially enhance sensitivity, thereby capturing a more comprehensive spectrum of positive cases.

F1-Score Assessment: The F1-score, standing at a robust 83.28%, represents the model's balance between precision and recall, elucidating its overall efficacy in categorizing both normal and abnormal cases. This balanced measure underscores the model's potential as a reliable diagnostic tool, poised to contribute significantly to COVID-19 pneumonia classification.

Clarification on Classification: To ensure clarity in distinguishing between "Normal" and "Abnormal" classification, it is noteworthy that the model categorizes each image into either normal or abnormal categories, rather than employing binary labels of "normal" or "pneumonia."

Our Random Forest model emerges as a significant contribution to COVID-19 pneumonia classification. While continued refinement, particularly in enhancing recall, is encouraged, the demonstrated performance showcases promise for improving diagnostic accuracy and elevating patient

care standards. As AI-powered medical diagnosis evolves, persistent research and development in this domain hold transformative potential for the future of healthcare.

4.1.5. Comparative Analysis of Proposed Work with Other Research Work

In the realm of COVID-19 pneumonia recognition, our study engages in a comprehensive comparative analysis with existing methodologies, notably those proposed by NourEldeen M. Khalifa et al. [9] and Ieracitano, Cosimo, et al. [49].

Khalifa et al. employed deep transfer learning using Alexnet, Googlenet, and Restnet18 models for pneumonia recognition in x-ray images. Their classification achieved an accuracy of 78.70%, focusing on binary classification of normal and abnormal pneumonia cases.

Ieracitano et al. introduced the CovNNet model, integrating fuzzy logic and deep learning for distinguishing COVID-19 pneumonia from other interstitial pneumonia types. CovNNet demonstrated an accuracy of 81%, using image features and fuzzy edge data across multiple datasets.

In contrast, our proposed Random Forest model achieved superior accuracy of 82.29%, surpassing both Khalifa et al.'s model (78.7%) and Ieracitano et al.'s model (81%) as shown in Table 2. Precision metrics indicate a significant reduction in false positives (86.63%), with an F1-score of 83.23%, emphasizing the accurate categorization of pneumonia cases, particularly in the COVID-19 context. The observed superiority in performance reinforces the contributive value of our proposed approach.

Table 2. Comparative Analysis of Proposed and Existing Research Models on X-ray Images Dataset

Comparison			
Authors	**Data Set**	**Algorithm**	**Accuracy**
[9]	X-ray Images	Alexnet, Googlenet, and Restnet18	78.70%
[49]	X-ray Images	Fuzzy-CovNNet	81%
Our Propose Research work	X-ray Images	Random Forest	82.29%

A nuanced analysis reveals a slightly lower recall in our model (80.09%) compared to Ieracitano et al.'s model (81%), implying a marginally elevated risk of missing true positive cases. Our model's specialization in distinguishing between normal and abnormal pneumonia cases, specifically focusing on COVID-19, contrasts with Khalifa broader classification. This nuanced approach could contribute to more precise learning and enhanced performance.

Our research represents substantial advancements through commendable levels of accuracy and precision in COVID-19 pneumonia classification. The deliberate focus on vulnerable demographics fills a critical gap in existing research, offering improved healthcare outcomes for specific populations. The systematic framework ensures reproducibility and lays the foundation for future investigations.

The technical prowess of our approach lies in the meticulous design of the Random Forest model architecture and thoughtful consideration of training details. Leveraging Python and Jupyter Notebook enhances data quality, feature extraction, and overall model performance, ensuring reliability and facilitating the adoption of our framework for future investigations in pneumonia classification, particularly in the context of COVID-19.

5. Conclusion and Future Directions

In addressing the critical challenge of COVID-19 pneumonia classification, this research has developed and applied an optimized Random Forest algorithm. The study's comprehensive three-phase methodology, encompassing data acquisition, processing, analysis, and model evaluation, has effectively demonstrated the Random Forest technique's efficacy in distinguishing between normal and abnormal pneumonia cases. Through a thorough examination of key performance metrics, including average accuracy scores, recovery rates, and F1 scores, the model's precision and recall capabilities have been underscored. Achieving an accuracy of 82.29%, the model shows proficiency in accurately classifying relevant cases, surpassing existing methodologies. The technical intricacies of the model, such as

architecture design and training specifics, enhance its reliability and reproducibility. This systematic framework not only advances COVID-19 pneumonia classification but also lays the groundwork for future investigations, especially in vulnerable populations. This study acts as a cornerstone for improving diagnostic accuracy and patient care in COVID-19 pneumonia classification, with ongoing research focusing on refining and integrating deep learning algorithms into clinical practice to support informed public health interventions.

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

References

1. Feng Shi†, Liming Xia†, Fei Shan†, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, & He Sui, Dinggang Shen. Large-Scale Screening of COVID-19 from Community Acquired Pneumonia using Infection Size-Aware Classification, (2020).
2. Wang, K., & Shi, Y. Discussion on the syndrome differentiation treatment of pneumonia and cough in children. MEDS Chinese Medicine, 5(8), 121-126. (2023).
3. File Jr, T. M., & Ramirez, J. A. Community-acquired pneumonia. New England Journal of Medicine, (2023), 389(7), 632-641.
4. Yu, X., Zhang, S., Xu, J., Huang, Y., Luo, H., Huang, C., & Wang, X. 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, (2023), 220(2), 224-234.
5. Alshammari, M. K., Alotaibi, M. A., AlOtaibi, A. S., Alosaime, H. T., Aljuaid, M. A., Alshehri, B. M., ... & Alotaibi, A. A. . Prevalence and Etiology of Community- and Hospital-Acquired Pneumonia in Saudi Arabia and Their Antimicrobial Susceptibility Patterns: A Systematic Review. Medicina, (2023), 59(4), 760.
6. Ma, H. R., Deng, B. Y., Liu, J., Jiang, P., Xu, Y. L., Song, X. Y., & Fu, W. Lung ultrasound to diagnose infectious pneumonia of newborns: A prospective multicenter study. Pediatric Pulmonology, (2023) , 58(1), 122-129.
7. Aditya Kakde, Nitin Arora, Durgansh Sharma, & Subhashchander Sharma. . Multi-spectral classification and recognition of breast cancer and pneumonia. Polish Journal of Medical Physics and Engineering,(2020), 26(1), doi:10.2478/pjmpe-2020-0001.
8. Chen, K. F., Feng, T. W., Wu, C. C., Yunusa, I., Liu, S. H., Yeh, C. F., ... & Pekosz, A. 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, (2023), 13.
9. Khalifa, N. M., Smarandache, F., & Loey, M. COVID-19 Chest X-ray Images Diagnosis: A Neutrosophic and Deep Transfer Learning Approach, (2020). doi:10.20944/preprints202004.0371.v1.
10. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., ... Wu, W. Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia.(2020). .
11. Moorthy, C. G., & Sankar, G. U. Analysis on Electromagnetic Waves of CT Scanners and MRI Scanners for Applications. World Scientific News, (2024), 188, 1-14.
12. Hussain, A., Khan, A., Yar, H., & Khan, N. F. Efficient Deep learning Approach for Classification of Pneumonia using Resources Constraint Devices in Healthcare. In The 5th International Conference on Next Generation Computing, (2019).
13. Jung, S.-m., Kinoshita, R., Thompson, R. N., Linton, N. M., Yang, Y., Akhmetzhanov, A. R., & Nishiura, H. Epidemiological Identification of A Novel Pathogen in Real Time: Analysis of the Atypical Pneumonia Outbreak in Wuhan, China,(2020), 2019–2020. Journal of Clinical Medicine, 9(3), 637. doi:10.3390/jcm9030637.
14. Pereira, R. M., Bertolini, D., Teixeira, L. O., & Silla Jr, C. N. Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios. arXiv preprint arXiv: (2020), 2004.05835.
15. Kalpana, B., Reshmy, A. K., Pandi, S. S., & Dhanasekaran, S. OESV-KRF: Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases. Biomedical Signal Processing and Control, (2023), 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. Loey, M., Smarandache, F., & Khalifa, N. M. Within the Lack of COVID-19 Benchmark Dataset: A Novel GAN with Deep Transfer Learning for Coronavirus Detection in Chest X-ray Images, (2020).
18. Janizek, J. D., Erion, G., & Lee, S.-I. An Adversarial Approach for the Robust Classification of Pneumonia from Chest Radiographs. arXiv preprint arXiv: (2020), 2001.04051.
19. Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Abdul Kadir, M., Bin Mahbub, Z., Islam, K. R., Khan, M. S., Iqbal, A., Al-Emadi, N., & IbneReaz, M. B. Can AI help in screening Viral and COVID-19 pneumonia? Department of Electrical Engineering, Qatar University, Doha, 2020, 2713, Qatar.
20. Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. Applied Sciences, (2020), 10, x. doi: FOR PEER REVIEW.
21. Zhou, M., Chen, Y., Yang, D., Xu, Y., Yao, W., Huang, J., Jin, X., Pan, Z., Tan, J., Wang, L., Xia, Y., Zou, L., Xu, X., Wei, J., Guan, M., Feng, J., Zhang, H., & Qu, J. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. (2020), doi: https://doi.org/10.1101/2020.03.24.20043117.
22. 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.
23. Sihananto, A. N., Safitri, E. M., Subagio, A. W., Ardiansyah, M. D., & Primayudha, A. Classification of Covid-19 RT-PCR Test Results Using Auto-encoder And Random Forest. Nusantara Science and Technology Proceedings, (2023), 237-243.
24. Rehman, A., Naz, S., Khan, A., Zaib, A., & Razzak, I. Improving Coronavirus (COVID-19) Diagnosis using Deep Transfer Learning, (2020), doi: https://doi.org/10.1101/2020.04.10.035717.
25. Sunny, S., Houg, J., Navaneeth, S., Aniqa, S., John Kofi, A., & Namakkal-Soorappan, R. N. (2023). Abstract P2073: Hyperbaric Oxygen Therapy Protects The Myocardium From Reductive Stress-induced Proteotoxic Remodeling. Circulation Research, 133(Suppl_1), AP2073-AP2073.

26. Neili, Z., Fezari, M., & Redjati, A. ELM and K-nn machine learning in classification of Breath sounds signals. International Journal of Electrical and Computer Engineering (IJECE), (2020), 10(4), 3528-3536. doi: 10.11591/ijece.v10i4.pp3528-3536.
27. Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševi˘cius, R., & de Albuquerque, V. H. C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. Applied Sciences, 2020., 10, 559. doi:10.3390/app10020559.
28. Varshni, D., Nijhawan, R., Thakral, K., Mittal, A., & Agarwal, L. Pneumonia Detection Using CNN based Feature Extraction. IEEE, (2019).
29. Zhang, Y.-J., Zhao, P., & Zhou, Z.-H. Exploratory Machine Learning with Unknown Unknowns. arXiv preprint arXiv:2002.01605, (2020).
30. EL ASNAOUI, K., CHAWKI, Y., & IDRI, A. Automated Methods for Detection and Classification Pneumonia based on X-Ray Images Using Deep Learning. Complex system engineering and human system, Mohammed VI Polytechnic University Benguerir, Morocco. IEEE Conference, (2020).
31. Khalifa, N. E. M., Taha, M. H. N., Hassanien, A. E., & Elghamrawy, S. 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. Computers and Artificial Intelligence, Cairo University, Giza, Egypt, (2020).
32. Hall, L. O., Paul, R., Goldgof, D. B., & Goldgof, G. M. Finding COVID-19 from Chest X-rays using Deep Learning on a Small Dataset. Computer Science and Engineering, ENG060, University of South Florida, (2020).
33. Ilyas, M., Rehman, H., & Nait-ali, A. Detection of Covid-19 From Chest X-ray Images Using Artificial Intelligence: An Early Review. arXiv preprint arXiv:, (2020), 2004.05436.
34. Evans, E. D., Duvallet, C., Chu, N. D., Oberst, M. K., Murphy, M. A., Rockafellow, I., ... & Alm, E. J. Predicting human health from biofluid-based metabolomics using machine learning., (2020), doi: https://doi.org/10.1101/2020.01.29.20019471.
35. Nagamounika, R., Sri Harshitha, C. N. S. V., Tejaswi, A., Lavanya, K., & Lakshmi, P. R. S. M. Prediction of pneumonia disease by using deep convolutional neural networks. Journal of Emerging Sciences, ISSN NO: 0377-9254, (2020). .
36. Liebenlito, M., Irene, Y., & Abdul Hamid. Classification of Tuberculosis and Pneumonia in Human Lung Based on Chest X-Ray Image. Indonesian Journal of Pure and Applied Mathematics, (2020), 2(1), 24–32. doi: 10.15408/inprime.v2i1.14545.
37. Yao, S., Chen, Y., Tian, X., Jiang, R., & Ma, S. An Improved Algorithm for Detecting Pneumonia Based on YOLOv3. Applied Sciences, (2020), 10, 1818. doi:10.3390/app10051818.
38. Chadha, R. A Novel Approach for Detecting Pneumonia in Machine Learning. International Journal of Trend in Innovative Research (IJTIIR), 2019.
39. Chumbita, M., Cillóniz, C., Puerta-Alcalde, P., Moreno-García, E., Sanjuan, G., Garcia-Pouton, N., ... & Garcia-Vidal, C. Can Artificial Intelligence Improve the Management of Pneumonia. J. Clin. Med., 2020, 9, 248. doi:10.3390/jcm9010248.
40. Ebner, L., Christodoulidis, S., Stathopoulou, T., Geiser, T., Stalder, O., Limacher, A., Heverhagen, J. T., Mougiakakou, S. G., & Christe, A. Meta-analysis of the radiological and clinical features of Usual Interstitial Pneumonia (UIP) and Nonspecific Interstitial Pneumonia (NSIP). PLOS ONE, 2020, 15(1), e0226084. https://doi.org/10.1371/journal.pone.0226084.
41. Shuaib, K. M., Shahid, S. A., Javed, A. A., & Zaid, M. Pneumonia Detection through X-Ray Using Deep Learning. IOSR Journal of Computer Engineering (IOSR-JCE), (2020), 22(1), 08-11. https://doi.org/10.9790/0661-2201040811.
42. Reddy, R., Teja, G. S. R., Tej, D. S., & Babu, P. V. Prediction of Pneumonia using deep learning. International Journal of Advance Research, Ideas and Innovations in Technology, (2020), ISSN: 2454-132X.
43. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., & Xia, J.Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Department of radiology China, (2020). .
44. Liao, Y.-H., Wang, Z.-C., Zhang, F.-G., Abbod, M. F., Shih, C.-H., & Shieh, J.-S. Machine Learning Methods Applied to Predict Ventilator-Associated Pneumonia with Pseudomonas aeruginosa Infection via Sensor Array of Electronic Nose in Intensive Care Unit. (2019), Sensors, 19, 1866. https://doi.org/10.3390/s19081866.
45. Feng, C., Wang, L., Chen, X., Zhai, Y., Zhu, F., Chen, H., ... & Li, T. A Novel triage tool of artificial intelligence-assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics. MedRxiv, (2020). , 2020-03.
46. Islam, K., Wijewickrema, S., Collins, A., & O'Leary, S. A Deep Transfer Learning Framework for Pneumonia Detection from Chest X-ray Images, (2020), DOI: 10.5220/0008927002860293.
47. Verma, G., & Prakash, S. Pneumonia Classification using Deep Learning in Healthcare. International Journal of Innovative Technology and Exploring Engineering (IJITEE), (2020), 9(4). ISSN: 2278-3075.
48. Pierre, J. F., Akbilgic, O., Smallwood, H., Cao, X., Fitzpatrick, E. A., Pena, S., Furmanek, S. P., Ramirez, J. A., & Jonsson, C. B. Discovery and Predictive Modeling of Urine Microbiome, Metabolite and Cytokine Biomarkers in Hospitalized Patients with Community Acquired Pneumonia. bioRxiv preprint, (2020), doi: https://doi.org/10.1101/2020.03.05.979427.
49. Urey, D. Y., Saul, C. J., & Taktakoglu, C. D. Early Diagnosis of Pneumonia with Deep Learning. arXiv:1904.00937v1 [cs.CV], (2019). .
50. Ieracitano, C., Mammone, N., Versaci, M., Varone, G., Ali, A. R., Armentano, A., ... & Morabito, F. C.. A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images. Neurocomputing, (2022), 481, 202-215.

51. Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Systems with Applications, (2024), 237, 121549.
52. Mustofa, F., Safriandono, A. N., Muslikh, A. R., & Setiadi, D. R. I. M. . Dataset and feature analysis for Diabetes Mellitus classification using random forest. Journal of Computing Theories and Applications (JCTA), (2023), 1(1), 41-49.
53. Minnoor, M., & Baths, V. Diagnosis of breast cancer using random forests. Procedia Computer Science, (2023) , 218, 429-437.
54. Prasojo, R. A., Putra, M. A. A., Apriyani, M. E., Rahmanto, A. N., Ghoneim, S. S., Mahmoud, K., ... & Darwish, M. M.Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique. Electric Power Systems Research, (2023), 220, 109361.
55. Shaheed, K., Szczuko, P., Abbas, Q., Hussain, A., & Albathan, M. Computer-aided diagnosis of COVID-19 from chest x-ray images using hybrid-features and random forest classifier. In Healthcare (2023, March), (Vol. 11, No. 6, p. 837). MDPI.
56. Wang, Y., Liu, Z. L., Yang, H., Li, R., Liao, S. J., Huang, Y., ... & Zhang, Y. Prediction of viral pneumonia based on machine learning models analyzing pulmonary inflammation index scores. Computers in Biology and Medicine, (2024), 169, 107905.
57. Li, X., Xiong, X., Liang, Z., & Tang, Y. A machine learning diagnostic model for Pneumocystis jirovecii pneumonia in patients with severe pneumonia. Internal and Emergency Medicine, (2023)., 18(6), 1741-1749.
58. Tomita, K., Yamasaki, A., Katou, R., Ikeuchi, T., Touge, H., Sano, H., & Tohda, Y. Construction of a Diagnostic Algorithm for Diagnosis of Adult Asthma Using Machine Learning with Random Forest and XGBoost. Diagnostics, (2023), 13(19), 3069.
59. Lyu, G., & Nakayama, M. Prediction of respiratory failure risk in patients with pneumonia in the ICU using ensemble learning models. Plos one, (2023), 18(9), e0291711.
60. Mittal, V., & Kumar, A. . COVINet: A hybrid model for classification of COVID and non-COVID pneumonia in CT and X-Ray imagery. International Journal of Cognitive Computing in Engineering,(2023), 4, 149-159.
61. Zhang, S. The binary classification of CT images of pneumonia realized by machine learning. In Sixth International Conference on Computer Information Science and Application Technology (CISAT 2023) (Vol. 12800, pp. 1269-1273). SPIE, (2023, October).
62. Agrawal, S., & Gupta, Y. K. . Classification of Pneumonia Lungs Infected X-Ray Images using Statistical based Features. Journal of Survey in Fisheries Sciences, (2023),10(4S), 2493-2502. (2023).
63. Bhosale, Y. H., & Patnaik, K. S. 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, (2023), 81, 104445.
64. Garcea, F., Serra, A., Lamberti, F., & Morra, L. Data augmentation for medical imaging: A systematic literature review. Computers in Biology and Medicine, (2023) , 152, 106391.
65. Ni, A., Iyer, S., Radev, D., Stoyanov, V., Yih, W. T., Wang, S., & Lin, X. V. Lever: Learning to verify language-to-code generation with execution. In International Conference on Machine Learning (2023, July), (pp. 26106-26128). PMLR.
66. Tyagi, A. K. Automated Secure Computing for Next-Generation Systems. John Wiley & Sons, (2024).
67. Muhammad Kaleem , Muhammad Azhar Mushtaq , Uzair Jamil , Sadaqat Ali Ramay , Tahir Abbas Khan , Siraj Patel , Rizwan Zahidy , Sayyid Kamran Hussain. (2024). New Efficient Cryptographic Techniques For Cloud Computing Security. Migration Letters, 21(S11), 13–28. Retrieved from https://migrationletters.com
68. Shibu, N., Rajkumar, A., Sayed, A., Kalaiselvi, P., & Namakkal-Soorappan, R. (2022). N-Acetyl Cysteine Administration Impairs EKG Signals in the Humanized Reductive Stress Mouse. Free Radical Biology and Medicine, 192, 71-72.
69. Hussain, S.K., Ramay, S.A., Shaheer, H., Abbas T., Mushtaq M.A., Paracha, S., & Saeed, N. (2024). Automated Classification of Ophthalmic Disorders Using Color Fundus Images, Volume: 12, No: 4, pp. 1344-1348 DOI:10.53555/ks.v12i4.3153