

RESEARCH ARTICLE


OPEN ACCESS

Manuscript received July 25, 2024; revised August 10, 2024; accepted August 20, 2024; date of publication November 11, 2024
Digital Object Identifier (DOI): <https://doi.org/10.35882/jeeemi.v7i1.529>

Copyright © 2024 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-Share Alike 4.0 International License ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).

How to cite: Sravan kiran Vangipuram, Rajesh Appusamy, "A Novel Image Feature Extraction based Machine Learning Approach for Disease Detection from Chest X-Ray Images", Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 7, no. 1, pp. 47-55, January 2025.

A Novel Image Feature Extraction Based Machine Learning Approach for Disease Detection from Chest X-Ray Images

Sravan Kiran Vangipuram^{1,2} , Rajesh Appusamy¹ 

¹ Department of Computer Science & Engineering, Jain (Deemed to be University), Bangalore, India

² Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

Corresponding author: Sravan Kiran Vangipuram (e-mail: shravankiran12@gmail.com).

The authors are thankful to the Department of Computer Science and Engineering, Jain University and IT Department, VNR Vignana Jyothi Institute of Engineering, Hyderabad for providing valuable inputs and computing resources for carrying this research.

ABSTRACT The limitation of feature selection is the biggest challenge for machine learning classifiers in disease classification. This research proposes a novel feature extraction method to extract representative features from medical images, combining extracted features with original image pixel features. Additionally, we propose a new method that uses data values from Andrews's curve function to transform chest x-ray images into spectrograms. The spectrogram images are believed to aid in distinguishing near-similar medical images, such as COVID and pneumonia. The study aims to build an efficient machine learning system that applies the proposed feature extraction method and utilizes spectrogram images for distinguishing near-similar medical images. For experimental analysis, we have used the award winning Kaggle Chest Radiography image dataset. The test results show that among all machine learning classifiers, the logistic regression classifier could correctly distinguish COVID and pneumonia images with a 97.18% test accuracy, a 98.34% detection rate, a 97.8% precision rate, and an AUC value of 0.99 on the test dataset. The machine learning model has learned to distinguish between medical images that appear similar using features found through the proposed feature extraction and spectrogram images. The results also proved that the proposed approach using XGBoost has outperformed state-of-the-art models in recent research studies when (i) binary classification is performed using COVID-19 and Normal Chest x-ray images and (ii) multiclass classification is performed using Normal, COVID and Pneumonia Chest x-ray images.

INDEX TERMS Feature extraction, Linear transformation, Classification, Prediction, Balanced accuracy.

I. INTRODUCTION

Generally, a virus is a microscopic organism that relies on the host cell, be it human, animal, or another virus, for its development. It then uses the host's cell to produce new virus particles, thereby making the host sick. The SARS-CoV-2 virus causes the serious infectious disease coronavirus, often referred to as COVID-19. Its impact ranges from a mild respiratory illness to organ failure and death. This virus emerged in Wuhan, China, in November 2019 and later spread all over the world. This outbreak resulted in around 34, 23,217 confirmed cases in India alone, the second-highest number

following the US until October 2021. Pneumonia, an infectious disease, affects one or both lungs. Generally, the lungs contain alveoli (small sac-like structures). When a normal person breathes in, they are full of air. However, when a person with pneumonia receives a diagnosis, their alveoli fill with fluid, leading to a restricted oxygen intake. It is a fact that pneumonia contributes to one-third of global deaths. A comprehensive overview of the global burden of these diseases, including recent statistics and trends is presented below.

A. CONTEXTUALIZATION OF THE PROBLEM – GLOBAL BURDEN

COVID-19 has disrupted healthcare systems globally, leading to delays in routine medical care and increased demand for healthcare resources. The high number of cases and fatalities indicates that COVID-19 has significantly impacted global health. Vaccination efforts have reduced the severity and mortality of the disease, but new variants and long COVID present ongoing challenges. Also, Pneumonia continues to be a major health issue, particularly affecting children under 5 and vulnerable adults. Despite vaccination efforts, pneumonia remains a leading cause of morbidity and mortality globally. A detailed and comprehensive overview of the global burden of pneumonia and COVID-19, including the most recent statistics and trends is presented below.

1. GLOBAL BURDEN OF PNEUMONIA

The air sacs in one or both lungs can fill with fluid or pus in pneumonia. Each year, it impacts millions of people, making it a significant global health concern. According to the World Health Organization (WHO), in 2022, there were approximately 150 million new cases of pneumonia among children under 5 years old globally. Pneumonia was responsible for approximately 600,000 deaths among children under 5 years old in 2022. This makes it one of the leading causes of child mortality worldwide. In 2022, there were about 1 million hospitalizations for pneumonia in the United States alone, among adults, as reported by the Centers for Disease Control and Prevention (CDC). In the United States, pneumonia caused around 50,000 deaths in adults in 2022. Pneumonia remains a critical global health issue, particularly for children under 5 and vulnerable adults. Despite progress through vaccination and improved treatments, pneumonia continues to be a leading cause of morbidity and mortality, especially in low-resource settings.

2. GLOBAL BURDEN OF COVID-19

The SARS-CoV-2 virus, which caused COVID-19, first appeared as a pandemic in early 2020 and has significantly impacted daily life, economies, and global health. As of August 2024, there have been over 700 million confirmed cases of COVID-19 globally, according to Johns Hopkins University. The first half of 2024 saw approximately 10 million new cases reported worldwide. COVID-19 has resulted in over 6.9 million deaths globally as of August 2024. In the first half of 2024, there were about 200,000 deaths attributed to COVID-19, reflecting a reduced but ongoing impact compared to earlier in the pandemic.

Both COVID-19 and Pneumonia have had a negative impact on people's lives and continue to pose significant challenges. In general, we diagnose these diseases using two methods: laboratory techniques and medical image analysis.

Laboratory techniques entail the analysis of nasal swabs to detect the SARS-CoV-2 virus's genetic material (RNA). This is known as the RT-PCR test, often considered the golden standard for detecting COVID-19 at early stages. However, the test has a few drawbacks. Since it is a manual process, it is

time-consuming, prone to human errors (technique variations and incorrect labeling) and may also be expensive for a few individuals. We perform blood tests to determine the presence of any infection in the body, specifically for pneumonia.

The second method involves the analysis of medical images, especially chest X-rays and CT scan images. Although these are the best ways to detect respiratory diseases, there are a few drawbacks to completely relying on them. Skilled radiologists must analyze these X-rays and scan images, and it's possible that different radiologists may interpret the images differently, potentially leading to an inconsistent diagnosis. Moreover, expert radiologists may not be present in remote areas, which is a serious concern.

To overcome RT-PCR's limitations in viral detection, machine learning with chest X-rays involves leveraging the strengths of each method to complement and enhance diagnostic capabilities. Here, we outline how this machine learning-based disease diagnosis could potentially address some of the limitations of RT-PCR.

For example, by using large datasets consisting chest X-ray images and also having the respective RT-PCR results, we can train machine learning algorithms, enabling the development of models that integrate knowledge from both modalities. When we combine features from chest X-rays, such as viral pneumonia patterns, with RT-PCR results, we can improve the overall sensitivity and specificity of finding viral infections.

Chest X-ray images can sometimes show characteristic signs of viral pneumonia before RT-PCR results become positive, especially in the early stages of lung infection. By utilizing machine learning models to identify changes in chest X-ray features over time and correlating these findings with RT-PCR results, we can detect viral infections earlier and respond to them more promptly.

RT-PCR provides molecular confirmation of viral presence, while chest X-rays offer structural and morphological information about lung tissue. By combining these datasets through machine learning, clinicians can obtain a more comprehensive picture of the infection status and severity, aiding in clinical decision-making.

We can train machine learning models to identify patterns in chest X-rays linked to various viral strains or mutations that could influence disease presentation. This could potentially provide insight into variant-specific lung pathology and help adapt treatment strategies accordingly. While chest X-rays may not be as widely available or inexpensive as RT-PCR tests in all settings, advancements in portable X-ray technology and automated image analysis through machine learning could improve accessibility over time.

Machine learning algorithms can measure small changes in chest X-rays over time, showing how the disease is getting worse and how well the treatment is working. This may add to the positive or negative results of RT-PCR by providing more clinical information. Hence, by integrating machine learning-enhanced chest X-ray analysis into existing clinical workflows, we can streamline diagnostic processes, potentially reducing turnaround times and enhancing overall

patient care. The challenges in using medical images for disease classification and pros of machine learning are outlined in subsections below.

B. CHALLENGES IN USING MEDICAL IMAGES FOR DISEASE CLASSIFICATION

Handling medical images for classification poses several challenges due to the complexity and unique characteristics of medical imaging data. Some of the key challenges are as follows (a) Medical imaging data evolves over time with advances in technology and changes in clinical practices. Machine learning models must therefore be adaptable and capable of continual learning in order to remain effective. (b) Medical images often require extensive preprocessing, such as image normalization, image enhancement, image noise reduction, and geometric transformations, to improve classification accuracy and standardize inputs to machine learning models. Some medical conditions may be rare, leading to imbalanced datasets where certain classes have fewer examples. For instance, the recent pandemic has resulted in a scarcity of data regarding COVID-19 images. This imbalance has the potential to bias the model training, which, in turn, can impact the overall classification performance. Handling medical data raises ethical concerns regarding patient privacy, patient consent, and patient data security. Compliance with regulatory standards adds complexity to data access and usage (d). Labeling medical images for supervised learning often requires expert knowledge and can be time-consuming and subjective. Inter-observer variability among experts can also impact the quality and consistency of labels, which can affect the performance of machine learning classifiers. (e) Medical images are typically highly dimensional and large, leading to challenges in storage, processing, and transmission. This necessitates efficient data management strategies and computational resources.

C. PROS OF USING MACHINE LEARNING FOR DISEASE CLASSIFICATION FROM MEDICAL IMAGES

Machine learning (ML) offers several advantages for disease detection from medical images, leveraging its capabilities to analyze complex image patterns and make effective predictions. Some of the advantages of machine learning-based disease diagnosis are: (a) The process of automating the analysis of medical images reduces the workload for radiologists and clinicians. The automation process speeds up diagnosis and treatment planning, leading to more efficient healthcare delivery (b) Machine learning models possess the capability to detect hidden patterns and salient features in medical images, which human eyes might find challenging to discern. This can lead to improved accuracy in disease detection and classification, potentially reducing false positives and false negatives (c) Machine learning techniques can effectively scale with large volumes of medical image data, enabling the analysis of diverse patient populations and rare conditions. This scalability is crucial for handling the growing volume and complexity of medical imaging data (d) ML models can identify early signs of disease or predict disease progression based on imaging biomarkers. This early

detection enables timely intervention and improved patient outcomes. (e) Machine learning methods enable continuous learning by continuously training the machine learning models. ML models can continuously learn from new data and adapt over time, improving their performance and robustness. This capability supports the ongoing refinement and optimization of disease classification algorithms (f) though initial implementation and training of ML models may require investment, they can ultimately lead to cost savings by optimizing resource allocation, reducing unnecessary procedures, and improving patient outcomes.

D. RESEARCH GAPS IDENTIFICATION

The research gaps that motivated for the proposed feature extraction presented in this paper are: (i) Feature extraction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA), which rely on linear correlations between features, may overlook intricate patterns in image data. For datasets with complex relationships, PCA may not be as effective since it is unable to capture complicated, non-linear dependencies (ii) Local Binary Patterns (LBP), a feature extraction method, captures local texture information by encoding pixel neighborhoods into binary patterns, but it is sensitive to changes in illumination and loses the spatial context. On the other hand, the Gray Level Co-occurrence Matrix (GLCM) analyzes the spatial relationships of image pixels for texture feature extraction, but it can be computationally intensive and sensitive to noise. Histogram of Oriented Gradients (HOG) focuses on the distribution of gradient orientations for object detection but is high-dimensional and sensitive to object orientation. All three methods are effective for texture analysis, each with specific strengths and limitations (iii) Standard feature extraction techniques like linear discriminant analysis (LDA) come with predefined assumptions and may not be easily adaptable to diverse datasets (iv) By comparing the image matrix structures, we can discover structural similarities that are not immediately apparent with conventional feature extraction methods. There is a scope to explore the potentiality of utilizing matrix structures in the context of COVID and pneumonia diagnosis. By analyzing the similarity between image matrices, we can capture intricate disease patterns and relationships between chest x-ray image features. This is particularly useful in scenarios where conventional methods might miss subtle correlations. There is a limited exploration of integrating matrix similarity methods with machine learning techniques, which could potentially enhance feature extraction capabilities (v) The research literature does not use spectrogram image representations of chest x-rays for COVID-19 and pneumonia diagnosis to the best of our knowledge.

E. OBJECTIVES AND HYPOTHESES

When investigating feature extraction using matrix conjugacy, testable hypotheses can guide the research and experiments. Here are two specific, testable hypotheses related to this

research. The hypotheses are: (i) Matrix similarity-based feature extraction improves the classification accuracy of disease diagnosis compared to traditional pixel-based features. (ii) Matrix similarity methods can effectively reduce the dimensionality of image features while retaining essential diagnostic information.

The testable aspect w.r.t first hypothesis is to compare the performance of machine learning models trained on traditional pixel-based features with those trained on features extracted using the proposed matrix conjugacy method by using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). In similar lines, the testable aspect w.r.t second hypothesis is to evaluate the impact of matrix similarity techniques on the number of features and assess whether the new feature set maintains diagnostic performance.

F. STUDY DESIGN

In this study, we have utilized the COVID-19 radiography dataset to classify disease. The dataset comprises three classes: normal, pneumonia, and COVID-19. We utilized the COVID-19 radiography dataset [24] that includes chest X-ray images, which belong to three classes. (a) X-ray images showing no signs of disease; (b) X-ray images indicative of pneumonia, which may include various types of pneumonia unrelated to COVID-19; and (c) X-ray images exhibiting features specific to COVID-19 infection. We standardized the size and format of chest X-ray images for image classification. We perform three types of scaling: (i) min-max scaling, (ii) standardization, and (iii) quantile transformation on the data obtained after feature extraction. The performance of ML classifiers (a) logistic regression, (b) decision trees, (c) naive bayes, (d) support vector machines (SVM), (e) linear discriminant analysis, (f) multilayer perceptron (MLP), (g) quadratic discriminant analysis, and (h) XGBoost models which are implemented using Python are assessed for classification and prediction. We have split the dataset into training and testing subsets with a 90%-10% split to evaluate the proposed method and have ensured that the chest x-ray images in the test dataset are not present in the training dataset. The performance of the learning models is evaluated using the metrics such as accuracy, precision, recall, F1-score, and ROC curves.

G. DATASET DESCRIPTION

For experimentation analysis in this research, we have used the COVID-19 Radiography dataset [24] which is available publicly at the Kaggle and can be accessed via <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>. This dataset is the Winner of the COVID-19 Dataset Award by Kaggle Community.

1. PURPOSE

The database was created to support the development and evaluation of diagnostic algorithms, particularly those using machine learning or deep learning techniques, to identify COVID-19 from radiographic images. The aim is to enhance the accuracy and speed of diagnosis during the pandemic. The

database typically includes thousands of chest X-ray images categorized into three main classes (a) Chest X-ray images of healthy lungs without any disease (b) Chest X-ray images showing signs of pneumonia, which may include various types not related to COVID-19 and (c) Chest X-ray images exhibiting characteristics specific to COVID-19 infections.

2. DATA COLLECTION

Images are collected from a variety of sources, including medical institutions, public health repositories, and research studies. The images are often accompanied by metadata that includes details about patient demographics and clinical findings.

3. QUALITY ASSURANCE

The images in the database are usually vetted for quality and relevance. Radiologists may review and annotate the images to ensure accurate labeling, which is crucial for training machine learning models.

4. ETHICAL CONSIDERATIONS

Efforts are made to anonymize patient information to protect privacy. Ethical guidelines and regulations regarding data use are followed to ensure compliance with legal standards in this dataset.

5. DESCRIPTION

Originally, the dataset comprised of 21,165 chest X-ray images which are distributed into four different classes: (i) COVID (3616 images), (ii) lung opacity (6012 images), (iii) viral pneumonia (1345 images), and normal (10192 images). Fig. 1 shows sample Chest X-ray images belonging to four classes from the COVID-19 Radiography dataset.

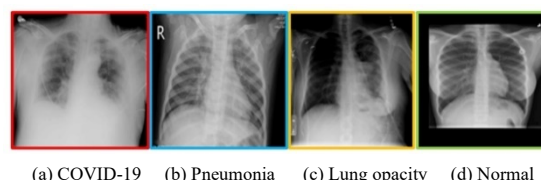


FIGURE 1. (A-D) Chest X-ray images collected from Kaggle dataset. Sample chest X-ray images of different classes in the dataset. The eyes of a non-expert human can hardly distinguish between the classes.

In this study, we considered three classes namely (i) Covid (ii) Pneumonia and (iii) Normal. We discard the lung opacity class in this paper as our objective is to discriminate between COVID, Pneumonia and Normal Chest X-ray images. For experimental analysis, we have split the resulting 3-class dataset into two subset datasets train and test datasets. The details of train and test datasets are provided in the results section of this paper. More details on training and testing subset datasets used for experimental analysis is described in experimental results section.

H. OUR CONTRIBUTIONS

The research gaps mentioned in the previous subsection forms the motivation to propose a new feature extraction method for machine learning from medical images such as chest X-ray images and a method to generate spectrogram images from chest radiography images which can not only detect COVID-19 with high detection rate but also differentiate between COVID-19 and Pneumonia with high detection rates. The contributions of this paper are as follows:

- (i) A new method is proposed for feature extraction from medical images which is based on equivalence relation on square matrices, indicating that they share all properties of their shared underlying operator with respect to different bases. This is based on the property, “Similarity is an equivalence relation on square matrices, indicating that similar matrices share all properties of their shared underlying operator with respect to different bases” which has strong mathematical basis.
- (ii) A method for generating spectrogram images from chest X-ray images is presented in this paper.
- (iii) We address the challenge in classifying COVID & Pneumonia by proposing an approach for generating spectrogram images from original chest X-ray images based on Andrews function. The new image representation is then used to classify COVID and Pneumonia which have near similarity between them.
- (iv) We have shown that better prediction rate can be achieved by the use of spectrogram images obtained from chest X-ray images for binary class classification and multi-class classification.

This paper is organized as follows: Section II describes the related works to the present research. Section III introduces the proposed methodology for spectrogram image generation from chest X-ray images. Section IV outlines the algorithm for spectrogram generation, feature extraction, and classification. Section V presents the proposed machine learning system for automated diagnosis; Section VI highlights the proposed method's performance results. Section VII highlights findings and insights derived from the experimental results along with a comparative study with state-of-the-art recent studies. Section VIII presents the future research directions and Section IX concludes the work carried in this research.

II. RELATED WORKS

In this section, we outline some of the state-of-the-art research contributions that forms the basis for the current research contribution. The study by Nikolaou et al. [1] employs a light convolution neural network to differentiate COVID-19 from other viral pneumonia and healthy lungs in chest X-ray images, making it the most efficient CNN for diagnosing COVID-19. They developed a hybrid CNN that combines a pre-trained EfficientNetB0 network with a dense 32-neuron layer to distinguish between COVID-19 and normal lung X-ray images. For experimental analysis, the test dataset consisted of 1020 normal and 362 COVID-19 chest X-ray

images, each measuring 150x150. The model achieved 91.53% accuracy after feature extraction and 94.93% accuracy after fine-tuning in two-class classification. The fine-tuned model effectively differentiated COVID-19 from normal lungs, with a positive predictive value of 91% and a specificity of 97%. Deep neural networks (DNN) have gained popularity for medical imaging, particularly in detecting COVID-19 cases from chest CT images. However, DNN models struggle to explain the reasoning behind diagnosis, a challenge that clinicians need. The study by Kanika Goel et al. [2] evaluates the quality of explanations for a deep learning model, comparing ground truth and machine learning explanations. Results show that while explanations improve clinicians' trust in automated diagnosis, their reliance on the diagnosis decreases as they are less likely to rely on algorithms that are not human-like. Clinicians desire higher recall for better understanding of automated diagnosis systems.

The research study carried out by Al-Zyoud et al. [4] aimed to develop an automatic diagnosis method for COVID-19 using binary segmentation of chest X-ray images. The research used frontal chest X-ray images from 27 infected and 19 uninfected individuals from the Kaggle COVID19 Radiography Database. Results showed that COVID-19 patients had higher attenuation in the lower lobes of the lungs compared to healthy individuals. The study supports the theory that COVID-19 primarily affects the lower and lateral fields of the lungs, with the virus accumulating mostly in the lower left quarter. In the research study by Abdullah et al. [5], a Hybrid Deep Learning CNN model is proposed for diagnosing COVID-19 using chest X-ray s. The model consists of a heading model and a base model, incorporating pre-trained deep learning structures and reducing feature dimensions. Experimental analyses were conducted to compare the model with existing transfer learning architectures. The model achieved an accuracy of 92%, helping radiologists and physicians avoid misdiagnosis rates and validate positive COVID-19 cases. Ismael et al. [6] used three deep CNN approaches for COVID-19 detection on chest X-ray images. Deep feature extraction, fine-tuning, and an end-to-end trained model were tested. The ResNet50 model and SVM classifier with linear kernel function achieved a 94.7% accuracy score, with fine-tuned ResNet50 model achieving 92.6% and developed CNN model achieving 91.6%. Results showed that deep learning approaches outperformed local descriptors, fine-tuning and end-to-end training took more time, and the Cubic kernel function outperformed other kernels in deep feature classification. In terms of accuracy scores, the suggested approach (ResNet 50 Features + SVM) performed better than a newly released method by Togacar et al. [7]. Transfer learning offers a promising solution for medical diagnosis, transferring knowledge from generic object recognition tasks to domain-specific tasks. Abbas et al. [8] presents a deep CNN called Decompose, Transfer, and Compose (DeTraC) for the classification of COVID-19 chest X-ray images. For this work, authors modified a deep CNN architecture called

DeTraC that uses a class decomposition method to classify COVID-19 images inside a large dataset of CXR images. DeTraC can handle irregularities in the dataset and achieves a high accuracy of 93.1% in detecting COVID-19 cases from a comprehensive image dataset collected from various hospitals worldwide. The COVID-19 outbreak has caused huge outbreaks all over the world, necessitating initial screening to control the spread. To reduce dependency on limited test kits, studies suggest using CT scans or chest X-ray s. Researchers [9], in their study used eight deep learning techniques (VGG16, InceptionResNetV2, ResNet50, DenseNet201, VGG19, MobilenetV2, NasNetMobile, and ResNet15V2) to detect COVID-19 symptoms using two datasets. Experiment results proved that NasNetMobile outperformed all models with an accuracy of 82.94% in CT scans and 93.94% in chest X-rays. The models can identify infectious regions and top features.

In [10], authors investigate the use of chest X-ray images in diagnosing COVID-19 disease. Researchers obtained 135 COVID-19 and 320 pneumonia cases and trained a pre-trained deep convolutional neural network, Resnet50, on 102 cases. The results showed an accuracy of 89.2% with a true positive rate of 0.8039 and an AUC of 0.95. An ensemble of CNN classifiers was applied to 33 unseen COVID-19 and 218 pneumonia cases, achieving an accuracy of 91.24%. The study [11], evaluates ML-COVID-19 management applications, focusing on imaging methods, survival analysis, forecasting, economic and geographical issues, monitoring methods, drug development, and hybrid apps. Python contributes 74% of the effort, with 20.4 percent of applications categorizing imaging techniques. Monitoring earns 16%, and 14% support prediction and forecasting apps. The research aims to serve as a reference for future research on ML and medical applications, as ML is recognized as a tool for developing intelligent ways to combat the epidemic. Challenges encountered include the lack of non-English publications, large gaps in dataset explanations, and a lack of availability for certain publishers' papers. The COVID-19 pandemic has significantly impacted human life, leading to the use of machine learning (ML) in medical applications like detecting and monitoring patients. Medical imaging systems like CT and X-ray provide ML platforms for combating the pandemic. A systematic literature review [11] reveals that CNNs, LSTMs, RNNs, GANs, autoencoders, and random forest are commonly used. However, challenges such as safety and flexibility are often overlooked. Keras is the most frequently used library, and medical imaging systems are used for diagnostic purposes.

The systematic review study [12] identifies artificial intelligence, machine learning methods and techniques for disease prediction, drug development, vaccines, existing models, and datasets for the COVID-19 pandemic. The most commonly used approaches for classification, prediction, and diagnosis are CNN, ResNet, SVM and Random forest. Other applications include risk assessment, workload reduction, social control, patient outcome prediction, and early warnings. The study found that the success of these methods varies widely, with 69% of studies measuring accuracy, sensitivity,

and specificity. Since it offers a solid foundation for further research and thorough information on AI's possible role in battling the pandemic, the study [12] is important for novice practitioners and researchers who want to create AI/ML models or medications for COVID-19.

The systematic literature study [13] reviewed 44 published research papers from 2013 to 2022 to understand the contributions and limitations of deep learning in pandemic control. The study aims to provide researchers with crucial research briefings to create more effective DL-based approaches for pandemic detection and prediction. When investigating the benefits of deep learning approaches in pandemic detection and prevention, issues such as feature selection, identification, optimization, and computing complexity are considered [13]. The intention is to provide recommendations for future lines of inquiry in this field.

The COVID-19 pandemic is causing significant damage to lung cells and potentially leading to death if not diagnosed early. To address this, researchers are developing advanced deep learning techniques to accurately diagnose and predict the virus. The research study [14] reviews recent research on diagnosing and predicting COVID-19 using deep learning networks and medical imaging techniques, including attentions, transformers, fusion, graphs, classification, segmentation, and forecasting techniques. It also discusses the challenges and future directions in COVID-19 diagnosis and prediction, including distribution shifts, fairness, and data privacy. In 2019, 2.5 million deaths were caused by pneumonia, with 14% of these occurring in children aged 0-5. Diagnosing pneumonia is crucial to prevent body failure. Deep learning techniques, such as convolutional neural networks, pre-trained models, and ensemble models, are preferred due to better performance and automatic feature extraction. The systematic literature review [15] evaluates these models' effectiveness in various medical domain challenges, highlighting research gaps and potential solutions for pneumonia detection tasks. The study [16] explores the evolution of deep learning methods for identifying lung diseases using chest X-ray images, highlighting the challenges of interpreting radiographs and the need for skilled interpretation. It also discusses the future path of research in detecting these diseases using these images. Mallick et al. [17], reviews 48 studies from 2020-2023 using TL-based models for COVID-19 detection and diagnosis. Challenges include model training difficulties, precision issues for lung segmentation, and the lack of un-annotated X-rays. The study [18] explores the use of deep learning architectures in lung disease diagnosis using CXR images, analyzing 129 articles and finding pre-trained networks enhance sensitivity and accuracy, while also discussing limitations and future research opportunities. Agrawal et al. [19] explores the use of deep learning in chest radiography for lung segmentation and detection using publicly available datasets, including Generative Adversarial Network models, to address medical data scarcity. Airway disease is a major healthcare issue causing 3 million fatalities annually and is expected to become

one of the leading causes of death globally by 2030. Advances in artificial intelligence algorithms are being used to detect airway disorders, but challenges remain. The systematic literature review [20] examines 150 articles on various airway diseases and their predictive capabilities. The study [21] explores the use of machine learning methods for automatic diagnosis of COVID-19 using X-ray images. Two commonly used classifiers, logistic regression and convolutional neural networks, were used. A dimensionality reduction approach was also explored. To increase training samples, generative adversarial network (GAN) was employed. The study found that both CNN and LR models showed high accuracy, with LR and CNN showing 95.2% and 97.6% with PCA for positive cases identification.

Pneumonia is the leading cause of mortality globally, causing 3.2 million deaths in 2015. Between 2000 and 2015, child pneumonia hospital admissions increased 2.9 times, with a more rapid increase in South East Asia. Early diagnosis and treatment are crucial for preventing lung damage and functional deficiencies. The scoping review [26] explores diagnostic techniques for community-acquired pneumonia, a lethal infectious disease. The review categorized techniques into lab-based methods, imaging-based techniques, acoustic-based techniques, and physiological-measurement-based techniques. The review found that imaging-based techniques are the most common, but there's a need for safer, non-invasive, and faster methods. The COVID-19 pandemic has led to the development of a new diagnostic method using artificial intelligence (AI). The study [27] proposes a deep learning AI-based system for automatic multiclass detection and classification of pneumonia from chest X-ray images. The system uses seven pre-trained convolutional neural networks, with the best results achieved using DenseNet201, VGG16, and VGG16. For the five-class model, the system outperforms existing methods by 1.2%. The advantage of the system is that it quickly detects COVID-19, providing results in seconds. It's cost-effective, requiring only a patient's chest X-ray scan. They have conducted six classification experiments with consistent accuracy, demonstrating its robustness for practical applications. The system's ability to detect infection severity is limited by collimator noise. We can use denoising methods for pre-processing and predict severity to aid in treatment selection and patient recovery. The system has not undergone k-fold cross-validation because of a large database of CXR images. Further advancements include heatmap images, extreme learning machines, pruning methods, and advanced image analysis solutions like stochastic imaging.

Deep learning algorithms for diagnosing pneumonia from chest X-rays require large, sparse training datasets, which can overfit. The work [28] proposes a domain adaptation and classification technique using a private-small dataset and a public-large labeled dataset. This involves data selection, image translation, and convolutional neural network exploration. Fine-tuning specific layers with selected-adapted images improves sorting accuracy and reduces trainable parameters, achieving a classification accuracy of up to

97.78%. The CX-DaGAN algorithm improves the classification accuracy of a small X-ray dataset by using information from a large public dataset. The algorithm selects images based on intra-class similarity and interclass dissimilarity and generates new images through a GAN network. The algorithm is a complete domain adaptation workflow consisting of three stages: selecting images, training the GAN-based image-to-image translation network, and fine-tuning the pretrained CNN classification network. The results show an overall accuracy of 88.36% without the domain adaptation workflow.

Chest radiography is a crucial diagnostic tool for detecting chest diseases [29]. Advancements in deep learning techniques have led to the development of automated systems for detecting pneumonia from chest X-rays. However, a comprehensive literature review is lacking, highlighting the need for more efficient methods. This study aims to help medical practitioners select the most effective methods, analyze available datasets, and understand the results. It also discusses the usability, goodness factors, and computational complexities of the algorithms used for intelligent pneumonia identification. The study reveals that most applied datasets are unbalanced and limited, making them unsuitable for large-scale use. Deep learning-based algorithms have been found to achieve the best results for pneumonia classification, with an accuracy of 98.7%, a sensitivity of 0.99, and a specificity of 0.98. The COVID-19 pandemic has prompted the development of advanced diagnostic and monitoring devices, including automated detection using chest X-ray (CXR) images. However, diagnosing pneumonia is challenging due to imaging similarities. A new classification model, PDMLP-Bi-LSTM [30], uses multi-source generated data to discriminate Normal, COVID-19, and Other pneumonia cases. The model uses parallel deformable multi-layer perceptron and Bi-directional Long Short-Term Memory modules to extract abstract features and analyze correlations. Extensive simulations on 4099 CXR images validated the method's performance, showing excellent accuracy, specificity, precision, recall, and F1-score of approximately 98% or above. The ultimate goal is to integrate this functionality into portable X-ray equipment and computer platforms. Generally, researchers focused on utilizing various approaches and classifiers to construct a more accurate model without considering the implementation of security measures to protect patient medical data. Consequently, our study aimed to fill this research gap by converting the input chest X-rays into spectrograms. This conversion transforms the images into a 2D heatmap-like representation, concealing the chest X-ray details and rendering them incomprehensible to an individual. Spectrograms are visual representations of the frequency components of an image. We ensure patient data privacy and improve model prediction accuracy by converting the chest X-ray images into spectrograms in our proposed work.

III. PROPOSED METHOD FOR SPECTROGRAM IMAGE GENERATION

This section outlines the suggested process for creating spectrogram images from X-ray images of the chest. These spectrogram images which are obtained from chest X-ray images are then used to extract new features which are better representatives of respective images by applying the proposed feature extraction method. The idea is to use the input images represented in terms of new image features for classification and prediction. For extracting features from chest xray images, in the present work, we apply the proposed feature extraction method presented at Section IV.

A. RESIZING CHEST XRAY IMAGES

Image preprocessing is a crucial step when handling of image data for classification, particularly in case of medical image data like chest X-rays. This is because chest X-rays are complex images that come in different sizes and formats and contain potential anomalies. To address the problem of varied image sizes, we first perform resizing of chest X-ray images to bring all the chest X-ray images into common size. We have performed experiments comparing model performance on 32x32 images versus larger sizes (like 64x64 or 128x128). We observed that the performance is acceptable and meets our application's requirements that supports our choice. Also, smaller images require less computational power and memory, making training faster and allowing for larger batch sizes. This is especially important as we are working with limited computational resources.

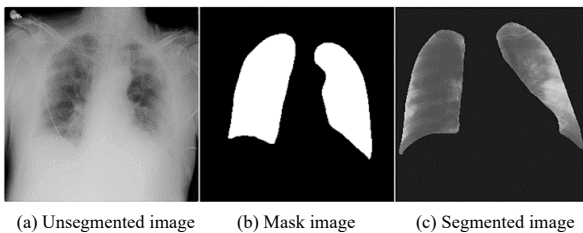


FIGURE 2. Figure (a-c) demonstrating unsegmented image, respective mask image and segmented chest x-ray image

B. LUNG SEGMENTATION

During the lung segmentation phase, we apply the bitwise AND function between each chest X-ray image and its corresponding mask images from the dataset. This operation compares each pixel in the chest X-ray image to its corresponding pixel in the mask image. Mask images are generally templates for highlighting the lung area. If both the original image and the mask image had a non-zero value, then the resulting pixel of the segmented image retained that value. If not, we set the pixel to zero, thereby eliminating it from the analysis. This process helps us obtain the lung part that is believed to be valuable for further analysis. Fig. 2 displays the process of segmenting the required part of the lung from the chest X-ray. We apply the bitwise AND operation between the raw chest X-ray image (Fig. 2a) and its corresponding mask image (Fig. 2.b) to obtain the segmented lung, as illustrated in Fig. 2c.

Mathematical Modeling for Lung segmentation

This mathematical modeling outlines the process of lung segmentation using an image and a mask. Here, we give a general method of how to create the mask based on a threshold [38], apply the bitwise AND operation to obtain the segmented image. Let I be the original chest X-ray image. M be the mask image.

Original Image, I

Let I be the original image represented as a matrix $R^{H \times W}$, where each element of $I(x, y)$ represents the pixel intensity at the coordinate (x, y) . In our case as the chest x-rays are grey scale images and hence the pixel values may range from 0 to 255.

Mask Image, M

Let M be the Mask image represented as a matrix $R^{H \times W}$, where $M(x, y) = 1$, if the pixel belongs to the lung region and $M(x, y) = 0$, if the pixel does not belong to the lung region.

Creating the Mask

To create a binary mask from the original image, we may apply a threshold ' T ' [38] to distinguish between lung and non-lung areas. This can be represented using Eq. (1) [38],

$$M(x) = \begin{cases} 1, & I(x, y) > T \\ 0, & \text{else} \end{cases} \quad (1)$$

Where T is a predefined threshold value between 0 and 255.

Bitwise AND operation

The segmentation process uses a bitwise AND operation to isolate the lung region from the original image. The mathematical representation of this operation can be expressed using Eq. (2) [39],

$$S(x, y) = I(x, y) \cdot M(x, y) \cdot 255 \quad (2)$$

where S is the segmented image, representing lung regions, the multiplication by 255 converts the binary mask back to a format compatible with the original image's pixel values.

Resulting Segmented Image

The resulting segmented image is given by Eq. (3) [39],

$$S(x, y) = \begin{cases} I(x, y), & M(x, y) = 1 \\ 0, & M(x, y) = 0 \end{cases} \quad (3)$$

If the mask indicates a lung pixel, i.e., $M(x, y) = 1$, the pixel value in S is the same as in I . If the mask indicates a non-lung pixel, $M(x, y) = 0$, the pixel value in S is set to 0.

C. ANDREW CURVES

Andrew's curves allow for the representation of high-dimensional features extracted from chest X-rays in a two-dimensional format, simplifying the analysis and interpretation. By transforming X-ray data into curves, it becomes easier to identify patterns, anomalies, and relationships within the data, aiding in the detection of conditions like pneumonia or COVID-19. The transformed data can be fed into machine learning models, improving classification performance by providing a more compact representation of the relevant features from the original

images. Fig. 3 depicts Andrew curves generated for COVID and Normal classes. In Fig.3, the x-axis (t) represents a cyclical or angular variable ranging from $-\pi$ (negative pi) to π (pi), and the y-axis (f(t)) represents the transformed data points.

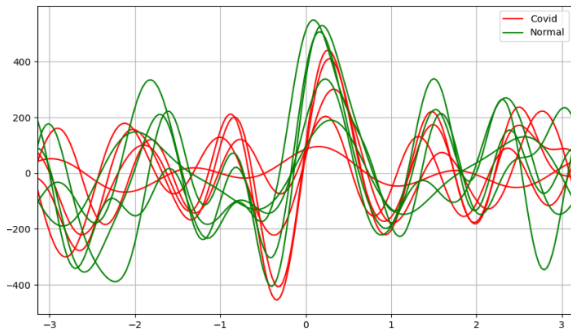


FIGURE 3. Andrew plot depicting Covid class against normal class

D. MUTUAL INFORMATION

In our work, we have used mutual information to extract higher-order features from existing data. Eq. (4), gives the mathematical expression for mutual information between two random states, X and Y. For the random states X and Y, $I(X, Y)$ represents mutual information as given by Eq. (4) [37],

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) * \log \frac{P(x, y)}{P(x) * P(y)} \quad (4)$$

where $p(x, y)$ represents the joint probability of these variables X and Y, which gives the probability of the event, which is defined as $X = x$, and $Y = y$. and $p(x)$ and $p(y)$ give the marginal probability mass (also called density) functions of the variables X and Y, respectively. A non-negative expression, $I(X, Y) \geq 0$, provides mutual information, and its value drops to zero when X and Y are independent. When we get higher values for mutual information, it shows a stronger relationship or dependence between the variables. In practical implementation, mutual information measures information gain and understands the relationships between variables in different applications, which include machine learning and data analysis.

We store the image pixel values in the CSV format, and then we apply the Andrew curve function defined by $T(n)$. We use $T(n)$, which ranges from 0 to 360 degrees with a 0.1 step increment, to upscale an image from any size to 60x60, resulting in 3600 features. We then store these computed values in a variable named 'andrew_csv' using Eq. (5) [36],

$$T(n) = \frac{x_1}{\sqrt{2}} + x_2 * \sin(n) + x_3 * \cos(n) + x_4 * \sin(2n) + x_5 * \cos(2n) + \dots \quad (5)$$

where x_i represents the i^{th} pixel value.

E. PERFORMING NORMALIZATION

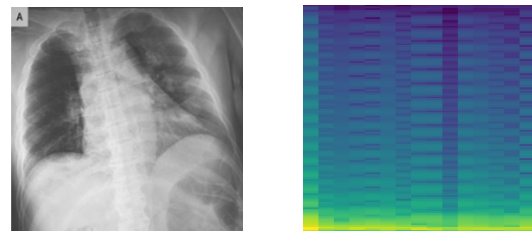
These Andrew_Csv Values may contain negative values as well, So, we need to normalize them to remove these negative values, as pixel values should not be negative. We ensure that

all these values are within the range of 0 And 1. For this, we apply Min-Max scaling. The formula for Min-Max scaling is given by Eq. (6) where X is the original feature value, X_{min} is the minimum value of the feature, X_{max} is the maximum value of the feature, and X_{scaled} is the scaled feature value which lies between 0 and 1.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

F. SPECTROGRAM GENERATION

A spectrogram is a 2D representation that displays the intensity of different frequencies in a signal over time. Signal processing and audio analysis commonly use them. We converted chest X-ray images into spectrogram images because abnormalities in lung sounds, such as wheezing, may produce different frequency patterns. The respiratory system produces sounds at various frequencies during normal and abnormal conditions. Fig. 4(a) represents chest X-ray image and Fig. 4(b) represents the spectrogram image generated from chest X-ray image in 4(a).



(a) Chest Xray

(b) Spectrogram

FIGURE 4. Chest Xray and Spectrogram image generated from Chest XRay

In this work, we have used the `scipy.spectrogram()` function in python to generate spectrograms.

`scipy.signal.spectrogram(x, fs=1.0, window='tukey', 0.25), nperseg=None, noverlap=None, nfft=None, detrend='constant', return_onesided=True, scaling='density', axis=-1, mode='psd')`

where x = input array of time-domain signals, window = Window function that is applied on each segment, nperseg= Number of samples per segment, nooverlap = Number of overlapping samples between segments.

IV. ALGORITHM FOR IMAGE FEATURE EXTRACTION AND DISEASE DETECTION

The algorithm outlines a comprehensive approach for processing chest X-ray images, extracting relevant features, and preparing data for machine learning. It combines several key tasks—image resizing, data normalization, feature extraction and classification into a streamlined workflow suitable for diagnostic applications.

Algorithm Feature_extraction (image, size)

```
{
image := GrayScaleMatrix(image)
```

```

resized_image := Resize(image, (size, size))
r := FindRank(resized_image)
d := FindDet(resized_image)
EV [ ] := FindDet(resized_image)
flatten_image:= Flatten (resized_image)
features := Append(flatten_image, r, d, EV[])
x, y:= Split(features)
return x, y {x, y represent features and output label}
}

```

Algorithm SpectrogramImageGeneration (image_path, outputcsv_path, size, step_size , range)

```

{
function_values = []
image := GrayScaleMatrix(image_path)
resized_image:= Resize (image, (size, size))
nr:= (resized_image – mean (resized_image)
standardize_image:= nr/ standardDeviation(resized_image)
flatten_image = Flatten(standardized_image)
num_pixels = len(flatten_image)
t_values = []
t:=0
i:=0
while (t<=range) do
{
t_values[i] = t
t:= t + step_size
i:= i+1
}
t_size := length(t_values)
for i:=0 to t_size do
{
f_t:=0
for j:=0 to num_pixels do
{
if (t_values[i] == 0) then f_t:= f_t + flatten_image [j]
else
{
factor:= sin (j π t_values[i]) / sqrt (pow (2, j))
value:= flatten_image[j] x factor
f_t:= f_t+value
}
}
append f_t to function_values
}
}
append label to function_values
store function_values to csv
function_values := Read function_values from csv
label:= function_values [: -1]
spectrogram:= Reshape (function_values, size)
normalized_spectrogram:= Normalize (spectrogram)

```

```

spectrogram_unit8:= unit8(normalized_spectrogram)
color_spectrogram:= ColorMap(spectrogram_unit8)
store color_spectrogram as image in output_path
}

```

Algorithm Classifiers(train_data_set, test_data_set)

```

{
for i:=0 to size(train_data_set) do
{
x_train, y_train := Feature_extraction (train_data_set[i])
}
for i:=0 to size(test_data_set) do
{
x_test, y_test := Feature_extraction (test_data_set[i])
}
Normalization(x_train, y_train, x_test, y_test)
Classifier(x_train, y_train, x_test, y_test)
}

```

The proposed image feature extraction method to extract features from chest x-ray images is based on the linear transformation representation. The image feature extraction process starts by converting all chest x-ray images in the input dataset to equivalent greyscale chest x-ray images if they are already not in the greyscale form. Then, these greyscale chest x-ray images are resized to a predefined image size, say $n \times n$ for consistency among all medical images in the dataset. This step is carried to make sure that all images in the dataset are of one standard size so that the features extracted from these images shall have same feature dimensionality in the derived feature vector. We then consider each grey scale image from the dataset and represent (or view) every image as a square matrix M of size $n \times n$. Each element of the matrix, $M(i,j)$ thus represents the intensity of the pixel (i,j) in the image, I . For each image matrix, we obtain the structural properties of the matrix. They are (i) rank of matrix denoted as $\text{rank}(M)$, (ii) determinant of matrix denoted as $\det(M)$, (iii) trace of matrix which can be computed as $\text{trace}(M) = \sum_{i=1}^n M_{ii}$ and (iv) eigen values for M , $\text{eig}(M) \rightarrow [\lambda_1, \lambda_2, \dots, \lambda_n]$. Then, we flatten the image, i.e. Convert the image matrix I into a vector V by flattening the matrix ($V=\text{flatten}(I)$) and then Combine original image pixels and structural features obtained. Thus, we create a new feature vector F that adds the derived structural features to the original pixel vector V , the feature vector can be represented as $F = V + [\text{rank}(M), \det(M), \text{trace}(M), \text{eig}(M)]$. When forming feature vector, we ensure that the dimensions match when combining, possibly repeating the structural features to fit the pixel vector length. The final augmented feature vector F can be used for classification tasks, input into machine learning models, or further analysis.

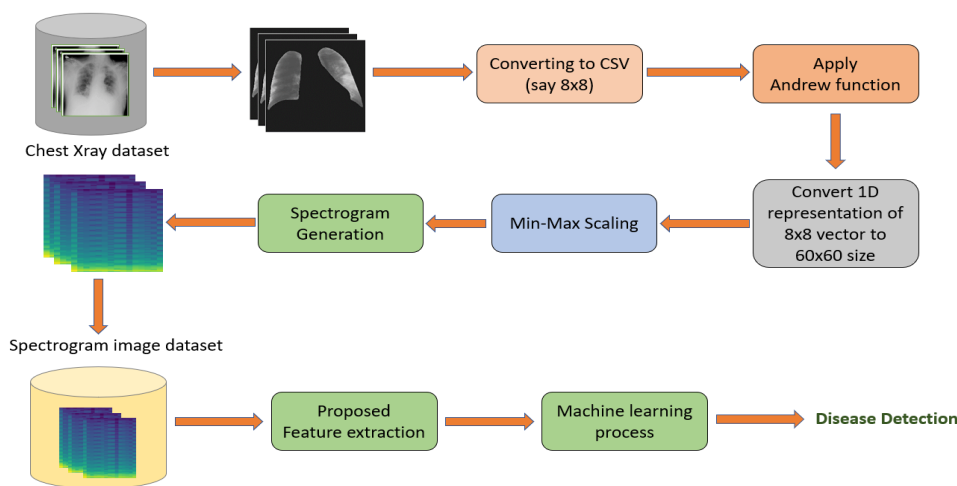


Figure 5. Proposed System for disease diagnosis using chest radiography images

V. PROPOSED MACHINE LEARNING SYSTEM FOR DISEASE DIAGNOSIS

FIGURE 5 represents the proposed system architecture for disease diagnosis using chest radiography images. The first step (is an optional step) which involves the segmentation of the lungs from these chest X-rays to obtain only the necessary lung section having important details about the abnormalities. The dataset we considered includes the corresponding mask images for each chest X-ray. To obtain the segmented lung image, we apply the bitwise AND operation between the two mask images and the raw chest X-rays.

Secondly, we resize the images to a common size, for example 32x32 image size using the `cv2.resize()` function. Further, we convert these resized images into CSV files for easier access and structured organization of the pixel values. Next, we create Andrew curves by mapping each observation onto a function, a common method for visualizing multivariate data. We use Andrew curves to identify similarities or differences between classes by analyzing the shape of the obtained curves. In our case, the curves overlap for multiple classes, indicating the presence of similar characteristics that could potentially lead to inaccurate predictions. We then apply the Andrew function, which allows for image upscaling or downscaling. After applying the Andrew function, we store the output values as `andrew_csv`.

We perform min-max normalization on this CSV file to eliminate any negative values, then scale the values from 0 to 1, storing them as `normalized_csv`. The spectrogram function receives this `normalized_csv` for each image as input to produce the corresponding spectrogram. Spectrograms are a visual representation of the frequency content of an image. The intensity, or color, at each point in the spectrogram represents the power of the frequency component in the image. Internally, the spectrogram function applies first windowing, which divides the large signal (in a spectrogram, the array values are considered signals) into smaller signals for easier calculations.

After forming the windows, the first step involves applying the Fast Fourier Transform function to compute the discrete Fourier transform value of each window's component. The next step uses the power spectral density to determine the distribution of a signal's strength across different frequencies. These values later form the spectrogram matrix, which in turn generates the required spectrogram. From the spectrogram images, we extract the matrix properties such as rank, trace, determinant, and Eigen values and store them as a vector, which will help to find the similarity degree between any two images. We compare the accuracies obtained by ML models by considering features with and without incorporating the above-mentioned matrix properties, and we aim to find the optimal configuration that gives us the best accuracy.

The entire process for disease diagnosis using proposed method is depicted in the FIGURE 5. Initially, we segment the lungs from the X-ray images if this step desired. Otherwise, we can use the chest xray images directly. Both the segmented and unsegmented images then undergo resizing to a standard size (say, 32x32, 64x64, 128x128, 224x224 etc.). We convert resized images into the CSV file format.

Subsequently, we explore two different methods to observe which one yields better accuracy and predictions. The first approach involves selecting higher-order features and utilizing mutual information gain to measure each pixel's value for classification. This helps us choose the most important features for model prediction. The second approach selects same pixel ordering and applies Andrew function for images from various classes. In our case, we have overlapping Andrew curves, which creates a problem for model prediction. Next, we input the entire image (segmented or unsegmented chest X-ray images) to the Andrew function and normalize the resulting values using min-max scaling. The spectrogram function then uses these normalized values as input, generating individual spectrogram images for each chest X-ray image.

Spectrograms provide a comprehensive view of the underlying patterns within the chest X-rays. Furthermore, we

feed these spectrogram images generated using chest X-ray images to various machine learning models. For feature extraction, in this work, we have applied the proposed feature extraction method outlined in section IV. We study the performance of various ML classifiers by considering two cases (i) by applying proposed feature extraction on chest x-ray images (ii) by applying proposed feature extraction on spectrogram images generated from chest xray images. We run the machine learning models on the same data distribution for all classifiers and obtain the classifier metrics of ML classifiers to compare their relative performance on the test dataset.

VI. EXPERIMENTAL RESULTS

This section outlines the data preparation which is essentially important in any machine learning task. Subsection A describes the dataset used in the study and the data preparation for experimental study. Subsection B presents a comprehensive overview of the sample size justification, Subsection C mentions the experimental setup details and Subsection D outlines the classification performance results for Binary and Multiclass classification.

A. DATA PREPARATION

Data preparation is a crucial step in any classification task. For experimental analysis in this paper, we have used the publicly available benchmark dataset, i.e. COVID-19 Radiography dataset [24]. We choose this dataset as this is the most widely used dataset in many state-of-the-art research studies.

1. COVID-19 Radiography Dataset Description and Division of the Dataset

COVID-19 Radiography dataset is the Winner of the COVID-19 Dataset award by Kaggle Community, is publicly available at Kaggle. This dataset and can be accessed at <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>. Originally, the dataset comprised of 21,165 chest X-ray images which are distributed into four different classes: (i) COVID (3616 images), (ii) lung opacity (6012 images), (iii) viral pneumonia (1345 images), viral pneumonia (1345 images), and normal (10192 images). We discarded lung opacity for multiclass classification as this study focused on detection of COVID-19 and Pneumonia classes. For experimental analysis, the dataset is divided into train and test datasets.

TABLE 1

Statistics of number of CXR images used from COVID-19 radiography dataset to obtain training and testing datasets for multiclass classification

Class	Training dataset	Testing dataset	Total
Normal	9172	1020	10192
Pneumonia	1210	135	1345
COVID-19	3254	362	3616
Total	13636	1517	15153

Thus, we form a three-class dataset using Normal, Covid and Pneumonia Chestxray images. For experimental analysis, we split the dataset into train dataset and test dataset. The train and test dataset details are depicted using the Table. I below which

are obtained by performing 90%-10% split using original dataset.

B. SAMPLE SIZE JUSTIFICATION

To ensure that the training dataset sample size is adequately powered to detect significant differences in classification accuracy among the three groups, thereby providing a strong statistical basis for the research conclusions, we have considered GPower statistical software tool. GPower is a widely used statistical software tool designed to perform power analysis for a variety of statistical tests. Power analysis helps researchers determine the minimum sample size required to detect an effect of a given size with a specified degree of confidence.

To determine the appropriate sample size for the multiclass classification task with the given dataset of Normal, COVID, and Pneumonia images, we perform a power analysis using the dataset available to assess if it's sufficient based on statistical power considerations.

Originally, the dataset has chest x-ray images distributed among three classes as Normal - 10192 images, COVID - 3616 images and Pneumonia - 1345 images. Since, we are looking to assess if this dataset is statistically sufficient, we consider the following settings for power analysis: The effect size is set as $f=0.25$ (medium effect size), the value of Alpha (α) is set at 0.05 (commonly used significance level), the power value denoted by $(1 - \beta)$ is set for 0.80 (80% chance of detecting a true effect), number of groups is set to 3 since we have three classes (COVID, normal, pneumonia) in the dataset.

Using G*Power analysis, we obtained total sample size as 159 and sample size per group as 53. This means that each class in the training dataset should contain at least 53 images. We performed training dataset analysis by comparing the number of images in three classes with the sample size given by power analysis. As per power analysis test, we need a sample size consisting at least 53 images representing each class in the dataset. In our case, we have 9172 Normal chest x-ray images, 1210 pneumonia chest x-ray images and 3254 covid chest x-ray images. Hence, as per power analysis test, the number of images w.r.t each class in the training dataset are sufficient. Similarly, in the testing dataset we have we have 1020 Normal chest x-ray images, 135 pneumonia chest x-ray images and 362 covid chest x-ray images. As per power analysis test, we need a sample size consisting at least 53 images representing each class in the testing dataset. Hence, as per power analysis test, the number of images w.r.t each class in the testing dataset are sufficient. Thus, the calculated sample size ensures that our study is adequately powered to detect significant differences in classification accuracy among the three groups, thereby providing a strong statistical basis for the research conclusions.

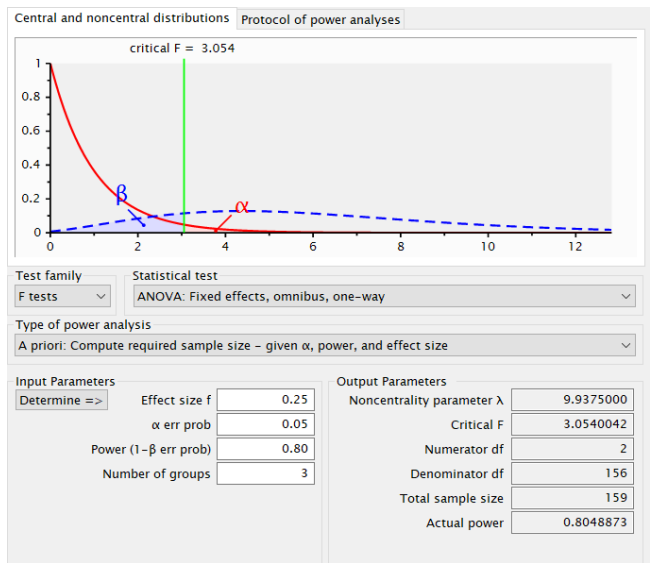


FIGURE 6. Input and Output parameters of power analysis test using to determine appropriate sample size for multi-class classification

FIGURE 6 depicts the screenshot representing input parameters and output parameters of power analysis carried to determine appropriate sample size for multiclass classification.

C. EXPERIMENTAL SETUP

To implement the proposed feature extraction method and perform machine learning, we have used Python programming language, Scikit-learn (sklearn) and OpenCV which are two widely used libraries in the field of machine learning and computer vision. All experiments are run on the system having following specifications. Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.70 GHz with 16.0 GB RAM.

D. CLASSIFICATION PERFORMANCE ANALYSIS

We assessed the proposed system for binary classification and multiclass classification. Two scenarios are considered for experimental study. In the first scenario, the chest x-ray images are converted into respective spectrogram images. These spectrogram images are then used to study the performance of classifiers when the proposed feature extraction method is applied. In the second case, we used the chest x-rays images without converting them to spectrogram images. For experimental study and analysis, we have split the original dataset into two subsets (i) training dataset and (ii) testing. For all experiments, the split ratio is 90%-10% w.r.t train and test datasets respectively. The images in the test dataset are unseen by the training model. We trained the machine learning classifiers on the training dataset and then tested them on the test dataset. Before feeding the classifier with data, we apply the proposed feature extraction method.

SCENARIO-1: BINARY CLASSIFICATION WITH PROPOSED FEATURE EXTRACTION USING SPECTROGRAM IMAGES GENERATED FROM CHEST XRAY IMAGES

In the first scenario, we consider the chest X-ray images in the dataset and convert them to their respective spectrogram

images. The spectrogram images are obtained by using the procedure outlined in Section III. We use these spectrogram images for feature extraction and machine learning. We evaluate the performance of state-of-the-art ML classifiers by applying proposed feature extraction and considering spectrogram images as input. We consider two use cases. The first use case involves binary classification, which uses spectrogram images to distinguish between COVID and Pneumonia. The second use case involves binary classification, which uses spectrogram images to distinguish between COVID and Normal.

Use Case 1: Binary Classification Using COVID and Pneumonia Spectrogram Images

We perform the first experiment to evaluate the detection performance of state-of-the-art ML classifiers using the proposed feature extraction method. We conduct this experiment to investigate how well ML classifiers distinguish between PNEUMONIA and COVID disease classes, using spectrogram images from chest X-ray images as input for machine learning classifiers. We obtained the training and testing datasets by carrying out a 90%-10% split from the original dataset. Table 2 mentions the class distribution of the training and testing datasets.

TABLE 2

Statistics of number of CXR images used from covid-19 radiography dataset to obtain training and testing datasets for binary classification of covid and pneumonia

Class	Training dataset	Testing dataset	Total
PNEUMONIA	1210	135	1345
COVID-19	3254	362	3616
Total	4464	497	4961

The dataset consisted of 1345 pneumonia and 3616 COVID chest X-ray images. For the experimental study, we have used the training dataset, which included 1210 pneumonia and 3254 COVID chest X-ray images, and the testing dataset, which comprised 135 pneumonia and 362 COVID chest X-ray images. The images from both the train and test datasets undergo feature extraction. We extract 35 new features for each image using the proposed feature extraction method. We augment these 35 extracted features with the original image pixel values to generate a new feature vector for each image. This process applies to every image in the training dataset, generating a new training dataset that expresses each image as a feature vector with 1059 features. We train state-of-the-art machine learning classifiers on the train dataset to build machine learning models, which we store as pickle files to prevent repeated training. We evaluate the performance of machine learning classifiers using the test dataset, which contains test images that the training model hasn't seen before. The machine learning classifiers considered for the evaluation are (i) logistic regression, (ii) decision trees, (iii) Naive Bayes, (iv) support vector machines (SVM), and (v). Linear Discriminant Analysis (vi) Multilayer Perceptron (MLP); (vii) Quadratic Discriminant Analysis; (viii) XGBoost.

Table 3 depicts ML classifiers performance using the proposed feature extraction method with three evaluation metrics: (i) accuracy, (ii) precision, and (iii) recall. We

observed that for the logistic regression model, the detection rate for the COVID class is 98.34%, and the overall accuracy is 97.18% . In this case, the detection rate for the Normal class is 94.07%, resulting in a balanced accuracy of 96.21%.

TABLE 3
Performance metrics of machine learning classifiers on the test dataset consisting of covid and pneumonia spectrogram images using the proposed feature extraction

S.No	ML Models	Accuracy (%)	Precision (%)	Recall (%)
1	Logistic Regression	97.18	97.8	98.34
2	Decision Tree	91.15	94.41	93.37
3	Naive Bayes	72.84	72.83	100
4	Support Vector Machine	96.58	96.49	98.89
5	Linear Discriminant Analysis	93.96	92.78	99.44
6	Multilayer perceptron	96.78	97.01	98.61
7	Quadratic Discriminant Analysis	72.84	72.83	100
8	XG Boost	96.78	96.51	99.17

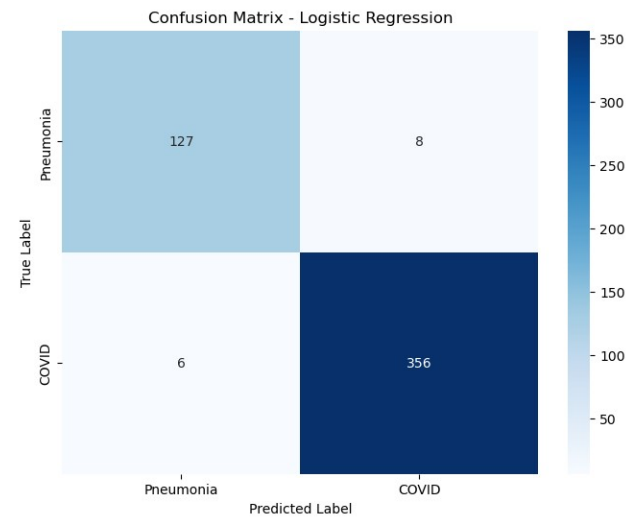


FIGURE 7. Confusion matrix representing binary classification results of logistic regression model for covid and pneumonia classes

FIGURE 7 depicts the confusion matrix obtained for the logistic regression model for the test dataset consisting of covid and pneumonia classes. **FIGURE 8** presents the obtained ROC curve for the logistic regression model when test dataset is used for prediction. We obtained an AUC area of 0.99 for the covid and pneumonia classes, respectively for logistic regression model. This shows that the logistic regression model is performing equally well with both covid and pneumonia classes. Following the logistic regression model, we observed that MLP classifier has achieved a disease detection rate of 98.61% for the covid class, with a balanced accuracy of 91.85%.

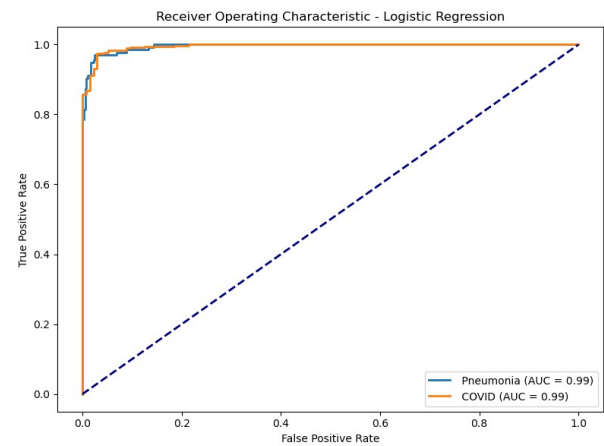


FIGURE 8. ROC plot representing binary classification results of logistic regression model for covid and pneumonia classes

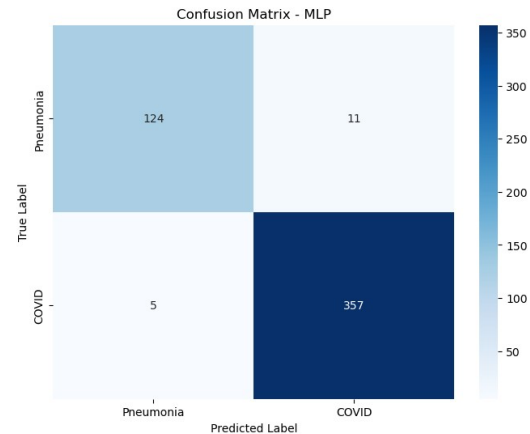


FIGURE 9. Confusion matrix representing binary classification results of MLP model for covid and pneumonia classes

FIGURE 9 depicts the confusion matrix for MLP classifier model for the test dataset. The MLP classifier has achieved a balanced accuracy of 95.23%. In terms of accuracy, although XGBoost has attained 96.78% accuracy, which is the same as MLP, its balanced accuracy is 94.77%, which is less than MLP. Overall, the logistic regression model outperforms all other classifier models, achieving 97.18% accuracy, 97.8% precision, 98.34% sensitivity, and 94.07% specificity.

Use Case 2: Binary Classification Using Normal and Pneumonia Spectrogram Images

We performed the second experiment to evaluate the detection performance of state-of-the-art ML classifiers using the proposed feature extraction method to investigate how well ML classifiers distinguish between pneumonia and normal classes, using spectrogram images from chest X-ray images as input for machine learning classifiers. We obtained the training and testing datasets by carrying out a 90%-10% split from the original dataset. Table 4 mentions the class distribution of the training and testing datasets.

TABLE 4

Statistics of number of CXR images used from COVID-19 radiography dataset to obtain training and testing datasets for binary classification of normal and pneumonia

Class	Training dataset	Testing dataset	Total
Normal	9172	1020	10192
Pneumonia	1210	135	1345
Total	10382	1155	11537

The dataset consisted of 1345 Pneumonia and 10192 Normal chest X-ray images. For the experimental study, we have used the training dataset, which included 1210 Pneumonia and 9172 Normal chest X-ray images, and the testing dataset, which comprised 135 Pneumonia and 1020 Normal chest X-ray images. The images from both the train and test datasets undergo feature extraction. We extract 35 new features for each image using the proposed feature extraction method. We augment these 35 extracted features with the original image pixel values to generate a new feature vector for each image. This process applies to every image in the training dataset, generating a new training dataset that expresses each image as a feature vector with 1059 features. We evaluated the performance of machine learning classifiers using the test dataset, which contains test images that the training model hasn't seen before. The machine learning classifiers considered for the evaluation are (i) logistic regression, (ii) decision trees, (iii) Naive Bayes, (iv) support vector machines (SVM), and (v). Linear Discriminant Analysis (vi) Multilayer Perceptron (MLP); (vii) Quadratic Discriminant Analysis; (viii) XGBoost.

Table 5 depicts ML classifiers performance using the proposed feature extraction method with three evaluation metrics: (i) accuracy, (ii) precision, and (iii) recall.

TABLE 5

Performance metrics of machine learning classifiers on the test dataset consisting of normal and pneumonia spectrogram images using the proposed feature extraction

S.No	ML Models	Accuracy (%)	Precision (%)	Recall (%)
1	Logistic Regression	95.50	82.17	78.52
2	Decision Tree	91.77	63.89	68.15
3	Naive Bayes	85.80	44.40	85.19
4	Support Vector Machine	96.54	85.71	84.44
5	Linear Discriminant Analysis	92.47	62.12	91.11
6	Multilayer perceptron	96.62	87.50	82.96
7	Quadratic Discriminant Analysis	88.31	NAN	0.00
8	XG Boost	96.02	83.46	82.22

The experimental results show that the MLP classifier model has achieved 96.62% accuracy, followed by SVM with 96.54% accuracy. When we consider the accuracy metric, MLP achieves the highest accuracy. The accuracy metric does not consider the imbalance effect. Therefore, we have also analyzed the balanced accuracy of classifiers to determine which classifier outperforms the others. We observe that the balanced accuracy of SVM, MLP, and logistic regression models is 91.29%, 90.69%, and 88.13%, respectively. Among all classifier models, SVM proved to be the best classifier,

which can effectively discriminate between pneumonia and normal classes in terms of balanced accuracy.

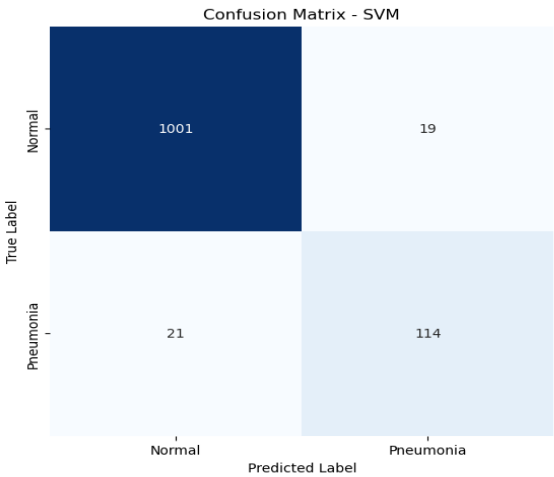


Figure 10. Confusion matrix representing binary classification results of SVM model for normal and pneumonia classes on the test dataset

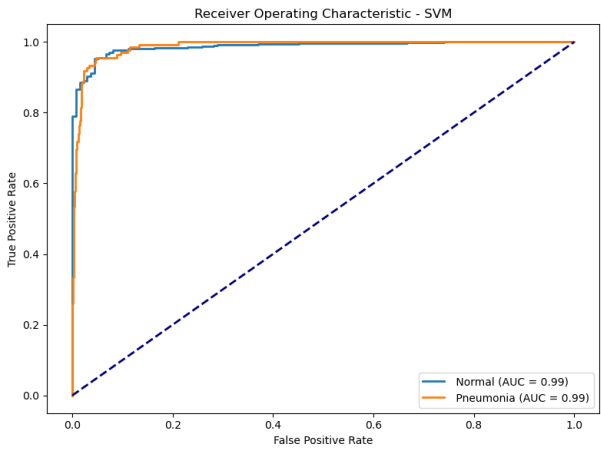


Figure 11. ROC plot representing binary classification results of SVM model for normal and pneumonia classes on the test dataset

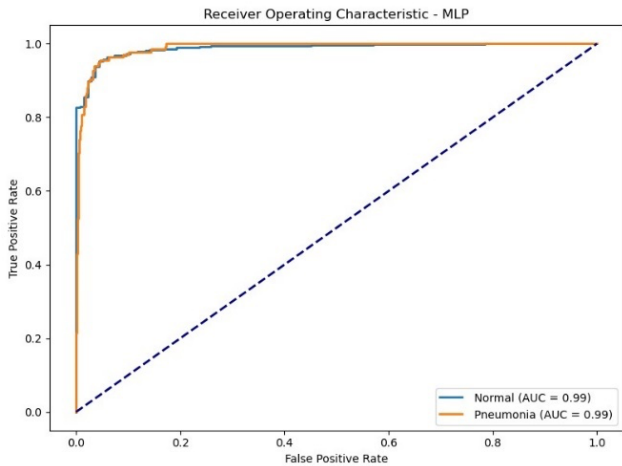


Figure 12. ROC plot representing binary classification results of MLP model for normal and pneumonia classes on the test dataset

FIGURE 10 displays the confusion matrix the SVM classifier produced when performing disease prediction using the test dataset containing the Normal and Pneumonia classes. Figure 11 depicts the ROC curve plot of the SVM classifier model obtained for the test dataset. The ROC curves for the SVM classifier show that the area under curve (AUC) values for the normal and pneumonia classes are 0.99 and 0.99 for the testing dataset respectively. We compared AUC values obtained for test dataset to train dataset and observed that the model is not overfitting.

Figure 12 depicts the ROC plot for the MLP classifier model for normal and pneumonia classes on the test dataset. The MLP classifier attained AUC values of 0.99 and 0.99 for the normal and pneumonia classes, respectively. We also observe that for the train dataset, the MLP classifier attained AUC values of 0.99 and 0.99 for the normal and pneumonia classes, respectively. The AUC values obtained for the train and test datasets in relation to the normal and pneumonia classes demonstrate that the model is not overfitting.

SCENARIO-2: BINARY AND MULTI-CLASS CLASSIFICATION WITH PROPOSED FEATURE EXTRACTION USING CHEST XRAY IMAGES FROM COVID-19 CHEST RADIOGRAPHY DATASET

In the second scenario, we considered the chest X-ray images present in the COVID-19 radiography dataset. We evaluate the performance of state-of-the-art ML classifiers by applying the proposed feature extraction. We consider two use cases. The first use case involves binary classification, which uses chest X-rays to distinguish between COVID and Normal classes. The second use case involves multiclass classification to distinguish between Normal, COVID, and Pneumonia classes.

The machine learning classifiers that are considered for performance evaluation are (i) decision tree, (ii) naive bayes, (iv) support vector machine (SVM), (v) linear discriminant analysis, and (vi) multilayer perceptron (MLP). (vii) Quadratic discriminant analysis; (viii) XGBoost. We present the results of binary classification and multi-class classification for the COVID-19 chest radiography dataset in the below mentioned use cases.

USE CASE 1: BINARY CLASSIFICATION USING NORMAL AND COVID CHEST X-RAY (CXr) IMAGES

For this experiment, we have separated Normal and COVID CXr images from the original dataset and formed the training and testing datasets. The training dataset used for experimental study consisted of 9172 normal and 3254 COVID CXr images, while the testing dataset consisted of 1020 normal and 362 COVID chest X-ray images, as shown in Table 6.

TABLE 6

Statistics of number of CXr images used from COVID-19 radiography dataset to obtain training and testing datasets for binary classification of normal and covid CXr images

Class	Training dataset	Testing dataset	Total
NORMAL	9172	1020	10192
COVID	3254	362	3616
Total	12426	1382	13808

We have carried power analysis for deciding the sample size of the dataset using G*power tool. Since, we are looking to assess if this dataset is statistically sufficient, we consider the following settings for power analysis: The effect size is set as $f=0.25$ (medium effect size), the value of Alpha (α) is set at 0.05 (commonly used significance level), the power value denoted by Power ($1 - \beta$) is set for 0.80 (80% chance of detecting a true effect), number of groups is set to 3 since we have two classes (COVID and Normal) in the dataset.

Using G*Power analysis, we obtained total sample size as 128 and sample size per group as 64. This means that each class in the training and testing datasets should contain at least 64 images. We performed training dataset analysis by comparing the number of images in three classes with the sample size given by power analysis. As per power analysis test, we need a sample size consisting at least 64 images representing each class in the dataset. In our case, we have 9172 Normal chest x-ray images and 3254 covid chest x-ray images. Hence, as per power analysis test, the number of images w.r.t each class in the training dataset are sufficient. Similarly, in the testing dataset we have we have 1020 Normal chest x-ray images and 362 covid chest x-ray images. Hence, as per power analysis test, the number of images w.r.t each class in the testing dataset are sufficient. Thus, the calculated sample size ensures that our study is adequately powered to detect significant differences in classification accuracy among the three groups, thereby providing a strong statistical basis for the research conclusions.

Table 7 depicts machine learning classifier model's performance using the proposed feature extraction method with three evaluation metrics: (i) Accuracy, (ii) Precision, and (iii) Area under curve (AUC). The experimental results reveal that the XGBoost model achieves 95.37% accuracy, while the MLP and SVM classifier models follow with 93.42% and 93.13% accuracy, respectively. The AUC values for XGBoost, SVM, and MLP classifier models are obtained as 0.99, 0.98, and 0.97, respectively, for the testing dataset.

Figure 13 shows the confusion matrix that XGBoost generates when it performs binary classification using the covid and normal CXr images from the test dataset. Figure 14 displays the ROC plot for the XGBoost classifier model. The XGBoost model yields AUC values of 0.99 and 0.99 for normal and COVID, respectively, highlighting the importance of the proposed feature extraction.

After XGBoost, the MLP classifier model achieved 93.42% accuracy. This is followed by SVM model with 93.13% accuracy. We have also compared the ROC plot for Normal and Covid classes obtained using SVM for train and test datasets. The ROC plots clearly indicated that the SVM model does not exhibit overfitting, as the marginal difference between the AUC values for the training and testing datasets is only 0.01. This is also true for MLP and XGBoost classifiers.

So, to discriminate between Normal and Covid classes, we can use SVM, MLP, and XGBoost models. But, the best choice is XGBoost model.

TABLE 7
Performance metrics of machine learning classifiers on test dataset consisting normal and covid classes using the proposed feature extraction method

S.No	ML Models	Accuracy (%)	Precision (%)	AUC
1	Logistic Regression	85.89	76.34	0.92
2	Decision Tree	85.89	72.51	0.82
3	Naive Bayes	70.98	46.52	0.74
4	Support Vector Machine	93.13	89.61	0.98
5	Linear Discriminant Analysis	84.88	73.39	0.90
6	Multilayer perceptron	93.42	85.56	0.97
7	Quadratic Discriminant Analysis	78.29	63.60	0.70
8	XG Boost	95.37	93.57	0.99

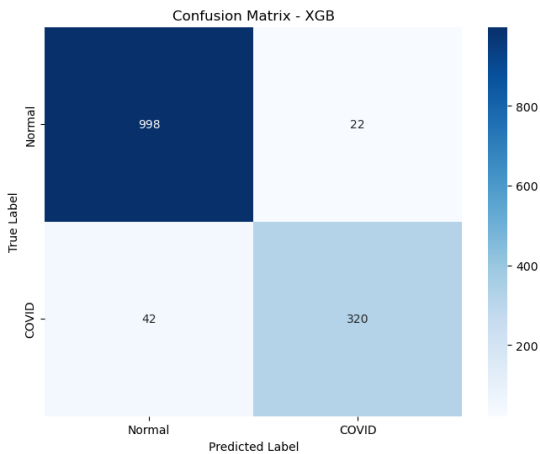


Figure 13. Confusion matrix representing the binary classification results of XGBoost model for Normal and COVID classes on the testing dataset

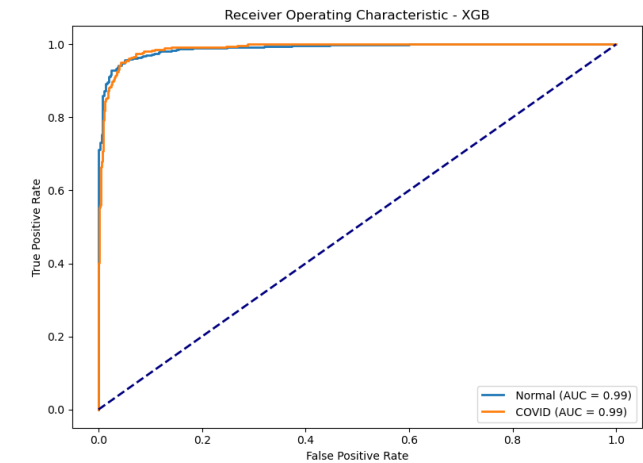


Figure 14. Confusion matrix representing the binary classification results of XGBoost model for Normal and COVID classes on the testing dataset

Table 8 illustrates how MLP, SVM, and XGBoost models perform when trained with features obtained using various feature extraction methods. Our study looks at how the new feature extraction method stacks up against four popular ones:

(i) Grey Level Cooccurrence Matrix (GLCM), (ii) Histogram of Oriented Gradients (HOG), (iii) Local Binary Pattern (LBP), and (iv) Principal Component Analysis (PCA). We evaluate the performance of three classifiers, MLP, SVM, and XGBoost, using the features extracted from both the proposed and conventional feature extraction methods. The results presented in Table 8 clearly demonstrate that the performance of all three ML classifiers, MLP, SVM, and XGBoost, when using the proposed feature extraction outperforms the traditional feature extraction methods. The results demonstrate the importance of the proposed method for feature extraction in COVID diagnosis, particularly when using normal and Covid CXR images for disease detection.

TABLE 8
Performance comparison of proposed feature extraction method to state-of-the-art feature extraction methods using MLP, SVM, XGBoost classifiers

S.No	Feature Extraction/ Feature Selection	MLP Accuracy (%)	SVM Accuracy (%)	XGBoost Accuracy (%)
1	GLCM	77.42	78.73	80.25
2	HOG	88.35	90.44	87.91
3	LBP	79.23	78.72	79.37
4	PCA	89.65	89.50	89.87
5	Proposed Feature Extraction Method	93.42	93.13	95.37

Figure 15 depicts the ROC plot obtained for MLP classifier when the model is trained by using the features extracted using GLCM feature extraction method. The AUC values for COVID and Normal classes are obtained as 0.82 and 0.82 respectively.

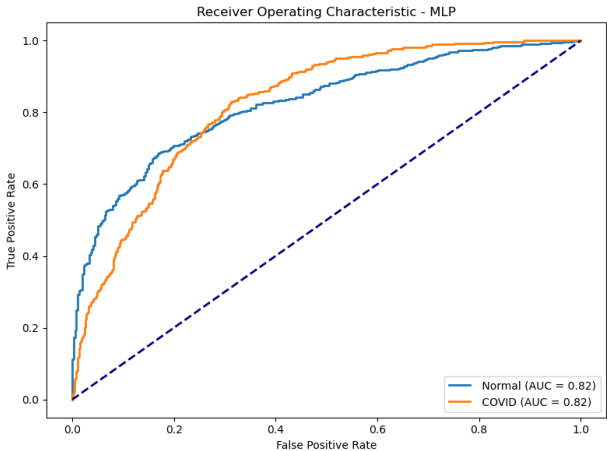


Figure 15. ROC plot obtained for MLP classifier using GLCM feature extraction for COVID detection w.r.t test dataset

Figure 16 depicts the ROC plot obtained for MLP classifier when the model is trained by using the features extracted using HOG feature extraction method. The AUC values for COVID and Normal classes are obtained as 0.93 and 0.93 respectively. Figure 17 depicts the ROC plot obtained for MLP classifier when the model is trained by using the features extracted using LBP feature extraction method. The

AUC values for COVID and Normal classes are obtained as 0.84 and 0.84 respectively.

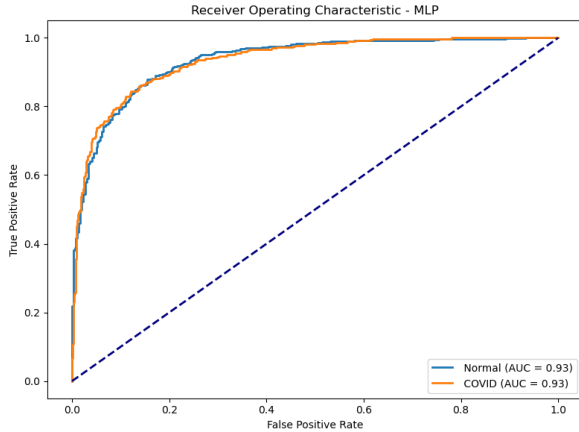


Figure 16. ROC plot obtained for MLP classifier using HOG feature extraction for COVID detection w.r.t test dataset

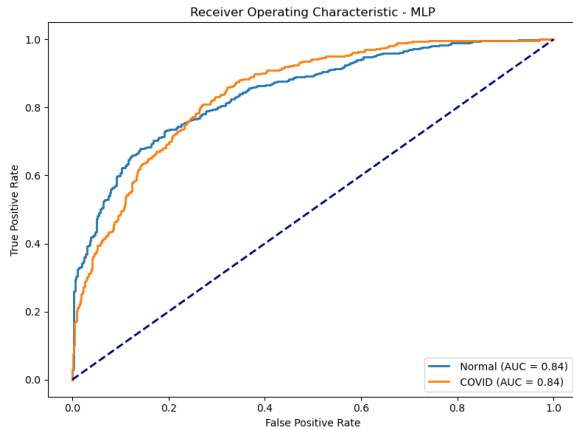


Figure 17. ROC plot obtained for MLP classifier using LBP feature extraction for COVID detection w.r.t test dataset

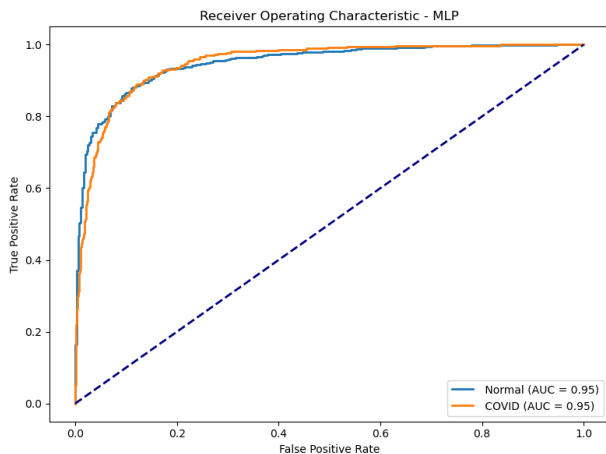


Figure 18. ROC plot obtained for MLP classifier using PCA feature extraction for COVID detection w.r.t test dataset

Figure 18 depicts the ROC plot obtained for MLP classifier when the model is trained by using the features extracted using PCA. In case of PCA, AUC values are obtained as 0.95 and 0.95 for COVID and Normal classes.

Figure 19 depicts the ROC plot obtained for MLP classifier when the model is trained by using the features extracted with proposed feature extraction method. The AUC values for COVID and Normal classes are obtained as 0.98 and 0.98 respectively. All these ROC results show that the performance of the MLP classifier is better when the model is trained and tested using the features extracted with proposed feature extraction when compared to the features extracted using the existing state-of-the-art feature extraction methods

This is also the case for SVM and XGBoost models. We now present the multi-class classification performance of the classifiers when the proposed feature extraction method is employed for discriminating between normal, covid and pneumonia classes.

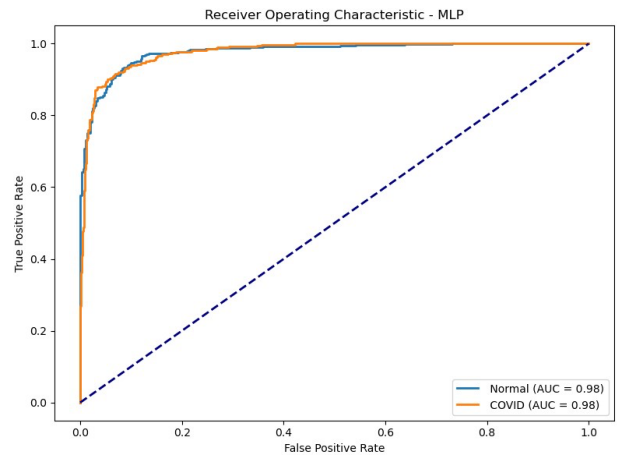


Figure 19. ROC plot obtained for MLP classifier using proposed feature extraction for COVID detection w.r.t test dataset

USE CASE 2: MULTICLASS CLASSIFICATION USING NORMAL AND COVID CHEST X-RAY (CXR) IMAGES

We now present the results of multiclass classification obtained when the classifiers are trained using the features extracted with the proposed feature extraction method and some of the widely used feature extraction and feature selection methods in the literature. For experimental analysis, the original dataset is split into two subsets of training and test datasets by making a 90%-10% split. So, 90% of the dataset is used for training the classifiers, and 10% of the dataset is used for testing the performance of classifiers trained using features extracted with and without the proposed feature extraction method. We have ensured that the images present in the test dataset are unseen CXR images during training. The details of training and testing datasets are already described in Table 1.

Table 9 depicts ML classifiers performance using the proposed feature extraction method for multiclass classification with three evaluation metrics: (i) accuracy, (ii) macro precision, and (iii) macro f-score. All ML classifiers are trained using the features obtained by applying the proposed feature extraction method and then the learned models are used to test the performance of these classifier models on the test dataset.

TABLE 9
Performance metrics of machine learning classifiers on test dataset consisting normal, pneumonia and covid classes using the proposed feature extraction method

S.No	ML Models	Accuracy (%)	Macro Precision (%)	Macro F-Score
1	Logistic Regression	84.4	82.78	0.8135
2	Decision Tree	82.86	77.42	0.7772
3	Naive Bayes	65.13	59.72	0.6577
4	Support Vector Machine	92.09	92.18	0.904
5	Linear Discriminant Analysis	83.52	80.44	0.7947
6	Multilayer perceptron	92.49	91.46	0.9028
7	Quadratic Discriminant Analysis	71.19	NAN	NAN
8	XG Boost	94.07	93.05	0.9247

From the experimental study, as depicted using Table 9, it is observed that the XGBoost model achieved 94.07% accuracy which is the highest among all classifiers. This is followed by MLP classifier having 92.49% and SVM classifier with 92.09% accuracy.

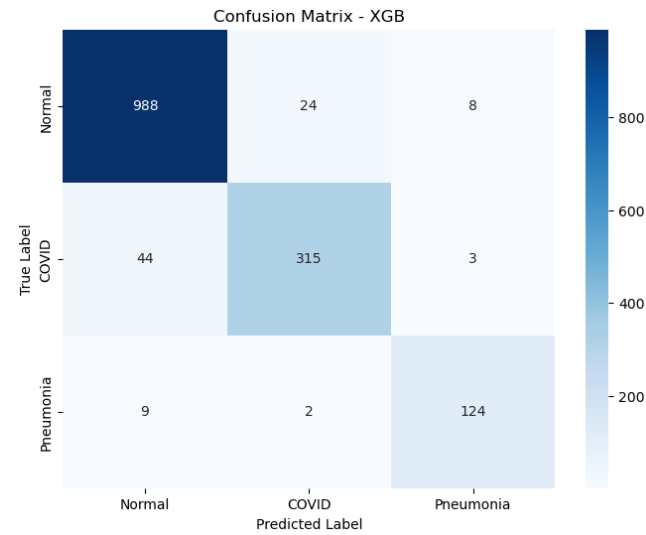


Figure 20. Confusion matrix obtained for XGBoost model for multiclass classification

Figure 20 and Figure 21 depicts the respective confusion matrix and ROC plot for XGBoost model. ROC analysis is less sensitive to class imbalances compared to accuracy. This makes it a reliable metric when evaluating models on datasets where some classes are significantly underrepresented. From the ROC plot depicted for test dataset in the Figure 21, it can be observed that the AUC values for Normal, Covid and Pneumonia classes are 0.99, 0.99 and 1.00 respectively. These AUC values and the curves for respective class approaching the top-left corner indicates better performance in distinguishing that class.

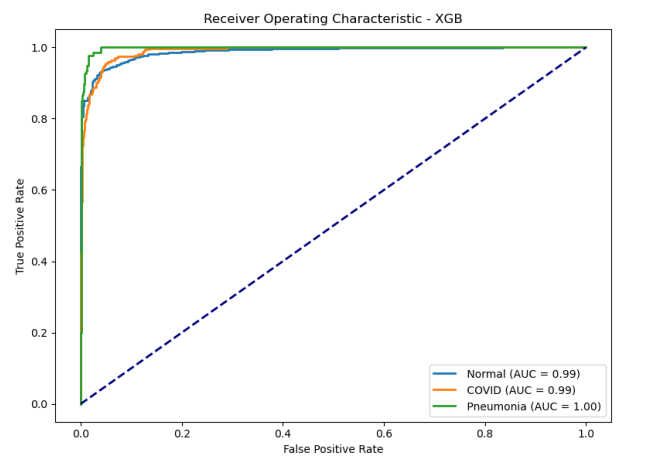


Figure 21. ROC plot obtained for XGBoost classifier using proposed feature extraction for multiclass classification w.r.t test dataset

The significance of the proposed feature extraction method is compared by considering four widely applied feature extraction methods: (i) Grey Level Cooccurrence Matrix (GLCM), (ii) histogram of oriented gradients (HOG), (iii) Local Binary Pattern (LBP), Linear Discriminant Analysis, principal component analysis (PCA), and three feature selection techniques (i) Information Gain (IG), (ii) Chi-Square (CHI-SQ), and (iii) Correlation Coefficient (Corr.Coeff). Table 10 depicts the performance of MLP, SVM, and XGBoost classifiers using the proposed feature extraction method and some of the widely used state-of-the-art feature extraction and feature selection methods in the literature.

TABLE 10
Multiclass classification performance comparison of proposed feature extraction method to state-of-the-art feature selection and feature extraction methods considering MLP, SVM, XGBoost classifiers

S.No	Feature Extraction/ Feature Selection	MLP Accuracy (%)	SVM Accuracy (%)	XGBoost Accuracy (%)
1	GLCM	72.45	72.25	72.91
2	HOG	89.65	87.61	87.54
3	LBP	74.42	75.48	74.49
4	PCA	88.60	87.15	89.06
5	LDA	81.74	81.34	81.67
6	IG	92.09	90.18	93.34
7	CHI-SQ	88.86	87.87	91.69
8	CORR. COEFF	91.89	92.22	93.67
9	Proposed Method	92.55	92.09	94.07

The accuracies achieved by MLP, SVM and XGBoost classifiers with the proposed feature extraction 92.55%, 92.09% and 94.07% respectively. It is evident from the results that the performance of all three ML classifiers (SVM, MLP, and XGBoost) using the proposed feature extraction is better when compared to the traditional feature extraction and feature selection methods. Figure 22 presents the comparison of accuracy values obtained for the MLP classifier using various feature extraction and selection methods. The results prove the

significance of the proposed method for feature extraction even when multiclass classification is considered.

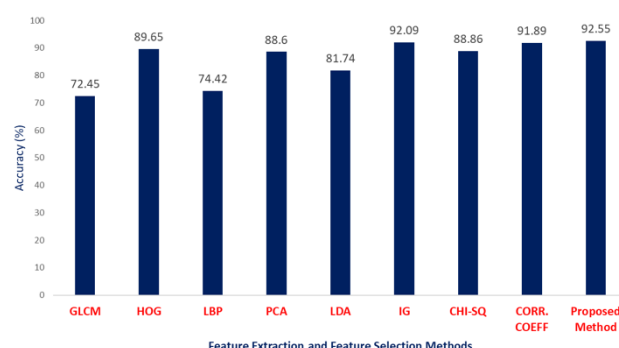


Figure 22. Performance comparison of MLP classifier for various feature extraction methods

VII. DISCUSSIONS

From the experimental analysis and study presented in section VII, the following insights are obtained. We present the findings and insights in detail for each use case in both scenarios considered.

When we have to discriminate between COVID and Pneumonia chest x-ray images for disease detection, logistic regression is the best model considering a balance of accuracy (97.18%), precision (97.80%), and good recall (98.34%). Other high-performing models: Multilayer Perceptron and XGBoost (96.78%). XGBoost and Multilayer Perceptron also perform well, making them suitable alternatives depending on the specific needs of the application (e.g., complexity vs. interpretability).

When we have to discriminate between Normal and Pneumonia chest x-ray for disease detection, the Multilayer Perceptron (96.62%) has the highest accuracy, closely followed by Support Vector Machine (96.54%) and XGBoost (96.02%). Multilayer Perceptron (87.50%) leads in precision, suggesting it has a high rate of true positives among the predicted positives. Also, Multilayer Perceptron has the best accuracy and precision, making it a robust option for tasks where both true positive identification and correct classification of positives matter.

When we have carried experimental analysis to discriminate between Normal and COVID chest x-ray for disease detection, XGBoost (95.37%) demonstrates the highest accuracy, indicating it effectively classifies most instances correctly. Multilayer Perceptron (93.42%) and Support Vector Machine (93.13%) also show strong accuracy. XGBoost (93.57%) leads in precision, suggesting it has a high rate of true positives among its positive predictions. In conclusion, XGBoost is the most effective model among those evaluated, exhibiting strong performance in accuracy, precision, and AUC. Support Vector Machine and Multilayer Perceptron provide solid alternatives.

The findings of comparative analysis of state-of-the-art and proposed feature extraction methods for discriminating

between Normal and COVID CXR images using MLP, SVM and XGBoost Models are as follows : The Proposed Feature Extraction Method stands out significantly, outperforming all other methods in all three models, with XGBoost achieving an accuracy of 95.37%. PCA and HOG also show robust performance, especially with SVM, but do not match the effectiveness of the proposed method. GLCM and LBP provide the least accuracy, suggesting that these methods may not be the best for this particular classification task. We recommend to utilize the proposed feature extraction method for maximum accuracy across all machine learning models. The analysis shows that the proposed feature extraction method significantly enhances model accuracy, particularly with XGBoost. HOG and PCA also provide strong performance and can be useful in various contexts, while GLCM and LBP are less effective for the given classification tasks.

The findings of the experimental study carried to discriminate between Normal, Pneumonia and COVID chest x-ray images for multiclass classification are as follows: XGBoost (94.07%) shows the highest accuracy, indicating that it correctly classifies the majority of instances. Multilayer Perceptron (92.49%) and Support Vector Machine (92.09%) also perform strongly. XGBoost (93.05%) leads in macro precision, indicating a high rate of true positives across classes. Support Vector Machine (92.18%) and Multilayer Perceptron (91.46%) also demonstrate robust precision. XGBoost (0.9247) excels in macro F-score, which balances precision and recall, making it a strong model for multi-class classification. The other notable models Support Vector Machine (0.904) and Multilayer Perceptron (0.9028) also achieve high F-scores. XGBoost stands out as the best model, excelling in all metrics (accuracy, macro precision, and macro F-score). In conclusion, XGBoost is the most effective model among those evaluated, demonstrating superior performance in accuracy, macro precision, and macro F-score. Support Vector Machine and Multilayer Perceptron also provide solid alternatives. Naive Bayes should be approached with caution, and further analysis is needed for Quadratic Discriminant Analysis.

The findings of the experimental study carried for comparative analysis of state-of-the-art and proposed feature extraction methods for discriminating normal, pneumonia and COVID CXR images using MLP, SVM and XGBoost Models are as follows: The proposed method stands out as the best feature extraction technique, achieving the highest accuracy in all models, particularly with XGBoost (94.07%). IG, CORR. COEFF, and CHI-SQ also provide high accuracy, making them effective choices for feature selection. PCA and HOG demonstrate robust performance but do not reach the highest levels of GLCM and LBP show the least effectiveness, suggesting they may not be suitable for this classification task. Optimal Choice is to use the proposed method for the best accuracy and performance across all models. In conclusion, the proposed feature extraction method significantly enhances model accuracy, especially with XGBoost. Information Gain, Chi-Squared, and Correlation Coefficient also provide strong performance.

In contrast, GLCM and LBP are less effective for the given classification tasks.

Now, we present a detailed comparison with some of the most recent research studies in the literature. The Table 11 presents a comparative study which considers the performance of the proposed method with state-of-the-art machine learning and deep learning studies that are carried on binary classification using COVID-19 and normal chest x-ray images. The attributes considered for comparative analysis are (i) ML/DL model (ii) Accuracy (iii) Precision (iii) Recall (iv) Specificity and (v) Balanced accuracy. Table 12 presents the performance comparison done by considering various studies w.r.t. pneumonia detection when

binary classification is performed using COVID-19 and pneumonia images. The comparison is done by comparing the accuracy values obtained in the respective research studies. From these studies, it is evident that proposed feature extraction method aided ML classifier to achieve better accuracy when compared to state-of-the-art research studies. In case of binary classification with COVID and Normal CXR images, XGBoost achieved 95.37% accuracy using the proposed feature extraction method. In case of binary classification w.r.t Pneumonia and Covid CXR images, logistic regression model attained 97.18% accuracy.

TABLE 11
Performance comparison with various state-of-the-art ML/DL models in the literature for discriminating between covid-19 and normal chest radiography images

S.No	Literature	Year	ML/DL model	Accuracy	Precision	Sensitivity	Specificity	Bal. Acc
1	Nikolaou et al. [1]	2021	Model with feature extraction	91.53%	81.81	87	93.13	90.06
2	Nikolaou et al [1]	2021	Model with Fine tuning - EfficientNetB0 (hybrid CNN pre-trained EfficientNetB0 network with a dense layer (32 neurons))	94.93%	91.01	89.5	96.86	93.18
3	Panwar et al. [3]	2020	CNN with nCovnet	88.10%	82%	97.62%	78.57%	88.09
4	Abdullah et al. [5]	2024	Hybrid deep learning model(average pooling layer)- SVM (linear) - 4096 features	92%	93%	89.55%	96.29%	92.92
5	Abdullah et al. [5]	2024	Hybrid deep learning model (average pooling layer)-NN with 4096 feature size	92%	93%	98.68%	77.77%	88.22
6	Ismael et al. [6]	2020	ResNet50 Features + SVM	94.70%	N/A	N/A	N/A	N/A
7	Manjurul Ahsan et al. [9]	2020	NasNetMobile	93.94%%	N/A	N/A	N/A	N/A
8	Lawrence et al. [10]	2020	VGG16 + Resnet50 + custom CNN	89.20%	N/A	N/A	N/A	N/A
9	Jawad Rasheed et al. [21]	2021	LR model at variance of 1	95.2%	1	90.47%	1	95.23
10	Chow LS et al. [22]	2023	VGG-16	94.3%	93.3%	95.2%	-	-
11	El Houby et al. [23]	2024	VGG-19	95%	-	96%	94%	95
12	Our Method	2024	XGBoost using proposed feature extraction	95.37%	93.57%	88.39%	97.84%	93.39

TABLE 12
Performance comparison with various state-of-the-art ML/DL models in the literature for discriminating between COVID-19 and pneumonia

S.No	Literature	Method	Accuracy
1	Oh et al. [31]	Patch based CNN	97.4%
2	Ronneberger et al. [32] [31]	U-Net	85.9%
3	S. Jegou et al. [33] [31]	FC-DenseNet67	81.8%
4	Lobo Torres et al [34] [31]	FC-DenseNet 013	88.9%
5	Our Method	Logistic Regression with spectrograms using proposed feature extraction (Also, (MLP or XGBoost)	97.18%
		MLP and XGBoost with spectrogram images and proposed feature extraction	96.78%
		SVM with spectrogram images and proposed feature extraction	96.58%

The practical applications of the proposed method in clinical settings and how these methods could be implemented in practice are presented below.

The proposed matrix conjugacy technique can extract features from chest X-rays for disease diagnosis. The method could enhance the detection of pneumonia by identifying

subtle patterns in lung images that traditional methods might miss. Advanced features could improve differentiation between COVID-19 pneumonia and other types of pneumonia. Thus, the proposed feature extraction method in clinical settings can improve diagnostic accuracy, improve workflow efficiency, and support research. By addressing challenges and utilizing advanced techniques, these methods contribute to better patient care, optimized clinical practices, and medical technology advancements.

Another application could be automated screening and workflow integration by developing automated diagnostic systems through integrating new features into automated diagnostic tools. The diagnostic tools can automatically screen large volumes of X-ray images to identify potential cases of pneumonia or COVID-19. Also, features can be used in decision support systems that provide radiologists with recommendations based on the analysis.

Feature extraction is essential for improving the efficiency of machine learning models in the field of image classification, especially in medical diagnostics like COVID-19 identification from chest X-ray images. A foundational aspect of this study is grounded in the principles of matrix conjugacy, which serves as a theoretical framework for understanding the behavior of images as mathematical entities.

Matrix conjugacy refers to the relationship between matrices that can be transformed into each other through a similarity transformation. This concept is pivotal in linear algebra, where it illustrates how different representations of data can yield equivalent properties while providing insights into underlying structures. In the context of image processing, matrices represent pixel values, and their properties such as rank, determinant, trace, and eigenvalues can significantly influence the interpretability and performance of feature extraction methods. By employing matrix conjugacy in our proposed feature extraction methods, we can harness the inherent relationships between different features derived from chest X-ray images. This approach allows for a more nuanced understanding of the image data, enabling the identification of crucial patterns that contribute to distinguishing between classes such as normal, pneumonia, and COVID-19. The anticipated impact of incorporating matrix conjugacy into feature extraction is twofold: first, it enhances the robustness of the extracted features by ensuring that they capture essential variations in the data. Second, it facilitates a deeper comprehension of the mathematical properties that govern the data, ultimately leading to improved diagnostic accuracy.

By grounding our study in this theoretical framework, we provide a solid foundation for understanding the significance of our proposed methods and their potential to enhance machine learning performance in medical imaging.

Potential class imbalance and the need for additional validation in diverse populations limit this study, even though it offers significant insights. Our feature extraction method is limited to extraction of $(n+3)$ properties from an image matrix. There is a scope to extract some more representative image features which could aid in improving detection performance. Because of the limitations in

capturing nonlinear interactions, it is necessary to create more advanced algorithms, including machine learning techniques that can better simulate the complexity inherent in medical imaging.

VIII. FUTURE RESEARCH DIRECTIONS

The future of research in medical imaging, particularly with the use of advanced machine learning (ML) and feature extraction techniques, holds immense promise. Here are some key areas where future research could have a significant impact. One of the key research areas which the researchers could focus is on the development of advanced feature extraction techniques. Improved feature extraction methods can lead to more accurate and early detection of diseases. Better features can help tailor treatments based on individual patient profiles. The future research directions are as follows: (i) There is a possibility for exploring new methods to extract high-dimensional and more abstract features from medical images that can capture subtle patterns and important variations which are not visible in conventional analysis. (ii) New ways for investigating the application of advanced matrix similarity and conjugacy concepts to derive novel features that can improve diagnostic accuracy can be proposed.

Building explainable AI models could be another research focus. Better interpretability will make ML models more acceptable in clinical settings. Clinicians will be able to understand the basis for model predictions, leading to better decision-making. Future Research Directions could be developing methods to make complex ML models more interpretable and understandable for clinicians, helping them trust and validate model predictions.

IX. CONCLUSIONS

This study aimed to explore the effectiveness of using matrix similarity properties for feature extraction in the binary classification and multiclass classification of chest X-ray images, specifically distinguishing between normal, COVID-19, and pneumonia cases. The results indicated that the application of matrix similarity properties significantly enhanced classification performance, achieving an overall accuracy of 95.37% when binary class classification is performed using XGBoost to discriminate between COVID and Normal chest xray images. XGBoost model also performed better in multiclass classification achieving an accuracy of 94.07% with a 93.05% macro precision. The model demonstrated strong precision, f-score highlighting its robustness in distinguishing between the different conditions. Logistic regression model achieved 97.18% accuracy, a precision of 97.8% and a 98.34% recall in discriminating between COVID and Pneumonia chest xray images. These findings suggest that incorporating matrix similarity techniques into diagnostic frameworks can improve the accuracy and efficiency of respiratory disease detection. This advancement could assist healthcare practitioners in making timely and informed decisions, particularly in critical situations such as the COVID-19 pandemic. Given the model's promising results, there is a strong case for healthcare policymakers to consider

integrating such analytical methods into routine diagnostic protocols. This could optimize resource utilization and enhance patient outcomes during respiratory disease outbreaks. While this study provides significant insights, it is limited by potential class imbalance and the need for further validation in diverse populations. Future research should focus on expanding the dataset size and exploring additional features to refine the classification model further. In conclusion, this study underscores the potential of matrix similarity properties as a powerful tool for feature extraction in multiclass classification tasks, ultimately contributing to improved diagnostic accuracy and more effective healthcare delivery in response to ongoing public health challenges. In future, we wish to extend the present work to carry multiclass classification and propose new feature extraction and selection algorithms as a future study.

REFERENCES

- [1] Nikolaou, V., Massaro, S., Fakhimi, M. et al., "COVID-19 diagnosis from chest x-rays: developing a simple, fast, and accurate neural network", *Health Inf Sci Syst*, Vol 9, No.36 pp. 1-11, 2021.
- [2] Goel, Kanika et al., "The effect of machine learning explanations on user trust for automated diagnosis of COVID-19", *Computers in biology and medicine*, Vol. 146, pp. 1-12, 2022.
- [3] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, V. Singh, "Application of deep learning for fast detection of COVID-19 in X-rays using nCoVnet", *Chaos Solitons Fractals*, Vol. 138, pp. 1-8, 2020.
- [4] Al-Zyoud W, Erekat D, Saraiji R., "COVID-19 chest X-ray image analysis by threshold-based segmentation", *Heliyon*, Vol. 10, No. 3, pp. 1-12, 2023.
- [5] Abdullah M, Abrha FB, Kedir B, Tamirat Tagesse T., "A Hybrid Deep Learning CNN model for COVID-19 detection from chest X-rays". *Heliyon*, Vol. 10, No. 3, pp. 1-13, 2024.
- [6] Ismael, A.M.; Şengür, A. "Deep learning approaches for COVID-19 detection based on chest X-ray images", *Expert Syst. Appl.*, Vol. 164, pp. 1-11, 2020.
- [7] M. Toğaçar, B. Ergen, Z. Cömert, "COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches", *Computers in Biology and Medicine*, Vol. 121, pp.1-12, 2020.
- [8] Abbas, A., Abdelsamea, M.M. and Gaber, M.M., "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network", *Applied Intelligence*, Vol. 51, pp. 854–864, 2021.
- [9] Ahsan MM, Gupta KD, Islam MM, Sen S, Rahman ML, Shakhawat Hossain M, " COVID-19 Symptoms Detection Based on NasNetMobile with Explainable AI Using Various Imaging Modalities", *Machine Learning and Knowledge Extraction*, Vol. 2, No. 4, pp. 490-504, 2020.
- [10] 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", *arXiv* 2020, arXiv: 2004.02060. (accessed on 15 June 2021).
- [11] Heidari, A., Jafari Navimipour, N., Unal, M. et al., "Machine learning applications for COVID-19 outbreak management", *Neural Computing & Applications*, Vol. 34, pp. 15313–15348, 2022.
- [12] Dogan, O., Tiwari, S., Jabbar, M.A. et al., "A systematic review on AI/ML approaches against COVID-19 outbreak", *Complex Intelligent. Systems* Vol. 7, pp.2655–2678, 2021.
- [13] Ajagbe, S.A., Adigun, M.O., "Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review", *Multimed Tools Appl*, Vol. 83, pp. 5893–5927, 2024
- [14] Afshin Shoeibi et al., "Automated detection and forecasting of COVID-19 using deep learning techniques: A review", *Neurocomputing*, Volume 577, 2024.
- [15] Sharma, S., Guleria, K., "A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images", *Multimed Tools Appl*, Vol. 83, pp. 24101–24151, 2024.
- [16] Koyyada, S.P., Singh, T.P., "A Systematic Survey of Automatic Detection of Lung Diseases from Chest X-Ray Images: COVID-19, Pneumonia, and Tuberculosis", *SN Computer Science*, Springer, Vol. 5, 2024.
- [17] Mallick, D., Singh, A., Ng, E.YK. et al., "Classifying chest x-rays for COVID-19 through transfer learning: a systematic review", *Multimedia Tools and Applications*, pp. 1-60, 2024. <https://doi.org/10.1007/s11042-024-18924-3>
- [18] Samira Sajed, Amir Sanati, Jorge Esparteiro Garcia, Habib Rostami, Ahmad Keshavarz, Andreia Teixeira, "The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review", *Applied Soft Computing*, Vol. 147, 2023.
- [19] Agrawal, T., Choudhary, P., "Segmentation and classification on chest radiography: a systematic survey", *The Visual Computer*, Vol. 39, pp. 875–913, 2023.
- [20] Koul, A., Bawa, R.K. & Kumar, Y., "Artificial Intelligence Techniques to Predict the Airway Disorders Illness: A Systematic Review", *Arch Computat Methods Eng*, Vol. 30, pp. 831–864 2023.
- [21] Rasheed, J., Hameed, A.A., Djeddi, C. et al., "A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images", *Interdisciplinary Sciences: Computational Life Sciences*, Vol. 13, pp. 103–117, 2021.
- [22] Chow, L. S. et al. Quantitative and qualitative analysis of 18 deep convolutional neural network (CNN) models with transfer learning to diagnose COVID-19 on chest X-ray (CXR) images. *SN Comput. Sci.* 4(2), 141 (2023).
- [23] El Houby, E.M.F. COVID-19 detection from chest X-ray images using transfer learning. *Sci Rep* 14, 11639 (2024).
- [24] COVID-19 Radiography Database. <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>. Accessed 3 May 2021.
- [25] Minaee, S. et al. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Med. Image Anal.* 65, 101794 (2020).
- [26] Kanwal K, Asif M, Khalid SG, Liu H, Qurashi AG, Abdullah S. Current Diagnostic Techniques for Pneumonia: A Scoping Review. *Sensors (Basel)*. 2024 Jul 1;24(13):4291.
- [27] Nillmani, Jain PK, Sharma N, Kalra MK, Viskovic K, Saba L, Suri JS. Four Types of Multiclass Frameworks for Pneumonia Classification and Its Validation in X-ray Scans Using Seven Types of Deep Learning Artificial Intelligence Models. *Diagnostics (Basel)*. 2022 Mar 7;12(3):652.
- [28] K. Sanchez, C. Hinojosa, H. Arguello, D. Kouamé, O. Meyrignac and A. Basarab, "CX-DaGAN: Domain Adaptation for Pneumonia Diagnosis on a Small Chest X-Ray Dataset," in *IEEE Transactions on Medical Imaging*, vol. 41, no. 11, pp. 3278-3288, Nov. 2022
- [29] W. Khan, N. Zaki and L. Ali, "Intelligent Pneumonia Identification From Chest X-Rays: A Systematic Literature Review," in *IEEE Access*, vol. 9, pp. 51747-51771, 2021
- [30] Y. Liu, W. Xing, M. Lin, Y. Liu and T. W. S. Chow, "A New Classification Method for Diagnosing COVID-19 Pneumonia via Joint Parallel Deformable MLP Modules and Bi-LSTM With Multi-Source Generated Data of CXR Images," in *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2794-2805, Feb. 2024
- [31] Y. Oh, S. Park and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets", *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688-2700, Aug. 2020.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, in *Lecture Notes in Computer Science*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [33] S. Jegou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19.
- [34] D. Lobo Torres et al., "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery," *Sensors*, vol. 20, no. 2, p. 563, 2020.
- [35] L. Wang and A. Wong, "COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," 2020, arXiv:2003.09871. [Online]. Available: <http://arxiv.org/abs/2003.09871>.

- [36] García-Osorio, César Fyfe, and Colin, "Visualization of High-Dimensional Data via Orthogonal Curves", *Journal of Universal Computer Science*, vol. 11, no 11, pp. 1806–1819.
- [37] Cover, T.M.; Thomas, J.A. (1991). *Elements of Information Theory* (Wiley ed.). John Wiley & Sons. ISBN 978-0-471-24195-9.
- [38] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.
- [39] Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing* (3rd ed.). Prentice Hall.

AUTHOR BIOGRAPHY



Mr. V. Sravan Kiran is currently pursuing a Ph.D. in computer science and engineering at Jain University, Bangalore. He is an assistant professor at the Information Technology Department, VNRVJIET, Hyderabad. He has 11 years of teaching experience. His research interests are in machine learning, algorithms, and software engineering.



Dr. Rajesh Appusamy is currently a professor of computer science and engineering at Jain (deemed to be university), Bangalore. He has earlier served as the HOD & Principal I/C at C. Abdul Hakeem College of Engg. & Tech., Melvisharam. He has more than 25 years of experience in academia. His research interests are in artificial intelligence, machine learning, data mining, web mining, and the semantic web. He is also a supervisor at Jain university for research scholars guiding towards their Ph.D.