Research Article

Application of Ensemble Learning for Respiratory Ailment Diagnosis: Case Studies on Biomedical and Chest X-ray Image Datasets

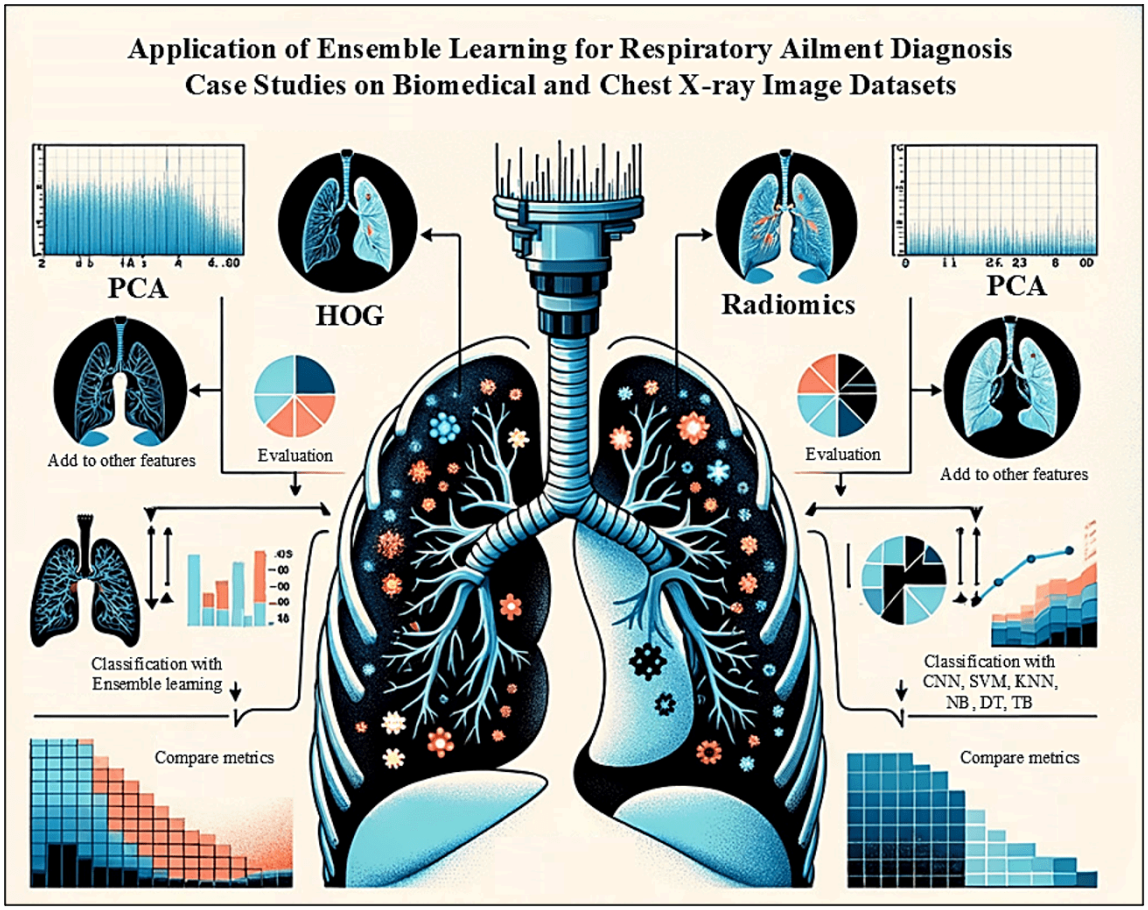
Zeinab Rahimi Rise[1], Mohammad Mahdi Ershadi[2]

1. Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Iran, Islamic Republic of; 2. Amirkabir University of Technology, Iran, Islamic Republic of

The rapid identification of respiratory ailments, such as lung cancer and COVID-19, is critical for timely intervention. Chest X-rays (CXR) serve as an accessible diagnostic tool; however, existing machine learning models often struggle with limited accuracy and sensitivity. This study proposes an ensemble learning-based approach for classifying respiratory ailments using both biomedical and image-based data. Three biomedical datasets and one CXR dataset are utilized as case studies. Histogram of Oriented Gradients (HOG) and Radiomics techniques are applied to extract features from CXR images, which are then processed using Principal Component Analysis (PCA) for dimensionality reduction. To enhance model performance, the Taguchi method is used to tune the parameters of multiple classifiers, including Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Tree Bagger (TB). The proposed ensemble learning approach outperforms individual classifiers by at least 10%, demonstrating significant improvements in accuracy, sensitivity, specificity, precision, recall, F-measure, and G-mean. Statistical tests, including the Wilcoxon Signed-Rank Test and ANOVA, are employed to determine the optimal train-test split and validate the efficiency of the applied methods. The results highlight the potential of ensemble learning in improving diagnostic accuracy for respiratory ailments.

Corresponding authors: Zeinab Rahimi Rise, zeinab.rahimi@aut.ac.ir; Mohammad Mahdi Ershadi, ershadi.mm1372@aut.ac.ir

Graphical Abstract



doi.org/10.32388/1NMNYE.3

Nomenclature

Abbreviation	*Description*	*Abbreviation*	*Description*	*Abbreviation*	*Description*
AI	*Artificial Intelligence*	*GANs*	*Generative Adversarial Networks*	*NLST*	*US National Lung Screening Trial*
BCET	*Balance Contrast Enhancement Technique*	*GLCM*	*Gray-Level Co-occurrence Matrix*	*NSCLC*	*Non-Small Cell Lung Cancer*
CLAHE	*Contrast Limited Adaptive Histogram Equalization*	*GLDM*	*Gray Level Dependence Matrix*	*PCA*	*Principal Component Analysis*
CNN	*Convolutional Neural Networks*	*GLRLM*	*Gray Level Run Length Matrix*	*ROI*	*Region of Interest*
COPD	*Chronic Obstructive Pulmonary Disease*	*GLSZM*	*Gray Level Size Zone Matrix*	*SAKHO*	*Self-Adaptive Kill Herd Optimization*
COVID	*Corona Virus*	*HOG*	*Histogram of Oriented Gradients*	*SMOTE*	*Synthetic Minority Over-sampling Technique*
CT	*Computed Tomography*	*KNN*	*K-Nearest Neighbors*	*SVM*	*Support Vector Machine*
CXR	*Chest X-rays*	*LBP*	*Local Binary Patterns*	*TB*	*Tree Bagger*
DSENet	*Deep Stacking Ensemble*	*LVP*	*Local Vector Patterns*	*TN*	*True Negatives*
DT	*Decision Tree*	*ML*	*Machine Learning*	*TP*	*True Positives*
FN	*False Negatives*	*NB*	*Naïve Bayes*	*WHO*	*World Health Organization*
FP	*False Positives*	*NGTDM*	*Neighboring Gray Tone Difference Matrix*		

1. Introduction

Respiratory ailments, including lung cancer, pneumonia, tuberculosis, chronic obstructive pulmonary disease (COPD), COVID19, etc. remain significant causes of morbidity and mortality worldwide. Timely and accurate diagnosis is crucial to improve patient outcomes, but diagnosing these conditions often involves complex processes requiring skilled specialists and advanced imaging techniques[1]. No country is immune to the threat of respiratory ailments such as COVID-19, which have demonstrated a global reach and profound impact on public health[2][3]. The COVID-19 pandemic, in particular, has highlighted how respiratory diseases can affect every nation, regardless of its economic standing or healthcare infrastructure[4][5]. As of now, the total number of COVID-19 cases reported to the World Health Organization (WHO) has reached approximately over 770,000,000. Figure 1 represents more details about it. [https://data.who.int/dashboards/covid19/cases?n=o]

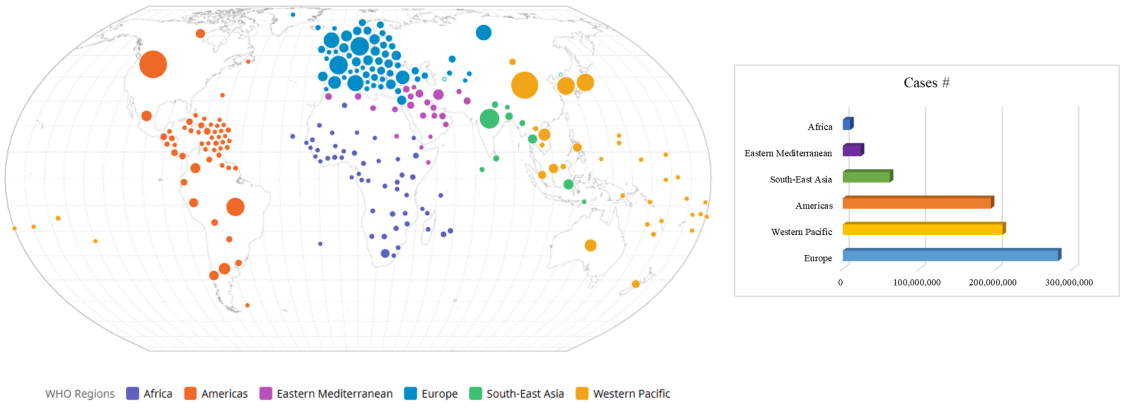


Figure 1. Number of COVID-19 cases reported to WHO (cumulative total - 26 January 2025)

This data reflects the continued global challenge posed by the virus, highlighting the need for ongoing vigilance, research, and public health interventions. Recent advances in data science and machine learning (ML) have brought forth powerful tools that can aid in early detection, reducing diagnostic errors and improving the speed of healthcare response[6][7]. Among these advancements, ensemble learning techniques have demonstrated exceptional potential by combining the strengths of multiple individual models to achieve higher accuracy and robustness[8][9]. Biomedical imaging, particularly chest X-ray (CXR) and computed tomography (CT), is widely used for diagnosing respiratory conditions. Chest X-ray images, due to their cost-effectiveness and quick turnaround, are often the

first line of diagnostic imaging for patients with suspected lung diseases. However, analyzing CXR images can be challenging due to the presence of subtle signs and the need for expert interpretation. This limitation, compounded by a shortage of trained radiologists, has prompted the exploration of automated methods to assist in the interpretation of these images. Similarly, biomedical datasets containing clinical and laboratory information about respiratory conditions offer valuable insights but require sophisticated analytical approaches to handle their complexity and high dimensionality.

This study explores the application of ensemble learning to improve the diagnosis of respiratory ailments using both biomedical and chest X-ray image datasets. Ensemble learning methods, which combine predictions from multiple ML models, have shown promise in boosting diagnostic performance by reducing overfitting and enhancing generalization. Specifically, this paper focuses on employing a combination of feature extraction techniques such as Histogram of Oriented Gradients (HOG) and Radiomics, along with classifiers like Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Tree Bagger (TB). These models are tuned using the Taguchi method for optimal parameter selection, and the proposed ensemble model is compared against individual classifiers to demonstrate its superior diagnostic accuracy. Accordingly, key highlights of this study are as follows:

- The study demonstrates the use of ensemble learning techniques to improve the diagnosis of respiratory diseases, such as lung cancer, COVID, and COPD, from both biomedical and chest X-ray image datasets. It outperforms other classifiers, demonstrating at least a 10% improvement in diagnostic accuracy, offering a more robust solution for early disease detection.
- The study employs Histogram of Oriented Gradients (HOG) and Radiomics for extracting key features from the chest X-ray images and reduces useless features using PCA method, contributing to more accurate disease identification.
- Several classifiers, including Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Tree Bagger (TB) are utilized, with their parameters optimized using the Taguchi method for improved accuracy.
- The study highlights the potential for integrating machine learning-based diagnostic tools into clinical workflows, enhancing decision-making, reducing diagnostic errors, and improving patient care in respiratory disease management.

The structure of this paper is as follows. In Section 1, the study introduces the objective of improving the diagnosis of respiratory ailments using ensemble learning, combining multiple ML classifiers to

enhance accuracy. Section 2 provides a literature review, examining existing methods for diagnosing respiratory diseases from biomedical and chest X-ray images, and identifies gaps in current approaches. Section 3 focuses on the background, detailing feature extraction techniques like HOG and Radiomics, as well as classifiers such as KNN, SVM, DT, NB, CNN, and PCA for feature reduction. In Section 4, the proposed ensemble learning model is outlined, with a discussion on the dataset used, preprocessing steps, and feature extraction methods applied to the images. Section 5 presents the results and discussion, evaluating the proposed method's performance through criteria such as train-test split analysis and efficiency. The results are compared with those of individual classifiers to demonstrate the improvement offered by the ensemble approach. Section 6 concludes the study, addressing the limitations of the approach and suggesting future research directions for enhancing the diagnosis of respiratory ailments using both biomedical and chest X-ray image datasets. Lastly, references are listed in Section 7.

2. Literature review

Artificial intelligence (AI) has significantly advanced the diagnosis of various medical conditions, including breast cancer, brain tumors, and respiratory ailments, through deep learning techniques applied to chest X-ray (CXR) images. However, many studies rely on limited datasets, affecting the generalizability and robustness of their models. This section reviews recent literature on AI-based respiratory ailments detection, focusing on methodologies, contributions, and challenges.

2.1. Machine learning advances in respiratory ailments detection

Different ML methods have been widely used for respiratory ailments detection from CXR images[10]. Afshar et al.[11] introduced COVID-CAPS, a capsule network that offers an alternative to CNNs for COVID-19 identification. Apostolopoulos et al.[12] explored transfer learning and pre-trained CNN architectures to classify COVID-19 cases from small datasets. Similarly, Nayak et al.[13] developed an automated deep neural network model for COVID-19 detection, while Taunk et al.[14] proposed COVID-Net, a deep CNN trained on over 14,000 CXR images for COVID-19 classification. To enhance performance, some studies employed optimization techniques. Arman et al.[15] integrated Bayesian optimization to fine-tune CNN hyperparameters, achieving 94% accuracy in COVID-19 detection. Das et al.[16] introduced a velocity-enhanced whale optimization algorithm hybridized with artificial neural networks for medical data classification. The work of Soares et al.[17] investigated the

effectiveness of transfer learning using pre-trained CNN models for COVID-19 detection, reporting substantial improvements in classification accuracy compared to traditional handcrafted feature extraction techniques. Similarly, Sareeta et al.[18] demonstrated that integrating multi-scale feature extraction with CNNs enhances model robustness when dealing with diverse CXR datasets. The study by Rahman et al.[5] explored an attention-based CNN model, highlighting its ability to focus on specific regions of infection within lung images, improving interpretability and classification performance. Other studies explored classification methods, such as Khan's SVM approach for COVID-19 detection from X-ray data[19] and Ko et al.'s random forest classifier with local wavelet-based CS-binary pattern features[20].

2.2. Hybrid and ensemble models for respiratory ailments detection

Several researchers have proposed hybrid and ensemble models to improve respiratory ailments detection accuracy. Mahin et al.[21] combined CNNs with feature fusion techniques for COVID-19 classification from CXR images. Lascu et al.[22] used transfer learning for multi-class classification of COVID-19, pneumonia, and healthy lung conditions, demonstrating the effectiveness of pre-trained networks. Similarly, Varma et al.[23] conducted a systematic review of ML-based COVID-19 classification techniques, emphasizing the need for benchmark datasets to enhance real-world applicability. Beyond CNNs, hybrid approaches integrating multiple classifiers have been explored. Hamed et al.[24] proposed a KNN variant for COVID-19 identification from incomplete datasets, while Kumar et al.[25] used deep features and correlation coefficients to enhance COVID-19 classification. COVID-19 classification is not solely reliant on deep learning; other ML techniques have also been explored. Ranganath et al.[26] introduced a pivot distribution approach for CXR-based COVID-19 identification. Additionally, Ershadi et al.[27] developed a hierarchical ML model that integrates clinical, biomedical, and image data for treatment planning, demonstrating the effectiveness of multi-modal data integration in medical diagnostics. The study by Balasubramaniam et al. [28] employed a hybrid ensemble approach combining CNNs with Gradient Boosting Decision Trees (GBDT), achieving higher precision than standalone models. Similarly, Shanmugavelu et al. [29] proposed a weighted averaging ensemble technique to integrate multiple deep learning models, demonstrating superior performance over individual networks. A novel stacking ensemble strategy was introduced by Khanna et al.[30], which combined deep and shallow learning models to capture both low-level and high-level features, resulting in improved sensitivity and specificity for COVID-19

detection. Additionally, Nikolaou et al.[31] explored an ensemble of Vision Transformers and CNNs, demonstrating that hybrid architectures can effectively capture spatial and contextual features for medical image classification. The work of Kaleem et al.[32] introduced a novel stacking-based ensemble that integrates multiple CNN architectures with a meta-classifier, improving classification robustness. In contrast, Win et al.[33] proposed an adaptive ensemble technique where model weights were dynamically adjusted based on prediction confidence. These ideas were extended in Das et al. study[34], where the authors developed a weighted average ensemble combining DenseNet201, ResNet50V2, and InceptionV3, achieving an accuracy of 91.62%. This study also introduced a GUI-based application, facilitating real-time COVID-19 detection using chest X-rays.

2.3. Challenges in AI-based respiratory ailments detection

Recent research has also highlighted the importance of dataset availability for AI-based respiratory ailments detection. Cohen et al.[35] contributed an open-access COVID-19 image dataset, supporting further research in AI-driven diagnosis. Despite the progress in AI-based COVID-19 detection, several challenges remain. One major limitation is dataset scarcity, which affects model generalizability. Many studies rely on small, imbalanced datasets, leading to overfitting. Soares et al.[17] attempted to address this issue by introducing a large dataset of real patient CT scans for SARS-CoV-2 identification. Additionally, Rahman et al.[36] demonstrated that image enhancement techniques, such as gamma correction, improve COVID-19 detection reliability by enhancing lung segmentation accuracy. Studies by Abbas S. et al.[9] proposed dynamic feature selection and clustering methods to enhance medical diagnosis interpretability and performance.

2.4. Enhancing respiratory ailments classification with augmentation

Data augmentation and class imbalance handling have also been crucial in improving model performance. The study by Singh et al.[37] highlighted the impact of synthetic data generation using Generative Adversarial Networks (GANs) to address class imbalance issues in CXR datasets. Likewise, Wang et al.[38] introduced a hybrid resampling technique combining oversampling and under sampling to mitigate the effects of skewed class distributions. These strategies were further extended in Win et al. study[33], where the authors applied multiple approaches such as weighted loss balancing, data augmentation, and hybrid resampling to improve classification performance on highly

imbalanced datasets. Their ensemble approach, which combined five different CNNs, achieved an accuracy of 99.23% and an AUC of 99.97%, outperforming many existing models.

2.5. Hybrid feature extraction for classification

Several studies have explored hybrid feature extraction techniques by integrating deep and handcrafted features. The study by Singh et al.[37] utilized a combination of deep CNN features and texture-based features such as Local Binary Patterns (LBP) and Gray-Level Co-occurrence Matrix (GLCM) to improve classification accuracy. Similarly, Balasubramaniam et al.[28] introduced an ensemble classification model incorporating Local Vector Patterns (LVP) along with deep features extracted using InceptionV3. Their proposed Self-Adaptive Kill Herd Optimization (SAKHO) technique was used to optimize neural network weights, resulting in improved classification precision.

2.6. Enhancing respiratory ailments detection with preprocessing

Image preprocessing techniques have also played a significant role in enhancing respiratory ailments detection performance. The work of Islam et al.[39] explored the impact of Contrast Limited Adaptive Histogram Equalization (CLAHE) for improving image contrast before feature extraction. This idea was further developed in[40], where the authors combined CLAHE with the Balance Contrast Enhancement Technique (BCET) to enhance CXR images before applying an ensemble learning model comprising Xception, ResNet50, InceptionV3, and VGG16. Their Deep Stacking Ensemble (DSENet) achieved a classification accuracy of 95%, outperforming individual models and conventional approaches.

Table 1 summarizes key studies in AI-based respiratory ailments detection, comparing their methodologies, contributions, and challenges with the proposed research.

Study	Methods	Contributions	Challenges
Afshar et al.[11]	COVID-CAPS (Capsule Network)	CNN alternative for COVID-19 detection	Limited dataset size
Apostolopoulos et al.[12]	Transfer Learning with CNNs	Improved performance on small datasets	Overfitting risk
Arman et al.[15]	Bayesian Optimization with CNN	Achieved 94% accuracy	Generalization to larger datasets
Balasubramaniam et al.[28]	InceptionV3 + Local Vector Patterns (LVP)	Optimization strategy improved precision	Complexity in feature selection
Das et al.[34]	DenseNet201, ResNet50V2, InceptionV3	GUI-based real-time COVID-19 tool	High computational demands
Hamed et al.[24]	KNN variant	Improved handling of incomplete data	Sensitivity to feature selection
Chaurasia et al.[40]	Xception, ResNet50, InceptionV3, VGG16	BCET + CLAHE preprocessing enhanced performance	Requires specialized preprocessing
Kaleem et al.[32]	ResNet50, DenseNet121, Xception	Big data-based ensemble improved scalability	Need for cloud-based deployment
Kumar et al.[25]	CNN + Gradient Boosting DTs	Boosted classification accuracy	Limited interpretability
Rahman et al.[36]	U-Net + Image Enhancement	Improved lung segmentation accuracy	High computational requirements
Sareeta et al.[18]	InceptionV3, EfficientNet, MobileNet	Multi-scale feature extraction enhanced robustness	Computationally expensive
Shanmugavelu et al.[29]	ResNet50, DenseNet121, VGG19	Weighted ensemble outperformed individual networks	Increased model complexity
Singh et al.[37]	Deep Features + Correlation Coefficient	Outperformed previous methods	Computational cost

Study	Methods	Contributions	Challenges
Soares et al.[17]	CT Scan Dataset	Large-scale COVID-19 dataset for AI research	Limited to CT scans
Taunk et al.[14]	COVID-Net (Deep CNN)	Trained on 14k CXR images	Need for real-world clinical validation
Win et al.[33]	DenseNet121, EfficientNet, Xception	Synthetic data addressed class imbalance	Overfitting risk
Proposed Study	*Ensemble Learning (CNN, SVM, DT, KNN, NB, TB) + Taguchi Optimization*	*10% performance improvement over individual models*	-

Table 1. Comparison table of respiratory ailments detection studies using ensemble learning

Overall, the literature demonstrates the effectiveness of ensemble learning in improving respiratory ailments detection from radiographic images. While individual CNN models such as ResNet, DenseNet, and VGG have achieved high accuracy, ensemble approaches consistently outperform single models by leveraging complementary strengths. The integration of hybrid feature extraction techniques, class imbalance handling, and advanced image preprocessing methods has further contributed to performance improvements. However, challenges such as model interpretability, computational efficiency, and generalizability across different datasets remain areas of ongoing research. Given the existing gaps in AI-based diseases detection, this study aims to develop an ensemble learning approach to improve diagnostic accuracy using features of biomedical and CXR images. Unlike previous works that rely on single classifiers, this research integrates multiple ML models, including CNNs, NBs, TBs, DTs, SVMs, and KNNs, optimized using the Taguchi method. By leveraging a diverse set of biomedical datasets and image-based datasets, this study seeks to enhance generalizability and clinical applicability.

3. Background

Before diving into the proposed methodology of this study, the relevant background information is presented as follows.

3.1. Histogram of Oriented Gradients

HOG is a popular method for extracting features from image data, focusing on object structure and shape. HOG identifies edge features by determining pixel edges and their directions, calculating gradients and edge orientations within localized sections of the image. These sections create histograms based on gradient orientations, producing distinct histograms for each region. Each image block overlaps by 50% and is divided into cells, with cells potentially appearing in multiple blocks due to overlap. For each pixel in each cell, x and y gradients (G_x and G_y) are computed. This process explains how gradients represent edges in two directions across an image (see Equation 1).

$$\theta = \arctan \frac{G_x}{G_y} \tag{1}$$

where r is the magnitude, and θ is the angle.

HOG feature extraction was applied to CXR images by dividing the images into small cells and computing gradient orientation histograms within each cell. These histograms capture local intensity gradients, providing valuable texture information. Subsequently, the histograms are normalized to improve robustness against variations in illumination and contrast. Finally, the normalized histograms are concatenated to form a feature vector, which serves as input to the ML classifiers. This process enables the extraction of discriminative features from CXR images, enhancing the performance of the classifiers in respiratory ailments diagnosis. The specific characteristics of the resulting data depend on the parameters chosen for the HOG algorithm, such as the size of the cells, the number of orientation bins, and any normalization techniques applied.

3.2. Radiomics Feature Extraction

Radiomics is an advanced feature extraction technique that quantitatively analyzes medical images to capture a vast array of textural, shape, and statistical features. This method converts images into high-dimensional data that can be used for predictive modeling in disease classification, treatment response assessment, and biomarker discovery. Radiomics extracts information beyond what is visible to the human eye, capturing details such as intensity distributions, shape descriptors, and texture

patterns. In CXR analysis, Radiomics plays a crucial role in identifying abnormalities, characterizing lung diseases, and supporting AI-driven diagnosis in medical imaging.

3.3. K-Nearest Neighbors Classifier

KNN supervised classification technique is employed for sample categorization. It operates by categorizing new data based on their features and labeled training data, without the need to fit a model, making it memory-based. Utilizing Euclidean distance, it identifies the k training points nearest to a query point, u_0. The new data point is assigned to a group based on the majority of its neighbors. The nearest neighbor classifier requires a dataset for accurate classification, with the training sample representing the existing dataset. Each training vector, u_{tp}, represents a point in the N-dimensional space, where N_v denotes all training patterns. The input test vector, up, is compared with the training data to determine its category, denoted by the class labels, i, and compared with the example vectors, m_{ik}, to ascertain the exact category (see Equation 2).

$$m_{ik} = u_p \tag{2}$$

In this context, m_{ik} signifies the example vector, while the input test vector is represented as up. We consider a collection of metric space points labeled 0 or 1. Given a query (S, T) and samples (S_1, T_1), (S_2, T_2),... represented as (S_n, T_n), the KNN classifier determines the label of the query based on the class with the highest prevalence among the k nearest points to s in the labeled sample. We employ an odd integer for k to avoid ties. Ties can occur either when multiple points at the same distance from the query fail to provide distinct answers or when multiple classes occur with the same frequency among the query's KNNs. To prevent distance ties, we demonstrate universal consistency without assuming density distributions. Various techniques, including random selection, are discussed in the literature to resolve ties in the voting process. The classifier variables, e.g., the number of neighbors (k), distance metric (e.g., Euclidean, Manhattan, Minkowski), weight function (e.g., uniform, distance-based), and algorithm type (e.g., brute-force, KD-tree, Ball-tree) are adjusted based on the Taguchi method for every dataset in this study.

3.4. Support Vector Machine Classifier

SVM operates by segmenting the search space to maximize distance to data points. It excels in text data analysis, allowing flexible feature selection. Its linear method suits high-dimensional text classification. However, excessive parameters hinder performance, mitigated by parameter reduction

and focused feature selection. SVM, a prominent kernel algorithm, employs hyperplane separation for classification based on maximizing margins between classes and nearest points. The classifier variables, e.g., kernel type, regularization parameter (C), gamma (γ) in the case of an RBF kernel, degree in polynomial kernels, and tolerance (ε) for stopping criteria, are adjusted based on the Taguchi method for every dataset in this study.

3.5. Decision Tree Classifier

DT serve as a versatile non-parametric supervised learning method for classification and regression tasks. Each internal node in a DT evaluates a specific attribute, with branches representing test outcomes and leaf nodes signifying examined features. The tree comprises decision nodes, chance nodes, and end nodes, with leaf nodes containing the final outcome. The path from root to leaf forms conjunctions in DT conditions, enabling the generation of decision rules. These rules can elucidate causal or temporal relationships, aiding in association rule building. DT's transparency as a white box model renders it easily interpretable, and it demonstrates efficacy even with limited training data, making it a valuable tool for various analytical tasks. DT methods, renowned for their widespread use in supervised learning, predict model accuracy. However, ensemble methods, such as bagging, boosting, and random forest, surpass individual DTs. These ensemble techniques combine multiple DTs to enhance predictive performance. DTs serve as graphical representations of complex decision scenarios, extracting knowledge from vast data. They efficiently classify new data and offer a concise and easily storable format. The classifier variables, e.g., criterion (Gini or entropy), max depth (the maximum depth of the tree), min samples split (the minimum number of samples required to split an internal node), min samples leaf (the minimum number of samples required to be at a leaf node), and max features (the number of features to consider for the best split), are adjusted based on the Taguchi method for every dataset in this study.

3.6. Naïve Bayes Classifier

NB classifier, rooted in Bayesian statistics, assumes strong independence between features, simplifying classification. It models each class feature independently, aiding in fruit classification, for instance. Trained via supervised learning, it estimates parameters using maximum likelihood, facilitating application with minimal training data. By assuming independence, only variable variances need be determined, not the entire covariance matrix. The classifier employs the maximal a posteriori choice rule, selecting the hypothesis with the greatest likelihood. This process involves

increasing conditional probabilities of features given the class label for each potential label. Overall, NB classifiers offer efficient classification, particularly suitable for scenarios with limited training data (see Equation 3).

$$\text{Classify } (t_1, t_2, \ldots, t_n) = \operatorname{argmax} p(C = c) \prod_{i=1}^{n} p\left(T_i = t_i \mid C = c \right) \tag{3}$$

where $p(C_j)$ is the conditional probability label, and $p(T_i, C_j)$ represents every label and feature. As a result, it appears that the only requirement to construct the classifier is to calculate every conditional probability, $p(T_i, C_j)$, for every label and feature before multiplying the results by the prior probability for that label, $p(C_j)$. The label for which the classifier gets best product is returned by the classifier. The classifier variables, e.g., prior probabilities, likelihoods (class conditional probabilities), smoothing parameters, and kernel functions are adjusted based on the Taguchi method for every dataset in this study.

3.7. Tree Bagger Classifier

In the decision-making process of a DT, progression occurs from a root node to a leaf node, with each step predicting the input variable. However, a single tree may overfit the model. To mitigate this, bootstrap aggregation, a bagging-based technique, is employed. It generates multiple learners by creating additional data points following the same uniform probability distribution. Typically, N learners are averaged to determine the final learning error (see Equation 4). Components of the tree are drawn using a bootstrap replica of the ensemble, growing independently. "Out of bag" observations refer to data elements excluded from computation. This approach helps reduce overfitting and enhances the robustness of the model.

$$e = \frac{1}{N} \sum_{i=1}^{i=N} e_i \tag{4}$$

where N is the learner, and e is the final error.

The classifier variables, e.g., the number of trees, maximum number of splits, minimum leaf size, and the criterion for splitting (such as Gini impurity or entropy), are adjusted based on the Taguchi method for every dataset in this study.

3.8. Convolutional Neural Network Classifier

A CNN is a powerful deep learning model designed specifically for image analysis and pattern recognition. Unlike traditional ML approaches that rely on handcrafted features, CNNs automatically learn hierarchical features from raw image data through multiple layers. A typical CNN consists of convolutional layers that extract spatial features, pooling layers that reduce dimensionality while preserving important information, and fully connected layers that perform classification. The model's ability to recognize patterns such as edges, textures, and complex shapes makes it highly effective in tasks like medical image classification, object detection, and facial recognition. In medical imaging, CNN classifiers play a crucial role in diagnosing diseases from X-rays, MRIs, and CT scans by learning to differentiate between normal and abnormal cases with high accuracy. The classifier variables, such as the learning rate, batch size, number of convolutional layers, number of filters per layer, kernel size, dropout rate, activation functions, optimizer type, and the number of epochs, are adjusted based on the Taguchi method for every dataset in this study.

3.9. Principal Component Analysis

PCA is a widely used dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional form while retaining as much variance (information) as possible. It works by identifying the directions (principal components) in which the data varies the most. PCA computes these components by performing an eigenvalue decomposition of the data's covariance matrix, where each principal component is a linear combination of the original features. The first few principal components often capture the most significant patterns in the data, allowing us to reduce the number of features while maintaining the essential structure of the dataset. PCA is particularly useful in situations where the original dataset has many correlated features, and reducing dimensionality can improve the efficiency of ML models while mitigating overfitting. However, it is important to note that PCA requires the data to be centered (mean-subtracted) and assumes linear relationships among features.

4. Proposed Methodology

This study explores the application of ensemble learning classification to enhance the diagnosis of respiratory ailments. Four datasets are used, consisting of three biomedical datasets and one CXR image-based dataset. Feature extraction methods, including HOG and Radiomics, were employed to

obtain key features from the CXR images. A comprehensive evaluation demonstrated the effectiveness of these features in enabling precise classification, particularly when combined with ensemble learning techniques. The proposed methodology leverages ML to efficiently classify respiratory ailments from both biomedical and CXR data, thereby streamlining the traditionally labor-intensive diagnostic process. The approach follows a two-stage strategy: the first stage involves preprocessing and feature extraction, while the second stage focuses on classification. Various classifiers, such as KNN, SVM, NB, DT, TB, CNN, and ensemble learning, are evaluated across the datasets. It is important to note that each classifier is individually tuned based on the dataset at hand and is trained and tested following the data preprocessing and feature extraction stages. Based on mentioned points, this study schema is represented in Figure 2.

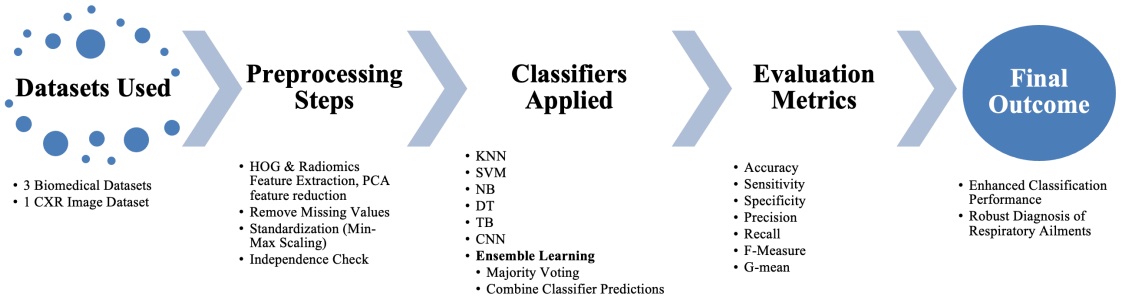


Figure 2. The study schema

4.1. The proposed ensemble learning model

The ensemble learning model combines the predictions from multiple classifiers to derive a final decision via majority voting. The following steps outline the process involved in implementing the ensemble learning algorithm:

Step 1: Load the Dataset: The dataset, which includes biomedical and CXR images, is loaded into the system for further processing.

Step 2: Prepare the Dataset: Perform data preprocessing, including removing non-numeric columns, converting columns to a numeric format, and handling missing values after feature extraction from image data and preforming an independence check on biomedical data.

Step 3: Define the Ensemble Classifier Function

- Create a function, 'ensemble_classifier,' which inputs the training data (X_{train}, y_{train}) and test data (X_{test}).
- Within this function, obtain predictions from five classifiers
- Combine these predictions using majority voting and return the ultimate ensemble prediction.

Step 4: Define Classifier Functions

- Establish separate functions for each classifier
- Each classifier possesses its unique architecture and hyperparameters.
- Compile and train each classifier on the training data, returning predictions for test data.

Step 5: Execute Ensemble Learning and Evaluation

- Specify the number of train-test splits to perform (num_{splits}).
- Initialize an empty list to collect evaluation results (results).
- Iterate over a range of num_{splits} for repeated train-test splits.
- Employ stratified sampling to divide the data into training and test sets.
- Train the ensemble classifier on the training data and predict on the test data.
- Calculate diverse evaluation metrics (e.g., accuracy, precision, recall, etc.).
- Append the evaluation results to the results list.

Step 6: Construct a Results DataFrame

- Create a DataFrame ($results_{df}$) to store the evaluation results, encompassing metrics, confusion matrices, and timing details.

The algorithm iterates through Steps 5 and 6 for each train-test split, generating multiple sets of evaluation metrics and confusion matrices. This repeated process allows for a robust assessment of the ensemble learning model's performance. In this study, the ensemble learning approach combines predictions from multiple classifiers, including SVM, DT, NB, KNN, CNN, and TB, to make the final classification decision. Unlike individual models, ensemble learning benefits from the diversity of various classifiers, leading to improved accuracy, robustness, and generalization. By aggregating predictions from different models, ensemble learning mitigates the weaknesses of individual classifiers, offering a powerful and reliable solution for CXR classification in the diagnosis of respiratory ailments. Based on mentioned steps, a view of proposed model is represented in Figure 3.

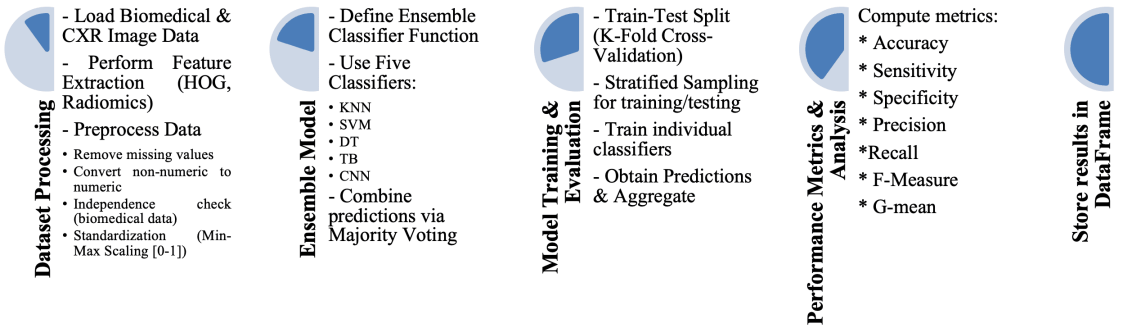


Figure 3. A view of proposed ensemble learning model

4.2. Used dataset

In this study, three biomedical datasets and a CXR dataset are used. Related details are as follows:

Biomedical dataset 1: This dataset is a subset of the US National Lung Screening Trial (NLST), which tracked current and former smokers over a 7-year period, conducting annual lung cancer screenings. Notably, non-smokers were excluded from the trial. The dataset captures key patient information, including age, gender, race, smoking status (current or former, with former defined as having quit within the last 15 years), the time in days until lung cancer was first observed, and the stage at which it was detected. [https://www.kaggle.com/datasets/raddar/smoking-related-lung-cancers]

Biomedical dataset 2: This dataset focuses on pathological lung cancer classification and contains 32 instances with 56 integer-based features that categorizes lung cancer into three pathological types. It is a multivariate dataset primarily used for classification tasks in health and medicine. Although no attribute definitions are provided and there're some missing values, the data has been used in research to explore optimal discriminant planes for classification. [https://archive.ics.uci.edu/dataset/62/lung+cancer]

Biomedical dataset 3: This dataset consists of 1,330 patient records collected from Modarres Hospital, a leading general hospital in Tehran, Iran, covering the period from March 2015 to September 2022. It is designed for classification tasks and includes 34 key features selected by medical experts to diagnose and stage non-small cell lung cancer (NSCLC). Patients are categorized into four stages of NSCLC, ranging from localized cancer to advanced metastasis. The dataset captures a comprehensive set of patient attributes, including demographics (age, gender, education, marital status), lifestyle factors (diet, smoking, occupational hazards), clinical symptoms (coughing, shortness of breath, chest pain),

and medical indicators (hemoglobin, WBC, RBC levels, tumor size, lymph node involvement, metastasis status, and survival time). [Available on request]

CXR dataset: The COVID-19 Radiography Database is a comprehensive collection of chest X-ray (CXR) images created by a team of researchers from Qatar University, the University of Dhaka, and collaborators from Pakistan and Malaysia, in partnership with medical professionals. This dataset, which won the COVID-19 Dataset Award by the Kaggle Community, is designed to support research in COVID-19 detection and classification. Initially, it contained 219 COVID-19, 1341 normal, and 1345 viral pneumonia CXR images, but it has undergone multiple expansions. The most recent update includes 3616 COVID-19 cases, 10,192 normal images, 6012 lung opacity (non-COVID lung infection) cases, and 1345 viral pneumonia images. [https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/data]

Figure 4 presents an illustrative example featuring CXR images of both normal individuals and those affected by viral respiratory conditions.

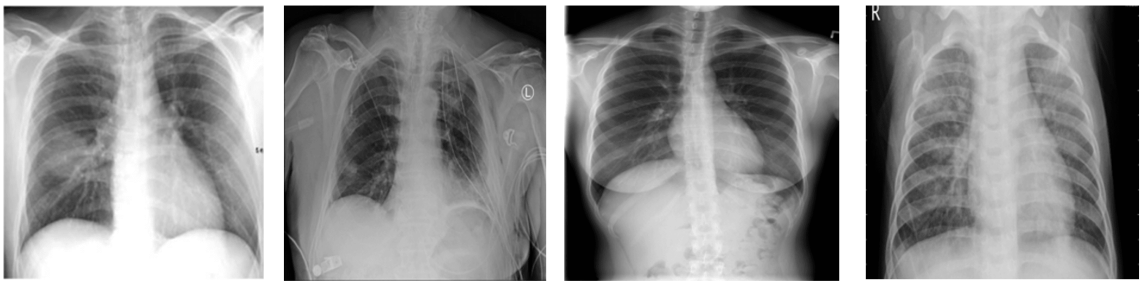


Figure 4. Sample image of COVID-19, lung opacity, normal and viral pneumonia CXR (left to right)

A summary of datasets is represented in Table 2.

Name of Dataset	Data Type	Usage	Type	Number of Cases	Number of Features	Classes	Year
US National Lung Screening Trial (NLST)	Multivariate	Classification	Integer	53428	5	8	2022
Lung Cancer Dataset (Hong and Young)	Multivariate	Classification	Integer	32	56	3	2010
Modarres Hospital Lung Cancer Dataset	Multivariate	Classification	Integer	1330	34	4	2022
COVID-19 Radiography Database	Image-based	Classification	Categorical	21,165	Only image	4	2022

Table 2. An overview on datasets in this study

4.3. Image preprocessing steps

To ensure the reliability and reproducibility of our study, rigorous preprocessing steps are applied to the CXR images. The preprocessing pipeline encompassed several key stages aimed at enhancing the quality and suitability of the dataset for classification tasks.

- Normalization: Prior to any further processing, pixel intensity normalization was performed on the CXR images. Normalization ensures that the pixel values across different images are scaled to a consistent range, typically between 0 and 1. This step is crucial for mitigating the effects of variations in illumination and exposure settings across images, thus enabling more robust model training.
- Augmentation Techniques: To augment the dataset and alleviate potential issues related to data scarcity and class imbalance, various augmentation techniques were employed. Augmentation techniques such as rotation, flipping, scaling, and random cropping were applied to generate additional synthetic training samples. These augmentations not only increase the diversity of the dataset but also enhance the model's ability to generalize to unseen data.
- Addressing Class Imbalance: Class imbalance, where certain classes have significantly fewer samples than others, is a common challenge in medical imaging datasets. To address this issue,

oversampling and/or under sampling techniques are implemented to ensure a more balanced distribution of samples across different classes. Techniques such as random oversampling, SMOTE (Synthetic Minority Over-sampling Technique), and class-weighted loss functions are utilized to mitigate the impact of class imbalance and prevent the model from being biased towards the majority class.

By adhering to these preprocessing practices, we aimed to enhance the quality of CXR dataset and improve the feature extraction/reduction steps to have better learning for ML methods.

4.4. Applying Histogram of Oriented Gradients

To apply the HOG to this image datasets and ensure consistent outputs for comparison, the following parameter are used in this study:

- Image size: 299×299
- Cell Size: 12x12 pixels
- Number of Orientation Bins: 4
- Block Normalization Technique: Block normalization with block sizes of 1x1 cells and a block stride of 1x1
- Orientations: 4

Length of the feature vectors (output of HOG) is calculated as follows:

1. Calculate the number of cells:
    - Cell size: 12×12 pixels
    - Image size: 288×288 pixels
    - Number of cells in horizontal direction: 299/12=24
    - Number of cells in vertical direction: 299/12=24

So, we have 24×24=576 cells in the image.

2. Calculate the number of blocks:
    - Block size: 1×1 cells
    - Block stride: 1x1 cells

To calculate the number of blocks, we need to see how many times a 1×1 block can slide across the 24×24 grid of cells

- $n_{blocks_{horiz}} = W(or\ width\ of\ picture)/width\ of\ cell\ size - width\ of\ block\ size + block\ stride$
- $n_{blocks_{vert}} = H(or\ Height\ of\ picture)/Height\ of\ cell\ size - height\ of\ block\ size + block\ stride$
  - Number of blocks in horizontal direction: 299/12−1+1=24
  - Number of blocks in vertical direction: 299/12−1+1=24

So, we have 24×24=576 blocks in total.

3. Calculate the number of features per block:
   - Each cell has 4 orientation bins (features).
   - Each block contains 1×1=1 cells.
   - $\text{number of features per block} = \text{width of cell size} * \text{height of cell size} * \text{number of orientation bins}$

Therefore, each block contributes 1×4=4 features.

4. Calculate the total feature vector length:
   - $Length\ of\ the\ feature\ vectors = \left(n_{blocks_{\text{horiz}}} \times n_{blocks_{\text{vert}}}\right) \times \text{number of features per block}$
   - Total number of blocks: 576
   - Features per block: 4
   - Total feature vector length: 576×4=2304

Based on mentioned steps, Figure 5 illustrates images after applying HOG feature extraction.

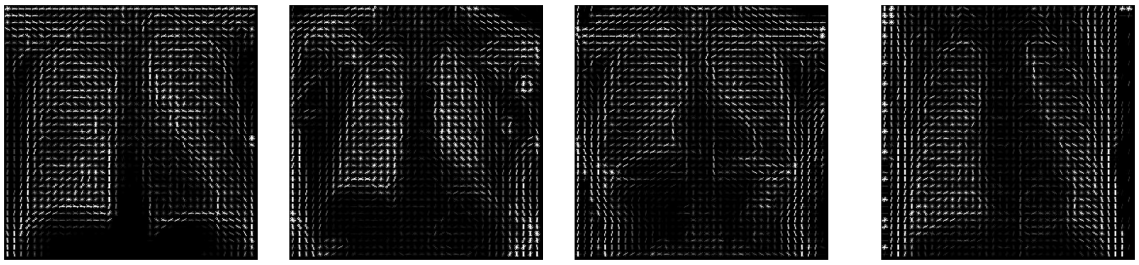


Figure (5). HOG feature extraction of image of COVID-19, lung opacity, normal and viral pneumonia CXR (left to right)

4.5. Applying Radiomics Feature Extraction

Based on this method, the following categories of features are extracted from every CXR image:

- First Order Statistics (18 features): These describe the intensity distribution of pixel values within the region of interest (ROI), e.g. Mean, Variance, Skewness, Kurtosis, Energy, Entropy.
- Shape Features (14 features): Capture the geometric properties of the segmented region, e.g. Volume, Surface Area, Compactness, Sphericity.
- Gray Level Co-occurrence Matrix (GLCM) (24 features): Measures texture by analyzing the spatial relationship of pixel intensities, e.g. Contrast, Correlation, Dissimilarity, Homogeneity, Energy, Entropy.
- Gray Level Run Length Matrix (GLRLM) (16 features): Captures the number of consecutive pixels with the same intensity, e.g. Short Run Emphasis, Long Run Emphasis, Run Entropy.
- Gray Level Size Zone Matrix (GLSZM) (16 features): Measures the size of uniform intensity regions, e.g. Small Area Emphasis, Large Area High Gray Level Emphasis.
- Gray Level Dependence Matrix (GLDM) (14 features): Analyzes the dependence of pixel intensities within the region, e.g. Dependence Variance, Dependence Entropy.
- Neighboring Gray Tone Difference Matrix (NGTDM) (5 features): Assesses the local contrast between a pixel and its neighbors, e.g. Coarseness, Strength, Contrast.

To have better Radiomics feature extraction, masks of every image is considered along with actual image. Figure 6 illustrates related CXR masks.

Figure 6. Related CXR masks of COVID-19, lung opacity, normal and viral pneumonia CXR (left to right)

4.6. Applying PCA Feature Reduction and data pre-processing

For extracted features from CXR data, a PCA feature reduction is applied to handle data size such that the reduced data cover more than 85% of initial data variance. Then, data preprocessing is applied on reduced data and biomedical data to ensure high-quality input for the classifiers. This includes:

- Removing any non-numeric columns and converting relevant columns into a numeric format where necessary.
- Handling missing values by employing imputation or removal as appropriate.
- Standardizing the features using Min-Max normalization to scale all values between 0 and 1, ensuring uniformity in data input.
- Performing an independence check to ensure no high correlation between features.
- Splitting the data into K-folds for train-test cross-validation, ensuring each model is evaluated on different subsets of the data to assess generalization and reduce overfitting.

5. Results and Discussion

Prior to presenting the results, the evaluation metrics are outlined as follows.

5.1. Evaluation Metrics

The confusion matrix is a fundamental tool for evaluating ML algorithms, comparing model predictions against actual reference data. It serves as the basis for key performance metrics, including accuracy, sensitivity, specificity, precision, recall, F-Measure, and G-Mean. The core statistical components of the confusion matrix include true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Table 3 illustrates form of a confusion matrix.

Real \ Prediction	***True***	***False***
True	*True Positives (TP)*	*False Negatives (FN)*
False	*False Positives (FP)*	*True Negatives (TN)*

Table 3. Confusion matrix for binary classification

A classifier's accuracy is measured as the ratio proportion of positive measures to all measures. It determines the degree of accuracy (see Equation 5).

$$\text{Accuracy } = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FN} + \text{FP}) \tag{5}$$

doi.org/10.32388/1NMNYE.3

The sensitivity of a classifier is evaluated as a ratio proportion of true positive measures to all positive measures (see Equation 6).

$$\text{Sensitivity } = \text{TP}/(\text{TP} + \text{FN}) = \text{TPR} \tag{6}$$

The specificity of a classifier is measured by the ratio of true negative measures to all negative measures[31] (see Equation 7).

$$\text{Specificity } = \text{TN}/(\text{FP} + \text{TN}) = \text{TNR} \tag{7}$$

The way in which the percent of all positives were correctly classified is by precision (see Equation 8).

$$\text{Precision } = \text{TP}/(\text{TP} + \text{FP}) \tag{8}$$

The words "recall" and "sensitivity" are interchangeable (see Equation 9).

$$\text{Recall} = \text{ Sensitivity} \tag{9}$$

Compared to the classic accuracy metric, the F1 score gives a more precise illustration of the classifier's performance (see Equation 10).

$$\text{F} - \text{ Measure } = 2 * ((\text{ Precision } * \text{ Recall })/(\text{ Precision } + \text{ Recall })) \tag{10}$$

G-Mean evaluates the rest of classification performance through greater and lesser classes. Despite the fact that negative situations are classified properly, a low G-Mean specifies poor performance in categorizing the positive data (see Equation 11).

$$\text{G} - \text{Mean} = \text{sqrt}(\text{TPR} \times \text{TNR}) \tag{11}$$

5.2. Train-Test split analysis

To determine the appropriate Train-Test split for this study, we apply repeated k-Fold Cross-Validation along with a statistical test (e.g., Paired t-Test for normal distributions or Wilcoxon Signed-Rank Test for non-normal distributions). Based on normality analysis (Skewness between -1 and 1, and Kurtosis near 0), we found that the performance metrics in this study follow non-normal distributions. We then compare significant differences between Train-Test splits (e.g., 60% train - 40% test and 65% train - 35% test) using the following steps:

Step 1: Perform Repeated k-Fold Cross-Validation

- Select multiple train-test splits (e.g., different percentages like 60-40, 65-35, etc.).
- For each train-test split, perform k-fold cross-validation (k = 10).
- Collect the performance metric (e.g., accuracy, F1-score) for each split.

Step 2: Compare Performance Across Different Splits

- Apply a Wilcoxon Signed-Rank Test.
- Null Hypothesis (H_0): "The classifier's performance is not significantly different between different train-test splits."
- Alternative Hypothesis (H_1): "The classifier's performance significantly changes based on the train-test split."

Step 3: Interpret Results

- If p-value < 0.05, reject H_0: the train-test split significantly affects performance.
- If p-value $\geq$ 0.05, fail to reject H_0: the classifier performs consistently across different splits, indicating stability.

Based on computational results, Table 4 is related to p-value acceptance or rejections.

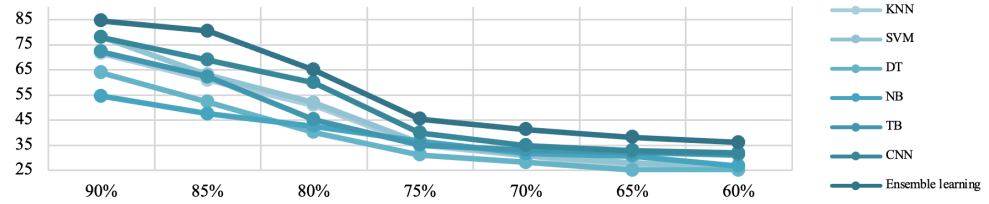
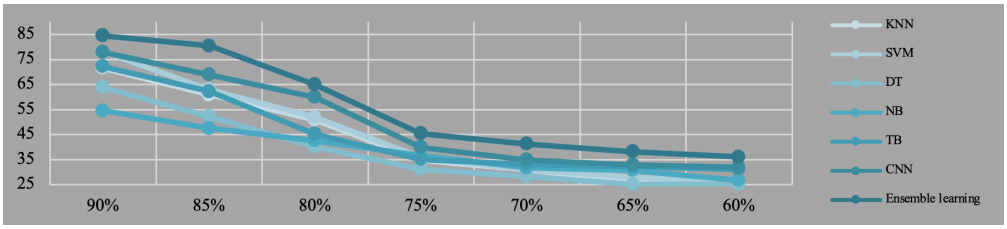
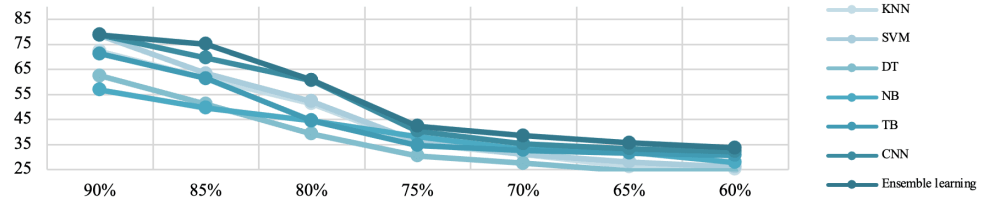
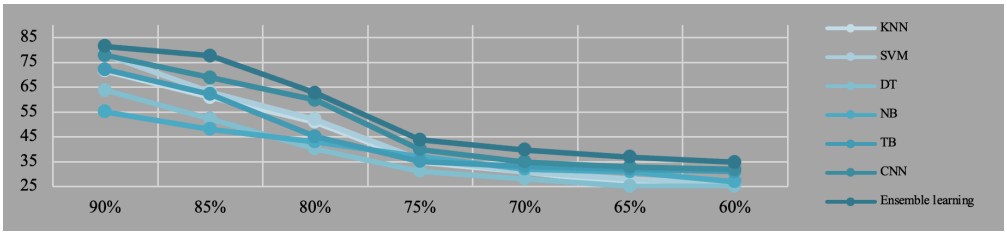
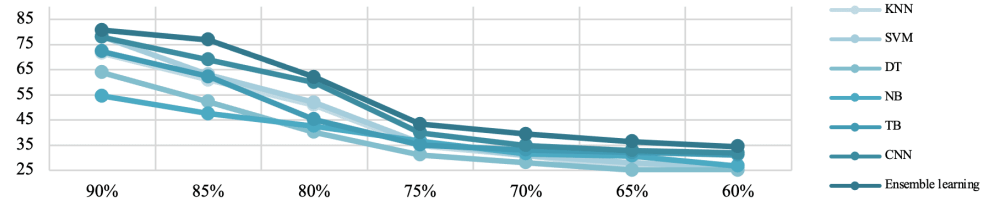
Datasets	Train splits	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
Dataset1	60%-65%	1	1	1	1	1	1	1
	65%-70%	1	1	1	1	1	1	1
	70%-75%	1	1	1	1	1	1	1
	75%-80%	1	1	1	1	1	1	1
	80%-85%	1	1	1	1	1	1	1
	85%-90%	1	1	1	1	1	1	1
	90%-95%	***0***	***0***	***0***	***0***	***0***	***0***	***0***
Dataset2	60%-65%	1	1	1	1	1	1	1
	65%-70%	1	1	1	1	1	1	1
	70%-75%	1	1	1	1	1	1	1
	75%-80%	1	1	1	1	1	1	1
	80%-85%	1	1	1	1	1	1	1
	85%-90%	1	1	1	1	1	1	1
	90%-95%	***0***	***0***	***0***	***0***	***0***	***0***	***0***
Dataset3	60%-65%	1	1	1	1	1	1	1
	65%-70%	1	1	1	1	1	1	1
	70%-75%	1	1	1	1	1	1	1
	75%-80%	1	1	1	1	1	1	1
	80%-85%	1	1	1	1	1	1	1
	85%-90%	1	1	1	1	1	1	1
	90%-95%	***0***	***0***	***0***	***0***	***0***	***0***	***0***
Dataset4	60%-65%	1	1	1	1	1	1	1
	65%-70%	1	1	1	1	1	1	1
	70%-75%	1	1	1	1	1	1	1
	75%-80%	1	1	1	1	1	1	1

doi.org/10.32388/1NMNYE.3

Datasets	Train splits	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
	80%-85%	1	1	1	1	1	1	1
	85%-90%	1	1	1	1	1	1	1
	90%-95%	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Table 4. Statistical results of train-test split analysis

The results indicate no significant improvement between train-test splits of 90%-10% and 95%-5%. Therefore, a 90%-10% train-test split is chosen for all datasets. It is noteworthy that 10% of train data is considered as validation data and it is separate from both the training and test sets. To have better understanding, Table 5 represents different metrics variance for dataset 4.

Metric (%)	*Train data*
Accuracy	
Sensitivity	
Specificity	
Precision	
Recall	

doi.org/10.32388/1NMNYE.3

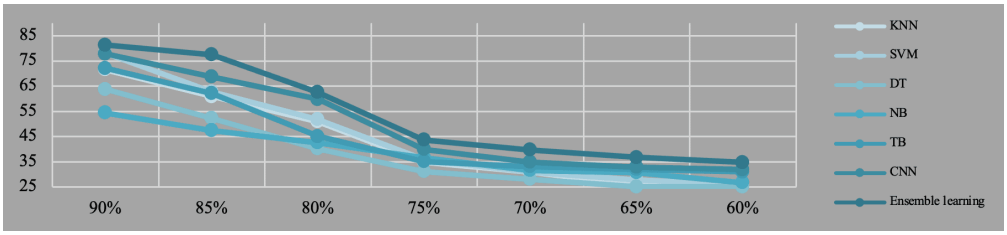
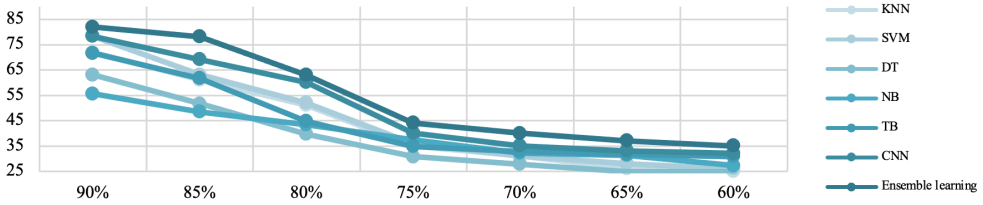
Metric (%)	*Train data*
F-Measure	
G-mean	

Table 5. Different metrics variance for dataset 4

5.3. Results Analysis and Discussion

In this study, all experiments are performed on a personal computer equipped with an Intel® Core™ i5-11400H CPU (2.70 GHz) and 16 GB of RAM, using Python 3.11.5 and Anaconda3 2024.10. The evaluation of classifiers was conducted using 10-fold cross-validation, with results averaged across ten distinct 90%-train, 10%-test splits, as summarized in Table 6 and Figure 7. This robust validation strategy helps reduce overfitting and ensures a more reliable assessment of the model's performance on unseen data, enhancing the study's credibility.

Datasets	Metrics	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
Dataset1	Accuracy	64.32	70.17	57.36	48.97	64.92	70.02	75.97
	Sensitivity	64.32	70.17	57.36	48.97	64.92	70.02	75.97
	Specificity	64.86	70.60	56.09	51.12	63.97	70.77	70.87
	Precision	64.34	70.19	57.29	49.50	64.85	69.97	73.23
	Recall	64.32	70.17	57.36	48.97	64.92	70.02	72.51
	F-Measure	64.32	70.10	57.32	48.94	64.84	69.95	73.11
	G-mean	64.59	70.39	56.72	50.03	64.44	70.39	73.70
Dataset2	Accuracy	81.20	88.60	72.42	61.83	81.96	88.40	95.91
	Sensitivity	81.20	88.60	72.42	61.83	81.96	88.40	95.91
	Specificity	81.89	89.13	70.81	64.53	80.76	89.35	89.48
	Precision	81.23	88.62	72.33	62.49	81.87	88.34	92.45
	Recall	81.20	88.60	72.42	61.83	81.96	88.40	91.54
	F-Measure	81.20	88.50	72.37	61.79	81.87	88.31	92.30
	G-mean	81.55	88.86	71.61	63.17	81.36	88.87	93.05
Dataset3	Accuracy	72.36	78.95	64.53	55.10	73.04	78.78	85.47
	Sensitivity	72.36	78.95	64.53	55.10	73.04	78.78	85.47
	Specificity	72.97	79.42	63.10	57.51	71.97	79.62	79.73
	Precision	72.38	78.97	64.46	55.69	72.96	78.72	82.38
	Recall	72.36	78.95	64.53	55.10	73.04	78.78	81.57
	F-Measure	72.36	78.86	64.49	55.06	72.95	78.69	82.25
	G-mean	72.66	79.18	63.81	56.29	72.50	79.19	82.92
Dataset4	Accuracy	71.55	78.07	63.82	54.48	72.23	77.90	84.52
	Sensitivity	71.55	78.07	63.82	54.48	72.23	77.90	84.52
	Specificity	72.16	78.54	62.40	56.87	71.17	78.73	78.85
	Precision	71.58	78.09	63.74	55.07	72.15	77.85	81.47

Datasets	Metrics	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
	Recall	71.55	78.07	63.82	54.48	72.23	77.90	80.67
	F-Measure	71.55	77.98	63.77	54.45	72.14	77.82	81.34
	G-mean	71.86	78.30	63.10	55.66	71.69	78.31	81.99

Table 6. Performance of different classifiers for every dataset

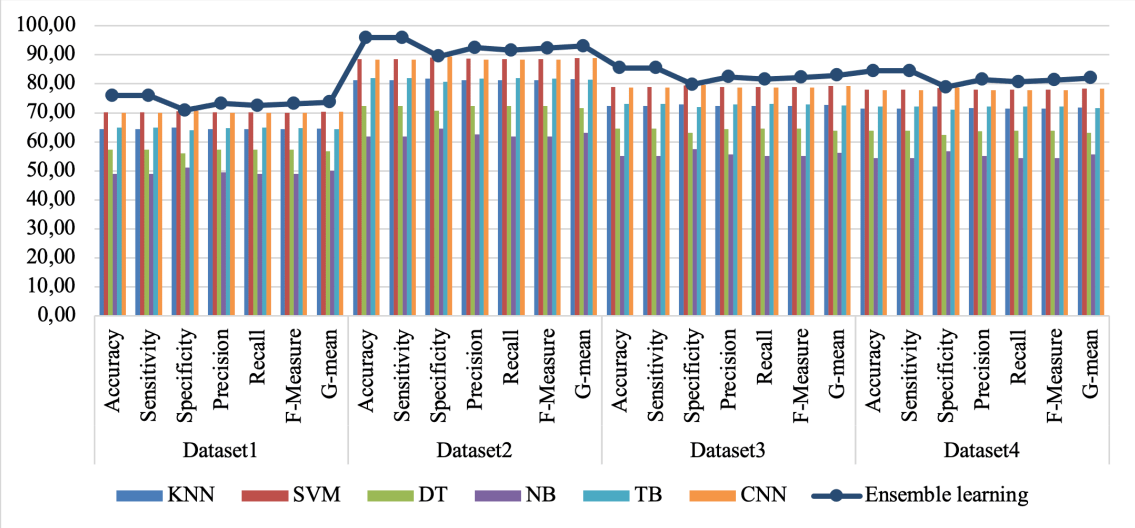


Figure 7. Performance of different classifiers for every dataset

After applying PCA to the extracted features from Dataset 4, both computational time and classifier performance improved significantly. Table 7 and Figure 8 provide a detailed overview of the impact of PCA-based feature reduction on Dataset 4.

doi.org/10.32388/1NMNYE.3

Metrics	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
Accuracy	80.397	87.718	71.705	61.218	81.152	87.529	94.963
Sensitivity	80.397	87.718	71.705	61.218	81.152	87.529	94.963
Specificity	81.079	88.248	70.109	63.895	79.964	88.465	88.593
Precision	80.423	87.739	71.617	61.876	81.062	87.467	91.534
Recall	80.397	87.718	71.705	61.218	81.152	87.529	90.637
F-Measure	80.396	87.62	71.656	61.178	81.056	87.434	91.389
G-mean	80.738	87.982	70.903	62.543	80.556	87.993	92.128

Table 7. Performance of different classifiers for dataset 4 after applying PCA

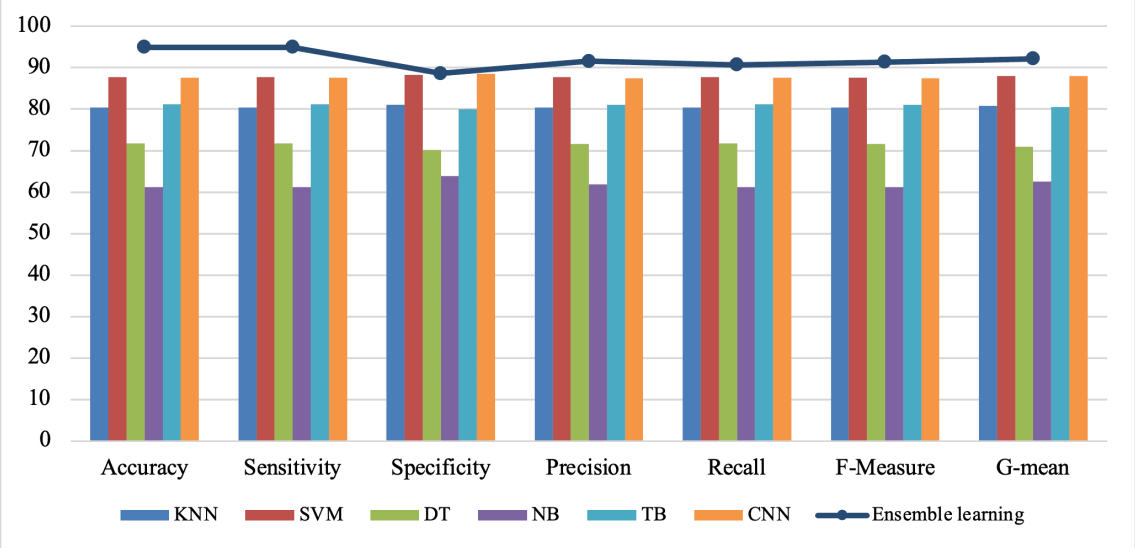


Figure 8. Performance of different classifiers for dataset 4 after applying PCA

The results clearly show that traditional ML classifiers, such as DT and NB, underperform in all datasets. DT achieves accuracy scores between 57.36% and 72.42%, while NB remains the weakest performer, with accuracy ranging from 48.97% to 61.83%. This suggests that DT and NB struggle to

doi.org/10.32388/1NMNYE.3

capture the complex patterns in biomedical and CXR data, likely due to their limited capacity to handle high-dimensional image-based features.

Conversely, more advanced models, such as SVM and CNN, demonstrate significantly higher accuracy, sensitivity, and specificity across all datasets. SVM achieves its best accuracy of 88.60% (Dataset 2), while CNN performs similarly with 88.40%. These models are more adept at learning the discriminative features from the HOG and Radiomics-based extractions, which explains their superior performance over simpler classifiers.

TB also performs consistently well, with accuracy scores ranging from 64.92% to 81.96%, outperforming KNN and DT in most cases. This suggests that TB's ensemble-based decision trees effectively mitigate overfitting and improve generalization.

Across all datasets, the proposed ensemble learning method consistently achieves the highest performance across all metrics. It outperforms individual classifiers by at least 10%, with improvements in accuracy, sensitivity, specificity, precision, recall, F-measure, and G-mean.

- Dataset 1: The ensemble learning method achieves an accuracy of 75.97%, outperforming the best individual classifier, SVM (70.17%), by 5.8%.
- Dataset 2: Ensemble learning reaches 95.91% accuracy, 7.31% higher than SVM (88.60%) and CNN (88.40%). This is the most significant improvement across datasets.
- Dataset 3: With an accuracy of 85.47%, the ensemble model surpasses SVM (78.95%) and CNN (78.78%) by approximately 6.5%.
- Dataset 4: The ensemble model scores 84.52% accuracy, which is 6.45% higher than SVM (78.07%) and CNN (77.90%).

These results indicate that combining multiple classifiers effectively leverages their strengths, leading to more robust and reliable predictions. The application of PCA to Dataset 4 results in significant improvements in accuracy and computational efficiency:

Ensemble learning accuracy increases from 84.52% to 94.96%, an improvement of 10.44%. SVM and CNN also benefit from PCA, with accuracy improvements from 78.07% to 87.72% (SVM) and 77.90% to 87.53% (CNN). DT and NB, which initially performed poorly, also see performance gains, though they remain weaker than other models.

This demonstrates that PCA effectively reduces dimensionality while preserving crucial information, leading to enhanced classifier performance and reduced computational overhead.

5.4. Analysis efficiency of applied method

Another statistical test to detect significant changes after applying a method is the Analysis of Variance or ANOVA. Specifically, One-Way ANOVA can be applied when we have more than two groups (e.g., different methods applied), and we want to test whether the means of these groups are significantly different. We use following steps in this study to compare significant differences after applying a method:

Step 1: Define the Groups

- Divide the data into multiple groups based on the variable you are testing (e.g., applying different methods).

Step 2: Hypothesis Formulation

- Null Hypothesis (H_0): "There is no significant difference between the means of the groups."
- Alternative Hypothesis (H_1): "At least one group mean is significantly different from the others."

Step 3: Perform One-Way ANOVA and interpret the results:

- If the p-value < 0.05, reject the null hypothesis (H_0) and conclude that there is a significant difference between the groups.
- If the p-value ≥ 0.05, fail to reject the null hypothesis and conclude that there is no significant difference between the groups.

According to this statistical method and related results in Table 8, PCA feature reduction on dataset 4 has significant effects on performances of all classifiers in this study. Same results are found for other datasets and 'If p-value < 0.05' condition is 1 for them as well.

	KNN	SVM	DT	NB	TB	CNN	Ensemble learning
p-value	2.56437E-5	4.70466E-5	8.21544E-6	4.07009E-4	6.05239E-5	1.38046E-5	4.73665E-4
If p-value < 0.05	1	1	1	1	1	1	1

Table 8. Statistical results about efficiency of applied PCA feature reduction method

A comparison between results of this paper and literature review is represent in Table 9.

doi.org/10.32388/1NMNYE.3

Study	Dataset	Models Used	Feature Extraction	Accuracy (%)	Key Findings
Win et al.[33]	COVID-19 CXR dataset	CNN	CNN-based feature learning	~88.1% (CoroNet)	Focus on using deep CNN for COVID-19 detection. High sensitivity and specificity but limited to COVID-19 detection only.
Kumar et al.[25]	COVID-19 CXR dataset	SVM, KNN, CNN	HOG	~75-85%	Application of HOG for feature extraction, achieving decent accuracy. Limited improvement from individual models.
Kaleem et al.[32]	Multiple respiratory datasets (including CXR)	CNN, Multi-CNN	CNN + Radiomics (image-based features)	~80-90%	Use of multi-CNN for better feature extraction and classification. The study focuses on a wide range of respiratory diseases but lacks ensemble methods for boosting performance.
Balasubramaniam et al.[28]	COVID-19 CXR dataset	CNN, RNN, SVM	Radiomics + HCF (Hierarchical Convolutional Features)	~85-92%	Combination of CNN, RNN, and SVM for COVID-19 detection, using advanced feature extraction but not utilizing ensemble learning for combining model strengths.
Proposed Study	COVID-19 Radiography Database	KNN, SVM, DT, NB, TB, CNN	HOG, Radiomics	~94.96%	Proposed ensemble learning method outperforms individual classifiers by at least 10%. PCA was applied for dimensionality reduction, improving performance significantly in every metric (accuracy,

Study	Dataset	Models Used	Feature Extraction	Accuracy (%)	Key Findings
					sensitivity, specificity, precision, recall, etc.).

Table 9. Comparison between literature review and this study

6. Conclusion, limitations and future works

This study explores the application of ensemble learning for respiratory ailment diagnosis using biomedical and chest X-ray (CXR) images datasets. Various machine learning classifiers, including CNN, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), and Tree Bagger (TB), demonstrated appropriate performance. However, the ensemble learning approach consistently outperformed individual classifiers, highlighting its potential as a more robust and reliable tool for respiratory disease identification.

By integrating Histogram of Oriented Gradients (HOG) and Radiomics features with Principal Component Analysis (PCA) for dimensionality reduction, the proposed method effectively enhances classification accuracy while reducing computational complexity. The statistical validation using Wilcoxon Signed-Rank Test and ANOVA confirms the significance of the improvements observed.

The findings of this study hold significant clinical implications, demonstrating how ensemble learning can enhance diagnostic accuracy and efficiency for respiratory ailments. Implementing this model in clinical settings could facilitate early detection, enabling timely intervention and treatment. Beyond COVID-19 diagnosis, the proposed approach can be applied to other respiratory diseases, contributing to broader advancements in medical imaging diagnostics.

The main strength of this study lies in its innovative approach to respiratory ailments diagnosis through ensemble learning, demonstrating impressive accuracy rates. However, its reliance on existing datasets and limited exploration of preprocessing techniques may pose limitations. Despite this, the study's contribution to the field is significant, offering a promising avenue for enhancing diagnostic precision in respiratory illnesses. Further research addressing dataset diversity and preprocessing rigor could strengthen its impact and applicability in real-world clinical settings.

6.1. Limitations

While our study demonstrates promising results in respiratory ailments diagnosis through ensemble learning on CXR images, limitations exist. These include the reliance on publicly available datasets, which may lack diversity or contain inherent biases. Additionally, the generalizability of the findings may be constrained by variations in biomedical and image acquisition protocols and population demographics. Although HOG and Radiomics features are effective, they may not fully capture complex patterns in CXR images. The inclusion of deep learning-based feature extraction could further enhance performance. Addressing these limitations could further enhance the robustness and applicability of our methodology in real-world clinical settings.

6.2. Future works

Moving forward, the exploration of deep learning techniques, transfer-based learning, and augmentation strategies presents avenues for further refinement and enhancement of classification accuracy. By delving deeper into these methodologies, researchers can potentially unlock new insights and improve the efficacy of respiratory ailments diagnosis.

Moreover, the scope of this study extends beyond respiratory ailments detection alone. There exists the potential to expand the capabilities of the existing model to not only ascertain the presence of respiratory ailments but also to identify other infectious diseases. This broader application could significantly contribute to the medical field's diagnostic capabilities, facilitating prompt and accurate identification of various illnesses.

Statements and Declarations

Acknowledgments: The authors gratefully acknowledge Dr. S. Ershadi for their invaluable expertise contributed to this paper.

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding: This paper received no financial support from any source.

Research involving human participants and/or animals: This article does not contain any studies with human or animal participants performed by any authors.

References

1. ^ *Shahabi Haghighi et al., 2020. Scenario-based analysis about COVID-19 outbreak in Iran using system atic dynamics modeling-with a focus on the transportation system. Journal of Transportation Research, 17(2), pp.33-48. https://dorl.net/dor/20.1001.1.17353459.1399.17.2.3.2*
2. ^ *Mohammad Sajad E. et al., 2022. "Logistic planning for pharmaceutical supply chain using multi-obje ctive optimization model". International Journal of Pharmaceutical and Healthcare Marketing. 16 (1): 7 5-100. doi:10.1108/IJPHM-01-2021-0004*
3. ^ *Zeinab R. et al., (2022), "Socioeconomic analysis of infectious diseases based on different scenarios usi ng uncertain SEIAR system dynamics with effective subsystems and ANFIS", Journal of Economic and A dministrative Sciences. doi:10.1108/JEAS-07-2021-0124.*
4. ^ *Mohammad M. et al. 2024. "Uncertain SEIAR system dynamics modeling for improved community hea lth management of respiratory virus diseases: A COVID-19 case study". Heliyon. 10 (3): e24711. doi:10.10 16/j.heliyon.2024.e24711*
5. a, b *Rahimi et al., 2022. Multidisciplinary analysis of international environments based on impacts of Co vid-19: state of art. International Journal of Industrial Engineering and Production Research, 33(1), pp.1 76-185. http://ijiepr.iust.ac.ir/article-1-1423-fa.html*
6. ^ *Rise et al., 2023. "Fusing clinical and image data for detecting the severity level of hospitalized sympto matic COVID-19 patients using hierarchical model". Research on Biomedical Engineering. 39 (1): 209-2 32. doi:10.1007/s42600-023-00268-w*
7. ^ *Seifi et al., 2020. "An efficient multi-classifier method for differential diagnosis". Intelligent Decision Technologies. 14 (3): 337-347. doi:10.3233/IDT-190060*
8. ^ *A. Seifi et al., 2020. "An efficient Bayesian network for differential diagnosis using experts' knowledg e". International Journal of Intelligent Computing and Cybernetics. 13 (1): 103-126. doi:10.1108/IJICC-1 0-2019-0112*
9. a, b *Abbas S. et al., 2022. "Applications of dynamic feature selection and clustering methods to medical d iagnosis". Applied Soft Computing. 126: 109293. doi:10.1016/j.asoc.2022.109293*
10. ^ *Zhou, M., Zhang, X. and Qu, J., 2020. "Coronavirus disease 2019 (COVID-19): a clinical update". Fronti ers of Medicine. 14: 126-135. doi:10.1007/s11684-020-0767-8.*
11. a, b *Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K.N. and Mohammadi, A., 2020. "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray imag*

es". Pattern Recognition Letters. 138: 638-643. doi:10.1016/j.patrec.2020.09.010

12. ^a, ^b*Apostolopoulos, I.D. and Mpesiana, T.A., 2020. "Covid-19: automatic detection from x-ray images ut ilizing transfer learning with convolutional neural networks". Physical and Engineering Sciences in Me dicine. 43: 635-640. doi:10.1007/s13246-020-00865-4*

13. ^^*Nayak, S.R., Nayak, J., Sinha, U., Arora, V., Ghosh, U. and Satapathy, S.C., 2023. "An automated lightwei ght deep neural network for diagnosis of COVID-19 from chest X-ray images". Arabian Journal for Scien ce and Engineering. 48 (8): 11085-11102. doi:10.1007/s13369-021-05956-2*

14. ^a, ^b*Taunk, K., De, S., Verma, S. and Swetapadma, A., 2019, May. "A brief review of nearest neighbor algo rithm for learning and classification". In 2019 international conference on intelligent computing and co ntrol systems (ICCS) (pp. 1255-1260). IEEE. doi:10.1109/ICCS45141.2019.9065747.*

15. ^a, ^b*Arman, S.E., Rahman, S. and Deowan, S.A., 2022. "Covidxception-net: A bayesian optimization-base d deep learning approach to diagnose covid-19 from x-ray images". SN Computer Science. 3 (2): 115. do i:10.1007/s42979-021-00980-3*

16. ^^*Das S, Mishra S, Senapati MR. "New Approaches in Metaheuristics to Classify Medical Data Using Artifi cial Neural Network". Arab J Sci Eng.2020; 45, 2459–2471. doi:10.1007/s13369-019-04026-y.*

17. ^a, ^b, ^c*Soares, E., Angelov, P., Biaso, S., Froes, M.H. and Abe, D.K., 2020. "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification". MedRxiv, pp.2020-04. doi:10.11 01/2020.04.24.20078584.*

18. ^a, ^b*Sareeta Mohanty, Manas Ranjan Senapati, 2024, "Chest X-Ray Image Classification for COVID-19 D etection Using Various Feature Extraction Techniques", Published in: Intelligent Systems; Publisher: Spri nger Nature. https://www.springerprofessional.de/en/chest-x-ray-image-classification-for-covid-19 -detection-using-va/26134002*

19. ^^*Khan, M.A., 2021. "An automated and fast system to identify COVID-19 from X-ray radiograph of the ch est using image processing and machine learning". International Journal of Imaging Systems and Tech nology. 31 (2): 499-508. doi:10.1002/ima.22564*

20. ^^*Ko, B.C., Kim, S.H. and Nam, J.Y., 2011. "X-ray image classification using random forests with local wa velet-based CS-local binary patterns". Journal of Digital Imaging. 24: 1141-1151. doi:10.1007/s10278-0 11-9380-3*

21. ^^*Mahin, M., Tonmoy, S., Islam, R., Tazin, T., Khan, M.M. and Bourouis, S., 2021. "Classification of COVID -19 and pneumonia using deep transfer learning". Journal of Healthcare Engineering, 2021. doi:10.115 5/2021/351482*

22. ^Lascu, M.R., 2021. "Deep learning in classification of Covid-19 coronavirus, pneumonia and healthy lungs on CXR and CT images". Journal of Medical and Biological Engineering. 41 (4): 514-522. doi:10.1007/s40846-021-00630-2

23. ^Varma, O.R., Kalra, M. and Kirmani, S., 2023. "COVID-19: A systematic review of prediction and classification techniques". International Journal of Imaging Systems and Technology. 33 (6): 1829-1857. doi:10.1002/ima.22905.

24. a, bHamed, A., Sobhy, A. and Nassar, H., 2021. "Accurate classification of COVID-19 based on incomplete heterogeneous data using a KNN variant algorithm". Arabian Journal for Science and Engineering. 46: 8261-8272. doi:10.1007/s13369-020-05212-z

25. a, b, cKumar, R., Arora, R., Bansal, V., Sahayasheela, V.J., Buckchash, H., Imran, J., Narayanan, N., Pandian, G.N. and Raman, B., 2022. "Classification of COVID-19 from chest x-ray images using deep features and correlation coefficient". Multimedia Tools and Applications. 81 (19): 27631-27655. doi:10.1007/s11042-022-12500-3

26. ^Ranganath, A., Sahu, P.K. and Senapati, M.R., 2021, August. "Detection of COVID from chest X-ray images using pivot distribution count method". In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 373-378). IEEE. doi:10.1109/SPIN52536.2021.9566114

27. ^Ershadi et al., 2022. "A hierarchical machine learning model based on Glioblastoma patients' clinical, biomedical, and image data to analyze their treatment plans". Computers in Biology and Medicine. 150: 106159. doi:10.1016/j.compbiomed.2022.106159

28. a, b, c, dBalasubramaniam, S., & Kumar, K. S. (2023). "Optimal Ensemble learning model for COVID-19 detection using chest X-ray images". Biomedical Signal Processing and Control. 81: 104392. doi:10.1016/j.bspc.2022.104392

29. a, bShanmugavelu, M. and Sannasy, M., 2023. "A scheme of opinion search & relevant product recommendation in social networks using stacked DenseNet121 classifier approach". Automatika. 64 (2): 248-258. doi:10.1080/00051144.2022.2140389

30. ^Khanna, M., Agarwal, A., Singh, L.K., Thawkar, S., Khanna, A. and Gupta, D., 2023. "Radiologist-level two novel and robust automated computer-aided prediction models for early detection of COVID-19 infection from chest X-ray images". Arabian Journal for Science and Engineering. 48 (8): 11051-11083. doi:10.1007/s13369-021-05880-5

31. a, bNikolaou V, Massaro S, Fakhimi M, Stergioulas L, Garn W (2021). "COVID-19 diagnosis from chest x-rays: developing a simple, fast, and accurate neural network". Health Information Science and Systems.

9 (1): 1–11. doi:10.1007/s13755-021-00166-4.

32. a, b, c *Kaleem S, Sohail A, Tariq MU, Babar M, Qureshi B (2023). "Ensemble learning for multi-class COV ID-19 detection from big data". PLOS ONE. 18 (10): e0292587. doi:10.1371/journal.pone.0292587.*
33. a, b, c, d *Win KY, Maneerat N, Sreng S, Hamamoto K (2021). "Ensemble Deep Learning for the Detection of COVID-19 in Unbalanced Chest X-ray Dataset". Applied Sciences. 11 (22): 10528. doi:10.3390/app1122 10528.*
34. a, b *Das AK, Ghosh S, Thunder S, Dutta R, Agarwal S, Chakrabarti A (2021). "Automatic COVID-19 detecti on from X-ray images using ensemble learning with convolutional neural network". Pattern Analysis a nd Applications. 24 (3): 1111–1124. doi:10.1007/s10044-021-00970-4.*
35. ^ *Cohen JP, Morrison P, Dao L (2020). COVID-19 image data collection. arXiv preprint arXiv:2003.11597. https://doi.org/10.48550/arXiv.2003.11597.*
36. a, b *Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Kashem SBA, Islam MT, Al Maadeed S, Zu ghaier SM, Khan MS, Chowdhury ME (2021). "Exploring the effect of image enhancement techniques o n COVID-19 detection using chest X-ray images". Computers in Biology and Medicine. 132: 104319. doi:1 0.1016/j.compbiomed.2021.104319.*
37. a, b, c *Singh D, Kumar V, Vaishali, Kaur M (2020). "Classification of COVID-19 patients from chest CT im ages using multi-objective differential evolution–based convolutional neural networks". European Jour nal of Clinical Microbiology & Infectious Diseases. 39: 1379–1389. doi:10.1007/s10096-020-03901-z.*
38. ^ *Wang L, Lin ZQ, Wong A (2020). "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images". Scientific Reports. 10 (1): 19549. doi:10.1038/s415 98-020-76550-z.*
39. ^ *Islam MT, Aowal MA, Minhaz AT, Ashraf K (2017). Abnormality detection and localization in chest x-r ays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850. https://doi.org/10.4855 0/arXiv.1705.09850.*
40. a, b *Chaurasia S, Gupta AK, Tiwari PP, et al. (2025). "DSENetk: An Efficient Deep Stacking Ensemble Appr oach for COVID-19 Induced Pneumonia Prediction Using Radiograph Images". SN Computer Science. 6: 68. doi:10.1007/s42979-024-03603-9.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.