Research Article

Application of Ensemble Learning in CXR Classification for Improving COVID-19 Diagnosis

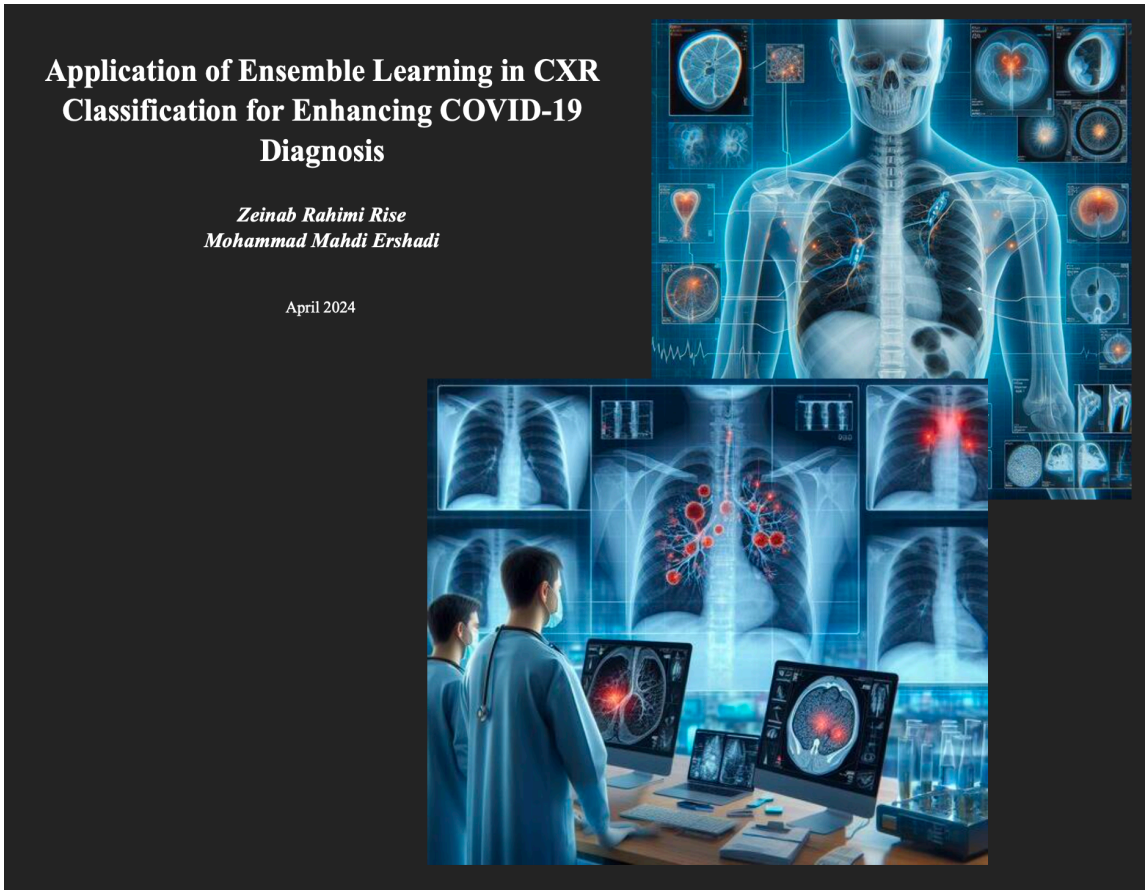
Zeinab Rahimi Rise[1], Mohammad Mahdi Ershadi[2]

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

This study delves into the vital task of classifying chest X-ray (CXR) samples, particularly those related to respiratory ailments, using advanced clinical image analysis and computer-aided radiology techniques. Its primary focus is on developing a classifier to accurately identify COVID-19 cases. Through the application of machine learning and computer vision methodologies, the research aims to enhance the precision of COVID-19 detection. It investigates the effectiveness of Histogram of Oriented Gradients (HOG) feature extraction techniques in conjunction with various classifiers, such as Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), K-nearest neighbor (KNN), and Tree Bagger (TB), alongside an innovative ensemble learning approach. Results indicate impressive accuracy rates, with KNN, SVM, DT, NB, and TB all surpassing the 90% mark. However, the ensemble learning method emerges as the standout performer. By leveraging HOG features extracted from CXR images, this approach presents a robust solution for COVID-19 diagnosis, offering a powerful tool to address the diagnostic challenges posed by the pandemic.

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

 doi.org/10.32388/1NMNYE.2



1. Introduction

Radiology, a medical discipline employing radiation and imaging technology, aids in diagnosing and treating diseases. Computer-aided diagnosis (CAD) serves as a valuable adjunct, offering radiologists a "second opinion" in interpreting chest X-rays (CXRs) to detect illness [1]. CAD assists in diagnosing various conditions such as atelectasis, consolidation, pneumothorax, and pneumonia, critical in infectious respiratory disorders diagnosis [2]. Enhancing CAD capabilities aims to automate illness identification and categorization during CXR interpretation, improving diagnostic precision and consistency. This advancement streamlines radiological workflows, allowing radiologists to work more effectively and efficiently [3]. Computer processing of medical images encompasses acquisition, generation, analysis, and visualization. Figure (1) illustrates the fundamental steps of image processing, underscoring its pivotal role in modern medical diagnostics.

Figure (1). Basic image processing steps

Pattern recognition, commonly known as image mining and machine learning, has emerged as a recent technique for analyzing clinical images, aiding in the automatic detection and classification of various diseases through Computer-Aided Diagnosis (CAD). With the World Health Organization declaring COVID-19 a pandemic on March 11, 2020, the urgent need for efficient screening and rapid clinical intervention for infected individuals became paramount. While real-time reverse transcription polymerase chain reaction (RT-PCR) remains the primary diagnostic method for COVID-19, its costliness and time-consuming nature necessitate the exploration of alternative diagnostic approaches. Chest X-ray (CXR) imaging provides a timely means of assessing suspected cases, yet the overlapping features of viral pneumonia with other lung infections underscore the need for improved diagnostic methods. This study aims to address this challenge by developing an advanced classifier capable of swiftly categorizing COVID-19 cases based on early X-ray findings as either positive or negative. Such a method holds promise for expediting patient treatment and ensuring accurate disease diagnosis, vital for saving lives amidst the ongoing pandemic. Observing a gap in existing literature, this study addresses the scarcity of research utilizing KNN, SVM, DT, TB, NB, and ensemble learning techniques for COVID-19 detection. It endeavors to construct a classifier employing these methods to discern COVID-19 cases as positive or negative. The underlying points are contributions of recommended technique:

- Introducing a novel classifier utilizing diverse machine learning techniques for COVID-19 classification.
- Leveraging X-ray imaging to detect COVID-19, focusing on the lungs as the primary site of infection.
- Proposing a machine learning-based approach using clinical data to identify COVID-19 in suspected cases.
- Developing a computer-aided design technique to analyze records from COVID-19 patients or suspects, employing machine learning for enhanced processing speed and accuracy.

The COVID-19 pandemic has underscored the crucial role of radiology and computer-aided diagnosis (CAD) in disease detection. With the rapid transmission of respiratory illnesses like COVID-19, the demand for accurate diagnostic tools has surged. Radiology, leveraging imaging modalities such as chest X-rays (CXRs), plays a pivotal role in disease assessment, yet interpreting CXR images, especially for COVID-19, presents challenges due to overlapping features with other respiratory conditions. CAD systems offer invaluable support to radiologists by automating image analysis. Through advanced machine learning and computer vision techniques, CAD enhances diagnostic precision and efficiency, crucial in the urgent diagnosis and management of COVID-19 cases. This study explores innovative methods like ensemble learning to classify COVID-19 cases from CXR images. By integrating diverse machine learning classifiers and advanced feature extraction techniques, such as Histogram of Oriented Gradients (HOG), the research aims to develop a robust and automated diagnostic approach. The significance of this study lies in its contribution to advancing CAD systems for respiratory illness diagnosis, particularly in the context of COVID-19. By addressing the need for precise and automated methods in CXR interpretation, the research endeavors to empower radiologists with effective tools for combating the pandemic and improving patient outcomes.

2. Related work

Artificial intelligence is revolutionizing the detection of medical conditions like breast cancer, brain tumors, and COVID-19 using deep learning methods on CXR images. Yet, many studies rely on limited COVID-19 datasets, making it hard to generalize results and ensure prototype efficacy on larger samples. Related works in this fields are reviewed as follows.

Instead of traditional convolutional neural networks (CNNs) for COVID-19 identification, Afshar et al. proposed the use of COVID-CAPS capsule network [1]. Arman et al. introduced a Bayesian optimization method, achieving a 94% accuracy in COVID-19 detection [2]. Apostolopoulos recommended transfer learning and CNNs for identifying COVID-19 from limited datasets [3]. Ranganath et al. suggested a pivot distribution approach for COVID-19 identification from chest X-ray (CXR) images [4]. Das et al. proposed the velocity-enhanced whale optimization algorithm hybridized by artificial neural networks for medical data classification [5]. Han suggested a support vector machine (SVM) classification approach for COVID-19 identification from X-ray data [6]. Ko et al. proposed a technique employing random forest and local wavelet-based CS-binary pattern for image classification [7]. Hamed et al. advocated using k-nearest

neighbor (KNN) variants to identify COVID-19 from incomplete heterogeneous data [8]. Nayak et al. introduced an automatic deep neural network for COVID-19 detection [9]. Khanna et al. recommended an automatic method for timely COVID-19 identification [10].

Kör et al. utilized transfer learning to develop a multi-class convolutional neural network model for automatic pneumonia identification and distinguishing between pneumonia with and without COVID-19 [11]. Meanwhile, Mahin et al. proposed a deep learning technique for COVID-19 identification from chest X-ray (CXR) data [12]. Singh et al. recommended a multi-objective approach for categorizing COVID-19 using computed tomography (CT) scan images [13]. In a similar vein, Islam suggested a CNN-based method to detect chest abnormalities indicative of COVID-19 [14]. Taunk et al. introduced COVID-Net, a deep CNN capable of analyzing 14k CXR images to detect COVID-19 cases [15]. Furthermore, Wang et al. presented a specialized CNN technique utilizing CXR images for COVID-19 recognition in patients [16]. Lastly, Zhang et al. outlined guiding concepts and medical interventions for COVID-19 [17]. Lascu MR utilized transfer learning to classify COVID-19, pneumonia, and healthy lungs from CXR and CT images, achieving reliable results. The study emphasizes the importance of accurate diagnosis and proposes a transfer learning model to aid medical professionals [18]. Varma, Kalra, and Kirmani conduct a systematic review of prediction and classification techniques for COVID-19, highlighting the urgent need for accurate diagnosis amidst the global health crisis. They survey machine learning and deep learning methods, identify challenges, and recommend the establishment of benchmark datasets to enhance effectiveness in real-time clinical settings [19]. Muhammad et al. introduced the COVID-19 Chest X-Ray Database, proposing AI-based rapid and accurate detection of COVID-19 pneumonia from chest X-ray images, achieving high classification accuracy [20]. Cohen et al. introduced the COVID-19 open image data collection, comprising 123 frontal view X-rays sourced from medical websites and publications [21][22]. Soares et al. introduced a large dataset of real patient CT scans for SARS-CoV-2 identification, aiming to aid research in AI methods for COVID-19 detection. The dataset contains 1252 positive and 1230 negative scans, sourced from hospitals in Sao Paulo, Brazil [23]. Sareeta and Manas represented a study about chest X-Ray image classification for COVID-19 detection using various feature extraction techniques [24]. Rahman TF investigated the impact of image enhancement on COVID-19 detection using chest X-rays. A large dataset (COVQU) was compiled, including 18,479 images. Gamma correction proved most effective. The proposed U-Net model achieved high accuracy (98.63%) for lung segmentation, enhancing COVID-19 detection reliability [25]. Kumar et al. proposed a COVID-19 classification method using deep features and

correlation coefficient. Their approach, tested on extensive datasets, outperformed previous methods, highlighting the potential of early detection via chest X-ray images [26]. Murugesan and Muthurajkumar proposed a deep learning approach for product recommendation in social networks, achieving a 92.22% positive score out of 2033 reviews, surpassing traditional methods in accuracy and quality [27].

COVID-19 is not only analyzed by machine learning approaches. Researchers explore the use of uncertain SEIAR system dynamics modeling for community health management, focusing on COVID-19. Their study employs Ensemble Kalman Filter and Metropolis-Hastings algorithms, offering insights into outbreak control scenarios and mortality rates [28].

Ershadi et al. introduce a hierarchical machine learning model for analyzing treatment plans of Glioblastoma Multiforme patients, integrating clinical, biomedical, and image data to improve decision-making efficacy. They employ Fuzzy C-mean clustering, Wrapper feature selection, and twelve classifiers to optimize outcomes [29]. Rahimi Rise et al. advocate for stronger environmental considerations post-COVID-19 to transition the global economy towards renewable energy and resilient public-health systems, emphasizing the need for institutional reforms within the United Nations System [30]. Rise et al. propose a hierarchical model combining expert knowledge, FCM clustering, and ANFIS classification to detect severity levels of hospitalized symptomatic COVID-19 patients, achieving high accuracy using both clinical and image data [31]. Rahimi Rise et al. analyze socioeconomic impacts of infectious diseases using an uncertain SEIAR model with scenario-based analysis, emphasizing future GDP and social impact predictions for policymaking [32]. Ershadi and Seifi proposed a dynamic multi-classifier method for disease diagnosis, combining feature reduction techniques and clustering selection to enhance accuracy and computation time efficiency [33]. Other authors proposed a multi-objective optimization model for pharmaceutical supply chain logistics in pandemic situations, aiming to minimize unsatisfied requests and transportation costs while considering various factors and employing a hybrid optimization approach [34]. Rahimi Rise et al. explored COVID-19 outbreak scenarios in Iran using system dynamics modeling, emphasizing the transportation system's impact and proposing strategies for government decision-making amid varying mortality rates and recovery scenarios [35]. Ershadi and Seifi represented dynamic feature selection and clustering methods to enhance medical diagnosis. Their novel approach combines feature selection, clustering, and deep learning to improve classification performance significantly [36]. They developed an efficient Bayesian network for differential diagnosis, integrating expert knowledge and data-driven methods, achieving up to 87% accuracy, as well [37].

According to the literature review, the motivation to develop a classifier for accurately identifying COVID-19 cases using chest X-ray (CXR) samples stems from the urgent need for alternative diagnostic methods amidst the COVID-19 pandemic. While RT-PCR testing is the gold standard, its limitations in terms of cost and turnaround time prompted exploration of supplementary diagnostic tools. CXR imaging, with its accessibility and potential for rapid assessment of lung pathology, emerged as a promising modality. Leveraging advancements in machine learning and image analysis, researchers aimed to train algorithms to automatically detect COVID-19 patterns in CXR images. Such a classifier could expedite diagnosis, inform treatment decisions, and aid in disease surveillance efforts. Ultimately, the goal is to improve patient care by providing timely and accurate identification of COVID-19 cases, thereby contributing to the management and containment of the pandemic.

3. Proposed Methodology

This method utilizes machine learning to efficiently identify and categorize COVID-19 from X-rays, streamlining the labor-intensive process of diagnosis. Employing a two-stage strategy, it extracts features in the first stage and classifies images in the second. Various classifiers, including KNN, SVM, decision tree (DT), Tree Bagger (TB), and ensemble learning, are evaluated post feature extraction. Histogram of Oriented Gradients (HOG) is employed for this purpose. It is noteworthy that each classifier was individually trained and tested using the extracted Histogram of Oriented Gradients (HOG) features from chest X-ray (CXR) images. The selection criteria for these classifiers were based on their established performance in medical image classification tasks, as well as their suitability for handling the characteristics of CXR data. Specifically, classifiers were chosen for their ability to handle non-linear relationships, interpretability, and scalability to large datasets. The subsequent sections detail the techniques employed in this method to enhance COVID-19 detection from chest X-ray images.

3.1. Histogram of Oriented Gradients

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 illustrates how gradients represent edges in two directions across an image (see Equation 1). Additionally, Figure (2) illustrates standard HOG and viral HOG. Gradients′ magnitudes and phases are then determined accordingly.

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

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

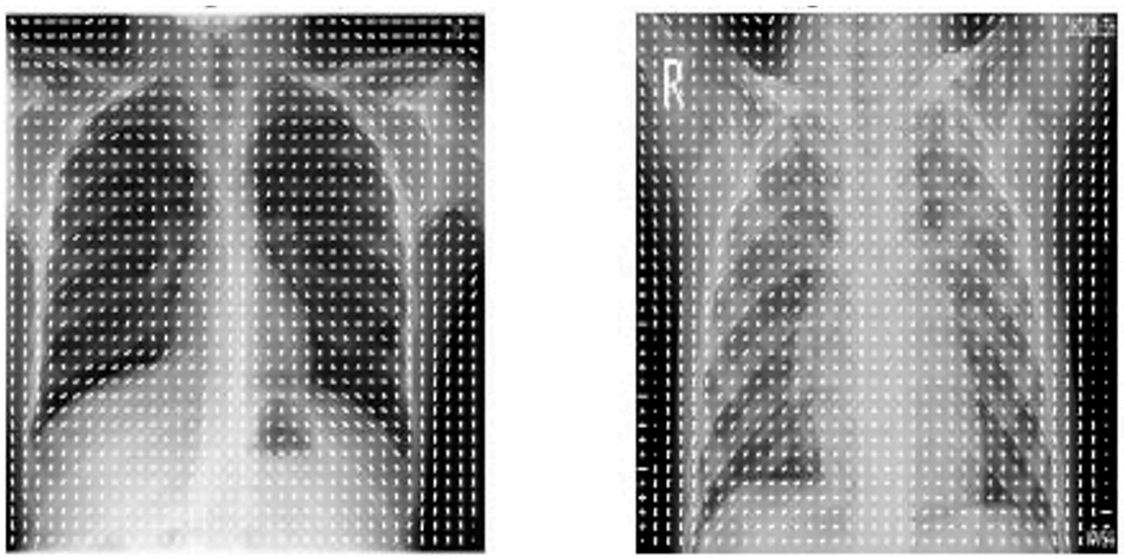


Figure (2). HOG feature extraction of normal (left picture) and COVID-19 (right picture-Viral) CXR images

Histogram of Oriented Gradients (HOG) feature extraction was applied to chest X-ray (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 machine learning classifiers. This process enables the extraction of discriminative features from CXR images, enhancing the performance of the classifiers in COVID-19 diagnosis.

In the study on the application of ensemble learning in CXR classification for enhancing COVID-19 diagnosis, a single feature selection method was utilized. Specifically, the research focused on the Histogram of Oriented Gradients (HOG) feature extraction technique, which was employed in conjunction with various machine learning classifiers and ensemble learning approaches to enhance the accuracy of COVID-19 diagnosis from chest X-ray images. This comprehensive evaluation demonstrated the efficacy

of HOG features in facilitating precise classification, particularly when combined with ensemble learning methodologies.

3.2. K-Nearest Neighbor Classifier

The k-nearest neighbor (k-NN) 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, utp, 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 k-nearest neighbor 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 k-nearest neighbors. 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.

3.3. Support Vector Machine Classifier

The Support Vector Machine (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.

3.4. Decision Tree Classifier

Decision trees (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 decision tree 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.

Decision tree methods, renowned for their widespread use in supervised learning, predict model accuracy. However, ensemble methods, such as bagging, boosting, and random forest, surpass individual decision trees. These ensemble techniques combine multiple decision trees to enhance predictive performance. Decision trees 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.

3.5. Naive Bayes (NB) Classifier

The Naive Bayes 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, Naive Bayes classifiers offer efficient classification, particularly suitable for scenarios with limited training data (see Equation 3).

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

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

doi.org/10.32388/1NMNYE.2

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.

3.6. Tree Bagger Classifier

In the decision-making process of a decision tree, 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.

3.7. The proposed ensemble learning model

The proposed ensemble method combines predictions from these five classifiers through majority voting to derive the final prediction. The steps involved in the proposed ensemble learning algorithm are outlined as follows:

Step 1: Load the Dataset

Step 2: Prepare the Dataset

- Perform data preprocessing, including removing non-numeric columns, converting columns to a numeric format, and handling missing values.

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 the test data.

Step 5: Execute Ensemble Learning and Evaluation

- Specify the number of train-test splits to perform ($nums_{plits}$).
- 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.

Step 7: Save the Results

This algorithm iterates through Steps 5 to 7 for each train-test split, yielding multiple sets of evaluation metrics and confusion matrices. By leveraging predictions from various classification models, this ensemble method enhances the robustness and accuracy of classifications.

The ensemble learning approach in this study combines the predictions of multiple individual classifier models, such as Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), K-nearest neighbor (KNN), and Tree Bagger (TB), to make a final decision. Unlike individual classifier models, ensemble learning leverages the diversity of multiple models to improve overall performance. By aggregating the predictions of diverse classifiers, ensemble learning mitigates the weaknesses of individual models and enhances accuracy, robustness, and generalization capability, making it a powerful technique for CXR classification, especially in the context of COVID-19 diagnosis.

3.8. Used dataset

Data collection is paramount in machine learning research, especially in medical imaging. This study requires a diverse set of chest X-ray (CXR) images encompassing pneumonia, COVID-19 positive and negative cases, and normal cases. Unfortunately, standalone datasets representing each category independently do not exist. Instead, samples are gathered from two sources: a dataset provided by Dr. Joseph Paul, a postdoctoral scholar, and CXR datasets available on Kaggle [20]. Dr. Paul's dataset includes CXR and CT scan samples not only of COVID-19 cases but also of other respiratory viruses such as ARDS, SARS, and MERS [21][22]. For this research, COVID-19 image samples from Dr. Paul's dataset are utilized. A total of 15,153 sample data points (from 512x512 pixels to 1024x1024 pixels) were selected for classification in this study. This dataset comprised 3615 COVID-19 positive cases, 10,192 Normal cases, 6012 Lung Opacity (Non-COVID lung infection) cases, and 1345 Viral Pneumonia cases. Additionally, Kaggle provides free access to relevant data for research purposes [23]. The collected images undergo organization, preprocessing, and conversion into NumPy arrays to facilitate the training process. Notably, the Kaggle dataset offers a variety of CXR images showcasing different chest perspectives of patients afflicted with COVID-19, ARDS, SARS, MERS, and other disorders [23]. Figure 3 presents an illustrative example featuring CXR images of both normal individuals and those affected by viral respiratory conditions. By leveraging these datasets, researchers can access a rich pool of CXR images crucial for training machine learning models. This comprehensive dataset not only aids in pneumonia and COVID-19 diagnosis but also contributes to advancing the broader field of medical imaging research.

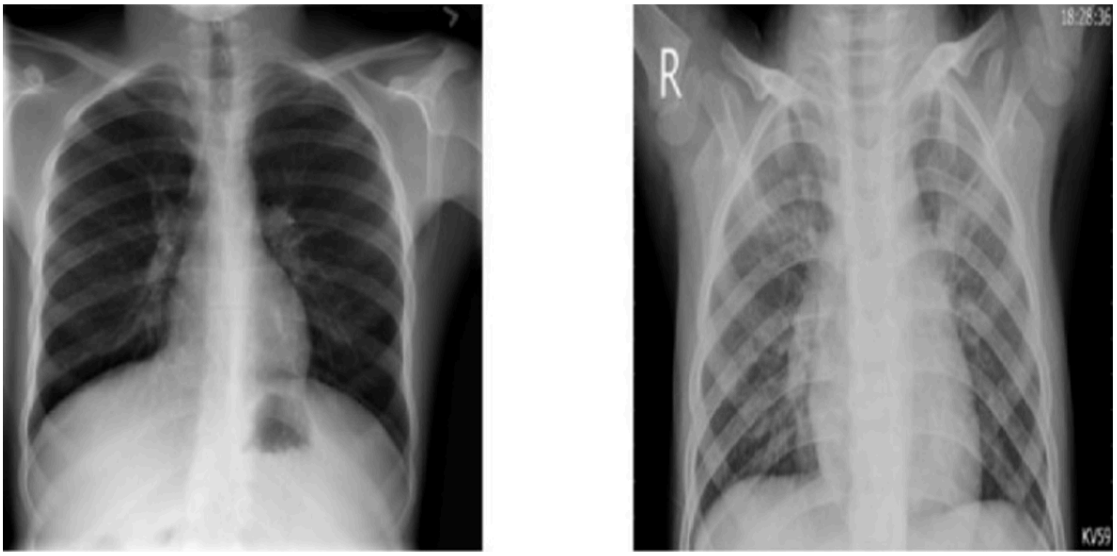


Figure (3). Sample image of normal (left picture) and viral (right picture) CXR

3.9. Image preprocessing steps

To ensure the reliability and reproducibility of our study, rigorous preprocessing steps were applied to the chest X-ray (CXR) images before training the machine learning models. 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 were implemented to ensure a more balanced distribution of

samples across different classes. Techniques such as random oversampling, SMOTE (Synthetic Minority Over-sampling Technique), or class-weighted loss functions were 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 our dataset and improve the robustness of our machine learning models for CXR classification, particularly in the context of COVID-19 diagnosis.

4. Results and Discussion

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

4.1. Evaluation Criteria

The confusion matrix, a pivotal metric in assessing machine learning algorithms [24], juxtaposes system outputs with reference data. Derived from it are accuracy, sensitivity, specificity, precision, recall, F-Measure, and G-Mean [25]. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are key statistical indices [25]. Figure 4 depicts a sample confusion matrix.

Real \ Prediction	True	False
True	TP	FN
False	FP	TN

Figure (4). 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 [24] (see Equation 5).

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

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

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

doi.org/10.32388/1NMNYE.2

The specificity of a classifier is measured by the ratio of true negative measures to all negative measures [24] (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 [26] (see Equation 8).

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

The words “recall” and “sensitivity” are interchangeable [27] (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 [26] (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 [26]. Despite the fact that negative situations are classified properly, a low G-Mean specifies poor performance in categorizing the positive data [26] (see Equation 11).

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

4.2. Analysis and Discussion

After preprocessing the data and employing classifiers through 10-fold cross-validation, the average results of 10 distinct test sets are shown in Table 1. In this study, employing a 10-fold validation or less may not necessarily lead to higher accuracy percentages. While fewer folds might yield slightly inflated accuracy rates due to the smaller validation sets, it could compromise the robustness of the model's performance assessment. A 10-fold cross-validation ensures more reliable estimates of the model's generalization ability by iteratively testing across diverse subsets of the data. This comprehensive evaluation approach helps mitigate overfitting and provides a more accurate representation of the model's true performance on unseen data, thus enhancing the study's credibility.

We conducted these experiments using Python 3.11.5 and Anaconda3 2023.03 on a personal computer equipped with an Intel® Core™ i5-11400H CPU running at 2.70GHz and 16.00 GB of RAM for all executions.

Related results with 60% training data samples, 10% validation and 30% testing data samples are represented in Table (1) and Figure (5).

60% train	KNN	SVM	DT	NB	TB	Ensemble learning
Accuracy	91.21	91.91	89.32	81.63	93.20	93.72
Sensitivity	94.62	96.27	93.42	79.72	97.66	98.14
Specificity	78.14	75.05	73.27	88.84	76.09	89.33
Precision	94.36	93.73	93.11	96.53	94.01	97.02
Recall	94.64	96.27	93.42	79.74	97.65	98.15
F-Measure	94.50	94.96	93.27	87.30	95.82	96.30
G-mean	86.01	84.97	82.70	84.18	86.21	86.73

Table (1). Results with 60% training data samples, 10% validation and 30% testing data samples

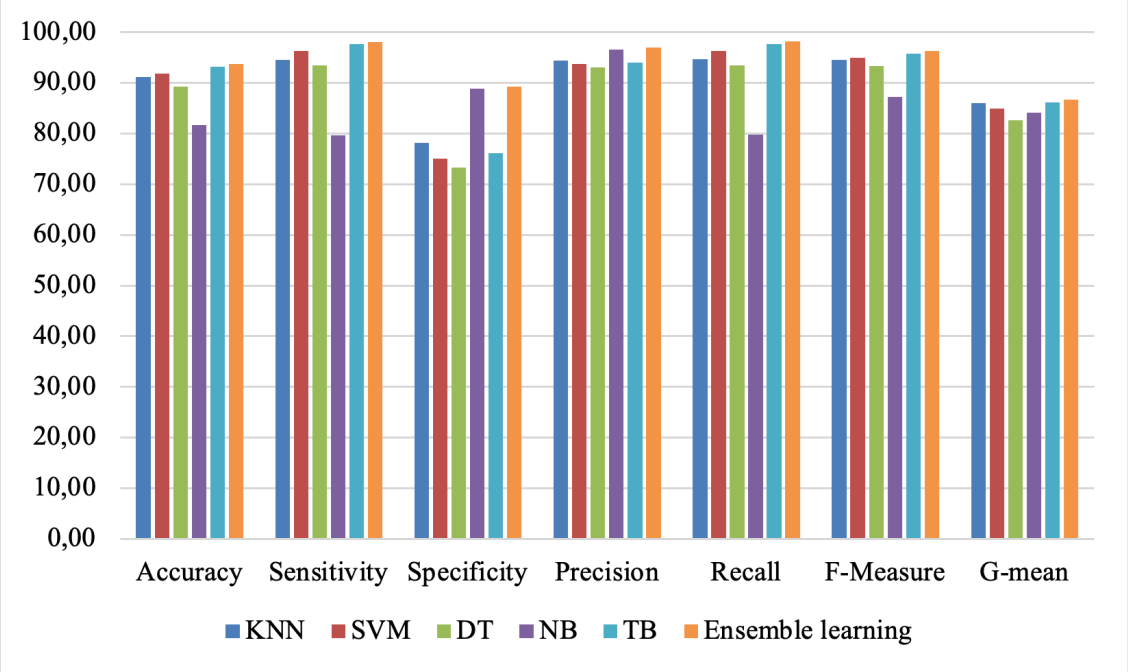


Figure (5). Bar plot of different classifier accuracies for 60% training data samples, 10% validation and 30% testing data sample

doi.org/10.32388/1NMNYE.2

Related results with 70% training data samples, 10% validation and 20% testing data samples are represented in Table (2) and Figure (6).

70% train	KNN	SVM	DT	NB	TB	Ensemble learning
Accuracy	91.63	92.01	89.62	81.70	93.39	93.88
Sensitivity	94.91	96.37	93.70	79.90	97.71	98.18
Specificity	78.98	75.25	73.81	88.72	76.71	89.22
Precision	94.58	93.79	93.24	96.46	94.21	97.00
Recall	94.92	96.36	93.68	79.88	97.68	98.19
F-Measure	94.72	95.03	93.45	87.43	95.92	96.40
G-mean	86.55	85.15	83.14	84.17	86.51	87.04

Table (2). Results with 70% training data samples, 10% validation and 20% testing data samples

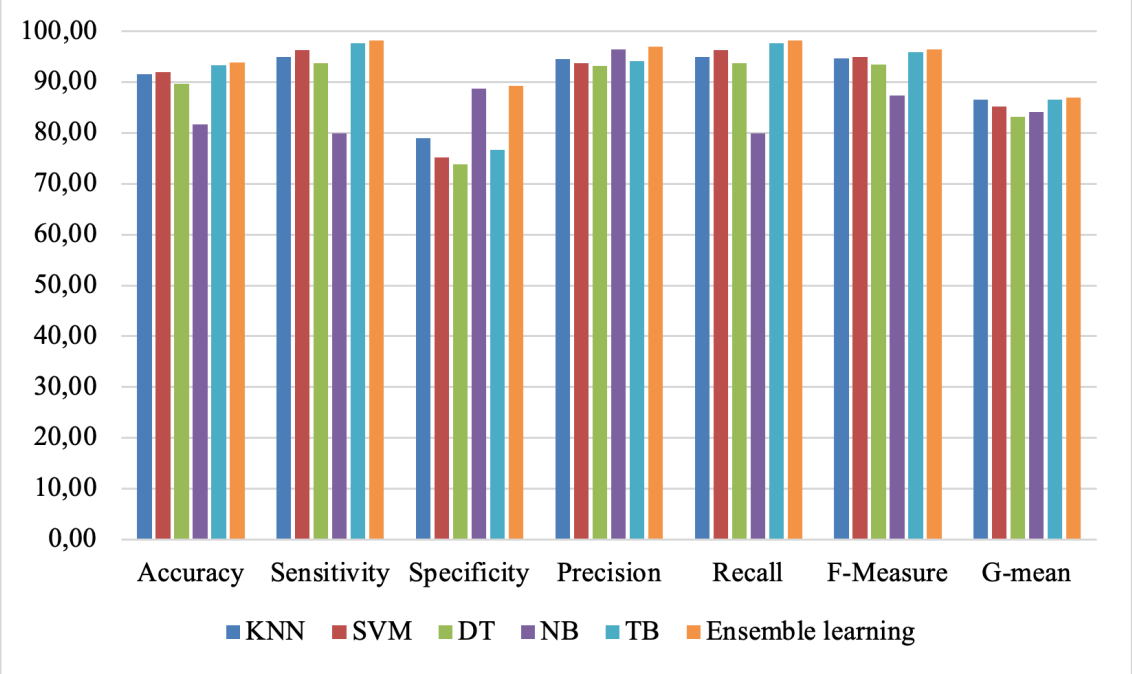


Figure (6). Bar plot of different classifier accuracies for 70% training data samples, 10% validation and 20% testing data samples

doi.org/10.32388/1NMNYE.2

Related results with 80% training data samples, 10% validation and 10% testing data samples are represented in Table (3) and Figure (7).

80% train	KNN	SVM	DT	NB	TB	Ensemble learning
Accuracy	91.43	91.92	89.22	81.87	93.53	94.07
Sensitivity	94.70	96.17	93.41	80.09	97.64	98.12
Specificity	78.91	75.70	73.08	88.74	77.78	89.27
Precision	94.53	93.84	93.04	96.51	94.47	96.96
Recall	94.68	96.14	93.41	80.10	97.64	98.12
F-Measure	94.61	94.98	93.20	87.51	96.01	96.52
G-mean	86.46	85.30	82.59	84.30	87.16	87.64

Table (3). Results with 80% training data samples, 10% validation and 10% testing data samples

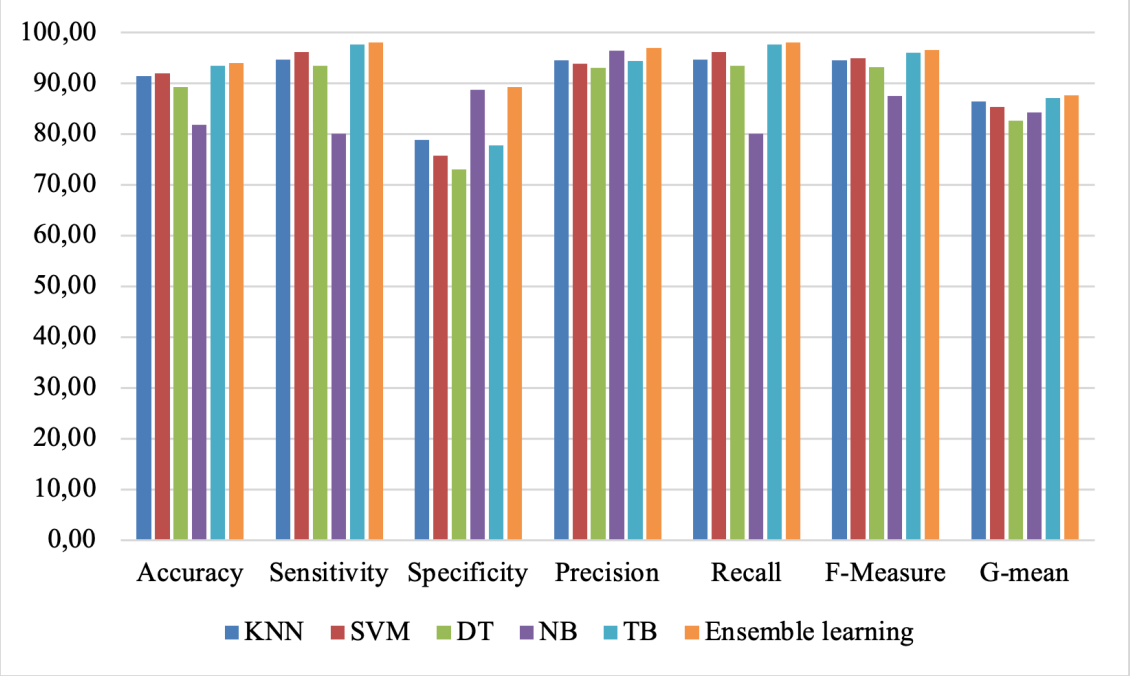


Figure (7). Bar plot of different classifier accuracies for 80% training data samples, 10% validation and 10% testing data samples

doi.org/10.32388/1NMNYE.2

Related results with 90% training data samples, 5% validation and 5% testing data samples are represented in Table (4) and Figure (8).

90% train	KNN	SVM	DT	NB	TB	Ensemble learning
Accuracy	91.66	91.96	89.54	81.97	93.96	94.43
Sensitivity	94.98	96.20	93.69	80.17	97.86	98.37
Specificity	78.87	75.52	73.39	88.99	78.69	89.46
Precision	94.57	93.85	93.20	96.60	94.67	97.11
Recall	94.98	96.19	93.72	80.15	97.90	98.36
F-Measure	94.80	95.01	93.42	87.59	96.22	96.74
G-mean	86.55	85.19	82.90	84.43	87.73	88.20

Table (4). Results with 90% training data samples, 5% validation and 5% testing data samples

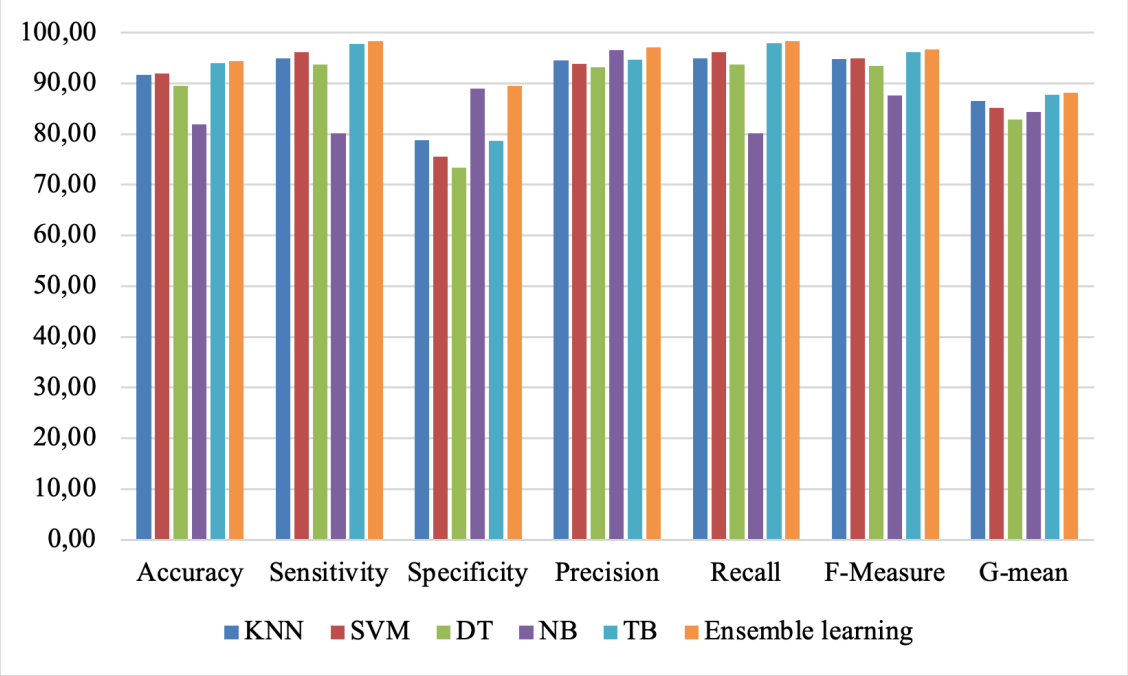


Figure (8). Bar plot of different classifier accuracies for 90% training data samples, 5% validation and 5% testing data samples

doi.org/10.32388/1NMNYE.2

It is shown in Tables (1-4) and Figures (5-8) that the proposed ensemble classifier has better performance among other classification methods. Figure 9 represents a radar chart to understand it better. The performances of the proposed ensemble learning classifier cover other performances in this chart and it is superior among other classifiers.

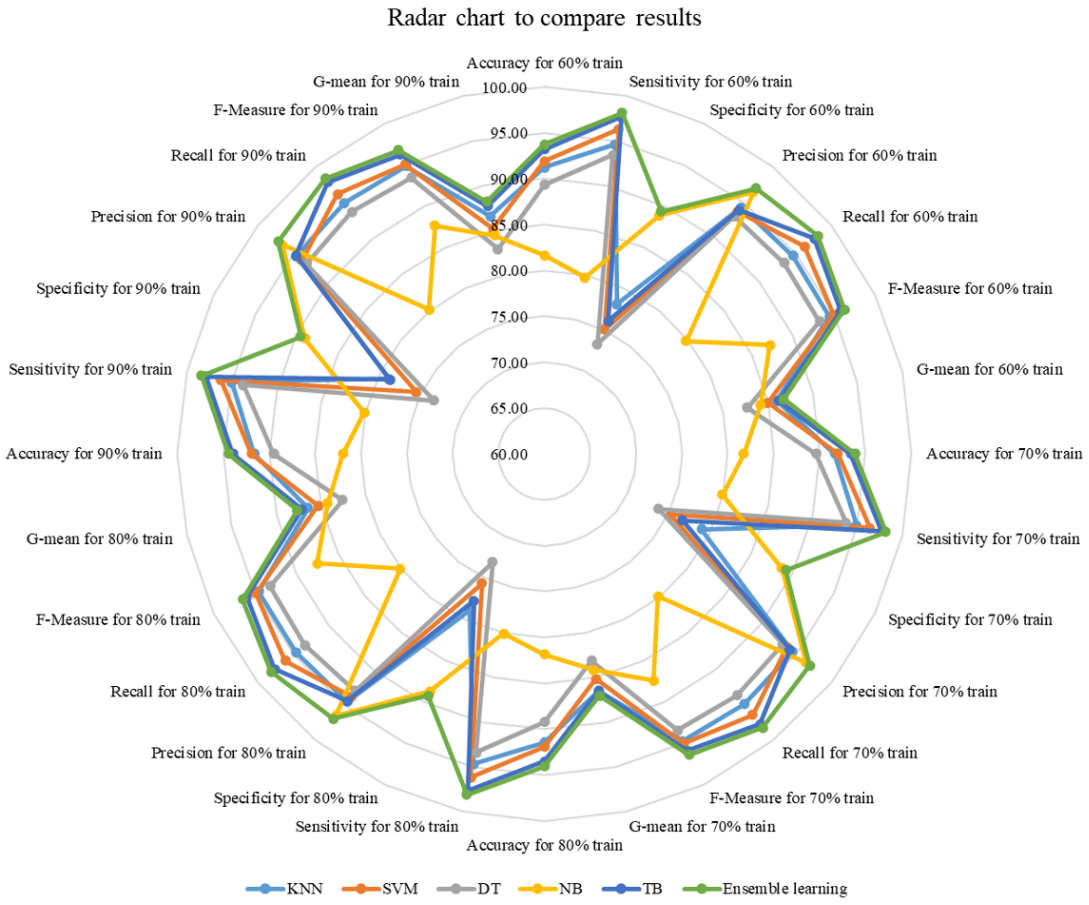


Figure (9). Radar chart of different metrics for various training/testing data samples

According to results, the ensemble learning method outperformed individual classifiers due to its ability to harness the collective wisdom of diverse models. By aggregating predictions from multiple classifiers, ensemble learning mitigates individual weaknesses and exploits complementary strengths, resulting in improved classification accuracy and robustness. This collaborative approach effectively leverages the diversity among classifiers, enhancing the model's ability to generalize and adapt to varying data characteristics, ultimately leading to superior performance in COVID-19 diagnosis from chest X-ray images.

doi.org/10.32388/1NMNYE.2

5. Conclusion, limitations and future works

This study delves into the realm of COVID-19 identification through the analysis of chest X-ray (CXR) images, employing various classification approaches. Key methods including K-nearest neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), and Tree Bagger (TB) demonstrated notable accuracies, all exceeding 90%. However, the standout performer proved to be ensemble learning, showcasing superior performance compared to the individual classifiers utilized. This suggests that the ensemble learning classifier, particularly when coupled with Histogram of Oriented Gradients (HOG) features extracted from CXR images, holds promise as a robust tool for COVID-19 identification.

The findings of this study hold significant implications for clinical practice and COVID-19 diagnosis. By demonstrating the effectiveness of ensemble learning techniques in CXR classification, particularly for COVID-19 detection, the study offers a promising avenue for enhancing diagnostic accuracy and efficiency. Implementing the proposed classifier in clinical settings could expedite the identification of COVID-19 cases based on early X-ray findings, enabling prompt clinical intervention and containment measures. Ultimately, the study's findings have the potential to improve patient outcomes by facilitating timely diagnosis and treatment in the ongoing battle against the COVID-19 pandemic. The proposed approach contributes to addressing the diagnostic challenges posed by the pandemic by offering a robust and automated method for COVID-19 diagnosis. By leveraging advanced machine learning techniques and ensemble learning, the classifier developed in this study demonstrates high accuracy in categorizing COVID-19 cases from chest X-ray images. This automated diagnosis not only expedites patient treatment but also ensures accurate disease detection, crucial for containing the spread of COVID-19 and mitigating its impact on public health systems.

The main strength of this study lies in its innovative approach to COVID-19 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.

5.1. Limitations

While our study demonstrates promising results in COVID-19 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 image acquisition protocols and population demographics. Moreover, the performance of the classifier may be influenced by the size and quality of the training dataset. Addressing these limitations could further enhance the robustness and applicability of our methodology in real-world clinical settings.

5.2. 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 COVID-19 diagnosis.

Moreover, the scope of this study extends beyond COVID-19 detection alone. There exists the potential to expand the capabilities of the existing model to not only ascertain the presence of COVID-19 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.

References

1. [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 images. Pattern Recognition Letters, 138, pp.638-643. https://doi.org/10.1016/j.patrec.2020.09.010.*
2. [a], [b] *Arman, S.E., Rahman, S. and Deowan, S.A., 2022. Covidxception-net: A bayesian optimization-based deep learning approach to diagnose covid-19 from x-ray images. SN Computer Science, 3(2), p.115. https://doi.org/10.1007/s42979-021-00980-3.*
3. [a], [b] *Apostolopoulos, I.D. and Mpesiana, T.A., 2020. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and engineering sciences in medicine, 43, pp.635-640. https://doi.org/10.1007/s13246-020-00865-4.*

4. ^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. https://doi.org/10.1109/SPIN52536.2021.9566114.
5. ^Das S, Mishra S, Senapati MR. New Approaches in Metaheuristics to Classify Medical Data Using Artificial Neural Network. Arab J Sci Eng.2020; 45, 2459–2471. https://doi.org/10.1007/s13369-019-04026-y.
6. ^Khan, M.A., 2021. An automated and fast system to identify COVID-19 from X-ray radiograph of the chest using image processing and machine learning. International journal of imaging systems and technology, 31 (2), pp.499-508. https://doi.org/10.1002/ima.22564.
7. ^Ko, B.C., Kim, S.H. and Nam, J.Y., 2011. X-ray image classification using random forests with local wavelet-based CS-local binary patterns. Journal of digital imaging, 24, pp.1141-1151. https://doi.org/10.1007/s10278-011-9380-3.
8. ^Hamed, A., Sobhy, A. and Nassar, H., 2021. Accurate classification of COVID-19 based on incomplete heterogeneous data using a K NN variant algorithm. Arabian Journal for Science and Engineering, 46, pp.8261-8272. https://doi.org/10.1007/s13369-020-05212-z.
9. ^Nayak, S.R., Nayak, J., Sinha, U., Arora, V., Ghosh, U. and Satapathy, S.C., 2023. An automated lightweight deep neural network for diagnosis of COVID-19 from chest X-ray images. Arabian journal for science and engineering, 48(8), pp.11085-11102. https://doi.org/10.1007/s13369-021-05956-2.
10. ^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), pp.11051-11083. https://doi.org/10.1007/s13369-021-05880-5.
11. ^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. https://doi.org/10.1155/2021/351482.
12. ^Nikolaou, V., Massaro, S., Fakhimi, M., Stergioulas, L. and Garn, W., 2021. COVID-19 diagnosis from chest x-rays: developing a simple, fast, and accurate neural network. Health information science and systems, 9, pp.1-11. https://doi.org/10.1007/s13755-021-00166-4.
13. ^Singh, D., Kumar, V., Vaishali and Kaur, M., 2020. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. European Journal of Clinical Microbiology & Infectious Diseases, 39, pp.1379-1389. https://doi.org/10.1007/s10096-020-03901-z.

14. ^Islam, M.T., Aowal, M.A., Minhaz, A.T. and Ashraf, K., 2017. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850. https://doi.org/10.48550/arXiv.1705.09850.

15. ^Taunk, K., De, S., Verma, S. and Swetapadma, A., 2019, May. A brief review of nearest neighbor algorithm fo r learning and classification. In 2019 international conference on intelligent computing and control systems (ICCS) (pp. 1255-1260). IEEE. https://doi.org/10.1109/ICCS45141.2019.9065747.

16. ^Wang, L., Lin, Z.Q. and 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), p.19549. https://doi.org/10.1038/s41598-020-76550-z.

17. ^Zhou, M., Zhang, X. and Qu, J., 2020. Coronavirus disease 2019 (COVID-19): a clinical update. Frontiers of m edicine, 14, pp.126-135. https://doi.org/10.1007/s11684-020-0767-8.

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

19. ^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), pp.1829-1857. https://doi.org/10.1002/ima.22905.

20. a, bMuhammad TR, Chowdhury EH, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, IslamKR, Khan MS, Iqbal A, Al-Emadi N, Reaz MBI. COVID-19 Chest X-Ray Database, 2020; Available at: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

21. a, bCohen, J.P., Morrison, P. and Dao, L., 2020. COVID-19 image data collection. arXiv preprint arXiv:2003.11597. https://doi.org/10.48550/arXiv.2003.11597.

22. a, bMonteral, J.C., 2020. COVID-Chestxray database. Available at: https://github.com/ieee8023/covid-chestxray-dataset.

23. a, b, cSoares, 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. https://doi.org/10.1101/2020.04.24.20078584.

24. a, b, c, dSareeta Mohanty, Manas Ranjan Senapati, 2024, Chest X-Ray Image Classification for COVID-19 Detection Using Various Feature Extraction Techniques, Published in: Intelligent Systems; Publisher: Springer Nature Singapore. https://www.springerprofessional.de/en/chest-x-ray-image-classification-for-covid-19-detection-using-va/26134002.

25. a, b, cRahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2021. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Computers in biology and medicine, 132, p.104319. https://doi.org/10.1016/j.compbiomed.2021.104319.

26. a, b, c, d, eKumar, 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), pp.27631-27655. https://doi.org/10.1007/s11042-022-12500-3.

27. 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), pp.248-258. https://doi.org/10.1080/00051144.2022.2140389.

28. ^MM Ershadi and Rise, Z.R., 2024. Uncertain SEIAR system dynamics modeling for improved community health management of respiratory virus diseases: A COVID-19 case study. Heliyon, 10(3). https://doi.org/10.1016/j.heliyon.2024.e24711.

29. ^Ershadi & Rise, Z.R. and Niaki, S.T.A., 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, p.106159. https://doi.org/10.1016/j.compbiomed.2022.106159.

30. ^Rahimi Rise, Z., Ershadi, M. and Ershadi, M.J., 2022. Multidisciplinary analysis of international environments based on impacts of Covid-19: state of art. International Journal of Industrial Engineering and Production Research, 33(1), pp.176-185. http://ijiepr.iust.ac.ir/article-1-1423-fa.html.

31. ^Rise, Z.R. and Ershadi, 2023. Fusing clinical and image data for detecting the severity level of hospitalized symptomatic COVID-19 patients using hierarchical model. Research on Biomedical Engineering, 39(1), pp.209-232. https://doi.org/10.1007/s42600-023-00268-w.

32. ^ZRahimi Rise, and MErshadi. (2022), "Socioeconomic analysis of infectious diseases based on different scenarios using uncertain SEIAR system dynamics with effective subsystems and ANFIS", Journal of Economic and Administrative Sciences. https://doi.org/10.1108/JEAS-07-2021-0124.

33. ^Ershadi, M. and Seifi, A., 2020. An efficient multi-classifier method for differential diagnosis. Intelligent Decision Technologies, 14(3), pp.337-347. https://doi.org/10.3233/IDT-190060.

34. ^Ershadi, M.M. and Ershadi, M.S., 2022. Logistic planning for pharmaceutical supply chain using multi-objective optimization model. International Journal of Pharmaceutical and Healthcare Marketing, 16(1), pp.75-100. https://doi.org/10.1108/IJPHM-01-2021-0004.

35. ^ *Rahimi Rise, Z., MM Ershadi and Shahabi Haghighgi, S.H., 2020. Scenario-based analysis about COVID-19 outbreak in Iran using systematic 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.*

36. ^ *Ershadi M and Seifi A, 2022. Applications of dynamic feature selection and clustering methods to medical diagnosis. Applied Soft Computing, 126, p.109293. https://doi.org/10.1016/j.asoc.2022.109293.*

37. ^ *Ershadi and Seifi, 2020. An efficient Bayesian network for differential diagnosis using experts' knowledge. International Journal of Intelligent Computing and Cybernetics, 13(1), pp.103-126. https://doi.org/10.1108/IJICC-10-2019-0112.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.