JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
SICAIET
2024

Research Article
Collection: Artificial Intelligence and Emerging Technologies

A Machine Learning-Based Classification of Chest Diseases Using Computed Tomography Images

Mubasher H. Malik[1*], Hamid Ghous[2], Maryem Ismail[3], Syed Ali Nawaz[4], and Hafiz Muhammad Ijaz[3]

[1]Vision, Linguistics and Machine Intelligence Research Lab, Multan, Pakistan.
[2]Department of Computer Science, Institute of Southern Punjab, Multan, Pakistan.
[3]Department of Information Technology, Institute of Southern Punjab, Multan, Pakistan.
[4]Department of Information Technology, The Islamia University of Bahawalpur (IUB), Bahawalpur, 63100, Pakistan.
[*]Corresponding Author: Mubasher H. Malik. Email: mubasher@isp.edu.pk

Academic Editor: Salman Qadri **Published:** February 01, 2024

Abstract: Chest diseases, particularly those caused by deadly viruses, substantially threaten human life worldwide. The coronavirus has emerged as one of the most widespread and lethal of these diseases. Medical professionals are lacking the accurate and timely detection of coronavirus. Artificial Intelligence (AI) provides such methods and techniques by using which accurate and timely detection of coronavirus possible. The primary goal of this research is to develop a system using Machine Learning (ML) methods which can detect coronavirus timely and accurately. Computed Tomography (CT) images of Lungs were taken. All images were processed using gray scale conversion, histogram equalization and image normalization pre-processing techniques. Region of Interest (ROI) based segmentation adopted to extracted features from pre-processed CT images dataset. Optimized features were collected using feature reduction method. ML classifiers such as Random Forest (RF), J.48, Naïve Bayes (NB), Bagging, Adaboost, Decision Tree (DT), SMO, Logitboost, Radom Tree (RT) and Bayes Net (BN) deployed to classify CT lungs images and produced 95.62%, 94.88%, 73.39%, 95.85%, 90.84%, 92.18%, 93.55%, 92.34%, 90.07% and 80.27% accuracy respectively. Experiments showed that Bagging attained highest accuracy of 95.85%. The proposed model can detect coronavirus from CT lungs images using ML methods more accurately and timely.

Keywords: Chest Disease; Machine Learning; Coronavirus; Covid-19; ROI segmentation.

1. Introduction

The coronavirus epidemic has placed the world's health at risk by spreading across many countries. According to research, those who have a history of diseases such as hypertension, diabetes, and coronary heart disease are at a higher risk of death from coronavirus. Furthermore, owing to the wide range of symptoms associated with coronavirus, it is critical to detect certain conditions and provide therapies in different medical situations (Sani, 2022). Coronavirus must be quickly identified by doctors to promptly separate the infected individuals from the general community and stop the spread of infections. Early in 2020, the diagnosis of coronavirus treatment will include nucleic acid testing and specific antibody detection. Since the initial screening indicates the window phase for antibody detection, it is ineffective. Nucleic acid testing primarily incorporates the ability to identify a bacterial species in blood or urine. However, the time required and poor identification rate are operating concerns (Bhargava, 2022) . When coronavirus started to spread, the Chinese authorities declared coronavirus exposure could be more precisely done with real-time reverse transcription polymerase chain reaction(RT-PCR) (Canayaz, 2022).

April 2020 The "Reverse Transcription - Polymerase Chain Reaction (RT-PCR)" test, a critical metric in identifying coronavirus-infected patients, requires obtaining RNA from sputum or a nasopharyngeal sample. (Mohammed, 2022). However, it takes roughly 10-15 hours to acquire the results. The rapid diagnostic test is a novel testing procedure (RDT) (Patel, 2022). However, the severe false-negative findings of RT-PCR and RDT testing raised concerns about their accuracy. Although RT-PCR produces good results, handling the sample of coronavirus patients requires highly trained individuals (Patel, 2022). The coronavirus test is used to identify viruses or antibodies. The diagnosis of coronavirus is done using two primary techniques. The first is laboratory-based techniques such as serology, antigen, and nucleic acid testing. The second technique involves lung imaging diagnostic technologies such as X-rays and CT scans (Canayaz, 2022).

Biomedical imaging has proven to be an excellent method for making non-invasive diagnoses of various illnesses. Medical imaging was developed as a result of the discovery of X-rays. The discovery of X-rays was the impetus for the development of medical imaging (Patel, 2022). Since coronavirus can emerge on X-ray and tomography images of the lungs like many diseases, such as pneumonia, a convincing coronavirus result may not be obtained based solely on observations from lung imaging without clinical diagnosis (Canayaz, 2022).

It is also discovered that X-ray offers more promising outcomes for coronavirus detection. However, CT imaging is preferable because it provides more information than X-rays (Patel, 2022). Regarding the efficacy of lung CT scans in diagnosing coronavirus, they are utilized as a priority tool in the clinical process when the tests' findings are analyzed (Canayaz, 2022). However, integrating distinctive characteristics seen in CT scan images has the potential to be used for systematically classifying the severity of lung problems in coronavirus patients (Ghashghaei, 2022). Several radiographic techniques, such as CT, diagnose lung aberrations, particularly in the initial stages (Farahat, 2022). The CT scans of the affected individuals illustrate the unique features of the coronavirus condition. Therefore, clinical professionals require lung CT imaging to make an early coronavirus diagnosis. Medical applications such as improving image quality, organ segmentation, and organ texture classification are supported by the development of computer vision systems. Some applications of ML in the field of biomedical image processing include the study of time series and tumor features (2), segmentation, and identification (3) of tumor modules (Barstugan, 2022). Several models based on ML with ensemble methods have been presented about CT images (Patel, 2022).

2. Literature Review

Many researchers contribute a lot to coronavirus research. Section 2 depicted the contribution of researchers to the classification, segmentation, and prediction of coronavirus. It showed different methods, techniques, and algorithms. The following section shows contributions, particularly ML and DL-based models.

2.1. Machine Learning Methods

In 2022, ML techniques were used to identify coronavirus in CT lungs using computer Aided diagnosis (CAD) system that produced and used different classification methods: DT, Support Vector Machine, K-means clustering, and Radial Basis Function. (Shahin, 2022). In 2022, an ML model proposed a novel technique that detects coronavirus from lung CT images using feature classification SVM (RBF) and feature extraction LBP and VAR. The proposed model achieves 93.47% accuracy (Patel, 2022). In 2022, an ML Hopefield neural network model will be used to find coronavirus symptoms from chest CT --scan images. The sensitivity and specificity of the model for detecting coronavirus were 97.4% and 98.6% from chest CT-scan images (Sani, 2022). In 2022, a proposed hybrid Method that integrates deep transfer learning and ML was introduced to classify coronavirus from CT images.Res-NET model with SVM and KNN classifier produced 97% accuracy (Al-jumaili, 2022). In 2022, an ML classification was used to detect coronavirus from chest CT images.GLCM features were classified on SVM and Random forest classifiers, producing 1% and 94% accuracy (Suguna, 2022). In 2022, an automated ML algorithm was proposed to detect coronavirus from CT images. Images were preprocessed by normalization; secondly, image segmentation was done by fuzzy-C mean clustering. Different classifiers were on images, but SVM produced 99.14% accuracy (Bhargava, 2022). In 2022, ML algorithms' BO-based approach diagnosed coronavirus from CT images. The BO approach had higher results, as 99.37% with KNN's higher performance. (Canayaz, 2022). In 2022, an ML classifier was tested and trained on lung CT images using two Matrices: Accuracy and Kappa. The Markov–Gibbs Random Field (MGRF) model was tested and trained on liaisons. It achieved 100% accuracy and 100% Kappa. (Farahat, 2022). In 2022, ML, DL, and AI techniques were used to diagnose and detect coronavirus from Lung CT images. CNN DL approach achieved a higher accuracy of 98.40%. (Butuner, 2022). In 2022, an ML method was used to detect coronavirus from lung CT images. The best classification accuracy was obtained as 99.68% with 10-fold cross-validation and the GLSZM feature extraction method (Barstugan, 2022). In 2020, the Proposed ML model implemented a sequence of Methods such as multi-thresholding, image separation using threshold filter, feature-extraction, feature-selection, feature-fusion, and Classification on CT scan slice to detect coronavirus. SVM with FFV (fused feature vector) helped to attain 89.90% accuracy. (Kadry, 2020). In 2020, an ML model was developed to detect coronaviruses by combining radiological outcomes with clinical biomedical indexes. Different validation models were applied to the dataset; the prediction model yielded cross-validation area AUROC, which achieved 93% accuracy. (Li, 2020). In 2020, different ML techniques were used to distinguish coronavirus lung CT scan images from another infected disease patient. In the proposed method, the Res-Net classification Model achieved 91% accuracy. (Sharma, 2020). In 2021, an ML model was used to detect the severity of coronavirus using lung CT scan images. Alex-net, Dense-Net-201, and ResNet-50 classification models were used for feature extraction. The SVM classification model achieved an overall 90% accuracy. (Aswathy, 2022). In 2022, ML technique hybrid CNN methods were used to classify the lung CT images.Alex-Net-SVM method achieved 97.56% accuracy from CT scan images (Kalayc{\i}, 2022). In 2021, the ML-based method proposed automatic coronavirus detection from CT scan segmented images, and in this process, three basic steps involved image preprocessing, segmentation, and classification. KNN classifier achieved 95.2% higher accuracy from mixed dataset CT images (Sayg{\i}l{\i}, 2022). In 2022, the radiomic method was used to classify the stages of coronavirus from lung CT images. The comparison of the RF and DT Radiomic models achieved 93.55% accuracy. (Mehrpouyan, 2022). In 2022, ML proposed we propose CoviNet, a deep three-dimensional convolutional neural network (3D-CNN), to diagnose coronavirus from CT images. Trained proposed model on two publicly available data sets achieved an accuracy of 75.00% and 94.12% (Mittal, 2021). In 2021, ML applications for COVID-19 diagnosis, detection, and the assessment of disease severity based on medical imaging have gained considerable attention. ML papers achieved remarkable predictive results in the detection of covid 19. (Rehouma, 2021). In 2021, an ML involved in detecting coronavirus from CT images .supervised learning showed better results than unsupervised learning. Supervised learning acquires 92.9% higher accuracy. (Kwekha-Rashid, 2021). In 2021, ML radiomic-based features were extracted from CGOs, and HRCT images were used to detect the coronavirus. In the early stages. The proposed method AUC acquires 86% higher accuracy. (Delli Pizzi, 2021).

2.2. Deep Learning (DL) Methods

In 2022, a classification model was proposed to classify coronavirus from lung CT images. Images were preprocessed and classified using Res-Net and VGG-16 models. The model produced 98%. Accuracy (Salama, 20). In 2021, a proposed ML-based method that differentiates coronavirus symptoms automatically from lung CT scans using radiomic features. The proposed ML method achieves 98% accuracy (velichko, 2022). In 2022, a DL-based model performed lung segmentation automatically using four feature selection algorithms: Analysis of Variance (ANOVA), Kruskal-Wallis (KW), Recursive Feature Elimination (RFE), and Relief and used seven classifiers Logistic Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Linear Discriminant Analysis (LDA), RF, AdaBoost, NB and Multilayer Perceptron (MLP). The proposed Model combination of ANOVA features selection at 81% and RF classifier at 72% acquires higher accuracy (Shiri, 2022). In 2021, a novel ensemble approach proposed various pre-trained models such as VGG16, VGG19, InceptionV3, ResNet50, ResNet50V2, InceptionResNetV2, Exception, and Mobile Net and fine-tune them using Lung CT Scan. Using a novel ensemble classifier VGG with 16 layers achieved 92% higher accuracy (Shaik, 2022). In 2022, a DL algorithm developed a novel Android application that detects coronavirus from CT chest images. Neural network pre-trained model classification accuracy achieved 99.58% (Verma, 2022). In 2022, a Blockchain DL model diagnosis of coronavirus from chest CT images produced a higher classification accuracy of 99.76%. (Heidari, 2022). In 2021, a conventional neural network proposed concepts of wavelet transformation for preprocessing to detect coronavirus from chest CT images. The proposed methodology produced 85% classification accuracy (Gaur, 2022). In 2022, a proposed DL model, EfficientNet-B1, was applied for classification to detect coronavirus from chest CT images. The proposed model consists of 16 mobile inverted bottleneck convolutions (MBConv), two convolutional (Conv) layers, one global average-pooling layer, and one fully connected layer, obtained a test accuracy of 94.24% (Liu, 2022). In 2022, hybrid DL-integrated CNN and SVM approaches were used to identify coronavirus from chest CT images. The proposed model CNN SoftMax,CNN SVM produced (98%) .(99.1%) accuracy after preprocessing, feature extraction, and classification (Mohammed, 2022). In 2022, a coronavirus was detected from lung CT images in emergencies, and eight DL models were used. Observed Mobile -Net V2 provides the highest accuracy at 99.12%. (Shamrat, 2022). In 2022, a proposed adversarial deep domain adaption-based approach for diagnosing coronavirus from lung CT scan images. ADA coronavirus achieved a significant classification improvement of 60%. Res-Net classification model produced 99.96% accuracy. (Aria, 2022) In 2022, a DL technique utilizes CNN, stack autoencoder, and deep neural network to develop a coronavirus diagnostic system for CT images. After the Classification of CT images, the proposed CNN model produced 88.30% Accuracy. (Abdulkareem, 2022). In 2022, a DL method was proposed for detecting coronavirus from CT images. The proposed CNN architecture VGG-16 gives better accuracy at 97.68%. (Kogilavani, 2022). In 2022, a DL technique to detect coronavirus from CT images using U-NET architecture based on CNN encoder and decoder produced 98% accuracy after precise image segmentation, fourfold cross-validation, and soft-max layer classification. (Mahmoudi, 2022). In 2021, a DL and ML technique was proposed to classify coronavirus from grey-scale CT images. CNN achieved 96.49% accuracy, and the algorithm involving just five grey scale image attributes with 99.81% accuracy was successfully used to distinguish the severity of coronavirus lung damaged (Ghashghaei, 2022). In 2022, a DL model Cimatec-CovNet-19 neural network was introduced inspired by VGG-16 architecture to support coronavirus identification from chest CT- images. After novel preprocessing on CT images, the ROC-AUC was 95 % accurate, and the PR-AUC was 95% accurate. (Furtado, 2022). In 2021, regional DL used a deep convolutional neural network to identify specific lung-infected areas of coronavirus. Fourfold cross-validation was applied to lung CT images, and specificity and Jaccard produced the best results on Coronal and Sagittal images. (Ahmed, 2022). In 2021, a] proposed a CNN model that clustered lung images using the K-mean Method. Our proposed method outperformed the classification of defective and non-defected lungs using CT images (Afshar, 2021). In 2021, a Novel CNN model extracted features from lung CT images. This extracted feature is then deployed on different ML algorithms: SVM, GNB, RF, LR, and DT. Our proposed model outperformed and, respectively, 100% accuracy. (Islam, 2022).

So, there is a need for an automatic system for the classification of coronavirus from CT images using ML algorithms. Section 3 shows the proposed methodology

3. Materials and Methods

The aim of this study is to produce a system that will detect and classify CT-lungs scanning and clinical specimens of coronavirus using ML methods. This method is used for early infectious disease detection. The proposed frame is divided into four phases: 1) The first phase is preprocessing, which begins use by converting CT images into grey-level CT images of equal size (256×256) (Mohammed, 2022). Then, contrast-limited adaptive histogram equalization (CLAHE) is used to enhance the channel contrast and consistency of pixel strength (Patel, 2022). Wavelet is a better technique for noise removal since it uses the threshold value at every decomposition phase. The second phase, the Feature Extraction GLCM, Histogram, Run length, Auto aggressive, wavelength transformation technique was applied as a deep feature extraction technique to identify CT samples as being either Covid or Non-Covid 3)Third phase is Feature Reduction, which is used for large dimension features data that can affect the process and accuracy of lungs classification 4) In the fourth phase RF, RT, J.48, NB, BN, DT, Ada Boost classifiers are employed to categories Patients with Covid pneumonia The data sets used for experimentation are collected from different available online resources.

3.1. Data Acquisition

The computer-extracted image features were identified with different prognoses for suspicious nodules on the lungs CT scan (velichko, 2022). To reduce motion artifacts during breath-hold, Image acquisition was performed (Shiri, 2022). Our proposed model is an essential task collection of Lungs Ct images from the dataset. Data had two categories 1)covid 2) non covid .We incorporate three scans of 100 patients' lung CT images. Each slice of The CT scan images has dimensions of 512 × 512. The data set contains 300 × 300 = 600 covid and non-covid images converted into a 24-bit BMP format to obtain quality results

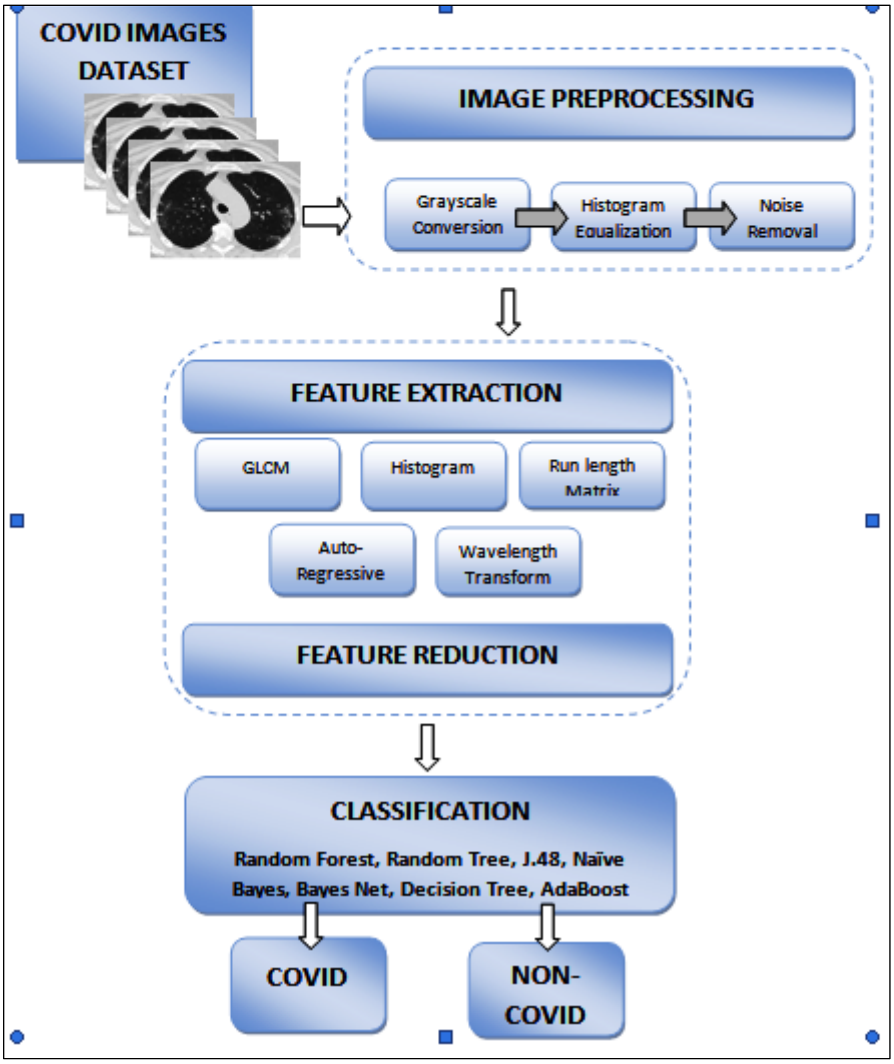

Figure 1. Chest Disease Classification Model

3.2. Image Preprocessing

Preprocessing images is essential for obtaining better results. For improving medical imaging, numerous techniques have been developed so far (Islam, 2022) . The coronavirus classification system employs a variety of foundational steps called "operations of preprocessing" to prepare data for the subsequent stage (the feature extraction stage) (Mohammed, 2022). Lung CT images have undergone additional preprocessing procedures to enhance their quality for more accurate diagnostic results. This step is important because the system cannot diagnose accurately due to several characteristics of the lungs (velichko, 2022).

Typically, non-linear digital cameras produce RGB-colored images, making extracting brightness information from color impossible. For this objective, The processing time is minimized using grayscale images transformed from RGB images. The conversion was performed by using RGB (average values) intensity value. Giant is the intensity value of grayscale, R is linear red, G is linear green, and B is linear blue (Bhargava, 2022).

$$\mathbf{G_{int}} = \frac{1}{3}\,(R + G + B) \tag{1}$$

A discrete function called a histogram counts the total number of pixels at various intensities. As a result, we can leverage it to read the data in the image (Chen Y.-Y. a.-L.-C.-H., 2021). Histogram equalization, often referred to as histogram flattening, is a non-linear stretching of images and redistribution of image pixel values so that the quantity of pixel values in a specific grey range is about equal. By elevating the peak portion's contrast and lowering the valley portions' contrast on both sides, the image histogram is made to appear relatively flat. In other words, the fundamental premise of histogram equalizations is that each grey level should appear with the same frequency, resulting in an evenly distributed probability for each grey level and a flat histogram. It also becomes evident in the image (Xiong, 2021).

$$cdf(x) = \sum_{k=-\infty}^{x} p\,(\mathrm{k}) \tag{2}$$

Although the low-light image's content information is revealed after being decomposed into a reflectance map, its noise substantially compromises its quality (Zhang F. a., 2021). Filtering techniques have gained much traction over the years among contemporary image-enhancing approaches. It is suitable for addressing noise removal and edge enhancement problems (Ullah, 2022) . we implemented wavelet filters, such as the gobar filter for Noise removal and image quality, which have been shown to improve the accuracy of a classification model. The grayscale, histogram Equalization, and Noise removal combine the dataset's grading (covid/Non-Covid).

3.3. Feature Extraction

The essential feature of the classification is feature engineering (Islam, 2022) in feature engineering, feature extraction cannot be done without the ROI process; in this technique, the moving window to the segment ROI is achieved using the RGB color. This concentrates primarily on where the moving object is and the implications of the moving object being rejected. The colors red, blue, and green indicate the system's detection. These hues could designate the regions under the coronavirus's influence (Shahin, 2022) For the CT images data set, five types of images were extracted. Grey level co-occurrence matrix, histogram, run length, autoregression, wavelet transformation.

3.3.1.Grey Level Co-occurrence Matrix (GLCM)

On the premise of the characteristics of grey-level oscillations in neighboring pixels, the GLCM feature extraction approach was chosen. Second-order statistical features, the most common feature extraction algorithm, can be extracted from images using this technique. (Mehrpouyan, 2022). To represent a co-occurrence matrix from an I image, use P=[p(i, j | d, Θ)]. The co-occurrence matrix compares the ith pixel frequency features to the jth neighbor pixel frequency features, taking into account the Θ direction and d length. This study chose d = 1. So, the Θ angle is 0° (Barstugan, 2022). Neighbor pixels can take a variety of types, but the following are the most typical ones: Single-pixel, double-pixel, and three-pixel steps, and consideration of 0-degree, 45-degree, 90-degree, and 135-degree neighborhoods. The GLCM was used to compute the second-order statistical features, which included autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares variance, sum average, intron variance, difference entropy, information measure of correlation 1, inverse difference (INV), and sum variance (Kim, 2021).

3.3.2.Histogram Features

A histogram depicting frequency is displayed. In image processing, histograms show the pixel intensity levels. For instance, the y-axis of a histogram declares the frequency of such intensities, whereas the x-axis proclaims the grey-level intensities. (Doken, 2021).

$$\mathrm{E}dgek \quad = \begin{cases} min\left(I_{inj} - n\right), \quad if^{k} = 0 \\ Edge_{k} - 1 + \Delta w, \dot{l}fk = 1,2,\ldots\ldots,\mathrm{m} \end{cases} \tag{3}$$

Where Dw = (Imax Imin)/m is the equispaced bin width, Imax and Imin, respectively, denote the maximum and minimum of Iinj_n, and Iinj_n indicates the luminance channel following detail injection. The parameter m controls the number of quantification levels in the histogram. A higher m number suggests that more intensity levels are used to produce a high-quality image. A lower m number, on the other hand, indicates a shorter computing time. The value of m was empirically chosen as 60 while considering the trade-off between time and quality (Chen Y.-Y. a.-L.-C.-H., 2021)

3.3.3.Wavelet Transformation Features

Histogram equalization improves the entire image, as seen in the example provided in the preceding subsection. However, using only equalization, a corrosion image with different scales and details cannot be effectively enhanced. For multiresolution signal analysis, wavelet transformations with good localization characteristics in both the time-spatial and frequency domains can be utilized (Xiong, 2021). Gobar wavelets compose a perfect filter for both orientation and spatial localization; a two-dimensional GWT can be stated by the convolution of the image I(a, b) as shown in Equation (Ertu{\u{g}}rul).

$$j(a, b)=\int \int I(a', b')g(a-a', b-b')da'\, db' \quad (4)$$

3.3.4.Auto Regressive Features

An AR model is simply a linear regression of the current observation of the series against one or more prior observations. The AR model is usually expressed as:

$$y(t)\bar{r}\sum_{i=1}^{p}\phi i^{y}(t-i)+\varepsilon_t \quad (5)$$

In the given context, ϵt denotes white noise that is not influenced by preceding data points, y(t) represents the time series to be modeled, ji signifies the model coefficients, and parameter p signifies the order of the AR model. An AR model of order p, represented by AR (p), implies that the current observation depends on p's previous observations (Zhang Y. a., 2017). The AR model can describe and model the characteristics and information inside a sign. (Hatamikia, 2014).

$$x(t)=-\sum_{i=1}^{p}a_i x\,(t-i) \quad (6)$$

3.3.5.Run-Length Features

Galloway proposed using a run-length matrix to extract texture features. A run-length matrix p(i; j) is defined as the number of runs with pixels of gray level I and run length j (Tang, 1998). A run length is defined as the number of contiguous pixels with the same grey intensity in a specific direction (Kairuddin, 2017) . Many numerical texture measurements can be obtained from the original run-length matrix, p(i;j). (Tang, 1998).

3.4. Feature Selection

The process of feature selection involves the identification of a subset of features that are pertinent to a specific application. It enhances classification accuracy by identifying the most optimal feature subset from the initial set of fixed features based on a specified criterion for feature evaluation. Classification accuracy is improved while data dimensionalities and computation time are decreased through optimized feature selection. A subset of all features must be chosen from the total number of features in the feature selection issue by a predetermined optimization criterion.

3.5. Classification

In the classification phase, training of the extracted features and performing the classification process are included (Sayg{\i}l{\i}, 2022). The classification step includes accuracy and comparing each model's accuracies to determine the optimal model for real-time coronavirus detection and classification. Most studies emphasize binary classification. Various authors employed various evaluation techniques to determine the severity of the lung infection. (Aswathy, 2022) . ML is one of the most promising tools in Classification (Kwekha-Rashid, 2021).In our proposed work. After feature engineering and selection, the classification uses J.48, RF, RT, NB, BN, DT, Bagging, Ada Boost, SMO, and Logitboost. The AdaBoost algorithm is an ensemble learning algorithm that addresses the limitations of individual sub-classifiers by utilizing many sub-classifiers to perform better classification. The fundamental concept entails training distinct sub-classifiers ht, t ∈ [1, 2, • •, T} using identical training samples (Chen S. a.-J., 2019). The logitBoost algorithm is generally used to best fit the linear logistic regression functions at the tree node (Pourghasemi, 2018). Bootstrap aggregating, often abbreviated as bagging, involves having each model in the ensemble vote with equal weight. It solves the over-fitting problems by building the Logit Boost algorithm. For example, the random forest algorithm combines random decision trees with bagging to achieve classification accuracy (Kabari, 2019). J.48 Algorithm is widely used in medical data analysis. Many studies have previously used the algorithm to predict disease using symptoms (Anjum, 2022). A supervised learning algorithm is called RF. It takes a set of DTs, usually trained using the "bagging" technique, and turns them into a "forest." The basic idea behind the bagging approach is to improve the result by mixing many learning models (Islam, 2022). The Bayes theorem is used to classify data, and the NB algorithm is based on it. The Bayes theorem allows you to compute the posterior probability of P(xy) by using p(x), p(y), and P(yx), as shown in equation 1 (Shahin, 2022)

$$P\left(\frac{x}{y}\right)=\frac{P\left(\frac{x}{x}\right)P(x)}{P(y)} \quad (7)$$

4. Results

In this work, we carried out a large-scale experiment with a dataset of COVID chest pictures to examine the effectiveness of several feature extraction and segmentation strategies combined with different classifiers for precise classification. Preprocessing techniques were used for the images, and techniques such as the Gray Level Co-occurrence Matrix (GLCM), Histogram, Run-Length Matrix, Auto-regressive, and wavelength features were used to extract features. A ROI-based segmentation method was used to improve the analysis even more. This produced four non-overlapping ROIs with the following dimensions: 13 x 13, 15 x 15, 17 x 17, and 19 x 19. Table 1 shows the experimental results produced by ML Classifiers.

The experimental results demonstrated differential performance metrics for each approach when different classifiers were applied to the COVID chest image categorization job. J.48 demonstrated its usefulness in decision tree-based categorization with an impressive accuracy of 94.88%. Popular ensemble technique Random Forest proved to be a strong performer, obtaining an accuracy of 95.62%, with RT coming in second at 90.07%. These findings highlight how well decision tree-based algorithms capture complex patterns in the dataset.

A probabilistic classifier called NB showed a somewhat lower accuracy of 73.39%, suggesting possible drawbacks in utilizing a primary probabilistic method to explain the intricate relationships in the COVID chest images. With modest accuracy rates of 80.27% and 92.18%, respectively, BayesNet and Decision Table demonstrated their capacity to reconcile interpretability and model complexity.

It was found that the ensemble approaches worked very well for improving classification performance. With an astounding accuracy of 95.85%, bagging was the classifier that performed the best out of all of them. This demonstrates how effective it is to combine predictions from several models to reduce overfitting and enhance generalization. With an accuracy of 95.62%, Random Forest—an additional ensemble technique—closely trailed, demonstrating how well ensemble approaches handle the complexities of the COVID chest image classification task.

Additionally, competitive accuracy was shown by AdaBoost, SMO, and Logitboost, with respective scores of 90.84%, 93.55%, and 92.34%. Together, our findings demonstrate the adaptability of several classifiers in tackling the difficulties presented by the COVID chest imaging collection.

Table 1. Experimental Results Showing Accuracy of Different ML Classifiers

Classifier	Accuracy	TP	FP	Precision	Recall	F-Measure	MCC	ROC	PRC
J.48	94.88%	0.949	0.051	0.949	0.949	0.949	0.898	0.947	0.925
Random forest	95.62%	0.956	0.044	0.956	0.956	0.956	0.913	0.992	0.992
Random tree	90.07%	0.901	0.099	0.901	0.901	0.901	0.801	0.901	0.861
Naïve Bayes	73.39%	0.734	0.267	0.755	0.734	0.728	0.488	0.823	0.774
BayesNet	80.27%	0.803	0.197	0.804	0.803	0.802	0.607	0.879	0.844
Decision Table	92.18%	0.922	0.078	0.922	0.922	0.922	0.844	0.976	0.975
Bagging	95.85%	0.959	0.041	0.959	0.959	0.959	0.917	0.992	0.992
AdaBoost	90.84%	0.908	0.092	0.909	0.908	0.908	0.818	0.972	0.972
SMO	93.55%	0.935	0.065	0.936	0.935	0.935	0.872	0.935	0.908
Logitboost	92.34%	0.923	0.077	0.924	0.923	0.923	0.847	0.979	0.979

The graphical representation of the results produced by ML classifiers is shown in Figure 2 below.

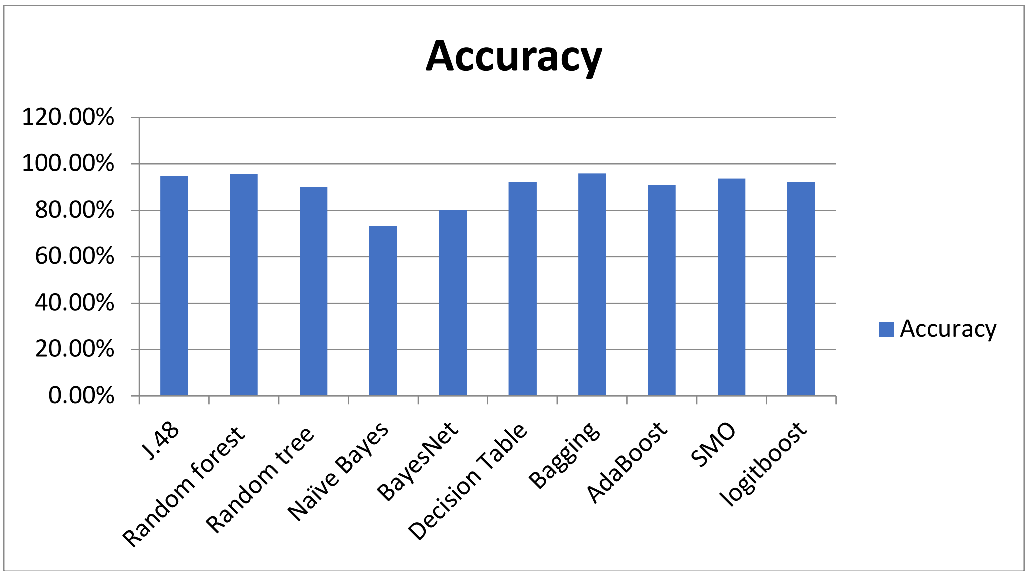

Figure 2. Graphical Representation of ML Classifiers Results

5. Discussion

The experiment results comprehensively analyze the classification performance on a collection of COVID chest imaging datasets. The dataset was preprocessed with great attention to detail, applying preprocessing methods to the images to improve pertinent information. After that, a wide range of features was retrieved, such as wavelength, run-length matrix, histogram, auto-regressive, and gray-level co-occurrence matrix (GLCM). This extensive feature set was created to capture various underlying visual features associated with coronavirus manifestations.

A segmentation technique based on ROI was used to hone the findings further. Strategically, four non-overlapping ROIs of different sizes—13x13, 15x15, 17x17, and 19x19—were constructed. This segmentation technique made it possible to examine several spatial scales within the images and gain a more complex understanding of the underlying patterns linked to coronavirus pathology.

Many different models were considered during the classifier selection process, and each was selected based on its unique ability to contribute to the image categorization problem. Using decision trees' natural capacity to capture intricate decision boundaries, decision tree-based models like J.48, RF, and RT were incorporated. Since probabilistic models work well with some kinds of data distributions, they were also used, such as NB and BayesNet.

The experiment strongly emphasized ensemble approaches, which are recognized for their capacity to improve predictive performance. The combined power of numerous base classifiers was utilized by applying Bagging, AdaBoost, and Logitboost. The experiments showed that ML methods such as RF, RT, DT, Bagging, AdaBoost, J.48, SMO and Logitboost produced highest auccuracy of 95.62%, 90.07%,92.18%, 95.85%, 90.84%, 94.88%, 93.55% and 92.34% respectively. While NB and BN produced poor frequency of 73.39% and 80.27% respectively. Among all Bagging produced higher accuracy of 95.85%.

The proposed model provided the higher accuracy to detect coronavirus using CT images of Lungs. Firstly, pre-processed images were used to extract features using ROI based segmentation approach. These features were optimized using feature reduction method. Finally ML classifiers deployed to classify images using optimized features list into two classes namely covid and non-covid. The proposed approach helps medical practitioners' to diagnose coronavirus efficiently and effectively.

6. Conclusions

The proposed model produced promising results using ML classifiers. This research focused on detection of coronavirus using CT lungs images. These images were processed using Image Procdessing techniques such as grayscale conversion, histogram equalization and image normalization for improved quality. ROI segmentation makes the detection of coronavirus more accurately. Bagging provides higher accuracy as compare to other classifiers. Most of the ML classifiers used during experiments produced good accuracy above 90%. This showed that proposed model can accurately and timely detect coronavirus using ML classifiers. In Future, this work could be extended using Lungs MRI images. The experiments could be deployed using Deep Learning (DL) algorithms. The combination of ML and DL algorithms on both CT and MRI lungs images can improve the diagnosis and detection of all variants of coronavirus.

References

1. Abdulkareem, K. H.-Q.-W. (2022). Automated System for Identifying COVID-19 Infections in Computed Tomography Images Using Deep Learning Models. *Journal of Healthcare Engineering, 2022*.
2. Afshar, P. a. (2021). COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. *Scientific Data, 8*, 1--8.
3. Ahmed, I. a. (2022). A Sustainable Deep Learning-Based Framework for Automated Segmentation of COVID-19 Infected Regions: Using U-Net with an Attention Mechanism and Boundary Loss Function. *Electronics, 11*, 2296.
4. Al-jumaili, S. a. (2022). Classification of Covid-19 Effected CT Images using a Hybrid Approach Based on Deep Transfer Learning and Machine Learning.
5. Anjum, M. S. (2022). Diastolic Dysfunction Prediction with Symptoms Using Machine Learning Approach. *Science and Technology, 4*, 714-727.
6. Aria, M. a. (2022). ADA-COVID: adversarial deep domain adaptation-based diagnosis of COVID-19 from lung CT scans using triplet embeddings. *Computational Intelligence and Neuroscience, 2022*.
7. Aswathy, A. a. (2022). COVID-19 severity detection using machine learning techniques from CT-images. *Evolutionary Intelligence*, 1-9.
8. Barstugan, M. a. (2022). Coronavirus (covid-19) classification using ct images by machine learning methods. *arXiv preprint arXiv:2003.09424*.
9. Bhargava, A. a. (2022). Machine learning-based automatic detection of novel coronavirus (COVID-19) disease. *Multimedia Tools and Applications, 81*, 13731--13750.
10. Butuner, R. a. (2022). Diagnosis and Detection of COVID-19 from Lung Tomography Images Using Deep Learning and Machine Learning Methods. *International Journal of Intelligent Systems and Applications in Engineering, 10*, 190--200.
11. Canayaz, M. a. (2022). COVID-19 diagnosis on CT images with Bayes optimization-based deep neural networks and machine learning algorithms. *Neural Computing and Applications, 34*, 5349--5365.
12. Chen, S. a.-J. (2019). A strong machine learning classifier and decision stumps based hybrid adaboost classification algorithm for cognitive radios. *Sensors, 19*, 5077.
13. Chen, Y.-Y. a.-L.-C.-H. (2021). Photographic reproduction and enhancement using HVS-based modified histogram equalization. *Sensors, 21*, 4136.
14. Delli Pizzi, A. a. (2021). Radiomics-based machine learning differentiates “ground-glass” opacities due to COVID-19 from acute non-COVID-19 lung disease. *Scientific reports, 11*, 1-9.
15. Doken, I. a.-S. (2021). Histogram Equalization Of The Image. *arXiv preprint arXiv:2108.12818*.
16. Ertu{\u{g}}rul, {. F. (n.d.). Detection of Covid-19 from X-ray Images via Ensemble of Features Extraction Methods Employing Randomized Neural Networks. *European Journal of Technique (EJT), 11*, 248--254.
17. Farahat, I. S. (2022). The Role of 3D CT Imaging in the Accurate Diagnosis of Lung Function in Coronavirus Patients. *Diagnostics, 12*, 696.
18. Furtado, A. a. (2022). A Light Deep Learning Algorithm for CT Diagnosis of COVID-19 Pneumonia. *Diagnostics, 12*, 1527.
19. Gaur, P. a. (2022). COVID-19 disease identification from chest CT images using empirical wavelet transformation and transfer learning. *Biomedical Signal Processing and Control, 72*, 103076.
20. Ghashghaei, S. a. (2022). Grayscale image statistics of COVID-19 patient CT scans characterize lung condition with machine and deep learning. *Chronic Diseases and Translational Medicine*, Wiley Online Library.
21. Hatamikia, S. a. (2014). The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. *Journal of medical signals and sensors, 4*, 194.

22. Heidari, A. a. (2022). A privacy-aware method for COVID-19 detection in chest CT images using lightweight deep conventional neural network and blockchain. *Computers in Biology and Medicine, 145*, 105461.
23. Islam, M. R. (2022). Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach. *Expert Systems with Applications, 195*, 116554.
24. Kabari, L. G. (2019). Comparison of bagging and voting ensemble machine learning algorithm as a classifier. *International Journals of Advanced Research in Computer Science and Software Engineering, 9*, 19-23.
25. Kadry, S. a. (2020). Development of a machine-learning system to classify lung CT scan images into normal/COVID-19 class. *arXiv preprint arXiv:2004.13122*.
26. Kairuddin, W. N. (2017). *Texture feature analysis for different resolution level of kidney ultrasound images* (Vol. 226). IOP Publishing.
27. Kalayc{\i}, M. a. (2022). Can laboratory parameters be an alternative to CT and RT-PCR in the diagnosis of COVID-19? A machine learning approach. *International journal of imaging systems and technology, 32*, 435--443.
28. Kim, D.-H. a.-Y. (2021). Classification of chronic kidney disease in sonography using the GLCM and artificial neural network. *Diagnostics, 11*, 864.
29. Kogilavani, S. a. (2022). COVID-19 detection based on lung CT scan using deep learning techniques. *Computational and Mathematical Methods in Medicine, 2022*.
30. Kwekha-Rashid, A. S. (2021). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience*, 1-13.
31. Li, D. a. (2020). Prediction of COVID-19 severity using chest computed tomography and laboratory measurements: evaluation using a machine learning approach. *JMIR Medical Informatics, 8*, e21604.
32. Liu, C. a. (2022). Automatic detection of COVID-19 and community-acquired pneumonia on CT images with artificial intelligence.
33. Mahmoudi, R. a.-Z. (2022). A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging. *A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging, 12*, 4825.
34. Mehrpouyan, M. a.-K. (2022). Detection of stage of lung changes in COVID-19 disease based on CT images: a radiomics approach. *Physical and Engineering Sciences in Medicine*, 1-9.
35. Mittal, B. a. (2021). CoviNet: Covid-19 diagnosis using machine learning analyses for computerized tomography images. In *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)* (Vol. 11878, pp. 340--349). SPIE.
36. Mohammed, A. F. (2022). The diagnosis of COVID-19 in CT images using hybrid machine learning approaches (CNN \& SVM). *Periodicals of Engineering and Natural Sciences (PEN), 10*, 376--387.
37. Patel, R. K. (2022). Automated diagnosis of COVID stages from lung CT images using statistical features in 2-dimensional flexible analytic wavelet transform. *Biocybernetics and Biomedical Engineering, 42*(3), 829--841.
38. Pourghasemi, H. R.-W. (2018). Assessment of landslide-prone areas and their zonation using logistic regression, logitboost, and na{\"\i}vebayes machine-learning algorithms. *Sustainability, 10*, 3697.
39. Rehouma, R. a.-P. (2021). Machine learning for medical imaging-based COVID-19 detection and diagnosis. *International Journal of Intelligent Systems, 36*, 5085--5115.
40. Salama, W. M. (20). Framework for COVID-19 segmentation and classification based on deep learning of computed tomography lung images. *Journal of Electronic Science and Technology, 3*, 100161.
41. Sani, S. a. (2022). A novel algorithm for detection of COVID-19 by analysis of chest CT images using Hopfield neural network. *Expert Systems with Applications, 197*, 116740.
42. Sayg{\i}l{\i}, A. (2022). Computer-aided detection of COVID-19 from CT images based on Gaussian mixture model and Kernel support vector machines classifier. *Arabian Journal for Science and Engineering, 47*, 2435--2453.

43. Shahin, O. R.-A. (2022). Machine Learning Approach for Autonomous Detection and Classification of COVID-19 Virus. *Computers and Electrical Engineering, 101*, 108055.
44. Shaik, N. S. (2022). Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans. *Computers in Biology and Medicine, 141*, 105127.
45. Shamrat, F. a. (2022). Analysing most efficient deep learning model to detect COVID-19 from computer tomography images. *Indonesian Journal of Electrical Engineering and Computer Science, 26*, 462--471.
46. Sharma, S. (2020). Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients. *Environmental Science and Pollution Research, 27*, 37155--37163.
47. Shiri, I. a. (2022). COVID-19 prognostic modeling using CT radiomic features and machine learning algorithms: Analysis of a multi-institutional dataset of 14,339 patients. *Computers in Biology and Medicine, 145*, 105467.
48. Suguna, G. a. (2022). A Machine learning Classification approach for detection of Covid 19 using CT images. *EMITTER International Journal of Engineering Technology*, 183--194.
49. Tang, X. (1998). Texture information in run-length matrices. *IEEE transactions on image processing, 7*, 1602--1609.
50. Ullah, A. a. (2022). Experimental and Numerical Research of Paved Microcrack Using Histogram Equalization for Detection and Segmentation. *Mathematical Problems in Engineering*.
51. velichko, E. a. (2022). Development of computer-aided model to differentiate COVID-19 from pulmonary edema in lung CT scan: EDECOVID-net. *Computers in Biology and Medicine, 141*, 105172.
52. Verma, A. a. (2022). Detecting COVID-19 from chest computed tomography scans using AI-driven android application. *Computers in biology and medicine, 143*, 105298.
53. Xiong, J. a. (2021). Application of histogram equalization for image enhancement in corrosion areas. *Shock and Vibration, 2021*.
54. Zhang, F. a. (2021). Unsupervised low-light image enhancement via histogram equalization prior. *arXiv preprint arXiv:2112.01766*.
55. Zhang, Y. a. (2017). Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neural Processing Letters, 45*, 365--378.