Systems and Soft Computing 6 (2024) 200077

Contents lists available at ScienceDirect

Systems and Soft Computing

journal homepage: www.journals.elsevier.com/soft-computing-letters

COVID-19 detection from chest CT images using optimized deep features and ensemble classification

Muhammad Minoar Hossain [a,b,*], Md. Abul Ala Walid [c], S.M. Saklain Galib [d], Mir Mohammad Azad [e], Wahidur Rahman [a,f], A.S.M. Shafi [a], Mohammad Motiur Rahman [a]

[a] Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh
[b] Department of Computer Science and Engineering, Bangladesh University, Dhaka, Bangladesh
[c] Department of Data Science, Bangabandhu Sheikh Mujibur Rahman Digital University, Kaliakair, Gazipur-1750, Bangladesh
[d] Department of Biomedical Engineering, Khulna University of Engineering and Technology (KUET), Khulna, 9203, Bangladesh
[e] Independent researcher, Dhaka, Bangladesh
[f] Department of Computer Science and Engineering, Uttara University, Dhaka, 1230, Bangladesh

ARTICLE INFO

Keywords:
Feature optimization
Fusion
Ensemble
Classification

ABSTRACT

Diagnosis of COVID-19 positive patients is the eventual move to impede the expansion of coronavirus. Variations of coronavirus make it tough to recognize COVID-19 positive patients through symptoms. Hence, this research aims at a faster and automatic detection approach of COVID-19 disease from the chest Computed tomography (CT) scan images. For the composition of the system, this approach constructs a feature vector from the CT images through the features fusion of two Convolutional neural network (CNN) models namely VGG-19 and ResNet-50. Before the feature fusion, preprocessing techniques are applied to gain more accurate outcomes. Moreover, pertinent features are identified from the feature vector by using several feature optimization methods namely Recursive feature elimination (RFE), Principal component analysis (PCA), and Linear discriminant analysis (LDA), and among them, we have observed PCA as the best preference. Classification is performed on the optimized feature utilizing the Max voting ensemble classification (MVEC). The fused features of VGG-19 and ResNet-50, processed with PCA and MVEC, provide the best outcomes of accuracy, specificity, sensitivity, and precision at 98.51 %, 97.58 %, 99.49 %, and 97.47 %, respectively, after 5-fold cross-validation for the proposed method.

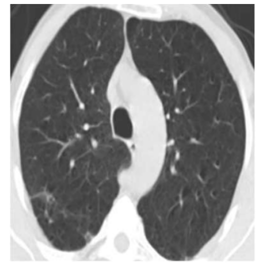
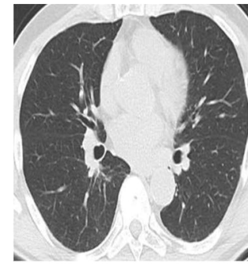
1. Introduction

The Coronavirus disease 2019 in short COVID-19 has expanded expeditiously all over the earth with a demolishing sequel on the population's soundness and well-being. It can be transmitted even with minimal or no symptoms but is typically associated with infection in individuals exhibiting acute respiratory symptoms caused by the coronavirus-2 (SARS-CoV-2).On January 30th, 2020 World health organization (WHO) announced this widespread an emergency public health crisis all over the world. Effective monitoring of infected people is a crucial step against COVID-19, so that infected patients can accept proper treatment as well as superintendence, and be isolated to generate a barrier for spreading the virus further. The real-time Polymerase chain reaction (PCR) test is the main screening method for recognizing COVID-19 disease. It can identify Ribonucleic acid (RNA) of SARS-CoV-2 from sputum, nasopharyngeal or oropharyngeal swabs. However, it has been stated that PCR is a time-spending, tedious, and complex manual procedure due to the availability of limited materials in hospitals. Moreover, the low sensitivity of PCR could not be good enough for the recognition and treatment of COVID-19 infected patients. It is also observable that the strong false-negative rates, as well as the test technique variabilities, resulted in a low-positive rate of 60–70 % in PCR [1]. Chest CT or Chest X-ray imaging can be an alternate screening method to identify COVID-19 patients. CT images produce a much more detailed and expeditious window than conventional X-rays of a body region overlaying the different body structures. Although CT and PCR are frequently identical, CT images can identify early COVID-19 infected cases with a negative PCR test in patients without symptoms, or after

* Corresponding author at: Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail, 1902, Bangladesh.
E-mail address: minoarhossain16005@gmail.com (M.M. Hossain).

https://doi.org/10.1016/j.sasc.2024.200077
Received 27 September 2023; Received in revised form 16 December 2023; Accepted 23 January 2024
Available online 4 February 2024

Table 1
Presentation of the dataset used in this method.

	Positive		Negative	
Total images	1252		1229	
samples				

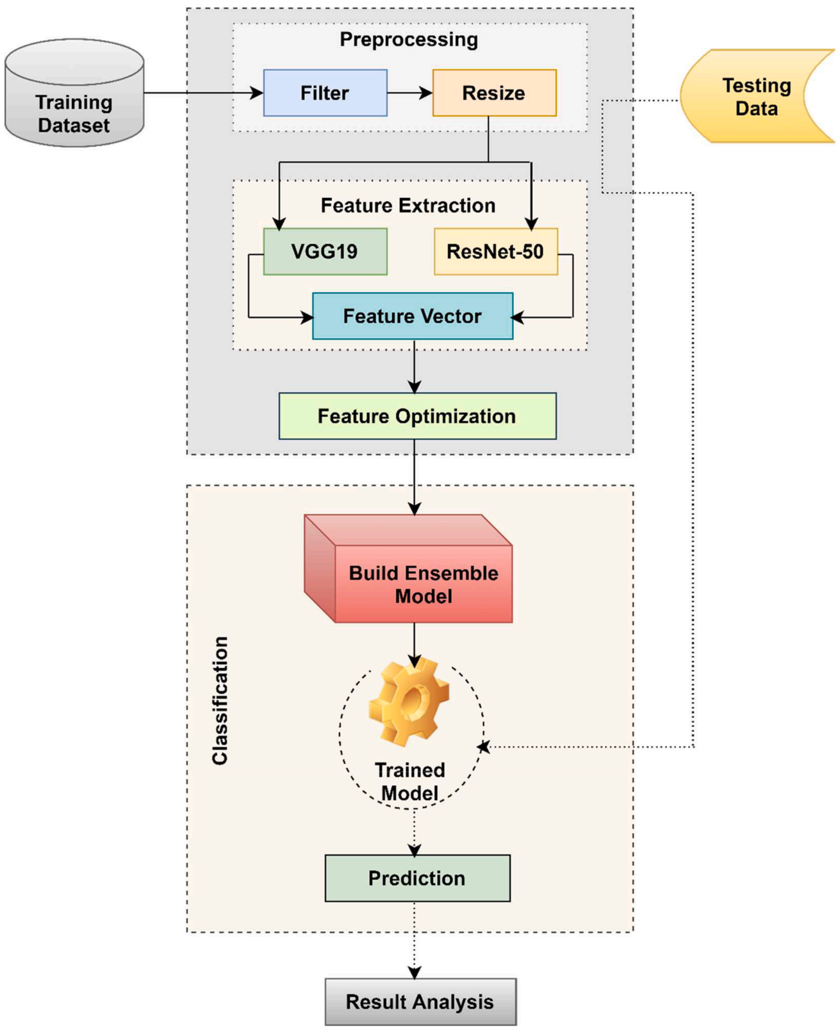


Fig. 1. The prime mechanism of the proposed method.

Table 2
MSE and PSNR values of several filtering techniques for the images in Fig. 2.

Filter	MSE	PSNR
average filter	104.31	27.94
Gaussian filter	38.78	32.24
Median filter	36.29	32.53
Bilateral	33.61	32.86

symptoms arise [2,3]. Since the number of recent COVID-19 suspected patients has been rising exponentially, therefore, artificial intelligence (AI) techniques can be a compatible approach for the recognition, classification, and characterization of COVID-19 by using chest CT images. In medical imaging systems, the advancement of AI specifically in deep learning (DL) has shown massive success in the automatic diagnosis of diseases owing to its high ability of feature extraction. Lung cancer, colon cancer, prostate cancer, pneumonia, COVID-19, tuberculosis, breast cancer, brain cancer, and other diseases are now efficiently

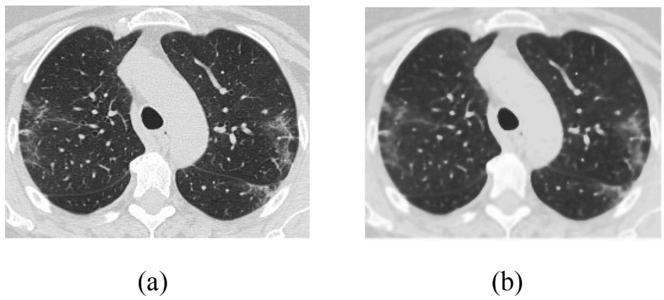


Fig. 2. Effect of BF on a CT scan image: (a)before filtering; (b)after filtering.

diagnosed utilizing DL strategies like CNN and machine learning (ML).

In this study, a COVID-19 diagnosis approach is developed from CT images with the assistance of deep features, feature optimization, and classification techniques. To develop the model, we have performed smoothing to a set of CT images. Two CNN models: VGG-19 and ResNet-50 are used to fuse the features from the preprocessed images and then feature optimization is used to apprehend the most unerring features. Several feature optimization techniques are inquired into this research to hold the supreme one. Finally, a max voting ensemble classifier is tooled to classify the optimized feature set. In our investigation, we have connected five distinct classifiers to implement the max voting technique. The primary contributions of this study are summarized below-

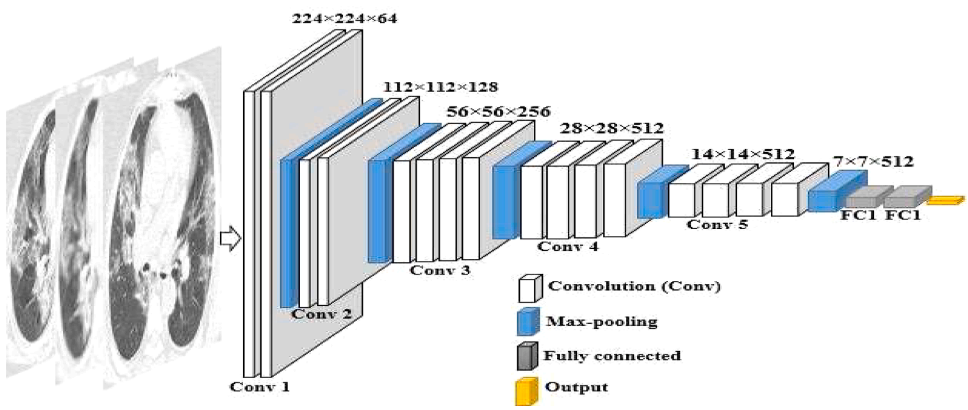


Fig. 3. The architecture of the VGG-19 model.

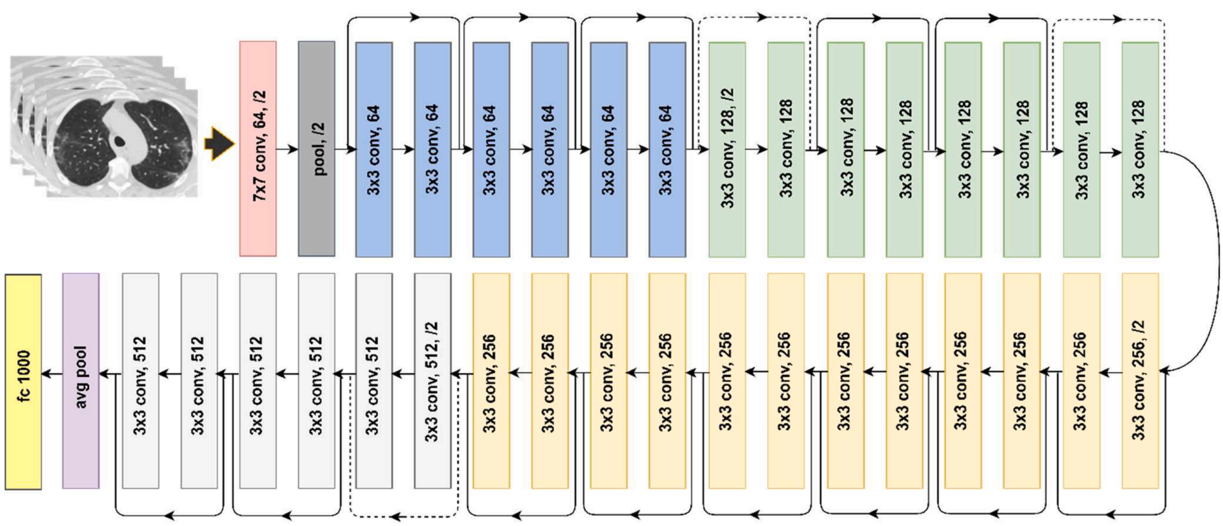


Fig. 4. The architecture of the ResNet-50 model.

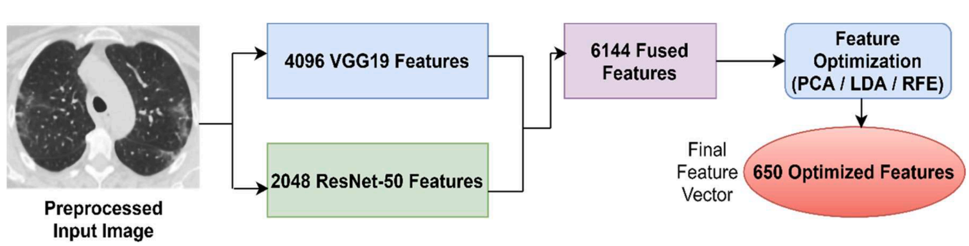


Fig. 5. Feature minimization process of this research.

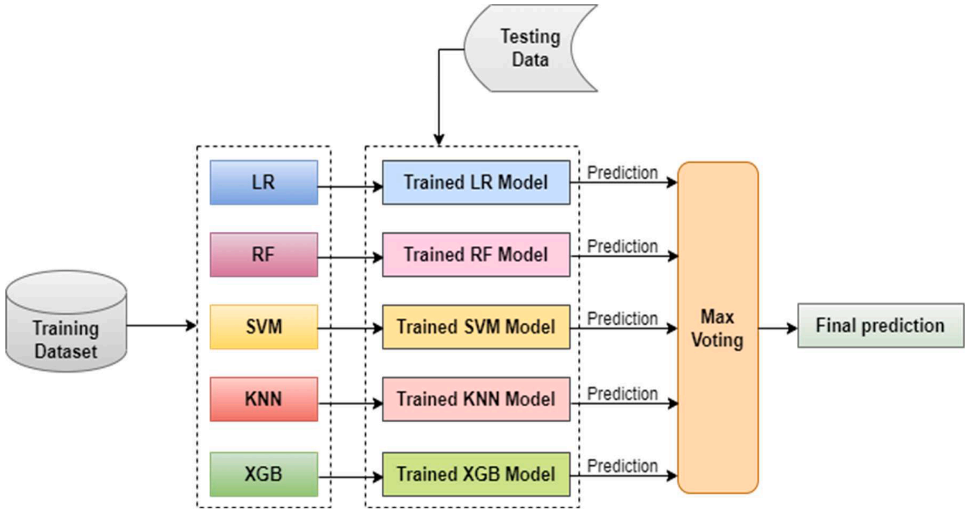
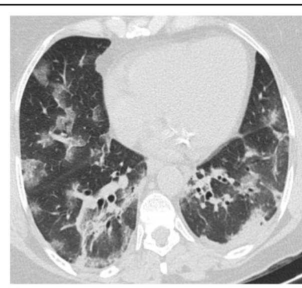
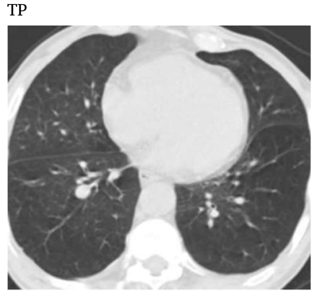
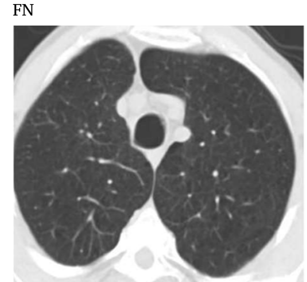


Fig. 6. The working mechanism of MVEC.

Table 3
Illustration of true and false detection by the proposed scheme.

Actual class	Predicted class *Positive (P)*	*Negative (N)*
Positive (P)	 TP	 FN
Negative (N)	 FP	 TN

- This research presents a technique to automatically detect COVID-19 infected patients. The formation can assist the physician in concluding decisions in the faster and easiest way.
- The proposed method can detect COVID-19 infected cases from CT scan images with a lower misdetection rate compared to the existing approaches because of features fusion by two CNN models namely VGG-19 and ResNet-50.
- The findings of this research offer an encouraging analysis of various classification techniques, CNN models, and feature optimization methods.

This research is organized into five sections. Section 2 delves into existing works related to our research, while Section 3 outlines the methodology and dataset. Experimental results and corresponding discussions are detailed in Section 4. Lastly, Section 5 concludes the

Table 4
Description of performance measurement metrics.

Metrics	Formula	Description
Accuracy	$\frac{TP+TN}{FP+TP+FN+TN}\times 100$	Total correct detection rate by the system.
Precision	$\frac{TP}{TP+FP}\times 100$	The correct detection rate of positive cases from the total predicted positive cases.
Sensitivity	$\frac{TP}{TP+FN}\times 100$	Total correct detection rate by the system for the positive cases.
Specificity	$\frac{TN}{TN+FP}\times 100$	Total correct detection by the system for the negative cases.

Table 5
Performance of proposed scheme.

Accuracy	Specificity	Sensitivity	Precision
98.51	97.58	99.49	97.47

Table 6
Normalized CM of different classifiers for each fold.

Fold number	Classifier	TP	FN	FP	TN
	LR	0.98	0.02	0.03	0.97
Fold 1	RF	0.87	0.13	0.19	0.81
	SVM	0.97	0.03	0.02	0.98
	KNN	0.96	0.04	0.01	0.99
	XGB	0.99	0.01	0.03	0.97
	Ensemble	0.996	0.004	0.01	0.99
	LR	0.99	0.01	0.03	0.97
Fold 2	RF	0.85	0.15	0.23	0.77
	SVM	0.98	0.02	0.03	0.97
	KNN	0.97	0.03	0.004	0.996
	XGB	0.97	0.03	0.04	0.96
	Ensemble	0.992	0.008	0.2	0.98
	LR	0.97	0.03	0.007	0.993
Fold 3	RF	0.88	0.12	0.20	0.80
	SVM	0.97	0.03	0.006	0.994
	KNN	0.94	0.06	0	1
	XGB	0.97	0.03	0.08	0.92
	Ensemble	0.99	0.01	0.007	0.993
	LR	0.98	0.02	0.04	0.96
Fold 4	RF	0.83	0.17	0.32	0.68
	SVM	0.98	0.02	0.03	0.97
	KNN	0.95	0.05	0.008	0.992
	XGB	0.97	0.03	0.95	0.05
	Ensemble	1	0	0.04	0.96
	LR	0.98	0.02	0.03	0.97
Fold 5	RF	0.91	0.09	0.26	0.75
	SVM	0.98	0.2	0.03	0.97
	KNN	0.93	.07	0.007	0.993
	XGB	0.92	0.08	0.0	0.94
	Ensemble	1	0	0.03	0.97

research and suggests future directions.

2. Related works

Diverse research has been initiated and ongoing to generate an automated accurate COVID-19 identification approach by using DL and ML. This section presents these works in a nutshell.

2.1. Methods based on DL

Jain et al. [4] developed DL-based CNNs, including Xception, ResNeXt, and Inception V3, to identify COVID-19 by utilizing 6432 chest X-ray images. Xception yielded the highest classification rate at 97.97 %. Wang et al. [5] employed AI methods using CT scans for COVID-19 analysis. In this study, segmentation used U-Net, FCN-8 s, 3D U-Net++, and V-Net, while classification involved Inception, DPN-92, ResNet-50, and Attention ResNet-50. The combination of 3D U-Net++ and ResNet-50 excelled in their experiments. In [6], a DL model identified COVID-19 from CT scans, using 3D CNN to segment infection regions. Combining ResNet with location-attention achieved 86.7 % accuracy in this work. In the paper [7], the authors presented a COVID-19 identification scheme with the assistance of CT images where the prediction was done by a details relation extraction neural network (DRE-Net). inaccuracy of this model was 6 %. Zheng et al. [8] merged ResNet50 and SEnet CNN architectures by employing squeeze and excitation blocks to differentiate COVID-19 patients from Bacterial as well as Viral Pneumonia from CT images. The accuracy of this scheme was 94 %. In [9], researchers utilized 630 CT images for COVID-19 identification using a 3D deep CNN (DeCoVNet) with pre-trained Unet-generated 3D lung masks. The model achieved a 90.1 % accuracy at a threshold of 0.5. In [10], authors employed transfer learning with a fine-tuned ResNet50 for COVID-19 patient classification based on CT images. Their approach achieved training and testing inaccuracies of 3.78 % and 6.98 %, respectively. In [11], a COVID-19 detection method used two subsets (16 × 16 and 32 × 32) CT images. GoogleNet,

Table 7
Performance of different classifiers for each fold.

Fold number	Classifier	Accuracy	Specificity	Sensitivity	Precision
	LR	97.78	97.41	98.11	97.74
Fold 1	RF	83.87	80.6	86.74	83.58
	SVM	97.38	97.84	96.97	98.08
	KNN	96.98	98.71	95.45	98.82
	XGB	98.39	97.41	99.24	97.76
	Ensemble	99.19	98.76	99.60	98.82
	LR	97.98	97.24	98.76	97.15
Fold 2	RF	81.05	76.77	85.54	77.82
	SVM	97.38	97.24	97.52	97.12
	KNN	98.19	99.61	96.69	99.57
	XGB	96.57	96.65	96.48	96.86
	Ensemble	98.19	97.24	99.17	97.17
	LR	98.19	99.26	96.88	99.09
Fold 3	RF	83.67	79.78	88.39	78.26
	SVM	98.19	99.26	96.88	99.09
	KNN	97.38	100	94.2	100
	XGB	94.55	92.31	96.55	93.33
	Ensemble	98.99	99.26	98.66	99.1
	LR	96.98	95.97	97.98	96.05
Fold 4	RF	75.4	68.15	82.66	72.18
	SVM	97.58	97.18	97.98	97.2
	KNN	96.98	99.19	94.76	99.16
	XGB	95.75	94.67	96.81	94.92
	Ensemble	97.98	95.97	100	96.12
	LR	97.38	97.05	97.78	96.49
Fold 5	RF	82.06	74.54	91.11	74.82
	SVM	97.38	97.05	97.78	96.49
	KNN	96.57	99.26	93.33	99.06
	XGB	92.92	93.85	92.03	93.90
	Ensemble	98.19	96.68	100	96.15

Table 8
Performance of various feature extractors and feature minimizers used in this research.

Technique	Accuracy	Specificity	Sensitivity	Precision
VGG-19	94.35	91.46	97.23	92.01
VGG-19+PCA	96.37	95.77	97.03	95.42
VGG-19+LDA	90.92	92.34	89.47	91.80
VGG-19+RFE	95.97	96.14	95.82	96.55
ResNet-50	95.56	92.19	99.56	91.5
ResNet-50+PCA	96.17	94.78	97.37	95.57
ResNet-50+LDA	66.16	66.23	66.05	66.51
ResNet-50+RFE	96.09	93.82	98.26	94.29
VGG-19+ ResNet-50	97.17	96.03	98.36	96.0
VGG-19+ ResNet-50+PCA	98.51	97.58	99.49	97.47
VGG-19+ ResNet-50+LDA	95.65	96.30	95.02	96.25
VGG-19+ ResNet-50+RFE	97.58	96.83	98.35	96.76

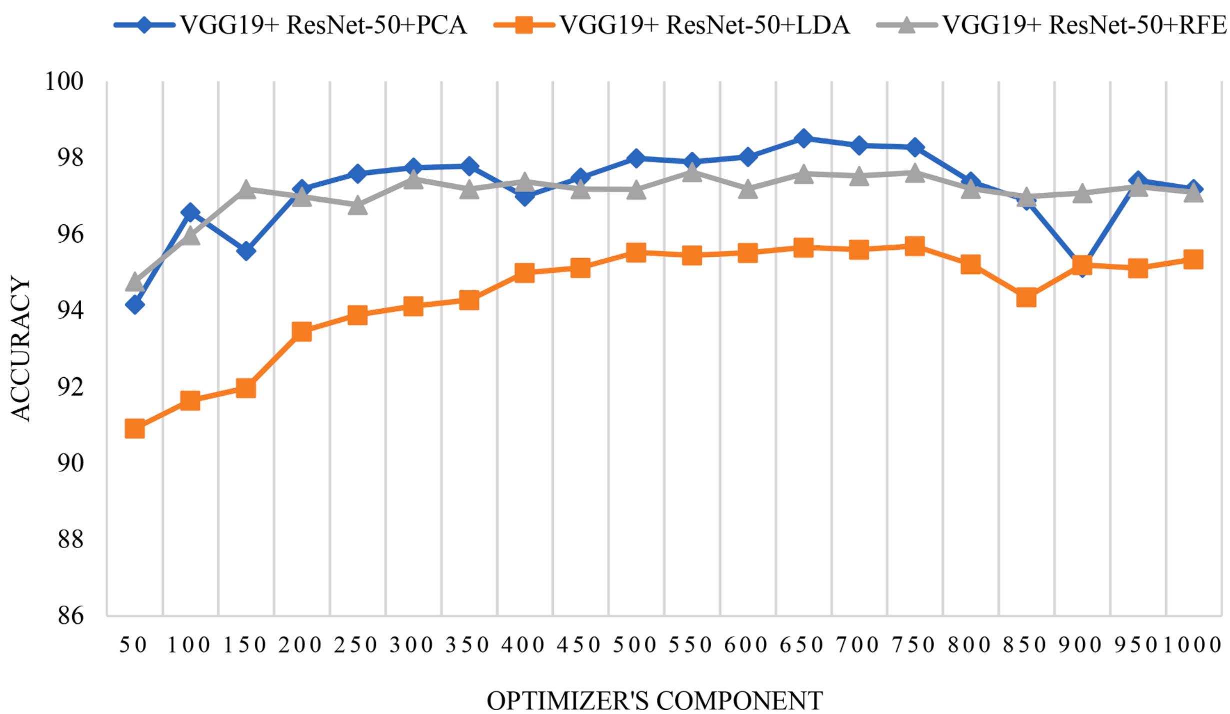


Fig. 7. Performance of feature optimizer regarding the number of components.

Table 9
Performance of different feature fusion approaches.

Technique	Accuracy	Specificity	Sensitivity	Precision
VGG-19+ ResNet-50+VGG-16+PCA	97.37	97.40	97.33	96.90
VGG-19+ ResNet-50+InceptionV3+RFE	96.76	95.27	98.33	95.16

Table 10
Comparison among existing methods and the proposed method.

Methods	Total images Positive	Negative	Accuracy	Specificity	Sensitivity
Wang et al. [5]	877	541	–	92.2	97.4
Pathak et al. [7]	413	439	93.02	94.78	91.45
Ozkaya et al. [11]	3000	3000	98.27	97.60	98.93
Eduardo et al. [15]	1252	1229	97.38	–	95.53
Smadi et al. [28]	1252	1229	98.79	–	–
Gupta and Bajaj [32]	1252	1230	98.91	98.86	98.96
Saygılı [37]	1252	1230	98.11	97.40	98.80
Proposed	1252	1229	98.51	97.58	99.49

ResNet-50, and VGG-16 were used to generate fused features from the images then features were ranked with the *t*-test, and SVM did classification. Subset-2 achieved 98.27 % accuracy in this work. A multitask deep learning-based classification and segmentation technique was designed in [12] to identify COVID-19 patients using CT images. They attained an 88 % dice coefficient for image segmentation and 94.67 % accuracy for multiclass classification. Matteo Polsinelli et al. [13] detected COVID-19 chest CT scan images with the assistance of the light CNN technique. Their CNN model was based on SqueezeNet and attained an accuracy of 85.03 %. An effective DL-based scheme was sketched in [14] for COVID-19 identification using chest X-ray images. They proposed an architecture where EfficientNet B0 was used as a base model. They analyzed the features of MobileNet, VGG, and ResNet models to their base model and discovered that their model had a lower computational cost. Smadi et al. [15] presented a Covid-19 diagnosis scheme called SEL-COVIDNET by utilizing nine CNN models. This technique was able to diagnose COVID-19 from both CT scans and X-ray images. In the paper [16] Smadi et al. proposed an Inception-v3 CNN-based technique to diagnose COVID-19. From X-ray images, this model was able to distinguish among normal, viral pneumonia, COVID-19, and lung opacity cases. To diagnose COVID-19 from X-ray scans Mehmood et al. presented different CNN-based techniques in papers [17,18]. Gupta and Bajaj [19] developed a lightweight CNN architecture to recognize COVID-19 cases from CT scan images. Ullah et al. [20] presented a holistic technique based on ShuffleNet CNN to diagnose COVID-19 from ECG, CT-Scan, and Chest radiograph images. Agnihotri and Kohli [21] presented a broad overview of the diagnosis of COVID-19 by using deep learning.

2.2. Methods based on ML

Akhtar et al. [22] developed an ML-based method to diagnose COVID-19 from complete blood count (CBC) data. Saygılı [23] proposed ML based method to detect COVID-19 from both X-ray and CT images using different handcrafted features. Bakheet and Al-Hamadi [38] developed a technique to recognize COVID-19 from X-ray images. This method used texture features to find the pattern in the X-ray image. The accuracy of the method was 95.88 %. Godbin and Jasmine [39] developed a COVID-19 diagnosis technique from CT images through the classification of texture features. This method experimented with various ML classifiers to recognize the pattern of CT images and found SVM as the best option. Najjar et al. [40] presented an ML-based COVID-19 screening method from an X-ray image. This technique used texture features as a core approach to identify the COVID-19 cases.

3. Materials and methodology

This section outlines the proposed method, including relevant techniques and the dataset used.

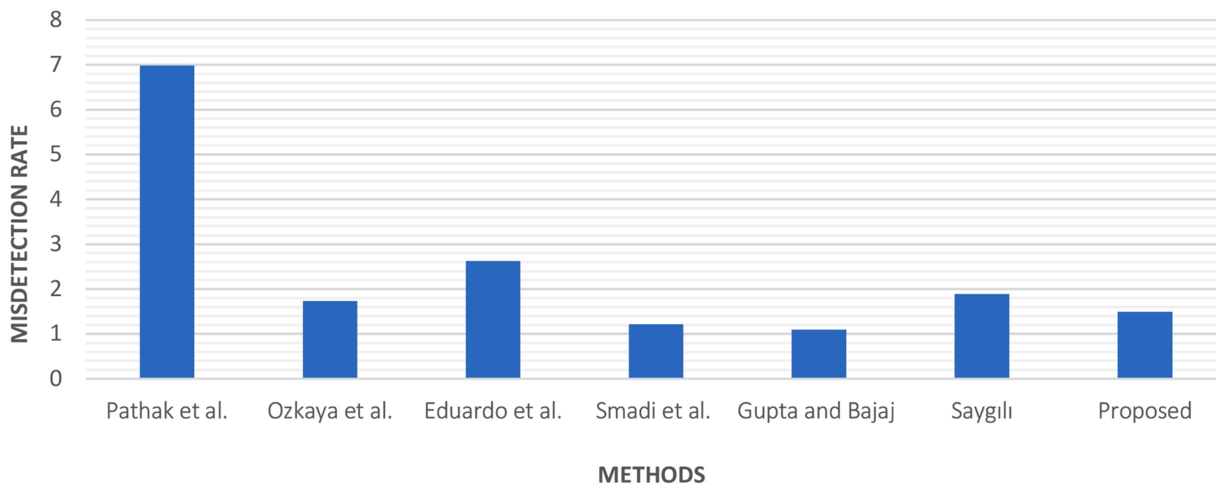


Fig. 8. Misdetection rate of different methods.

3.1. Dataset

This research utilizes a dataset provided by Eduardo, Plamen, et al. [24]. The dataset comprises a total of 2481 CT scan images collected from actual patients in hospitals across Brazil. It includes CT scans for both COVID-19 infected and non-infected patients. Within the non-infected category, various cases of pulmonary illnesses are also covered. Table 1 provides an overview of the dataset.

3.2. Methodology

Fig. 1 illustrates the main mechanism of the proposed method, and Sections 3.2.1 to 3.2.4 provide a comprehensive description of the figure.

3.2.1. Preprocessing

In this section, two operations are performed on each CT scan image: filtering and resizing. The scheme employs various filtering techniques, including average, Gaussian, median, and bilateral filters, to enhance image smoothness and reduce noise. By analyzing Mean square error (MSE) and Peak signal-to-noise ratio (PSNR) values for several images [25] we have found bilateral filter (BF) as the ideal technique. Besides BF is a noniterative, non-linear edge-preserving smoothing filter [26] that allows holding the most concerning features from each filtered image through CNN networks. After applying bilateral filtering, each image is resized to 224 × 224 to facilitate feature extraction by the CNN models. Table 2 shows the MSE and PSNR values for several filtering techniques for an arbitrary CT scan image and the result of BF for this image is shown in Fig. 2.

3.2.2. Feature extraction

To fuse the feature vector (FV) this method uses two CNN models namely VGG-19 and ResNet-50 as the feature extractor. For extracting features from any CT image Firstly, this method uses a pre-trained VGG19 CNN model. Fig. 3 showcases the architecture of the VGG-19 model, characterized by a sequence of convolutional layers succeeded by single or multiple fully connected layers. The model integrates 5 max-pooling layers between the fully connected and convolutional layers. The model can be defined into two parts. The first part is the feature extraction part which is the input layer to the last max-pooling layer and the residual network of the model is defined as the feature classification part [27]. Since this method utilizes the VGG19 for feature extraction hence the classifier part is excluded from the model. Thus, the model accepts any CT scan image of 224 × 224 × 3 and provides 4096(The output from the final layer of the feature extraction segment) features for each image.

Secondly, this method uses a pre-trained ResNet-50 CNN model. ResNet-50 is a 50-layer highly deep architecture residual network that has approximately 26 M parameters [28]. The ResNet-50 model architecture can be partitioned into many parts. The initial segment incorporates a convolutional layer with 64 kernels, succeeded by a max-pooling layer, another convolution layer, a fully connected layer, and the output layer. The proliferation of layers in neural models introduces a degradation issue. Resolving the vanishing gradient problem, the ResNet-50 model utilizes skip connections, functioning akin to a super-pathway. Since ResNet-50 is used in this method as a feature extractor hence the classifier part is excluded from the model. Therefore, the model accepts any CT scan image of 224 × 224 × 3 and provides 2048(Output from the last layer of the feature extraction part) features for each image. Fig. 4 shows the general architecture of ResNet-50.

The FV is generated by combining features from two models for each image. Therefore, the FV for an individual image comprises a total of 6144 values, obtained from the fusion of 4096 and 2048 values.

3.2.3. Feature optimization

To eliminate unnecessary and irrelevant features, feature optimization is employed. This study utilizes two types of feature optimization: feature reduction and feature selection. For feature reduction, LDA and PCA are used. PCA is a linear dimensionality reduction approach that identifies patterns in large data based on correlation between features [29]. When transferring the $(N \times D)$ dimension dataset into an $(N \times L)$ dimension data $(D >> L)$ PCA does not eliminate any features rather it gives new features by transforming them linearly. LDA is a supervised linear dimensionality reduction technique that finds feature subspace that optimizes class separability [30]. It performs separation by computing linear discriminants that enhance multiple access separability. For the selection of features, RFE is used. RFE is a feature selection method that selects significant features and removes the weakest features to increase the model performance until a specified number of features is reached [31].

In this scheme, each feature optimization technique reserves a total of 650 components. By varying the components of optimizers from 50 to 1000 with an interval of 50, we observe 650 components as the best option. So, 6144 features are reduced to 650 features, and comparing the outcomes in the results section, we observe that PCA provides the utmost performance. Fig. 5 illustrates the procedure of feature optimization.

3.2.4. Classification

Classification is performed on the optimized feature set to classify positive and negative COVID-19 patients from CT scan images. MVEC [32] is used to classify this method. MVEC is constructed by the conjunction of several individual classifiers. The outcome of MVEC can be acquired by the combination of the highest votes, calculated for each sample from the prediction of all base models. It can be called a meta-algorithm that enhances the performance of a predictive model. Logistic regression (LR) [33], Random forest (RF) [34], Support vector machine (SVM) [35], K-nearest neighbors (KNN) [36], and Xtreme gradient boosting (XGB) [37] classifiers are attached to construct the MVEC. Fig. 6 shows the core mechanism of MVEC.

4. Results and discussion

The prime aim of this research is to identify the positive and negative cases of COVID-19. To, assess the pertinence of this research 5-fold cross-validation (5-FCV) strait is used. The complete dataset is partitioned into two segments. The first segment apprehends 80 % of the data for the training stage and the second segment apprehends 20 % of the data for the testing stage. Therefore, the system uses a total of 1985 and 496 CT scan images for training and testing respectively. Four performance measurement metrics namely accuracy, specificity, sensitivity, and precision that are calculated from the parameters of the confusion matrix (CM) are used to evaluate the outcomes. The CM presents the actual and wrong detections of a classifier in a tabular form. Table 3 presents the formation of CM for this scheme where T and F indicate the true and false detection respectively. Table 4 presents the description of several performance measurement metrics.

The highest performance of this research is gained by using PCA with the max voting classifier. Table 5 shows this performance after 5-FCV.

This research offers excellent results because of MVEC. MVEC performs highest in our research as it works with a combination of several individual classifiers. Table 6 shows the CM (in a normalized form) of each fold related to Table 5 for different classifiers used in this research. Table 7 shows the performance measurement metrics of each fold related to Table 6.

This research uses several methods namely VGG-19, ResNet-50, PCA, LDA, RFE, etc., and by analyzing all the methods we have observed that PCA with MVC performs excellently. The performance of several methods with MVEC after 5-FCV is summarized in Table 8. From Table 8 it has been observed that the second-highest performance of our model is gained when RFE is used with VGG-19 and ResNet-50. From Table 8 it is also observable the superb performance of this research is possible because of the combination of two feature extractors- VGG-19, and ResNet-50.

In this scheme, we have used 650 components for each feature optimization technique. With the variation of optimizer components between 50 and 1000 by maintaining interval 50 we have chosen 650 components. Fig. 7 illustrates the accuracy of several feature optimizers along with different optimizer components. Fig. 7 shows LDA and RFE can't gain maximal accuracy with 650 components. Nevertheless, the maximal accuracy attained by LDA and RFE, as well as the accuracy obtained with 650 components, are nearly identical, with a very slight difference that can be considered negligible. Moreover, we have received the best accuracy due to 650 PCA components. Hence for better comparison and generalization, we have picked 650 components for all optimizer techniques.

Since the proposed method achieves satisfactory results due to the feature fusion by two CNN models, we are also attempting to perform feature fusion by using more than two CNN models with the assistance of two other CNN models, namely VGG-16 and InceptionV3. After 5-fold cross-validation, the highest results we have gained due to the combination of three feature extractors are summarized in Table 9. Table 9 presents that no technique outperforms the proposed system.

Table 10 presents the comparison between existing works and the proposed method based on performance measurement metrics. By investigating several existing methods, we have picked the most compatible works related to our method for Comparison. The comparison reveals that our method performs quite well when compared to existing systems.

Fig. 8 presents a comparison of existing works and the proposed method, focusing on the overall misdetection rate for each system. The figure highlights that Pathak et al. [10] recorded the highest misdetection rate at 6.98, while Gupta and Bajaj [19] achieved the lowest rate at 1.09, Smadi et al. [15] demonstrated the second lowest misdetection rate at 1.21, and this research follows with the third lowest rate at 1.49. Therefore, Fig. 8 establishes the compatibility of the proposed method with existing approaches.

5. Conclusion

This study investigates the potential of deep learning and machine learning to develop a diagnostic model that can identify COVID-19 cases from CT scan images. The model incorporates features from VGG-19 and ResNet-50 CNN models and employs preprocessing techniques to improve accuracy. Relevant features are extracted using Recursive feature elimination (RFE), Principal component analysis (PCA), and Linear discriminant analysis (LDA), with PCA proving to be the most effective. The Max voting ensemble classification (MVEC) technique, combining VGG-19 and ResNet-50 features processed with PCA, produces superior results. The proposed diagnostic system accurately distinguishes COVID-19 positivity or negativity in 496 cases, with overall inaccuracies of 0.51 % and 2.42 % respectively. However, the method has not been validated for diseases resembling COVID-19, highlighting identification challenges for individuals with unrelated pulmonary illnesses. Future modifications aim to improve the system's ability to recognize various diseases with symptoms similar to COVID-19 based on CT scan images.

CRediT authorship contribution statement

Muhammad Minoar Hossain: Methodology. **Md. Abul Ala Walid:** Investigation. **S.M. Saklain Galib:** Formal analysis, Visualization. **Mir Mohammad Azad:** Writing – original draft. **Wahidur Rahman:** Validation. **A.S.M. Shafi:** Writing – review & editing. **Mohammad Motiur Rahman:** Supervision.

Declaration of competing interest

The authors declare that they have no conflicts of interest regarding publishing the paper.

Data availability

I have shared the source of data in paper

References

[1] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, et al., Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology 296 (2020) E115–E117.
[2] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing, Radiology 296 (2020) E41–E45.
[3] S. Inui, A. Fujikawa, M. Jitsu, N. Kunishima, S. Watanabe, Y. Suzuki, et al., Chest CT findings in cases from the cruise ship diamond princess with coronavirus disease (COVID-19), Radiol.: Cardiothorac. Imaging 2 (2020) e200110.
[4] R. Jain, M. Gupta, S. Taneja, D.J. Hemanth, Deep learning based detection and analysis of COVID-19 on chest X-ray images, Appl. Intell. 51 (2021) 1690–1700.
[5] B. Wang, S. Jin, Q. Yan, H. Xu, C. Luo, L. Wei, et al., AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system, Appl. Soft. Comput. 98 (2021) 106897.
[6] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, et al., A deep learning system to screen novel coronavirus disease 2019 pneumonia, Engineering 6 (2020) 1122–1129.
[7] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, Y. Yang, Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (6) (2021) 2775–2780.
[8] F. Zheng, L. Li, X. Zhang, Y. Song, Z. Huang, Y. Chong, et al., Accurately discriminating COVID-19 from viral and bacterial pneumonia according to CT images via deep learning, Interdiscip. Sci.: Comput. Life Sci. 13 (2021) 273–285.

[9] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, et al., Deep learning-based detection for COVID-19 from chest CT using weak label, MedRxiv (2020), 2020.03.12.20027185.
[10] Y. Pathak, P.K. Shukla, A. Tiwari, S. Stalin, S. Singh, Deep transfer learning based classification model for COVID-19 disease, Irbm 43 (2022) 87–92.
[11] U. Özkaya, Ş. Öztürk, M. Barstugan, Coronavirus (COVID-19) classification using deep features fusion and ranking technique. Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach, Springer Nature, 2020, pp. 281–295.
[12] A. Amyar, R. Modzelewski, H. Li, S. Ruan, Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation, Comput. Biol. Med. 126 (2020) 104037.
[13] M. Polsinelli, L. Cinque, G. Placidi, A light CNN for detecting COVID-19 from CT scans of the chest, Pattern. Recognit. Lett. 140 (2020) 95–100.
[14] E. Luz, P. Silva, R. Silva, L. Silva, J. Guimarães, G. Miozzo, et al., Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images, in: Research on Biomedical Engineering Springer, 2021, pp. 1–14.
[15] A. Al Smadi, A. Abugabah, A.M. Al-Smadi, S. Almotairi, SEL-COVIDNET: an intelligent application for the diagnosis of COVID-19 from chest X-rays and CT-scans, Inf. Med. Unlock 32 (2022) 101059.
[16] A. Al Smadi, A. Abugabah, S.F. Mahenge, F. Shahid, Information systems in medical settings: a COVID-19 detection system using X-ray scans, in: 2022 26th International Computer Science and Engineering Conference (ICSEC), 2022, pp. 72–77.
[17] A. Mehmood, A. Abugabah, A. Al Smadi, Smart health care system for early detection of COVID-19 using X-ray scans, in: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2022, pp. 1–4.
[18] A. Mehmood, A. Abugabah, A.A. Smadi, R. Alkhawaldeh, An intelligent information system and application for the diagnosis and analysis of COVID-19, in: Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021) 3, 2022, pp. 391–396.
[19] K. Gupta, V. Bajaj, Deep learning models-based CT-scan image classification for automated screening of COVID-19, Biomed. Signal Process Control 80 (2023) 104268.
[20] N. Ullah, J.A. Khan, S. El-Sappagh, N. El-Rashidy, M.S. Khan, A holistic approach to identify and classify COVID-19 from chest radiographs, ECG, and CT-scan images using shufflenet convolutional neural network, Diagnostics 13 (2023) 162.
[21] A. Agnihotri, N. Kohli, Challenges, opportunities, and advances related to COVID-19 classification based on deep learning, Data Sci. Manag. 6 (2023) 98–109.
[22] A. Akhtar, S. Akhtar, B. Bakhtawar, A.A. Kashif, N. Aziz, M.S. Javeid, COVID-19 detection from CBC using machine learning techniques, Int. J. Technol. Innov. Manag. 1 (2021) 65–78.
[23] A. Saygılı, A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods, Appl. Soft. Comput. 105 (2021) 107323.
[24] E. Soares, P. Angelov, S. Biaso, M.H. Froes, D.K. Abe, SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification, MedRxiv (2020), 2020.04. 24.20078584.
[25] M.M. Hossain, M.A. Rahim, A.N. Bahar, M.M. Rahman, Automatic malaria disease detection from blood cell images using the variational quantum circuit, Inf. Med. Unlock 26 (2021) 100743.
[26] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision, IEEE, 1998, pp. 839–846. IEEE Cat. No. 98CH36271.
[27] J. Xiao, J. Wang, S. Cao, B. Li, Application of a novel and improved VGG-19 network in the detection of workers wearing masks, in: , 2020 012041.
[28] H.A. Khan, W. Jue, M. Mushtaq, M.U. Mushtaq, Brain tumor classification in MRI image using convolutional neural network, Math. Biosci. Eng. (2021).
[29] G. Chao, Y. Luo, W. Ding, Recent advances in supervised dimension reduction: a survey, Mach. Learn. Knowl. Extr. 1 (2019) 341–358.
[30] A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, Linear discriminant analysis: a detailed tutorial, AI Commun. 30 (2017) 169–190.
[31] V. Wottschel, D. Alexander, P. Kwok, D. Chard, M. Stromillo, N. De Stefano, et al., Predicting outcome in clinically isolated syndrome using machine learning, Neuroimage Clin. 7 (2015) 281–287.
[32] A.S. Assiri, S. Nazir, S.A. Velastin, Breast tumor classification using an ensemble machine learning method, J. Imag. 6 (2020) 39.
[33] S. Ambesange, A. Vijayalaxmi, S. Sridevi, B. Yashoda, Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques, in: 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 827–832.
[34] K. Açıcı, Ç.B. Erdaş, T. Aşuroğlu, M.K. Toprak, H. Erdem, H. Oğul, A random forest method to detect Parkinson's disease via gait analysis, in: Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, 2017, pp. 609–619.
[35] S. Ding, B. Qi, H. Tan, Review on Theory and Algorithm of Support Vector Machine, J. Univ. Electron. Sci. Technol. China 40 (2011) 2–10.
[36] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1255–1260.
[37] A.I.A. Osman, A.N. Ahmed, M.F. Chow, Y.F. Huang, A. El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, Ain Shams Eng. J. 12 (2021) 1545–1556.
[38] S. Bakheet, A. Al-Hamadi, Automatic detection of COVID-19 using pruned GLCM-Based texture features and LDCRF classification, Comput. Biol. Med. 137 (2021) 104781.
[39] A.B. Godbin, S.G. Jasmine, Screening of COVID-19 based on GLCM features from CT images using machine learning classifiers, SN Comput. Sci. 4 (2) (2022) 133.
[40] F.H. Najjar, K.A. Kadhim, M.H. Kareem, H.A. Salman, D.A. Mahdi, H.M Al-Hindawi, Classification of COVID-19 from X-ray images using GLCM features and machine learning, Malays. J. Fundam. Appl. Sci. 19 (3) (2023) 389–398.