

OWAE-Net: COVID-19 detection from ECG images using deep learning and optimized weighted average ensemble technique

Kunwar Prashant^a, Prakash Choudhary^{*,b}, Tarun Agrawal^c, Eam Kaushik^d

^a Department of Computer Science & Engineering, NIT Hamirpur, Hamirpur, Himachal Pradesh, India

^b Computer Science & Engineering, Central University of Rajasthan, Rajasthan, India

^c School of Computer Science Engineering & Technology, Bennett University, Greater Noida, Uttar Pradesh, India

^d Symbiosis Centre for Medical Image Analysis, India



ARTICLE INFO

Keywords:

COVID-19

CNN

Weighted average ensemble model

Transfer learning

ECG

ABSTRACT

COVID-19 is an infectious disease that has cost millions of lives all over the world. A faster and safer diagnosis of COVID-19 is highly desirable in order to stop its spread. An electrocardiogram (ECG) signal-based diagnosis has shown its potential in the diagnosis of cardiac, stroke, and COVID-19. In this study, an ensemble of three deep learning models are used for COVID-19 detection in ECG images for multi-class classification. The results obtained with the weighted average ensemble technique have been improved by using the grid search technique. For multi-class classification, an optimized weighted average ensemble (OWAE) model classifies the ECG images with an accuracy of 95.29%, an F1-score of 95.4%, a precision of 95.5%, and a recall of 95.3%. In case of binary classification, VGG-19, EfficientNet-B4, and DenseNet-121 performed comparatively well with 100% accuracy. These results show that deep learning can be used in the diagnosis of COVID-19 disease using ECG images.

1. Introduction

Covid-19 is a human-to-human transmissible disease, which makes it a severe problem (Shereen, Khan, Kazmi, Bashir, & Siddique, 2020). It affects people of all ages, but the non-vaccinated and people with weak immunity are at greater risk (Rutten et al., 2021). There are 507.9 million registered cases and more than 6.2 million life casualties due to COVID-19. The patients who have survived the COVID-19 have reported fatigue, dyspnea, and respiratory problems (Kamal, Abo Omirah, Hussein, & Saeed, 2021; Fernández-de Las-Peñas et al., 2021). This disease has been reported in almost all countries around the globe after its outbreak (Parry, 2020; Riou & Althaus, 2020).

The pathogenetic virus of this disease is SARS-CoV-2 and it mainly attacks the respiratory system in the body (Al-Aalim, Hamad, & AL-ledani, 2020). When a person breathes air contaminated by particles containing the SARS-CoV-2, it causes the COVID-19 infection. The proximity of infected people plays a vital role in the virus transmission as it can occur over longer distances too (Wang et al., 2021). Among the infected people, at least one-third do not show any noticeable symptoms (Oran & Topol, 2020). However, among those who develop symptoms, 81% develop mild to moderate symptoms, 14% develop severe symptoms that necessitate hospitalization, and 5% require ICU (Wu &

McGoogan, 2020).

Due to the exponential growth in COVID-19 cases, there is a huge burden on the medical infrastructure for diagnosing and treating a large number of people. Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is generally used for the COVID-19 detection. RT-PCR is costly and takes a minimum of four hours to detect the presence of a virus (Agrawal & Choudhary, 2021). This study proposes the use of electrocardiogram electrical impulses (in voltage) generated by the human heart (Lilly, 2012). Reading is taken at a certain frequency depending upon the device used. The ECG reading is collected by attaching the electrodes to the skin of the patient. The sensors attached to the skin are capable of detecting the electrical changes produced by the cardiac muscle's repolarization and depolarization during each cardiac cycle. There are three main components of an ECG plot: (1) P-wave represents the depolarization of the atria; (2) QRS-complex represents the depolarization of the ventricles; and (3) T-wave represents the repolarization of the ventricles (Lilly, 2012). Various studies show that the ECG plots tend to change in such a way that they could be utilized to classify the COVID-19 infected patient from other cardiac disease patients and the normal person (Ozdemir, Ozdemir, & Guren, 2021; Sobahi, Sengur, Tan, & Acharya, 2022). Various such studies are discussed in Section 2. A normal ECG plot is shown in Fig. 1. The

* Corresponding author.

E-mail addresses: prakash.choudhary@curaj.ac.in (P. Choudhary), tarun.agrawal@bennett.edu.in (T. Agrawal).

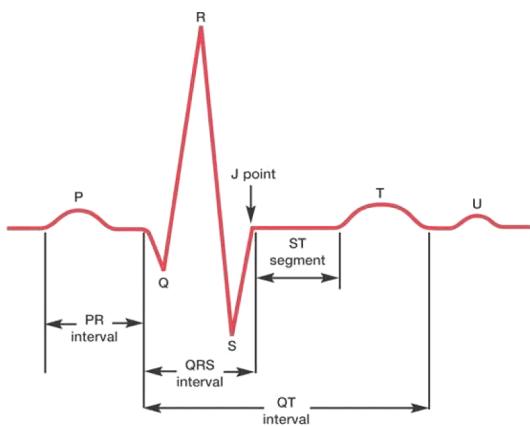


Fig. 1. A Plot showing normal ECG pattern and its different components. Image source- (Ashley and Niebauer, 2004).

Table 1
Brief summary on related work for COVID-19 detection.

Paper	Dataset	Technique	Limitation
Attallah (2022)	ECG Images	Five CNN models are used for feature extraction, and feature selection is done by symmetrical uncertainty method	Class imbalance problem is not addressed.
Ozdemir et al. (2021)	ECG Images	GLCM method is used for feature extraction and then these features are fed to the CNN	Hexaxial feature mapping is used which is sensitive to the resolution of the ECG images.
Sobahi et al. (2022)	ECG Images	Attention-based 3D CNN with residual learning is used for feature extraction	The manual placement of windows to isolate the area of interest.
Rahman et al. (2022)	ECG Images	Six CNN models are used for feature extraction	Limited dataset is used for the experiments.
Abirami et al. (2021)	CT-Scan Images	CNN and GAN model is used for feature extraction	Dice coefficient obtained for lesion segmentation is very low.
Abirami et al. (2022)	Chest X-ray Images	styleGAN2 is used for the image generation and CNN for the feature extraction	Multi-class classification is not performed.

COVID-19 diagnosis should be accurate and cost-efficient, with the least possible false positive rate (FPR) and false negative rate (FNR). A deep learning technique can be used to diagnose COVID-19 in patients with ECG images.

The main contributions are summarized below:

- In this paper, we have employed pre-trained deep learning models for binary classification and the OWAE model for multi-class classification using ECG images.
- The weighted average ensemble technique has been used to obtain the results for the multi-class classification.
- The weights assigned to deep models have been optimized using a grid search technique. The obtained results show that the grid search technique has improved the obtained results.

The paper is organized as follows: **Section 2** discusses the related work done for COVID-19 using the ECG images. While in **Section 3**, the ECG dataset and the methodology are discussed. **Section 4** discusses the evaluation metrics, experimental steps, and performance evaluation of the deep learning models for binary and multi-class classification. In **Section 5**, the limitations of the proposed work are discussed, and in

Section 6, the conclusion is presented.

2. Related work

Related work is divided into three subsections. In **Section 2.1**, the research work diagnosing the covid- 19 using the ECG is discussed. In **Section 2.2**, the research works diagnosing the COVID-19 in ECG using the deep learning models are discussed. Moreover, the research studies using the weighted average ensemble (WAE) model for diagnosing the diseases are discussed in **Section 2.3**.

2.1. COVID-19 diagnosis using ECG images

ECG has been used for the diagnosis of various cardiac diseases. It can be used for the diagnosis of stroke diseases, as first reported by Byer, Ashman, & Toth (1947). Recent studies show that it is also useful in the detection of COVID-19. Wang et al. (2020) show in their study that changes in the ECG pattern of COVID-19 patients can play a vital role in the diagnosis of myocardial injury and cardiac insufficiency. Li et al. (2020) found in their study that 50 COVID-19 patients died, out of 113 patients, who showed symptoms of ventricular arrhythmia in their ECG. Santoro et al. (2021) found QT-prolongation in 14% of patients out of 110 COVID-19 patients in their study, which can be easily detected in their ECG. Lam, McClelland, & Dallo (2020) conducted a study involving COVID-19 patients and found cardiac abnormalities in 63% of them. Bertini et al. (2020) conducted a study on 431 COVID-19 patients and found that 93% of patients have cardiac abnormalities, 22% of patients diagnosed with atrial fibrillation (AF), and 30% of patients have acute right ventricular pressure overload (RVPO). Moreover, Attia, Kapa, Noseworthy, Lopez-Jimenez, & Friedman (2020) have shown in their work that artificial intelligence can be used to diagnose COVID-19 using the ECG.

2.2. COVID-19 diagnosis using deep learning models

Attallah (2022) proposed an ECG-BiCoNet model for COVID-19 diagnosis using ECG images. ECG-BiCoNet uses deep learning models for feature extraction and then features are fused using discrete wavelet transform with lower layers. Then, these features are selected for the final prediction using three different machine learning classifiers. Ozdemir et al. (2021) uses the gray level co-occurrence matrix (GLCM) method for the feature extraction and generate hexaxial mapping images. These generated images are used as input for the convolutional neural network for the COVID-19 classification. Sobahi et al. (2022) proposed a 3DCNN a model consisting of residual connection and attention mechanism to classify COVID-19 from ECG images. Rahman et al. (2022) employed several different deep CNN models for COVID-19 detection in ECG trace images. DenseNet-201 outperformed the other deep models for the two-class and three-class classifications. The author(s) also used the Score-CAM visualization to track whether the models learn relevant information from the ECG images. Abirami, Vincent, & Kadry (2021) proposed the CNN and GAN-based model (P2P-COVIDSEG) to classify and segment the COVID-19 lesion in CT scans. The P2P-COVID-SEG model obtained the 98.10% accuracy for classification and 81.11% dice similarity coefficient for lesion segmentation. Abirami, Vincent, Rajinikanth, & Kadry (2022) have used chest X-rays to detect the COVID-19 using the styleGAN2 model. The model obtained 99.78% accuracy for the binary classification.

2.3. Weighted average ensemble model for disease diagnosis

Shashvat, Basu, Bhondekar, & Kaur (2019) predicted the incidence of infectious diseases like typhoid using an ensemble of models like mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The obtained results indicated the WAE model performed better. Similarly, Mahendran et al. (2019) used logistic

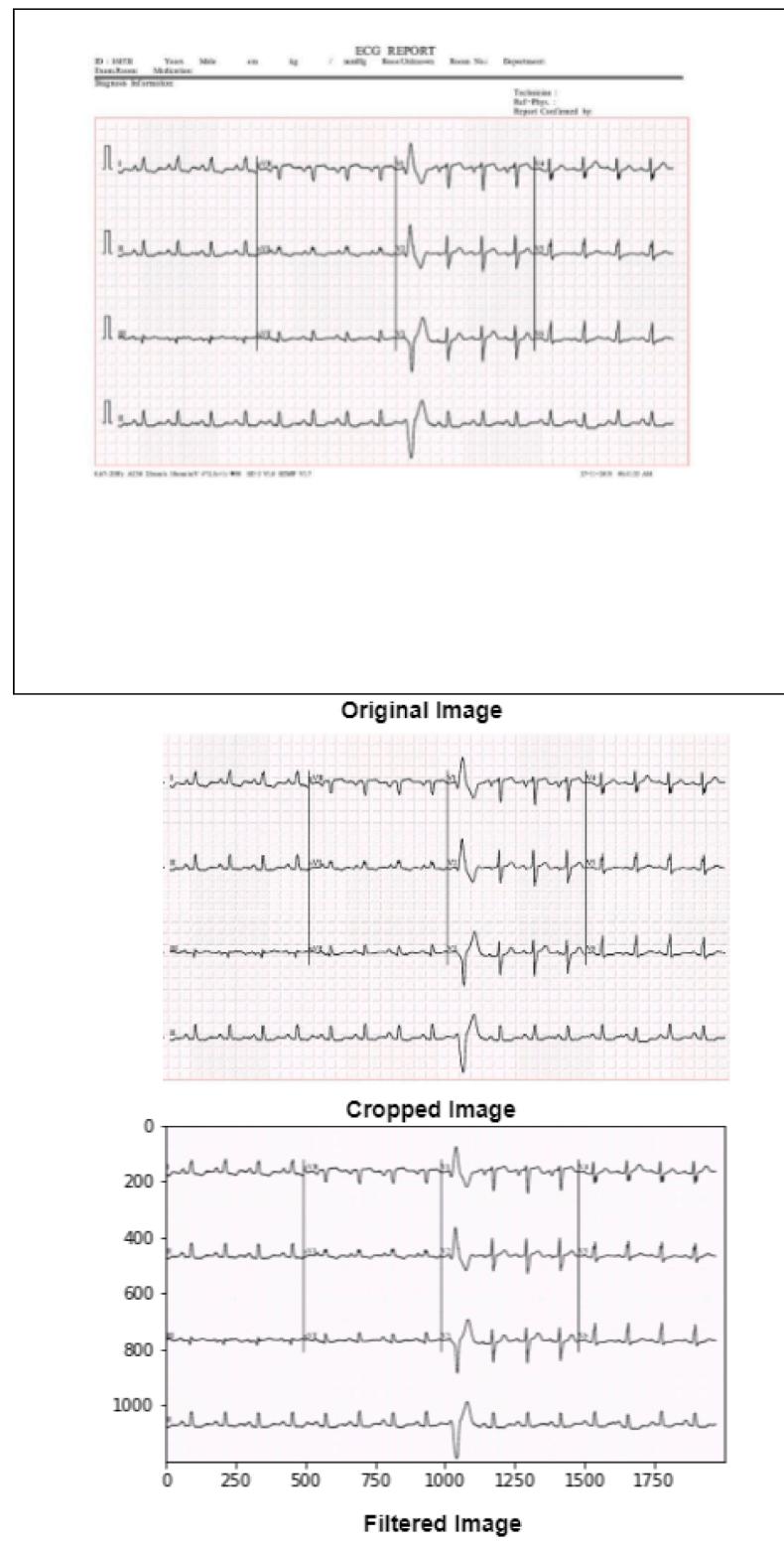


Fig. 2. Pre-processing of the ECG images.

regression, random forest, and WAE models for detecting major depressive disorder. The obtained results show that the WAE model outperformed the logistic regression and random forest models. The results obtained in these two research papers show that the WAE models obtained better results than individual models. The summary of the related works is presented in Table 1.

3. Materials and methodology

The ECG dataset and the pre-processing are discussed in Section 3.1, while the methodology is discussed in Section 3.2.

Table 2

Hyperparameters values used in training the model.

Hyperparameters	Values
Learning Rate	0.001
Epochs Set	200
Early Stopping used	Yes
Patience	3
Optimizer	Adam
Loss function	Binary cross-entropy and categorical cross entropy

Table 3

Binary class classification results.

Metric	VGG-19	EfficientNet-B4	DenseNet-121
Accuracy	100%	100%	99.09%
F1-score	100%	100%	99.07%
Precision	100%	100%	99.20%
Recall	100%	100%	98.95%

Table 4

Multi-class classification results.

Metric	VGG-19	EfficientNet-B4	DenseNet-121	OWAE
Accuracy	92.94%	92.35%	92.95%	95.29%
F1-score	91.64%	90.77%	90.76%	95.40%
Precision	89.97%	89.78%	89.79%	95.51%
Recall	95.94%	93.82%	93.20%	95.32%

3.1. Dataset and pre-processing

The ECG dataset used in this study is public (Khan, Hussain, & Malik, 2021). This dataset contains RGB images of scanned ECG reports of patients for five different classes. There are 250 ECG tracing images of COVID-19 patients; 77 ECG tracing images of myocardial infarction patients; 548 ECG tracing images of patients that have abnormal heartbeats; 203 ECG tracing images of patients that have a history of myocardial infarction; and 859 normal people's ECG images. Each image contains 12 LED ECG tracings and is reviewed by medical experts

using a telehealth ECG diagnostic system.

The dataset contains scanned copies of ECG reports and is pre-processed before being fed to the deep learning model. The images are in different resolutions, containing some text in the header and footer of the report, and a grid in the background, which is not favorable for the classification process. Therefore, first, ECG images are cropped such that the part containing the ECG report gets extracted only. Second, a filter is applied to each cropped image such that the grid in the background is removed while ECG traces remain in the image. Finally, all the images are resized to $256 \times 256 \times 3$ resolution. An example of a preprocessed image is shown in Fig. 2. The dataset is highly imbalanced, therefore, 300 images are selected from the abnormal heartbeat patient and the normal patient class, and 250 images from the COVID-19 class.

3.2. Methodology

Methodology is discussed in detail in the 3.2.1, 3.2.2, and 3.2.3 subsections.

3.2.1. Transfer learning technique

Deep learning models have a large number of parameters and need millions of images for training. If there is not enough data, then there is a chance of overfitting the model. In medical image analysis, there is no large dataset available due to privacy issues to train the deep learning models. Therefore, the researchers use the transfer learning technique to train their models. Different research groups have released pre-trained models for reuse (Agrawal & Choudhary, 2022). Most of these pre-trained models available are trained on the ImageNet dataset (Deng et al., 2009). In this study, we have used VGG-19, DenseNet-121, and EfficientNet-B4 pre-trained models for binary and multi-class classification. The fully-connected layers are removed from the pre-trained models to reduce the number of parameters and the dense layers have been added with a smaller number of filters. A common approach is to freeze the weights of the pre-trained model so there is no change in the weight of the model. In this study, we have also freezed the pre-trained layers and trained only the newly added dense layers for the final classification results.

```

1: Input: ECG images
2: Output: covid-19 classification
3: begin
4: 1 Image  $\leftarrow$  Read ECG training images
5: 2 R - Image  $\leftarrow$  Resize the Image
6: 3 Imagetrain, Imagetest  $\leftarrow$  Split R - Image
7: 4 Train all three CNN models separately
8: for (model i = 1 to 3)
9: begin
10: 4.1 Apply the three-fold cross-validation
11: for (k= 1 to 3)
12: begin
13: 4.1.1 Train - Imagetrain, Test - Imagetest  $\leftarrow$  Split Imagetrain
14: 4.1.2 CNN  $\leftarrow$  Load the deep CNN model
15: 4.1.3 Model  $\leftarrow$  Train (CNN, Train - Imagetrain)
16: 4.1.4 confusion - matrix  $\leftarrow$  Prediction(Model, Test - Imagetest)
17: 4.1.5 Print results
18: 4.1.6 Model  $\leftarrow$  Save (Model)
19: end
20: end
21: 5 Load and append VGG-19, EfficientNet-B4, DenseNet-121 model for ensemble average
22: 6 Model  $\leftarrow$  Load (Model)
23: 7 ModelAppend  $\leftarrow$  Append (Model)
24: 8 Apply the grid search to find the optimized weights
25: 9 Select the weights giving highest performance
26: 10 Evaluate the ModelAppend on Imagetest
27: 11 Print results
28: end

```

Algorithm 1. Steps for COVID-19 detection using the OWAE model.

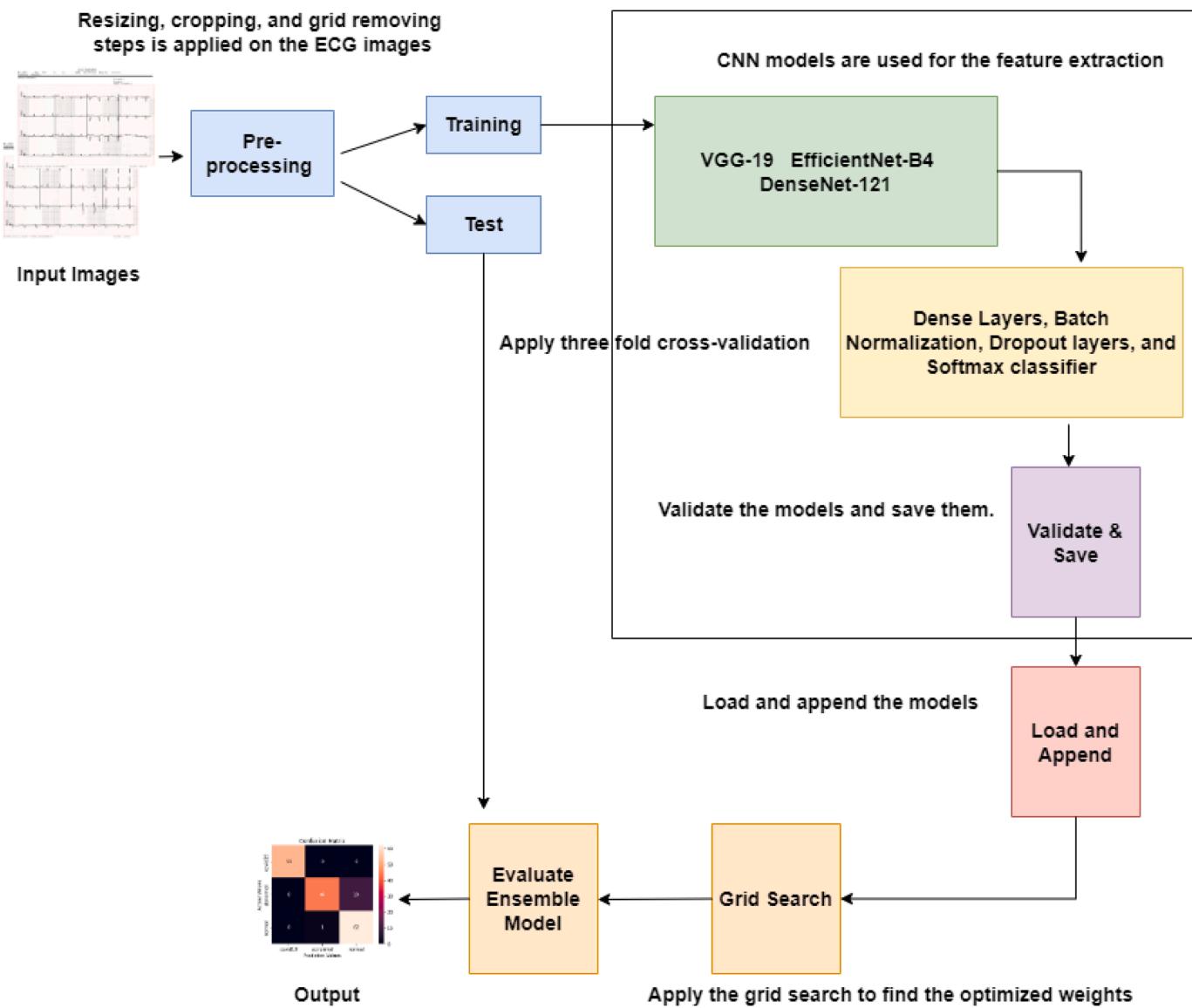


Fig. 3. Framework for COVID-19 detection in ECG images using the OWAE model.

3.2.2. Optimized weighted average ensemble method

In most research studies, the classification results are based on a single model. While there is a lot of research that shows the ensemble model can outperform the single model (Ekbal & Saha, 2013; Shahhosseini, Hu, & Pham, 2022). A single model can not extract all the features of the dataset. Therefore, researchers use an ensemble of different models to enhance the performance. In this paper, the optimized weighted average ensemble method is used for multi-class classification. This ensemble technique is an enhancement of the average ensemble method in which all the models contribute equally to generating predictions. In the weighted average ensemble method, the amount of contribution of each model is determined by assigning a weight to each model. Furthermore, in this study, the assigned weights are optimized using the grid search technique.

3.2.3. Model architecture

The CNN model can be divided into the feature extraction block and the fully-connected block. In this study, the feature extraction blocks of pre-trained CNN models such as VGG-19, DenseNet-121, and EfficientNet-B4 are used for the classification. The fully-connected layers of these pre-trained models are removed to reduce the trainable

parameters. After the feature extraction block, global average pooling (GAP), dense, batch normalization, and dropout layers are used. The two dense layers are added, with 128 and 256 neurons. While the dropout layer has been characterized by 20%. This means that 20% of neurons will be dropped during the forward and backward passes. The dropout layer adds regularisation and prevents the overfitting of the model. The weights of feature extraction blocks are freezed and the remaining layers are trained on the ECG dataset. Finally, the dense layer with a softmax classifier is used to predict the binary or multi-class classification. In the following, the pre-trained models are described in brief.

VGG-19 VGG-19 (Simonyan & Zisserman, 2014) was introduced in the ILSVRC-14 challenge. This model has repetitive convolutional layers, activation layers, and pooling layers. VGG-19 consists of 19 layers (16 convolution layers, three fully connected layers, and five max pooling layers). In this study, we have replaced the two fully-connected layers consisting of 4096 neurons with 128 and 256 neurons. Therefore, VGG-19 has 20.1 million parameters.

DenseNet-121 DenseNet-121 (Huang, Liu, Van Der Maaten, & Weinberger, 2017) consists of one convolution layer with 7×7 kernel, 58 convolution layers with 3×3 kernel, 61 convolution layers with 1×1 kernel, and one dense layer with 1000 neurons. DenseNet-121 also

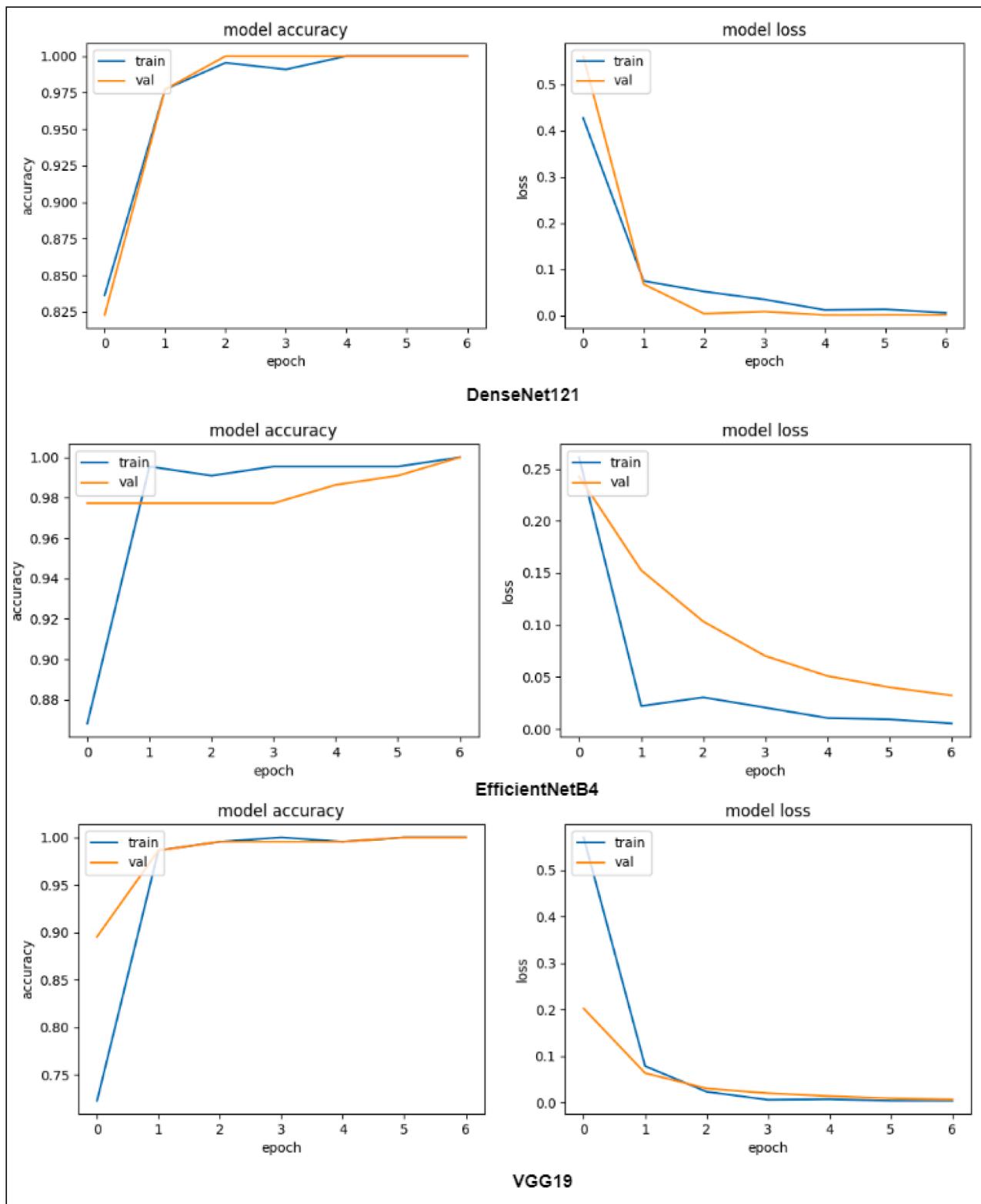


Fig. 4. Binary class classification: models accuracy and loss graph during training with no. of epochs.

has four average pooling layers to reduce the resolution. The layers are densely connected in a feed-forward manner to ensure the maximum flow of extracted features. In this study, the last dense layer with 1000 neurons is replaced with two dense layers of 128 and 256 neurons. Therefore, the number of parameters is 7.2 million.

EfficientNet-B4 EfficientNet-B4 (Tan & Le, 2019) have seven mobile

inverted bottleneck convolution (MBConv) as well as attention layers (Hu, Shen, & Sun, 2018). EfficientNet-B4 is 8.4 times smaller than the best existing ConvNet models and 6.1 times faster than the other best models in making inferences. In this study, the pre-trained dense layers are replaced with two dense layers having 128 and 256 neurons. In this study, the number of parameters in modified EfficientNet-B4 is 17.9

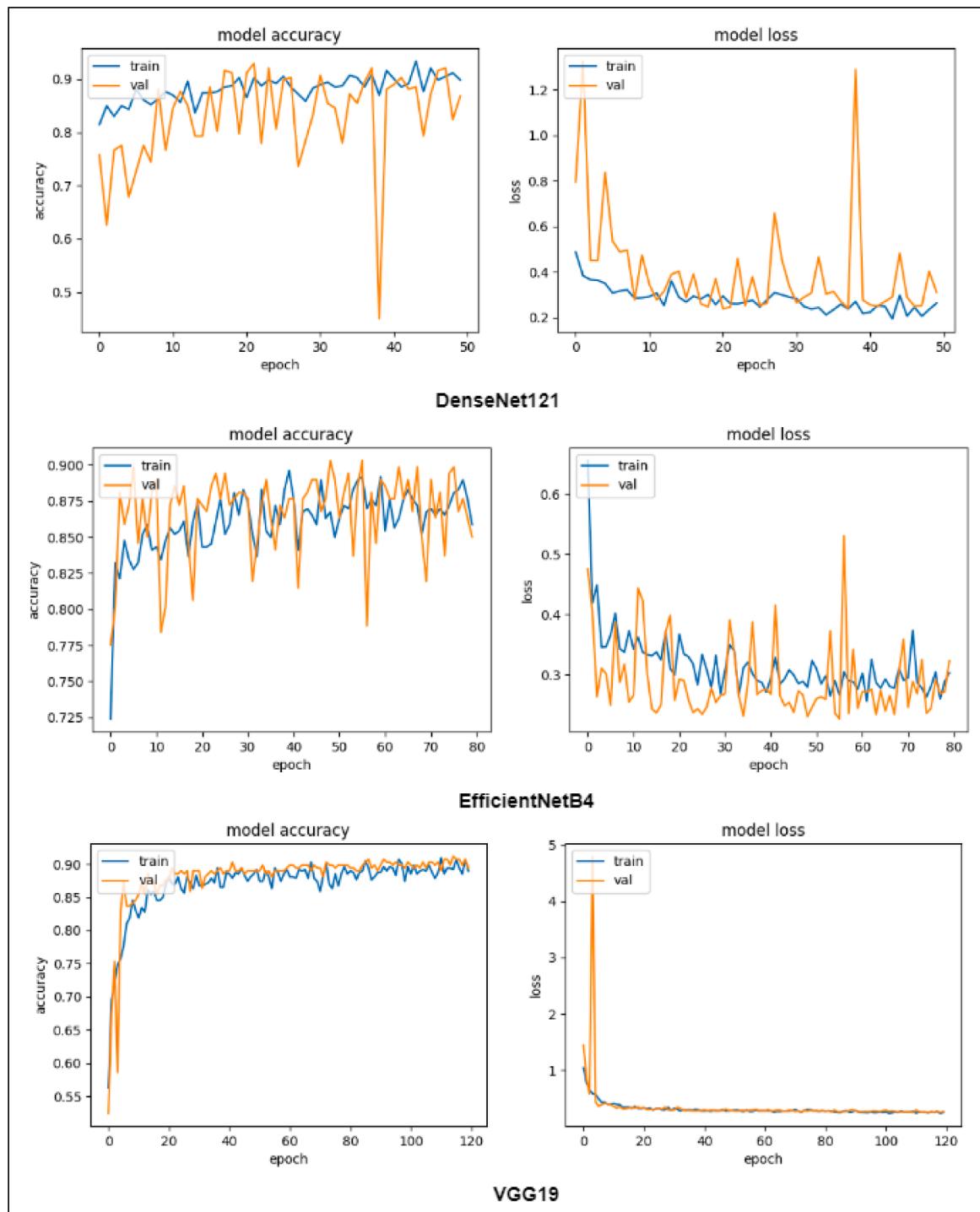


Fig. 5. Multi-class classification: models accuracy and loss graph during training with no. of epochs.

million.

At last, VGG-19, EfficientNet-B4, and DenseNet-121 models are ensemble and optimized weighted average is used for the final multi-class classification results for COVID-19 detection. The grid search technique used in finding the optimized weighted is discussed in Section 4.3.1. Fig. 7 shows the OWAE architecture.

4. Performance evaluation

Different evaluation metrics, as well as the formulation, are described in Section 4.1. The experimental setup is described in Section

4.2. While in Section 4.3, the results obtained for the binary and multi-class classification are discussed.

4.1. Evaluation metrics

The confusion matrix is used to evaluate the performance of the deep learning models for classification. The confusion matrix represents the true label and the predicted label in the tabular form. From the confusion matrix, the sensitivity or recall, precision, F1-Score, and accuracy parameters can be found. In the following, the formulas for calculating the above-mentioned parameters are given:

Short Title of the Article

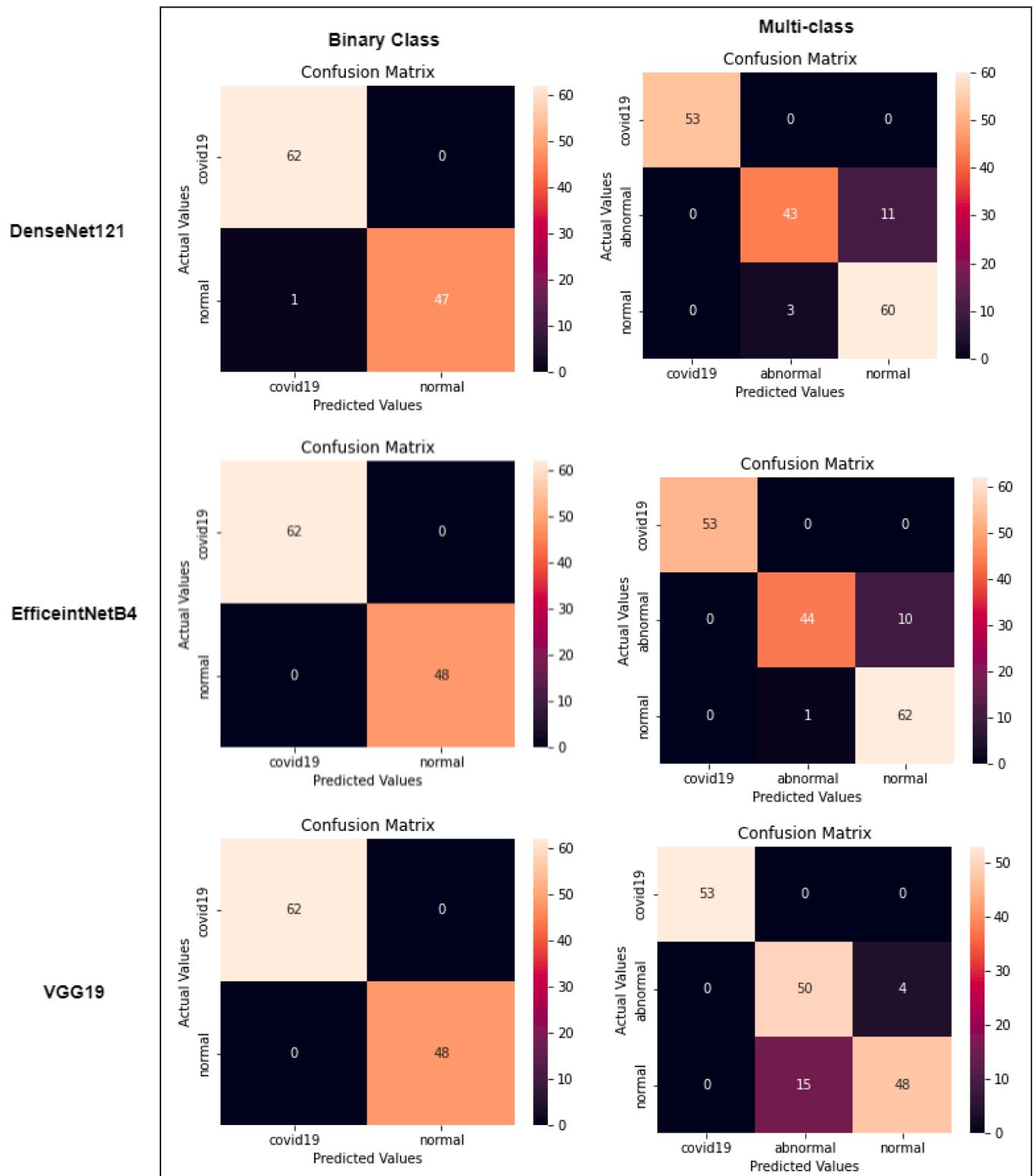


Fig. 6. Confusion matrix for binary class and multi-class classification using DenseNet-121, EfficientNet-B4, and VGG-19.

- Accuracy: Accuracy of any model shows how often it classifies the unseen data instance correctly.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision: The proportion of positive predictions made by the model that actually belongs to positive class is called Precision.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Short Title of the Article

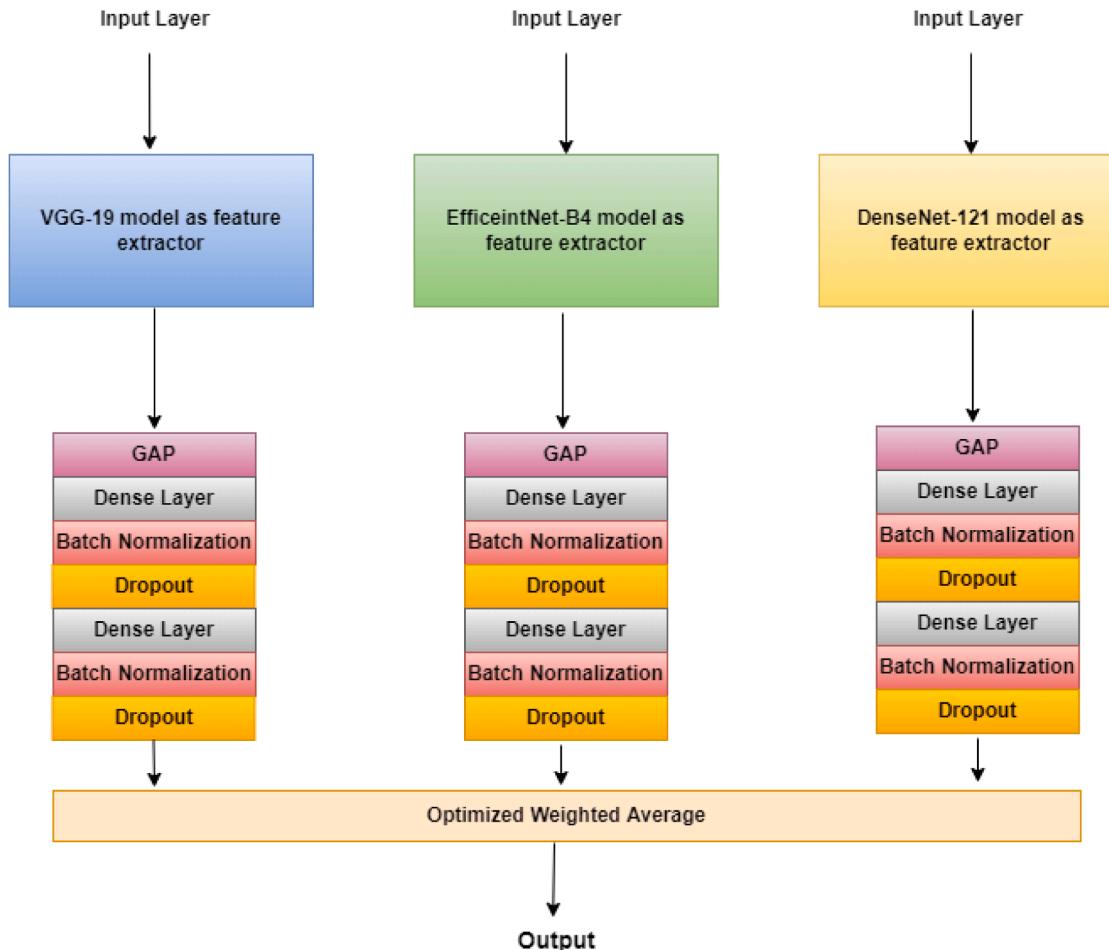


Fig. 7. Architecture of OWAE model.

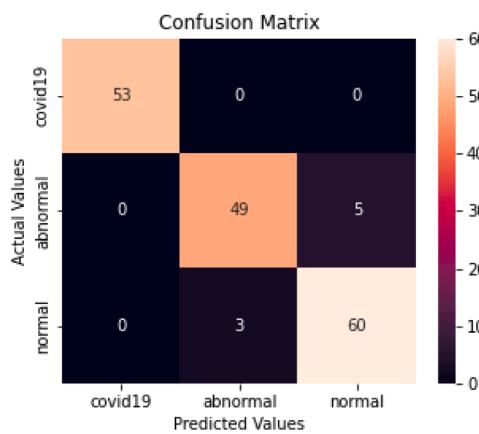


Fig. 8. Confusion matrix for multi-class classification using OWAE model.

3. Recall or sensitivity: The proportion of correct positive predictions made by the model out of all positive predictions that could have been made is called Recall.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. F1-score: F1-score is the harmonic mean of precision and recall.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

In the above equations, TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively. TP and TN represent the correct prediction of whether a subject has the particular disease or not. However, FP and FN determine the number of incorrect predictions made by the model.

4.2. Experimental setup

The Google Colab platform is used to train and test the models. It provides 12.68 GB RAM and NVIDIA Tesla K80 GPU for 12h. In this study, the Adam optimizer is used with a learning rate of 0.001. Binary cross-entropy and categorical cross-entropy loss functions are used for binary and multi-class classification, respectively. The number of epochs is set to 200 with a batch size of eight. Early stopping is used to prevent the model from overfitting. The patience in early stopping is fixed at three. It means that the model will stop training if the validation loss does not improve for three continuous epochs. Performance for the binary class and multi-class classification on all evaluation metrics is given in Tables 3 and 4, respectively. The accuracy and loss graphs for binary classification and multi-class classification are shown in Figs. 4 and 5,

Table 5

Comparative analysis of proposed method with existing similar work on same dataset.

Paper	Data used	Model	Performance
Attallah (2022)	ECG images	ECG-Bi-CoNet	For Binary class Accuracy- 98.8%, For Multi-class Accuracy- 91.73%
Ozdemir et al. (2021)	ECG images	CNN based model, with hexaxial feature mapping technique	For binary class Accuracy: 96.3%, For Multi-class Accuracy: 93.2%
Sobahi et al. (2022)	ECG images	3D-CNN model With attention mechanism	For binary class Accuracy: 99.0%, For Multi-class Accuracy: 92.0%
Rahman et al. (2022)	ECG Images	Transfer Learning models (ResNet18, ResNet50, ResNet101, InceptionV3, DenseNet201, and MobileNetv2)	For binary class, Accuracy: 99.1%, For three class, Accuracy: 97.36%
Proposed study	ECG images	Transfer learning (VGG-19, DenseNet121, EfficientNetB4)	For binary class, Accuracy: 100%, F1-score: 100%, Precision: 100%, Recall: 100%
Proposed study	ECG image	Transfer learning and OWAE	For multi-class, Accuracy: 95.29%, F1-score: 95.4%, Precision: 95.51%, Recall: 95.32%

respectively. The confusion matrix for both binary class classification and multi-class classification is shown in Fig. 6. The values of hyperparameters are shown in Table 2.

The dataset is divided into the training and test set. A three-fold cross-validation scheme has been applied to validate the model on the training dataset. The training dataset is divided into three sets. Two sets are used for the training, and the third is used to validate the model. The results obtained for the binary class and multi-class classification are shown in Tables 3 and 4. These results are recorded on the test set. The experimental setup used for the study is shown in Algorithm 1. Moreover, the framework of the method is shown in Fig. 3.

4.3. Results

In this section, we have discussed the results obtained in this study. First, we have evaluated three state-of-the-art models; VGG-19, DenseNet-121, and EfficientNet-B4 for multi-class classification individually. Then, these three models are appended and the optimized weighted average technique is used to enhance the performance. The same three deep learning models are further used for binary classification.

4.3.1. Multi-class classification

VGG-19, DenseNet-121, and EfficientNet-B4 are evaluated separately for multi-class classification. The VGG-19, EfficientNet-B4, and DenseNet-121 models obtained an accuracy value of 92.94%, 92.35%, and 92.95%, respectively. Further, VGG-19 reported an F1 score of 91.64%, a precision value of 89.97%, and a recall value of 95.94%. EfficientNet-B4 obtained an F1-score of 90.77%, a precision value of 89.78%, and a recall value of 93.82%. While the DenseNet-121 obtained 90.76% F1-score, 89.79% precision, and 93.20% recall. Fig. 6 shows the confusion matrix of multi-class classification of all three models.

As discussed in Section 3.2.2, the optimized weighted average is used to evaluate the performance of an ensemble model for multi-class classification. The grid search has been used to find the optimized weights. The architecture of the ensemble model is shown in Fig. 7. The domain for weights is defined from 0.0 to 1.0 in steps of 0.01, then by using the grid search method, all possible combinations of weights are generated. The combinations of weights which gave the highest performances are

[0.25, 0.50, 0.25] with an accuracy value of 95.29%. The OWAE model obtained an F1-score value of 95.40%, a precision value of 95.51%, and a recall value of 95.32%. This method outperformed all three models when they are evaluated separately. The accuracy and various performance evaluation metrics values are shown in Table 4. The confusion matrix of the OWAE model is shown in Fig. 8.

4.3.2. Binary classification

The COVID-19 and the normal class are used for the binary classification. The VGG-19 model reported 100% accuracy, F1-score, precision, and recall on the test set. Similarly, the EfficientNet-B4 also achieved 100% accuracy, F1-score, precision, and recall. The DenseNet-121 reported the 99.09% accuracy, 99.07% F1-score, 99.20% precision, and 98.95% recall. Table 3 and Fig. 6 shows the results obtained and the confusion matrix for binary classification, respectively. Individually, VGG-19 and EfficientNet-B4 have obtained the maximum results on all performance metrics. Therefore, no ensemble model is applied for the binary class classification.

5. Discussion

In this study, the deep learning models are used for predicting the COVID-19 in binary and multi-class classification using ECG images. The VGG-19 and EfficientNet-B4 obtained 100% accuracy, and DenseNet-121 reported 99.09% accuracy for the binary classification. VGG-19, EfficientNet-B4, and DenseNet-121 are evaluated separately as well as with the optimized weighted average ensemble method for multi-class classification. The OWAE model improved the performance in multi-classification by classifying the ECG images with 95.29% accuracy. The OWAE model outperformed the individual deep learning models, as can be observed by comparing the results given in Table 4. The summary of related studies done by researchers on the ECG dataset to predict COVID-19 is given in Table 5. As given in Table 5, the results reported in this study are better than those of related studies for binary and multi-class classification.

There are some limitations to overcome, which opens the door to improving the proposed work. The data used in this study was published in November 2020. Since then, many new variants of COVID-19 have been found, and certainly, the symptoms and effects of the infection have also varied. The ECG data from patients who have recently been diagnosed can be collected for further fine-tuning and testing of the model. The data used in this study is collected from one particular geographical location. As COVID-19 patients have been diagnosed all over the world, a more comprehensive data collection encompassing different geographical locations, age groups, and genders will ensure the effectiveness and robustness of the model. This study does not consider the ongoing treatment and medications of the patients that might alter the ECG signal during the recording of the data. In incremental learning, the input dataset is continuously used to increase the model knowledge, but in the proposed method, the models are not continuously learning. If there is any change in the dataset, then the proposed model needs to be trained again and this will affect the final results.

6. Conclusions

In this study, we have modified three pre-trained deep learning models for COVID-19 detection in ECG images. Deep learning techniques can be used to detect COVID-19 as it needs fewer human-to-human interactions in comparison to RT-PCR. In this study, we have classified the COVID-19 using ECG images with an accuracy of 100% for binary classification and 95.29% for multi-class classification. An optimized weighted average technique is used for the multi-class classification. This technique improved the classification results in comparison to when the deep learning models are evaluated separately. The limitation of the OWAE model is that it is trained and tested on a limited dataset. The proposed model needs to be validated on large ECG

datasets. Automatic analysis can potentially reduce the time spent on diagnosis because an ECG sample can be collected in a few minutes and a deep learning model can process the ECG image in less than a second. Moreover, it can also help a lot in the diagnosis process in remote areas, as ECG samples can be transmitted over the internet, which will reduce the cost involved.

CRediT authorship contribution statement

Kunwar Prashant: Conceptualization, Methodology, Software. **Prakash Choudhary:** Supervision, Writing – original draft. **Tarun Agrawal:** Investigation, Visualization, Validation, Writing – original draft. **Ebam Kaushik:** Data curation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

- Abirami, N., Vincent, D. R., & Kadry, S. (2021). P2P-COVID-GAN: Classification and segmentation of COVID-19 lung infections from CT images using GAN. *International Journal of Data Warehousing and Mining (IJDWM)*, 17(4), 101–118.
- Abirami, R. N., Vincent, P., Rajinikanth, V., & Kadry, S. (2022). COVID-19 classification using medical image synthesis by generative adversarial networks. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 13, 385–401.
- Agrawal, T., & Choudhary, P. (2021). Focuscovid: Automated COVID-19 detection using deep learning with chest X-ray images. *Evolving Systems*, 13, 1–15.
- Agrawal, T., & Choudhary, P. (2022). Segmentation and classification on chest radiography: A systematic survey. *The Visual Computer*, 1–39. <https://doi.org/10.1007/s00371-021-02352-7>
- Al-Alim, A. M., Hamad, M. A., & AL-ledani, A. A. (2020). Some insights of novel COVID 19 virus: Structure, pathogenicity and immunity aspects. *Iraqi Journal of Veterinary Sciences*, 34(2), 287–293.
- Ashley, E. A., & Niebauer, J. (2004). Arrhythmia. *Cardiology explained*. Remedica.
- Attallah, O. (2022). ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using bi-layers of deep features integration. *Computers in Biology and Medicine*, 142, 105210.
- Attia, Z. I., Kapa, S., Noseworthy, P. A., Lopez-Jimenez, F., & Friedman, P. A. (2020). Artificial intelligence ECG to detect left ventricular dysfunction in COVID-19: A case series, vol. 95. *Mayo clinic proceedings* (pp. 2464–2466). Elsevier.
- Bertini, M., Ferrari, R., Guardigli, G., Malagù, M., Vitali, F., Zucchetti, O., et al. (2020). Electrocardiographic features of 431 consecutive, critically ill COVID-19 patients: An insight into the mechanisms of cardiac involvement. *EP Europace*, 22(12), 1848–1854.
- Byer, E., Ashman, R., & Toth, L. A. (1947). Electrocardiograms with large, upright T waves and long QT intervals. *American Heart Journal*, 33(6), 796–806.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.
- Ekbal, A., & Saha, S. (2013). Stacked ensemble coupled with feature selection for biomedical entity extraction. *Knowledge-Based Systems*, 46, 22–32.
- Fernández-de Las-Peñas, C., Palacios-Ceña, D., Gómez-Mayordomo, V., Florencio, L. L., Cuadrado, M. L., Plaza-Manzano, G., et al. (2021). Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: A systematic review and meta-analysis. *European Journal of Internal Medicine*, 92, 55–70.
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
- Kamal, M., Abo Omirah, M., Hussein, A., & Saeed, H. (2021). Assessment and characterisation of post-COVID-19 manifestations. *International Journal of Clinical Practice*, 75(3), e13746.
- Khan, A. H., Hussain, M., & Malik, M. K. (2021). ECG images dataset of cardiac and COVID-19 patients. *Data in Brief*, 34, 106762.
- Lam, K., McClelland, S., & Dallo, M. J. (2020). ECG: Essential in care of patients with COVID-19. *The Medical Journal of Australia*, 213(10), 476.
- Li, L., Zhang, S., He, B., Chen, X., Wang, S., & Zhao, Q. (2020). Risk factors and electrocardiogram characteristics for mortality in critical inpatients with COVID-19. *Clinical Cardiology*, 43(12), 1624–1630.
- Lilly, L. S. (2012). *Pathophysiology of heart disease: A collaborative project of medical students and faculty*. Lippincott Williams & Wilkins.
- Mahendran, N., Vincent, D. R., Srinivasan, K., Chang, C.-Y., Garg, A., Gao, L., et al. (2019). Sensor-assisted weighted average ensemble model for detecting major depressive disorder. *Sensors*, 19(22), 4822.
- Oran, D. P., & Topol, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection: An narrative review. *Annals of Internal Medicine*, 173(5), 362–367.
- Ozdemir, M. A., Ozdemir, G. D., & Gurel, O. (2021). Classification of COVID-19 electrocardiograms by using hexaxial feature mapping and deep learning. *BMC Medical Informatics and Decision Making*, 21(1), 1–20.
- Parry, J. (2020). China coronavirus: Cases surge as official admits human to human transmission.
- Rahman, T., Akinbi, A., Chowdhury, M. E., Rashid, T. A., Şengür, A., Khandakar, A., et al. (2022). COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network. *Health Information Science and Systems*, 10(1), 1–16.
- Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4), 2000058.
- Rutten, L. J. F., Zhu, X., Leppin, A. L., Ridgeway, J. L., Swift, M. D., Griffin, J. M., et al. (2021). Evidence-based strategies for clinical organizations to address COVID-19 vaccine hesitancy, vol. 96. *Mayo clinic proceedings* (pp. 699–707). Elsevier.
- Santoro, F., Monitillo, F., Raimondo, P., Lopizzo, A., Brindicci, G., Gilio, M., et al. (2021). QTc interval prolongation and life-threatening arrhythmias during hospitalization in patients with coronavirus disease 2019 (COVID-19): Results from a multicenter prospective registry. *Clinical Infectious Diseases*, 73(11), e4031–e4038.
- Shahhosseini, M., Hu, G., & Pham, H. (2022). Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, 7, 100251.
- Shashvat, K., Basu, R., Bhondekar, A. P., & Kaur, A. (2019). A weighted ensemble model for prediction of infectious diseases. *Current Pharmaceutical Biotechnology*, 20(8), 674–678.
- Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 24, 91–98.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sobahi, N., Sengur, A., Tan, R.-S., & Acharya, U. R. (2022). Attention-based 3C CNN with residual connections for efficient ECG-based COVID-19 detection. *Computers in Biology and Medicine*, 143, 105335.
- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning* (pp. 6105–6114). PMLR.
- Wang, C. C., Prather, K. A., Sznitman, J., Jimenez, J. L., Lakdawala, S. S., Tufekci, Z., et al. (2021). Airborne transmission of respiratory viruses. *Science*, 373(6558), eabd9149.
- Wang, Y., Chen, L., Wang, J., He, X., Huang, F., Chen, J., et al. (2020). Electrocardiogram analysis of patients with different types of COVID-19. *Annals of Noninvasive Electrocardiology*, 25(6), e12806.
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese center for disease control and prevention. *JAMA*, 323(13), 1239–1242.