

Optimizing Multi Neural Network Weights for COVID-19 Detection Using Enhanced Artificial Ecosystem Algorithm



Hakan Koyuncu^{1*}, Munaf Arab²

¹ Computer Engineering Department, Altinbas University, Istanbul 34217, Turkey

² Electrical and Computer Engineering Department, Altinbas University, Istanbul 34217, Turkey

Corresponding Author Email: hakan.koyuncu@altinbas.edu.tr

<https://doi.org/10.18280/ts.400417>

ABSTRACT

Received: 30 January 2023

Revised: 16 July 2023

Accepted: 10 August 2023

Available online: 31 August 2023

Keywords:

COVID-19, multilayer perceptron,
convolutional neural network,
metaheuristics, optimization, enhanced
artificial ecosystem algorithm,
segmentation, pneumonia

The role of machine learning in medical research, particularly in addressing the COVID-19 pandemic, has proven to be significant. The current study delineates the design and refinement of an artificial intelligence (AI) framework tailored to differentiate COVID-19 from Pneumonia utilizing X-ray scans in synergy with textual clinical data. The focal point of this research is the amalgamation of diverse neural networks and the exploration of the impact of metaheuristic algorithms on optimizing these networks' weights. The proposed framework uniquely incorporates a lung segmentation process using a pre-trained ResNet34 model, generating a mask for each lung to mitigate the influence of potential extraneous features. The dataset comprised 579 segmented X-ray images (Anteroposterior and Posteroanterior views) of COVID-19 and Pneumonia patients, supplemented with each patient's textual medical data, including age and gender. An enhancement in accuracy from 94.32% to 97.85% was observed with the implementation of weight optimization in the proposed framework. The efficacy of the model in detecting COVID-19 was further ascertained through a comprehensive comparison with various architectures cited in the existing literature.

1. INTRODUCTION

In the early years of the 21st century, the World Health Organization (WHO) identified Severe Acute Respiratory Syndrome (SARS), a highly transmissible and severe virus, marking a pivotal moment in global health [1]. Nearly two decades later, a novel virus known as SARS-COV-2 or COVID-19 was discovered in Wuhan, China, sharing 80% of its genome with SARS. This new virus has since caused over 160 million confirmed cases and more than 4.8 million deaths by October 2021 [2]. Characterized by its droplet-based transmission and high mutation rate, COVID-19, declared as a pandemic by the WHO on March 11, 2020, has severely challenged global health infrastructures due to a lack of appropriate testing kits and increased pressure on intensive care units.

The reverse transcription-polymerase chain reaction (RT-PCR) test has been recognized as the standard method for detecting COVID-19 [3]. However, circumstances where RT-PCR tests are not viable have been observed, including cases where patients are physically unable to undergo such a test, such as those diagnosed with throat cancer [4]. The insufficient distribution and training for test kits in rural communities further exacerbates the challenges in containing the viral infection. Furthermore, the limitations of RT-PCR have been exposed by research indicating its inability to detect long-term effects of COVID-19 - a phenomenon highlighted in the work of Mallett et al. [5]. The need for alternative diagnostic methods, such as medical image analysis using X-ray scans, has thus become evident.

The last decade has witnessed significant advancements in

Artificial Intelligence (AI), with its applications spanning various fields and often outperforming traditional computing systems [6]. In particular, the potential of AI in medical image analysis has been realized, enabling automation of tasks previously exclusive to specialists [7]. Motivated by these developments, research efforts have shifted toward public health, especially since 2020. Numerous researchers have begun leveraging non-invasive methods like X-ray scans to detect COVID-19 and associated post-infection symptoms [8].

Deep learning (DL) algorithms, such as convolutional neural networks (CNNs) and multilayer perceptrons (MLPs), have demonstrated their impressive capabilities in computer vision tasks related to medical image analysis [9]. While CNNs excel at pattern and feature extraction, MLPs process textual clinical features, enabling more comprehensive decision-making. However, the importance of accuracy in these tasks often results in overlooking the incorporation of medical information during training, such as age and gender, which could potentially enhance classification results.

Training techniques for DL networks like gradient descent (GD) and backpropagation (BP) often face challenges that can hinder model accuracy [10]. These limitations can potentially be mitigated by employing metaheuristics optimization methods [11]. The use of such techniques can help navigate around local optima and improve performance. Moreover, the use of multiple layers can introduce a vanishing or exploding gradient problem during backpropagation, complicating the attainment of reliable training results.

Image segmentation plays a critical role in medical image data analysis by isolating relevant regions and extracting informative features [12]. Precise segmentation of X-ray

images can aid in identifying and localizing anatomical structures and abnormalities, eliminating non-essential information that could potentially influence the model's detection rate.

The primary objective of this study is to design and validate a novel weight-optimized Deep Learning model, leveraging CNN-MLP networks and integrating several image augmentation techniques such as image segmentation and histogram equalization for superior classification. The proposed framework utilizes medical images and textual clinical data as inputs to differentiate between COVID-19 and pneumonia in affected individuals.

2. RELATED WORK

Chest X-ray scans are utilized by many researchers, for instance, Apostolopoulos and Mpesiana utilized pre-trained CNN models to classify COVID-19 from X-ray images [13]; and due to the small dataset of 224 COVID-19 positive samples, the authors employed transfer learning. Wang et al. [14] proposed a deep learning model based on the design pattern of PEPX. The model was first pre-trained on the ImageNet dataset, and then the COVIDX dataset was used to retrain their model. Also, image augmentation was utilized. The authors reported an accuracy of 92.4% with 80% sensitivity for the COVID-19 class. Apart from this, Mukherjee et al. [15] concluded that a single CNN architecture could be trained on X-ray and CT scans. In their report, the proposed tailored, simple architecture achieved an accuracy of 96.28% on a dataset that has balanced classes (number of samples) of 672 samples for X-ray and CT scans.

Several studies that utilized machine learning (ML) algorithms using numerical/categorical data highlighted the importance of merging different data types in detecting COVID-19 in chest X-ray and CT images. For instance, Lassau et al. [16] used a deep learning-based CT scan model with textual clinical information. In their report, the authors found 12 variables that are significantly associated with severity such as age, gender, and oxygen saturation, highlighting the importance of combining several data types to achieve a lower misclassification rate. Likewise, the authors [17] proposed a multi-modal with the late fusion of textual clinical data to classify and assess the severity of COVID-19 infection.

Typically, network weights are trained using a variant of Gradient Descent (GD). The drawback with gradient descent is that the weight update at a given time (t) is determined solely by the learning rate and gradient at that time. It does not consider the previous actions performed when navigating the search space [18], in their pursuit of achieving global optima. A solution to this inherited drawback is performing significant update steps or using other searching methods such as Metaheuristics.

Metaheuristics algorithms can guide the search for near-optimal solutions, by efficiently exploring the search space without searching every possible solution. Classical Metaheuristics algorithms, such as Genetic Algorithms [19], Simulated Annealing [20], have shown their suitability to solve complex scheduling problems, space allocation problems, and learning problems.

Metaheuristic algorithms are less prone to get trapped in local optima. The use of such algorithms can be beneficial in training and optimization problems. For instance, Particle Swarm Optimization was used to optimize the weights and

architecture of the MLP network proposed in study [21]. The authors proposed the use of two separate PSOs (PSO and PSO with weight decay). The first PSO is used to search the number of units in a hidden layer, while the PSO with weight decay is responsible for optimizing the weights. Furthermore, the authors [22] applied an enhanced version of the black hole algorithm [23] to search for optimal weights and biases, this method achieved a better result than other MHAs that have been tested by the authors. Several research papers employed metaheuristics in training the network, in the study [24], used simulated annealing, which is a metaheuristic algorithm, for improving the network performance, after training, and fed to the optimization algorithm, the optimized weights are then returned to the network.

3. MATERIALS AND METHODS

3.1 Datasets

Table 1. The number of samples per class

Class	Number of Images	Age (Years)
COVID-19	257	53.29±15.23
Pneumonia	322	41.73±21.66

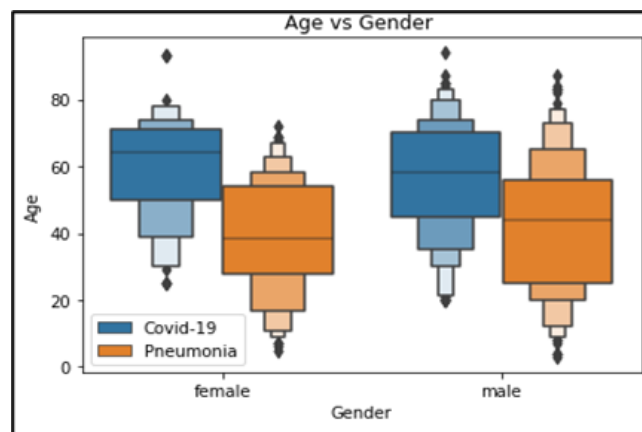


Figure 1. COVID-19 and pneumonia distribution in the dataset, considering age and gender

The datasets used in the experiments are compiled from two different public sources. While both datasets include over 122k samples, the utilization of only X-ray images with a single pathogen drastically limited the number of samples that could be utilized to train the neural network. The dataset is divided into two subsets training and validation (testing) sets with 70% training set and 30% testing set. Table 1 shows the sample size in the dataset, whereas Figure 1, provides an insight into the age and gender distribution in the dataset.

Cohen/IEEE 8023 dataset [25] is a collection of X-ray images that have been extracted from online publications, websites, as well as through indirect collection from hospitals and physicians. The dataset contains several pathogen types, including, but not limited to, COVID-19, Pneumonia, SARS. Furthermore, textual clinical data such as age and gender are also present in the dataset. Only COVID-19 was used in this study from the previously mentioned dataset due to the small number of samples in other classes such as pneumonia (less than 30 images).

The NIH Chest X-ray dataset [26] is used as a source for non-COVID-19 pneumonia images and comprises 112,120

images from 30,805 unique patients. The dataset contains X-ray images of 14 different pathogen types; the dataset also contains textual clinical data such as age and gender for each image. The employment of this dataset mitigates the lack of specific non-COVID-19 pneumonia data in Cohen/IEEE 8023 dataset.

Only X-ray images with PA and AP views are employed. Two images of each patient are processed if available; the difference in images is the time between each scan. The training set is rescaled while the augmentation is randomly applied to achieve higher generalization. The augmentation pipeline consists of six steps, initial transformation is first, in this step the X-ray image is scaled to 1024×1024 and cropped while maintaining the aspect ratio of the original image. After the transformation, the X-ray image is fed into the segmentation model and the lungs are extracted. After segmentation of the lungs, the color space transformation is carried out, which transform the images to grayscale, if the source image is not in the required color space (grayscale), thus standardizing the images in the dataset regarding the color space. Then a final scaling of the X-ray images is carried out to scale the image to 256×256 based on the network requirement. Contrast limited adaptive histogram equalization (CLAHE) [27] is applied next, which enhance the contrast of the X-ray image. Finally, in the last step an augmentation transforms the image randomly and generate images that differ from the original images, this step is only applied for the training set.

3.2 Proposed network

The proposed network architecture is a hybrid model that combines Convolutional Neural Networks (CNN) [28] and Multi-Layer Perceptron (MLP) [29] networks. This hybrid network architecture aims to leverage the strengths of CNNs in processing image data and MLPs in handling textual clinical data.

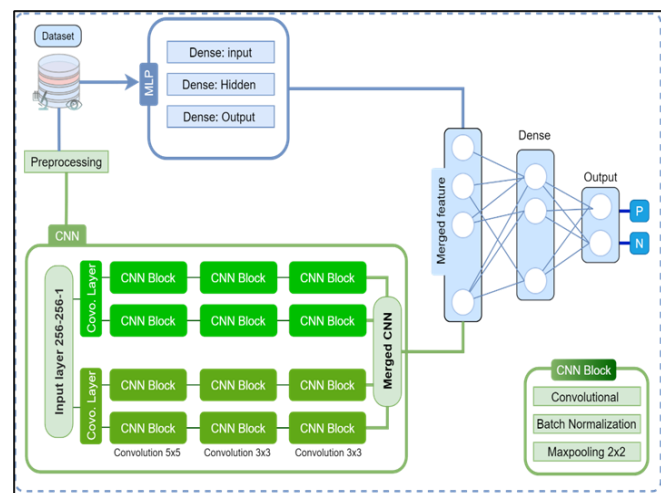


Figure 2. Proposed architecture for CNN-MLP network, illustrating both CNN network and MLP network

The architecture, as depicted in Figure 2, depicts the CNN network for processing image data and MLP for handling numerical/categorical data, such as textual clinical information. The model incorporates three input layers: two for the image

data and one for the textual clinical data. Using two input layers for the image data in the proposed network architecture allows for the incorporation of different types or sources of image data. While using multiple branches allows for the parallel processing and extraction of features from different aspects.

The features extracted from the X-ray images using CNN are concatenated with the MLP network’s output, the resulted data is flattened and fed to the fully connected layers for prediction. Focal loss for binary classification [30] is employed to address the class imbalance that is present in the dataset. Binary focal loss, which is an improved version of Cross-Entropy Loss (CE), introduced a parameter called focusing parameter (γ) which allows instances that are difficult to classify to be penalized more severely than instances that are easy to classify. The focal loss is defined as:

$$L(y, \hat{p}) = -\alpha y(1 - \hat{p})^\gamma \log(\hat{p}) - (1 - y)\hat{p}^\gamma \log(1 - \hat{p}) \quad (1)$$

where, $y \in \{0, 1\}$ is a binary class label, $\hat{p} \in [0, 1]$ is an estimate of the probability of the positive class, γ is the focusing parameter, α is a hyperparameter that governs the trade-off between precision and recall.

The neural network was built using the Keras library and TensorFlow. Adam optimizer is employed with learning decay: the learning rate decays if a monitored variable stagnates during training for a specified number of epochs (i.e., ‘patience’). The following are the network hyperparameters and are selected using grid search: epochs=50, batch size=16, initial learning rate=0.001, decay factor=0.96 and patience=5.

3.3 Using metaheuristics for enhancing classification performance

The use of metaheuristics algorithms is to explore the search space for a better value based on a set of pre-defined steps that systemically improves parts that are deemed to have poor performance based on calculating the prediction performance.

3.3.1 Enhanced artificial ecosystem optimization

Enhanced Artificial Ecosystem Optimization (EAEO) [31] is a nature-inspired metaheuristic based on Artificial Ecosystem Optimization (AEO) [32]. EAEO is a system-based optimizer that mimics the flow of energy in an ecosystem in three phases (production, consumption, decomposition) which are mimicked from the unique behaviour of the living organism in an ecosystem. A modified EAEO (MEAEO) algorithm Figure 3 is proposed to enhance the performance of the CNN network by optimizing the weights.

Weight optimization is carried out when the trained weights of the network are extracted and taken as an individual then vectorised for the initial population of ME-AEO. Subsequently, a super-individual is created that may hinder the search for optimal value; to solve this, EAEO employs randomness in the population initialization. Furthermore, after each iteration a random shuffle of data is employed to avoid misdirection of the search when given orderly data to improve the optimization efficiency of MEAEO which in turn further improves the classification accuracy.

Algorithm 1: MEAEO algorithm

```

Randomly initialize an ecosystem  $X_i$  ( solutions) and calculate the fitness  $Fit_i$ 
 $X_{best}$  = the extracted weights.
for  $i \leftarrow 0$  to  $Maxiteration$  do
  Production stage:
   $G = 2 \times (1 - i/Maxiteration)$ 
   $C = 0.5 v_1 / |v_2|$ 
   $X_2(t+1) = x_2(t) + G \times C \times [x_2(t) - x_1(t)]$ 
  Consumption stage:
  for  $i \leftarrow 2$  to  $MaxPopulation$  do
     $j = randi([2 \ i-1])$ 
     $X_{ij} = X_i(t) - x_j(t)$ 
     $r = randvector(0,1)$ ,  $r_2 = rand(0,1)$ ,  $r_3 = 2\pi \times rand$ ,  $r_4 = rand$ 
    Herbivore
    if  $r_{1j} < 1/3$  then
      if  $r_{2j} < 2/3$  then
         $x_1(t+1) = x_1(t) + sin(r_2) \times C \times [x_i(r_3) - x_1(t)]$ 
      else
         $x_1(t+1) = x_1(t) + cos(r_2) \times C \times [x_i(r_3) - x_1(t)]$ 
      end
    end
    Carnivore
    else if  $1/3 > r_{1j} > 2/3$  then
      if  $r_2 > 2/3$  then
         $x_i(t+1) = x_i(t) + sin(r_3) \times C \times x_{ij}(t)$ 
      else
         $x_i(t+1) = x_i(t) + cos(r_3) \times C \times x_{ij}(t)$ 
      end
    end
    Omnivore
    else
      if  $r_2 < 0.5$  then
         $x_1(t+1) = x_1(t) + sin(r_4) \times C [r_4 x_{i1}(t) + (1-r_4) x_{ij}(t)]$ 
      end
      else
         $x_1(t+1) = x_1(t) + cos(r_4) \times C [r_4 x_{i1}(t) + (1-r_4) x_{ij}(t)]$ 
      end
    end
  end
  Calculate the fitness of each individual.
end
Decomposition stage:
Update the best solution found so far  $X_{best}$ 
for  $i \leftarrow 0$  to  $MaxPopulation$  do
   $D = 3u$ ,  $u \sim N(0,1)$ 
   $e = r_3 \times randi([1, 2] - 1)$ ,  $h = 2 \cdot 3 \cdot 1$ 
   $x_1(t+1) = x_n(t) + D [e x_n(t) - h x_1(t)]$ 
  if  $X_{best} fitness < x_i fitness$  then
     $X_{best} = x_i$ 
  end
end
end
Return  $X_{best}$ 

```

Figure 3. Pseudo-code for the enhanced Artificial ecosystem algorithm

EAEAO consists of three phases: production, consumption, and decomposition. In the production phase, the worst individual is replaced by a new individual, which is mutated by the lower and upper search space and the best individual. The newly created individual is purposed to guide other individuals in searching different regions. In the consumption phase. The new individual is further mutated, to improve the search for a better solution. There are three types of consumers: Herbivore, Carnivore and Omnivore, which are chosen based on a random value. In the Decomposition phase, decomposition coefficients are utilized for a faster decay of the weights; as a result, the decomposition improves the exploration of the EAEAO algorithm. The network weights will be replaced if the optimized weights are better than the original network weights.

3.4 Experimental setup

The performance of the proposed framework was evaluated in terms of the impact of image augmentation and segmentation, and weight optimization. It is important to note that each experiment uses the same neural network discussed in section 2.2. And only differs in the pre-processing steps, which addresses the issues of superficial features that the network learns from. Image augmentation is used in all experiments, while image segmentation is applied only in experiment 3.

Metrics that were used to evaluate the performance are accuracy, precision, recall (sensitivity), and F1-score [33].

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1score = \frac{2 * precision * recall}{precision + recall} \quad (5)$$

where, TP =True Positives, TN =True Negatives, FP =False Positives, and FN =False Negatives.

Confusion Matrix is also employed, which is an important metric that describes the performance of the classifier on the test dataset.

4. RESULTS

The performance of the proposed hybrid network and the result from the different experiments are presented in this section.

An evaluation of the performance impact of merging images and textual data is investigated. While using the dataset, previously discussed in Section 3.1., the result showed an improvement in detection accuracy when using several data types, which are presented in Table 2. Images and textual data had an accuracy of 94% compared to using only images with an accuracy of 84%. The precision and sensitivity of using merged data types in the COVID-19 class averaged 93% while the average results while using images were 84%.

4.1 Optimization and segmentation experiments

The result in this section is obtained from the evaluation of the proposed network against the test subset, in other words, the result represents the performance of unseen data. A five-fold cross-validation technique has been utilized to average the network's performance using an independent testing set. Table 2 summarizes the averaged results obtained from the experiments that have been performed in this study.

Table 2. Performance of the different types of data used

Data Type	Class	Precision	Recall (Sensitivity)	F1	Acc.
Images	COVID-19	80%	86%	83%	84%
	Pneumonia	89%	83%	86%	
Images+TextualData	COVID-19	94%	95%	94%	94%
	Pneumonia	94%	93%	94%	

Likewise, the confusion matrices are presented in Figure 4, which compares the actual true values with those predicted by the network. The best accuracy obtained from the experiments while using the MEAEO for weight optimization is 97.85% compared to 94.23%. The result pointed out an even prediction distribution of the classes compared without optimizing the weights; the remaining metrics follow the same trend.

For experiment 1, the network is trained using image augmentation without segmentation or cropping, while CLAHE is employed regardless. Image and clinical information are provided as input to the neural network.

Without weight optimization, the detection performance varies for all classes (the Precision ranges from 91-93%), as shown in Table 3 while using the modified EAE0 for weight optimization, an increment is observed in sensitivity for detecting COVID-19 (the Precision ranges from 88-93%). Sensitivity in Table 3 for all classes reaches a value of more than 90% meanwhile, the non-optimized weights show lower accuracy, as shown in Table 3 and Figure 4 (a, d). Regarding general performance, sensitivity values reached 92% with the weight optimization and 86% without weight optimization, also supported by the sensitivity values. The confusion matrix, Figure 4 (a, d), shows that a small number of samples are misclassified in the COVID-19 class and can be attributed to

weight optimization. Figure 4 compares the result of weight optimization with the result obtained from not using the optimization in terms of F1-Score. The development of accuracy and loss during training and evaluation is depicted in Figure 5, while Figure 6 depicts the improvement gained in using the proposed method.

Experiment 2 applies image cropping and CLAHE, image augmentation was also used, which are employed as discussed in 3.1. The result from experiment 2 shows that experiment 1 outperformed this experiment in terms of accuracy. The detection performance increment while using weight optimization is in line with experiment 1.

Using the proposed optimization approach, the result in Table 3, shows the Precision values ranging from 86% to 91%, compared to 85% to 87% without utilizing the proposed approach for weights optimization. While Sensitivity values ranging from 85% to 92%. The overall performance while using the modified EAE0 for weight optimization reached an accuracy of 91.24% and 89.50% without optimizing the weights. Figure 4 (b, e) shows the prediction performance of the optimization algorithm used. The development of accuracy and loss during training and evaluation is depicted in Figure 7. While the F1-score increment is shown in Figure 8.

Table 3. The general performance of all experiments

Exp.	Class	CNN-MLP with Weight Optimization				CNN-MLP without Weight Optimization			
		Precision	Recall (Sensitivity)	F1	Acc.	Precision	Recall (Sensitivity)	F1	Acc.
Exp.1	Covid-19	92.15±1.58	92.33±0.18	92.22±1.32	91.21±2.98	93.21±1.02	86.52±0.11	89.68±0.28	91.37±3.68
	Pneumonia	91.34±0.43	91.25±1.28	91.41±0.73		88.88±0.73	94.45±1.00	91.61±0.67	
Exp.2	Covid-19	86.73±1.20	91.75±0.35	91.41±0.26	91.24±4.52	87.87±0.00	92.42±1.31	89.63±0.48	89.50±5.73
	Pneumonia	94.32±0.75	91.11±1.75	91.92±0.96		91.91±1.05	85.75±0.28	88.88±0.73	
Exp.3	Covid-19	95.42±0.88	99.05±0.50	97.07±0.65	97.85±3.86	94.62±1.23	95.65±1.25	94.85±1.82	94.23±2.25
	Pneumonia	99.50±0.50	93.37±0.75	97.17±0.80		94.44±0.50	93.13±0.74	94.44±0.18	

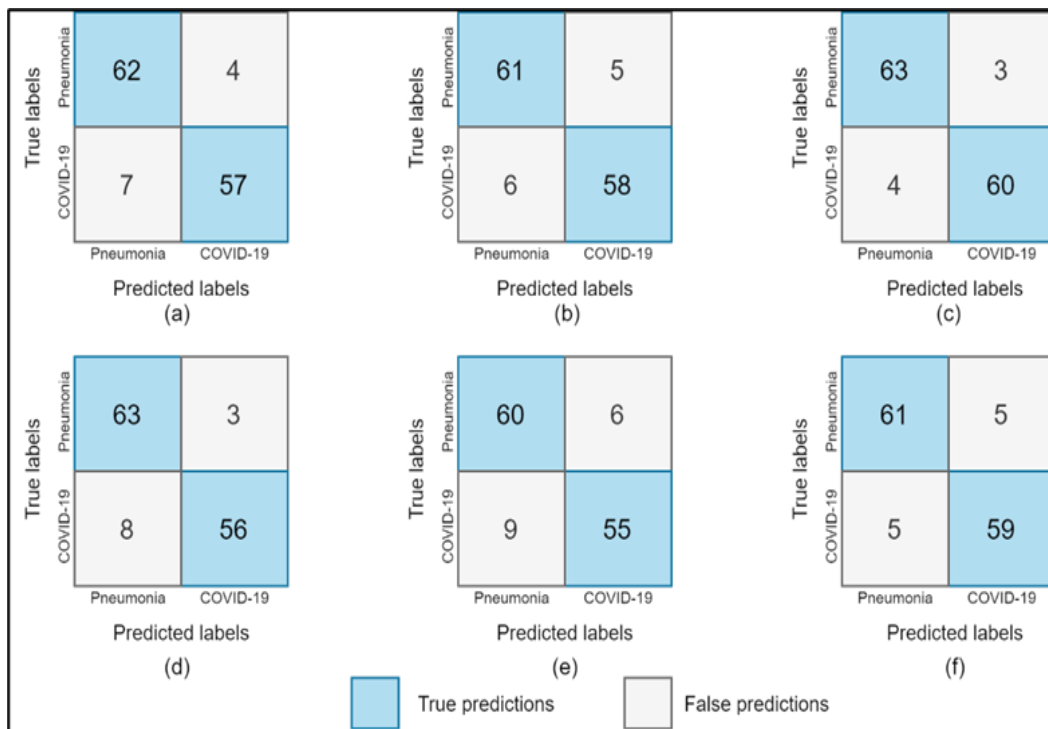


Figure 4. Confusion matrices for each of the experiments, considering each of the classes separately top: with weight optimization, Bottom, without Weight optimization

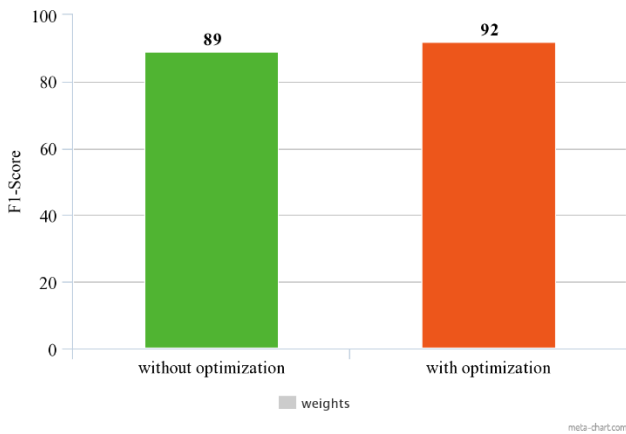


Figure 5. F1-score for experiment 1 considering the weight optimization

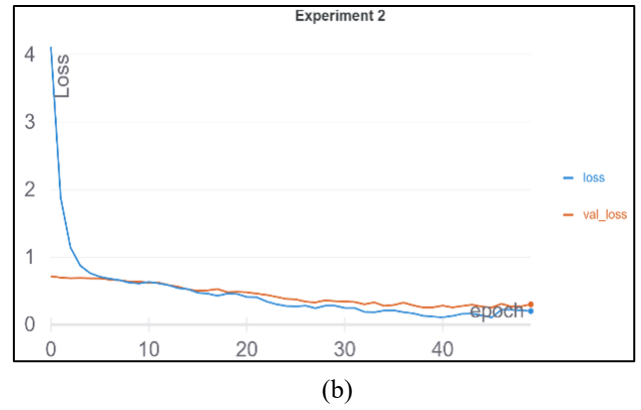
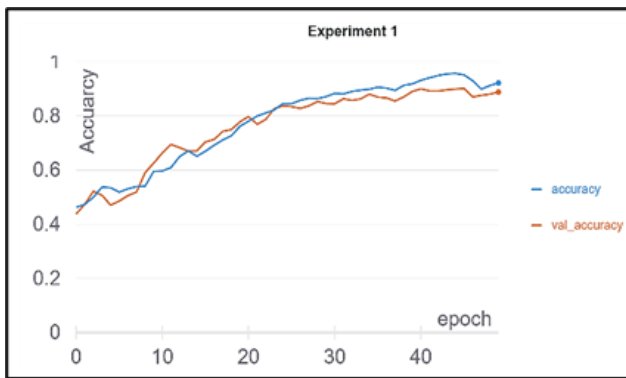
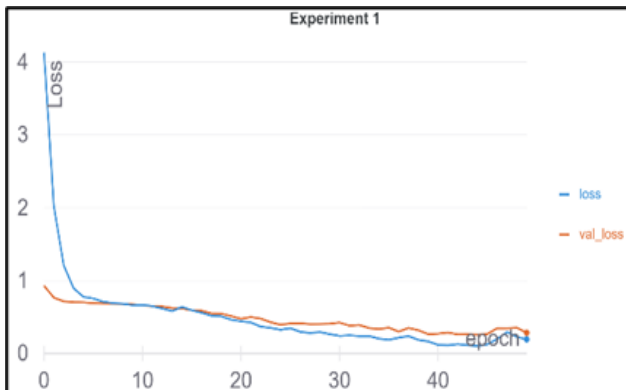


Figure 7. (a) Accuracy and validation accuracy during training for experiment 2; (b) loss and validation loss for experiment 2

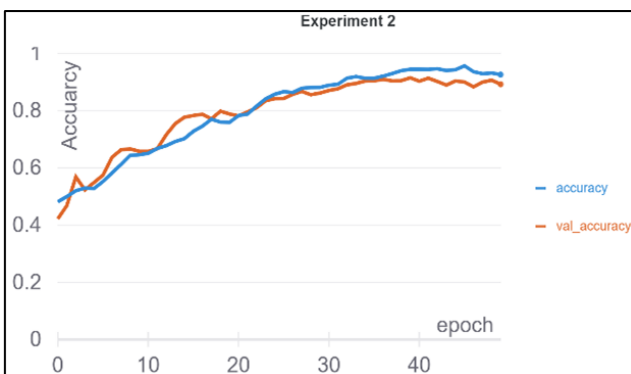


(a)



(b)

Figure 6. (a) Accuracy and validation accuracy during training for experiment 1; (b) loss and validation loss for experiment 1



(a)

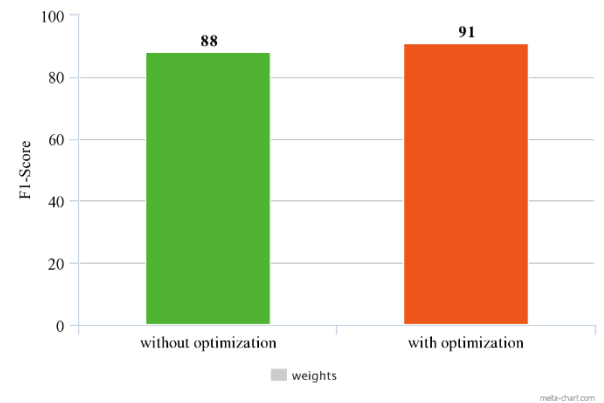
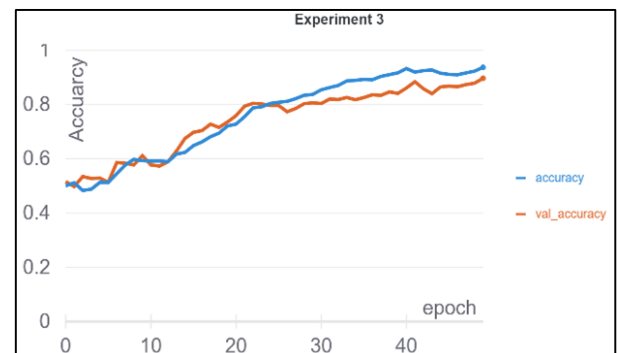


Figure 8. F1-score for experiment 2 considering the weight optimization

Finally, for Experiment 3, this experiment employed lung segmentation, image augmentation, and CLAHE, as discussed in section 3.1.

Using lung segmentation, the network was forced to focus on the lungs. The result shows increased sensitivity in the COVID-19 class compared to the other experiments. Experiment 3 showed the best result in sensitivity values in the range 95% to 97% while using the MEAEO for weight optimization and 95%-93% without the optimization approach, as shown in Table 3 and Figure 4 (c, f). The system's overall performance yielded an accuracy of 97.23%, with weight optimization using modified EAEO and 91.5% without optimization.

The development of accuracy and loss during training and evaluation is depicted in Figure 9. Figure 10 compares the result of weight optimization with the result obtained from not using the optimization in terms of F1-Score.



(a)

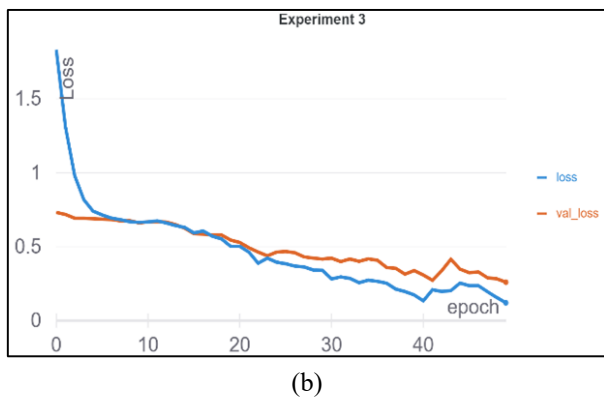


Figure 9. (a) Accuracy and validation accuracy for experiment 3; (b) loss and validation loss for experiment 3

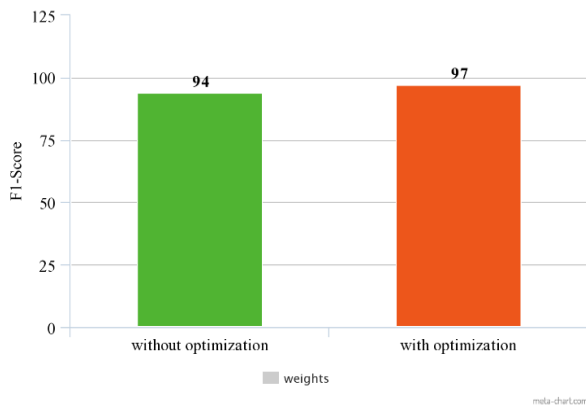


Figure 10. F1-score for experiment 3 considering the weight optimization

5. DISCUSSION

Due to the imbalanced classes (number of samples in each class) in the dataset, the focal loss has been employed to mitigate the bias from the inherited issue of the problem. As shown in the result that optimizing the weights using a metaheuristic algorithm is beneficial. As an outcome, the experiments that employ weight optimization have higher scores; utilizing weight optimization is advantageous in accurately predicting the classes. However, the gap between the optimized and non-optimized experiments is not significant in terms of accuracy.

5.1 Interpreting the improvements gained from applying modified EAEO

As shown earlier in Table 3, experiments using the proposed optimization technique improved prediction performance compared to those that did not employ weight optimization, this indicates a possible performance loss if exclusively trained with traditional training algorithms; however, when applied to the network, the suggested technique would gain an accuracy increase of 2-5% compared to utilizing typical training method. This increase is attributed to improved search methods, which can escape from local optima more effectively.

Furthermore, the MEAEO's capability to achieve higher performance is attributed to utilizing the trained weights as an individual in the EAEO population. This approach assisted the MEAEO to obtain better results in the optimization process. Figure 11 depicts the development of the fitness value. The

fitness function value (loss) continues to converge over the transformation process and seems relatively consistent in the end.

It should be noted that the cost function (loss) curve appears to have an inconsistency post-convergence since batches are employed with a relatively small number of samples while performing the optimization, in other words, the test set is altered in each cycle, resulting in oscillations in the fitness function output; however, the overall trend of the curve is positive.

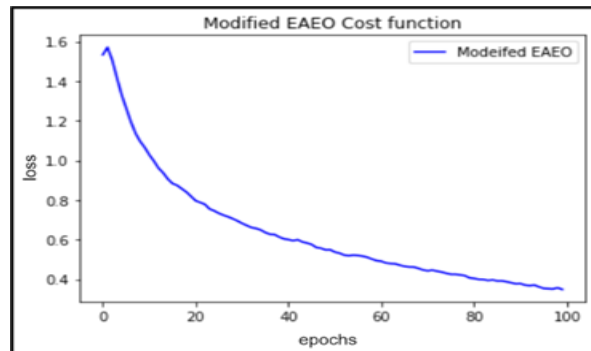


Figure 11. Loss function error curve in experiment 3

5.2 Comparing with other models

To determine how well the COVID-19 classification tests performed and to evaluate the effectiveness of the proposed method (which used medical data 579 samples). Table 4 contains a list of recent work in literature that has been referenced as a comparison to the performance of the proposed framework.

It is important to note that the current dataset and the problem formulation primarily differ from the mentioned studies, making a fair comparison with existing methods difficult. However, an attempt to provide an overview of these studies to determine the efficiency of the proposed approach in detecting COVID-19 has been conducted.

The proposed novel framework in this study performed better regarding accuracy than the recent studies Table 3. CNN model-based binary classifier is employed on 746 CT images containing COVID-19 and healthy patients, which achieved an accuracy of 86.9% [34]. The overall accuracy of 89.6% is achieved with 93% precision in the COVID-19 class while using transfer learning based on the ImageNet dataset [35]. However, the use of transfer learning based on the ImageNet dataset for biomedical image classification is not recommended. The experiments with image segmentation done by the authors achieved an accuracy of 91.5% when analysing three classes with an F-score of $78.57 \pm 1.15\%$ in the COVID-19 class [36]. The proposed network achieved an accuracy of 93.94% in classifying COVID-19 [37].

Furthermore, the authors used a Generative adversarial network to increase the training data. The authors [17] proposed a late fusion (merging) of CT scans and textual clinical data to assess the severity of COVID-19 in patients. They reported a classification accuracy between (95.4%-97.7%) and high class-specific accuracy (90.6%-99.9%). Ahsan et al. [38] used multiple data types to classify COVID-19 and achieved an accuracy of 95.38%; however, the dataset has few samples. To the best of our ability, a model reconstruction from the study [38], which is comparable to the proposed model in this study, achieved an accuracy of 80.5% using the dataset in section 2.1. A binary classification based

on the encoder-decoder model with random forest (RF) as a classifier, achieved 97.78% [39]. The dataset used in training the network consists of 2482 CT scans, of which 1230 images for patients not infected with COVID-19 (“but infected with other pulmonary diseases”). While the suggested approach in this paper offers the added benefit of being able to use both medical imaging scans and textual clinical data. Although the

improvement appears to be marginal (0.07%), there are advantages to using such a method. Such as the ability to handle information fusion, by combining different data types, we can demonstrate the viability of such method that can potentially provide a more comprehensive representation of the data, leading to improved accuracy in the classification task when compared to using X-rays or CT scans alone.

Table 4. Comparison of the performance of the proposed framework with current studies in the literature

Reference	Data Type	Approach	Accuracy
Shambhu et al. [34]	X-ray	CNN	86.9%
Khan et al. [35]	X-ray	Transfer Learning-Xception	89.6%
Arias-Londoño et al. [36]	X-ray	CNN+Lung Segmentation	91.53%
Sakib et al. [37]	X-ray	CNN+GAN	93.94%
Xu et al. [17]	CT Scans+Textual data	CNN+SVM	95.40-97.70%
Ahsan et al. [38]	X-ray Scans+Textual data	CNN+MLP	96.3%
Goel et al. [39]	CT scans	CNN	97.78%
Proposed Framework	X-ray Scans+Textual data	CNN+MLP	97.85%

5.3 Potential factors affecting the results

Several bias factors might affect the result. The lack of well documented images with a different detector that might use a different method in capturing the scan could affect the consistency of the results. In COVID-19 classification, it is necessary to use chest X-ray images that are adequate for this purpose. Most studies in the literature employed various techniques to handle biological data in their approaches. These proposed techniques are accepted in conventional imaging classifications. Biomedical images should be addressed with caution, and specific rules must be followed while processing them so that the modified images do not differ from the original image. Due to the newly developing knowledge in this area, it is impossible to recommend a method or a set of techniques that is more efficient in identifying COVID-19 from a chest X-ray image. Most studies have an accuracy rate of more than 85%, which is statistically a very high degree of accuracy. Aiming for 100% accuracy would be idealistic, as even a tiny percentage of misdiagnoses is problematic.

AI systems may have some utility in a supervised therapeutic context. We argue that putting these techniques into clinical practice requires more than just classification. As we show in experiment 1, it should be approached with caution. In contrast to what we saw in experiments 1 and 2, the result showed that the system recognizes areas as non-diagnostic regions. As a result, these findings are not in line with a clinical viewpoint. Lung segmenting is required to direct the neural network’s attention to the lung organ, as shown in experiment 3.

6. CONCLUSIONS

In this study, a COVID-19 framework was proposed, which incorporated several techniques such as, lung segmentation and augmentation, merging neural networks, and an optimization algorithm (modified EAEO). These techniques assist in creating a robust framework to classify COVID-19 and pneumonia using a different dataset for each pathogen. A hybrid network is a viable approach to analyse X-ray images that correspond to pneumonia, COVID-19 patients, and biomedical images in general. The results demonstrated that merging several CNNs with a multi-layer perceptron and optimizing the weights using the modified EAEO can be beneficial to increase the system robustness. Furthermore, the

employment of the lung segmentation technique aided in achieving a reliable classification result that is not affected by unnecessary features. The impact of merging several data types such as images and textual clinical data is beneficial in improving classification accuracy. The proposed framework obtained an accuracy of 97.85% using the data in section 2.1, which is considerably better than most systems in the literature that predict COVID-19 from pneumonia.

A modified version of EAEO was proposed to optimize the neural network’s weights and mitigate the performance loss from using typical training methods to train the neural network. In addition, weights of the hybrid network are incorporated in the initial population of the modified EAEO to guide the search to optimal values. The results obtained from all the experiments support the idea that an optimization technique can further enhance the performance of a neural network. Moreover, merging different data types such as images and textual clinical data improves classification accuracy.

REFERENCES

- [1] Cheng, V.C., Lau, S.K., Woo, P.C., Yuen, K.Y. (2007). Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, 20(4): 660-694. <https://doi.org/10.1128/cmr.00023-07>
- [2] World Health Organization. (2022). Who coronavirus (COVID-19) dashboard|who coronavirus (COVID-19) dashboard with vaccination data. <https://covid19.who.int/table>.
- [3] Hofman, P., Boutros, J., Benchetrit, D., Benzaquen, J., Leroy, S., Tanga, V., Bordone, O., Allégra, M., Lespinet, V., Fayada, J., Maniel, C., Griffonnet, J., Selva, E., Troncone, G., Portella, G., Lavrut, T., Chemla, R., Carles, M., Ilié, M., Marquette, C. (2021). A rapid near-patient RT-PCR test for suspected COVID-19: A study of the diagnostic accuracy. *Annals of Translational Medicine*, 9(11): 921. <https://doi.org/10.21037/atm-21-690>
- [4] Solis, R.N., Mehrzad, M., Faiq, S., Frusciant, R.P., Sekhon, H.K., Abouyared, M., Bewley, A.F., Farwell, D.G., Birkeland, A.C. (2021). The impact of COVID-19 on head and neck cancer treatment: Before and during the pandemic. *OTO Open*, 5(4). <https://doi.org/10.1177/2473974X211068075>

- [5] Mallett, S., Allen, A.J., Graziadio, S., Taylor, S.A., Sakai, N.S., Green, K., Suklan, J., Hyde, C., Shinkins, B., Zhelev, Z., Peters, J., Turner, P.J., Roberts, N.W., di Ruffano, L.F., Wolff, R., Whiting, P., Winter, A., Bhatnagar, G., Nicholson, B.D., Halligan, S. (2020). At what times during infection is SARS-CoV-2 detectable and no longer detectable using RT-PCR-based tests? a systematic review of individual participant data. *BMC Medicine*, 18(1): 1-17. <https://doi.org/10.1186/s12916-020-01810-8>
- [6] Kordon, A. (2020). Applied artificial intelligence-based systems as competitive advantage. In 2020 IEEE 10th International Conference on Intelligent Systems (IS), IEEE, 6-18. <https://doi.org/10.1109/IS48319.2020.9200097>
- [7] Hussain, S.A., Al Bassam, N., Zayegh, A., Al Ghawi, S. (2022). Prediction and evaluation of healthy and unhealthy status of COVID-19 patients using wearable device prototype data. *MethodsX*, 9: 101618. <https://doi.org/10.1016/j.mex.2022.101618>
- [8] Sekeroglu, B. Ozsahin, I. (2020). Detection of COVID-19 from chest X-Ray images using convolutional neural networks. *SLAS Technology*, 25(6): 553-565. <https://doi.org/10.1177/2472630320958376>
- [9] Zhang, S., Niu, Y.M. (2023). LcmUNet: a lightweight network combining cnn and mlp for real-time medical image segmentation. *Bioengineering*, 10(6): 712. <https://doi.org/10.3390/bioengineering10060712>
- [10] Rehmer, A., Kroll, A. (2020). On the vanishing and exploding gradient problem in gated recurrent units. *IFAC-PapersOnLine*, 53(2): 1243-1248. <https://doi.org/10.1016/j.ifacol.2020.12.1342>
- [11] Alqushaibi, A., Abdulkadir, S.J., Rais, H.M., Al-Tashi, Q. (2020). A review of weight optimization techniques in recurrent neural networks. In 2020 International Conference on Computational Intelligence (ICCI), IEEE, pp. 196-201. <https://doi.org/10.1109/ICCI51257.2020.9247757>
- [12] Sharma, N., Aggarwal, L.M. (2010). Automated medical image segmentation techniques. *Journal of Medical Physics/Association of Medical Physicists of India*, 35(1): 3-14. <https://doi.org/10.4103/0971-6203.58777>
- [13] Apostolopoulos, I.D., Mpesiana, T.A. (2020). Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43: 635-640. <https://doi.org/10.1007/s13246-020-00865-4>
- [14] Wang, L.D., Lin, Z.Q., Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *Scientific Reports*, 10(1): 19549. <https://doi.org/10.1038/s41598-020-76550-z>
- [15] Mukherjee, H., Ghosh, S., Dhar, A., Obaidullah, S.M., Santosh, K.C., Roy, K. (2021). Deep neural network to detect COVID-19: One architecture for both CT scans and chest x-rays. *Applied Intelligence*, 51: 2777-2789. <https://doi.org/10.1007/s10489-020-01943-6>
- [16] Lassau, N., Ammari, S., Chouzenoux, E., Gortais, H., Herent, P., Devilder, M., Soliman, S., Meyrignac, O., Talabard, M., Lamarque, J., Dubois, R., Loiseau, N., Trichelair, P., Bendjebbar, E., Garcia, G., Bendjebbar, C., Merad, M., Stoclin, A., Jegou, S., Griscelli, F., Tetelboum, N., Li, Y.P., Verma, S., Terris, M., Dardouri, T., Gupta, K., Neacsu, A., Chemouni, F., Sefta, M., Jehanno, P., Bousaid, I., Boursin, Y., Planchet, E., Azoulay, M., Dachary, J., Brulport, F., Gonzalez, A., Dehaene, O., Schiratti, J., Schutte, K., Pesquet, J., Talbot, H., Pronier, E., Wainrib, G., Clozel, T., Barlesi, F., Bellin, M., Blum, M.G.B. (2021). Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. *Nature Communications*, 12(1): 1-11. <https://doi.org/10.1038/s41467-020-20657-4>
- [17] Xu, M., Ouyang, L., Han, L., Sun, K., Yu, T.T., Li, Q., Tian, H., Safarnejad, L., Zhang, H.D., Gao, Y., Bao, F.S., Chen, Y.F., Robinson, P., Ge, Y.R., Zhu, B.L., Liu, J., Chen, S. (2021). Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: multimodal late fusion learning approach. *Journal of Medical Internet Research*, 23(1): e25535. <https://doi.org/10.2196/25535>
- [18] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02): 107-116. <https://doi.org/10.1142/S0218488598000094>
- [19] Goldberg, D.E., Holland, J.H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3: 95-99. <https://doi.org/10.1023/A:1022602019183>
- [20] Rere, L.M.R., Fanany, M.I., Arymurthy, A.M. (2015). Simulated annealing algorithm for deep learning. *Procedia Computer Science*, 72: 137-144. <https://doi.org/10.1016/j.procs.2015.12.114>
- [21] Carvalho, M., Ludermir, T.B. (2007). Particle swarm optimization of neural network architectures and weights. In 7th International Conference on Hybrid Intelligent Systems (HIS), IEEE, 336-339. <https://doi.org/10.1109/HIS.2007.45>
- [22] Abdulwahab, H.A., Noraziah, A., Alsewari, A.A., Salih, S.Q. (2019). An enhanced version of black hole algorithm via levy flight for optimization and data clustering problems. *IEEE Access*, 7: 142085-142096. <https://doi.org/10.1109/ACCESS.2019.2937021>
- [23] Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 222: 175-184. <https://doi.org/10.1016/j.ins.2012.08.023>
- [24] Bandyopadhyay, S., Saha, S., Maulik, U., Deb, K. (2008). A simulated annealing-based multiobjective optimization algorithm: AMOSA. In *IEEE Transactions on Evolutionary Computation*, 12(3): 269-283. <https://doi.org/10.1109/TEVC.2007.900837>
- [25] Cohen, J.P., Morrison, P., Dao, L. (2020). COVID-19 image data collection. *arXiv Preprint arXiv: 2003.11597*. <https://doi.org/10.48550/arXiv.2003.11597>
- [26] National Institutes of Health. (2017). NIH clinical center provides one of the largest publicly available chest x-ray datasets to scientific community.
- [27] Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E. (1990). Contrast-limited adaptive histogram equalization: Speed and effectiveness. *Proceedings of the First Conference on Visualization in Biomedical Computing*, 337-345. <https://doi.org/10.1109/VBC.1990.109340>
- [28] O'Shea, K., Nash, R. (2015). An introduction to convolutional neural networks. *arXiv Preprint arXiv: 1511.08458*. <https://doi.org/10.48550/arXiv.1511.08458>

- [29] Singh, G., Sachan, M. (2014). Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition. 2014 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-5. <http://dx.doi.org/10.1109/ICCIC.2014.7238334>
- [30] Lin, T.Y., Goyal, P., Girshick, R., He, K.M., Dollár, P. (2017). Focal loss for dense object detection. In 2017 IEEE International Conference on Computer Vision (ICCV), 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- [31] Eid, A., Kamel, S., Korashy, A., Khurshaid, T. (2020). An enhanced artificial ecosystem-based optimization for optimal allocation of multiple distributed generations. IEEE Access, 8: 178493-178513. <https://doi.org/10.1109/ACCESS.2020.3027654>
- [32] Zhao, W.G., Wang, L.Y., Zhang, Z.X. (2020). Artificial ecosystem-based optimization: A novel nature-inspired meta-heuristic algorithm. Neural Computing and Applications, 32: 9383-9425. <https://doi.org/10.1007/s00521-019-04452-x>
- [33] Hossin, M., Sulaiman, M.N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5: 01-11. <https://doi.org/10.5121/ijdkp.2015.5201>
- [34] Shambhu, S., Koundal, D., Das, P., Sharma, C. (2021). Binary classification of COVID-19 CT images using CNN: COVID diagnosis using CT. International Journal of E-Health and Medical Communications (IJEHMC), 13(2): 1-13. <https://doi.org/10.4018/IJEHMC.20220701.oa4>
- [35] Khan, A.I., Shah, J.L., Bhat, M.M. (2020). CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine, 196: 105581. <https://doi.org/10.1016/j.cmpb.2020.105581>
- [36] Arias-Londoño, J.D., Gomez-Garcia, J.A., Moro-Velazquez, L., Godino-Llorente, J.I. (2020). Artificial intelligence applied to chest x-ray images for the automatic detection of COVID-19. a thoughtful evaluation approach. IEEE Access, 8: 226811-226827. <https://doi.org/10.1109/ACCESS.2020.3044858>
- [37] Sakib, S., Tazrin, T., Fouda, M.M., Fadlullah, Z.M., Guizani, M. (2020). DL-CRC: deep learning-based chest radiograph classification for COVID-19 detection: A novel approach. IEEE Access, 8: 171575-171589. <https://doi.org/10.1109/ACCESS.2020.3025010>
- [38] Ahsan, M.M., E. Alam, T., Trafalis, T., Huebner, P. (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. Symmetry, 12(9): 1526. <https://doi.org/10.3390/sym12091526>
- [39] Goel, C., Kumar, A., Dubey, S.K., Srivastava, V. (2020). Efficient deep network architecture for COVID-19 detection using computed tomography images. MedRxiv. <https://doi.org/10.1101/2020.08.14.20170290>