Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Optimal Ensemble learning model for COVID-19 detection using chest X-ray images

Balasubramaniam S*, K Satheesh Kumar

Department of Futures Studies, University of Kerala, Thiruvananthapuram, Kerala 695581, India

ARTICLE INFO

Keywords:
COVID-19
Preprocessing
Feature Extraction
Classification
Optimization

ABSTRACT

COVID-19 pandemic is the main outbreak in the world, which has shown a bad impact on people’s lives in more than 150 countries. The major steps in fighting COVID-19 are identifying the affected patients as early as possible and locating them with special care. Images from radiology and radiography are among the most effective tools for determining a patient’s ailment. Recent studies have shown detailed abnormalities of affected patients with COVID-19 in the chest radiograms. The purpose of this work is to present a COVID-19 detection system with three key steps: (i) preprocessing, (ii) Feature extraction, (iii) Classification. Originally, the input image is given to the preprocessing step as its input, extracting the deep features and texture features from the preprocessed image. Particularly, it extracts the deep features by inceptionv3. Then, the features like proposed Local Vector Patterns (LVP) and Local Binary Pattern (LBP) are extracted from the preprocessed image. Moreover, the extracted features are subjected to the proposed ensemble model based classification phase, including Support Vector Machine (SVM), Convolutional Neural Network (CNN), Optimized Neural Network (NN), and Random Forest (RF). A novel Self Adaptive Kill Herd Optimization (SAKHO) approach is used to properly tune the weight of NN to improve classification accuracy and precision. The performance of the proposed method is then compared to the performance of the conventional approaches using a variety of metrics, including recall, FNR, MCC, FDR, Thread score, FPR, precision, FOR, accuracy, specificity, NPV, FMS, and sensitivity, accordingly.

1. Introduction

The SARS-CoV-2 (Corona Virus) started spreading from the Wuhan city of China and has shown its bad impact on several countries since December 2019 [1]. It is observed that early diagnosis is the only option to minimize the infection possibility in healthy people and should need sudden isolation of the infected person. Moreover, gene sequencing, blood specimens, or RT-PCR are the major screening techniques for COVID-19 [2 3]. However, the TPR of RT-PCR is reported as 30 to 60% for the throat swab samples [4 5].

Moreover, chest radiography images (i.e.), CT or X-ray imaging are routine tools to execute with the quick diagnosis for pneumonia patients. X-ray images provide visual indexes linked with COVID-19 Chest [6], and CT shows larger sensitivity to diagnose COVID-19 [7]. The chest imaging reports generate the peripheral airspace opacities and multi-lobar involvement. Diffuse airspace opacities or asymmetric patchy are also reported for COVID-19 [8]. As there are limited trained radiologists and the rate of suspected people increases, automatic approaches for early diagnosis are mandatory. AI/ML solutions are powerful tools to solve those issues.

Nowadays, the X-ray images related to COVID-19 are gathered as a tiny dataset that is more helpful for AI researchers for training the ML approach to execute the automatic COVID-19 detection from the images of Chest [9 10]. Furthermore, these images are taken from the academic report on the consequences of CT and X-ray images of COVID-19 [11]. Moreover, a subset of images is used as the -ve samples from the CheXpert dataset for the COVID-19 diagnosis. Likewise, around five thousand X-ray images related to Chest are combined as the dataset [12 13], known as COVID-Xray-5 k, and it is categorized into three thousand testings and two thousand training samples.

The ML [14 15 16] structure is employed for predicting the COVID-19 by the Chest images of X-ray [17 18 19 20]. Different classical schemes are used for the classification of the medical image that includes 2 step procedure like hand-crafted recognition and feature extraction; End-to-end DL structures are used for the direct prediction of COVID-19 disease without any additional features from raw images [21 22 23]. In recent years, the DL (more specifically CNN) based models [24 25 26 27] are attained better performance in the classical AI models based on medical image analysis tasks and computer vision [28 29 30].

* Corresponding author.
E-mail address: baluttn@gmail.com (S. Balasubramaniam).

https://doi.org/10.1016/j.bspc.2022.104392
Received 6 April 2022; Received in revised form 4 October 2022; Accepted 30 October 2022
Available online 21 November 2022
1746-8094/© 2022 Elsevier Ltd. All rights reserved.
The best possible environment that may be achieved for an aim under the present limitations is known as global optimization [54]. The largest metropolitan area of the Chinese province of Hubel, Wuhan, reports the first official coronavirus case, as certified by WHO. Millions of verified cases have been reported worldwide, and it has already claimed thousands of lives [52]. According to studies, this family of viruses exhibits a considerable manifestation in radiographic pictures [53].

The following list of the model’s main contributions:

- Proposes an ensemble classifier that includes SVM, RF, CNN, and optimized NN for the final classification of COVID-19 chest images.
- Presents the Self Adaptive Kill Herd Optimization (SAKHO) Method for determining the best weights for training the optimized NN model.

In this essay, section 2 presents an overview of COVID-19 detection. Section 3 provides an overview of the adopted COVID-19 detection model. Section 4 illustrates the pre-processing and suggested feature extraction method. Section 5 covers optimization-assisted ensemble integration of features from multiple CNN with the Bayesnet classifier and CFS model for identifying COVID-19. The proposed scheme was tested by two public datasets and attained better outcomes on both datasets. The simulation results have proven better effective during the detection of COVID-19in the pre-trained multi-CNN than the single CNN.

In 2020, Govardhan et al. [37] have introduced an alternative diagnostic tool and an advanced DL model for COVID-19 cases. The proposed method was introduced in 4 phases: “preprocessing, data augmentation, stage-I and stage-II deep network model.” Moreover, the network was executed in 2 phases that were implemented for differentiating the bacteria on chest X-ray images, COVID-19 with pneumonia from another virus-induced pneumonia, and healthy cases. The simulation outcomes of the proposed method have attained high classification accuracy, recall, and precision for detecting COVID-19.

In 2020, Suat et al. [38] determined a new ANN, Convolutional CapsNet, along with the capsule networks to identify the COVID-19 infection through chest X-ray images. To accomplish accurate and quick diagnostics with multi-class and binary classification, the adopted model was put into practice for COVID-19 disorders. The presented method has correspondingly attained better accuracy for multiple and binary classes. Further, the adopted method helped increase the diagnostic performance of COVID-19 disease. Further, the presented approach was believed to be an alternative scheme for diagnosing COVID-19 with a fast screening approach.

In 2020, Asif et al. [39] has implemented CoroNet, a Deep CNN scheme to diagnose the COVID-19 infection automatically. The presented approach was performed based on a trained end-to-end dataset while gathering the COVID-19 images, another chest pneumonia X-ray images from 2 various databases, and the Xception framework pre-trained on the ImageNet dataset. The CoroNet was tested and trained on the dataset. The simulation outcomes have shown that the presented approach achieved better overall accuracy, recall rate, and precision for COVID-19 cases for 4-class cases. The adopted method has produced better classification accuracy for 3-class classification. Table 1 shows the reviews on COVID-19 diagnosis by the chest X-ray images.

In 2022, Wang et al. [32] deployed an enhanced multiple-way data augmentation. Second, a rank-based pooling module (NRAPM) was presented that used rank-based average pooling (RAP) specifically to prevent over-tapping. Third, a deep rank-based average pooling network (DRAP Net) based on NRAPM and motivated by the VGG network was proposed. Heatmaps were produced by Grad-CAM, which also provided an understandable analysis for our AI model. After 10 cycles through the pre-trained CNNs from X-ray images. The adopted model has used the integration of features from multiple CNN with the Bayesnet classifier and CFS model for identifying COVID-19. The proposed scheme was tested by two public datasets and attained better outcomes on both datasets. The simulation results have proven better effective during the detection of COVID-19in the pre-trained multi-CNN than the single CNN.

## 2. Literature review

### 2.1. Related works

In 2021, Shankar et al. [33] have introduced a new fusion method hand-crafted with DL features known as theFM-HCF-DLF approach for the COVID-19classification and diagnosis. Furthermore, the proposed FM-HCF-DLF approach comprised three major processes: “Gaussian filtering-based preprocessing, FM for feature extraction and classification.” Moreover, the FM method has incorporated the handcrafted features fusion with the assistance of DL and LBP features and utilized the CNN-based Inception v3 schemes. The Adam optimizer used the backpropagation learning scheduler to improve the performance of the Inception v3 algorithm. The chest X-ray dataset evaluated the adopted FM-HCF-DLF method. The simulation results of the suggested model have, in the end, demonstrated improved performance with increased specificity, sensitivity, accuracy, precision, kappa value, and F score.

In 2020, Bejoy et al. [35] have determined the multi-CNN effectiveness for the automatic diagnosis of COVID-19 by mixing certain pre-trained CNNs from X-ray images. The adopted model has used the integration of features from multiple CNN with the Bayesnet classifier and CFS model for identifying COVID-19. The proposed scheme was tested by two public datasets and attained better outcomes on both datasets. The simulation results have proven better effective during the detection of COVID-19in the pre-trained multi-CNN than the single CNN.
3. Overall description of adopted COVID-19 detection model

This study intended to introduce a COVID-19 detection method that includes three steps: "(i) preprocessing (ii) Feature extraction (iii) Classification." First, the source image is given as the source of preprocessing step, and it is performed by using the median filtering process. The preprocessed image is then used to extract deep features and texture characteristics. Here, the deep features are extracted by inceptionV3. The features like proposed LVP and LBP are also extracted.

Moreover, the extracted features were fed as an input to the classification step, here, a proposed ensemble model is used to classify the image as diseased or not. The proposed ensemble model includes SVM, RF, Optimized NN, and CNN. The extracted characteristics were also used as individual inputs for each classifier, and the results from each classifier were averaged to produce the final result. SAKHO model ideally modifies the weight of NN to increase the classification's precision and accuracy. The broad structure of the study provided is shown in Fig. 1.

4. Preprocessing and proposed feature extraction process

4.1. Preprocessing via median filtering

The median filtering approach enhances the image during the preprocessing stage. Here, the input image is smoothed and denoised using median filtering [40]. And then, the sound score or digital image sequence was restored through the (neighborhood) mask of the median value. The neighboring pixels are sorted and saved with the aggregated median value, which takes the place of the noisy value, depending on the gray level. We can use linear or nonlinear filters to remove noise from an image that has been tainted by noise. In the features of the image are high-frequency components in the frequency domain, which are readily confused with high-pitched noises. Consequently, the challenge is to maintain visual features while successfully filtering random disturbances crucial to the processing of picture filters. Nonlinear filter known as the median filter is commonly utilized in digital due to its effective edge-keeping abilities and capacity to reduce impulsive noise. The rank-order filter is the median filter. Depending on the size and shape of the filters, it can reduce noise. In addition, \( f(c, d) = med \{h(c - u, d - v)u, v \in I \} \) is the output of the median filtering [40], where, \( f(c, d) \) and \( h(c, d) \) related to output image and real image, accordingly, \( I \) shows 2D mask, \( M \times M \) depicts the size of the mask (i.e.), \( M \) was an odd contains \( 3 \times 3 \), \( 3 \times 5 \), etc. The mask can also be cross-shaped, circular, square, or linear, among other shapes.

**Lowering the noise performance:** The median filter, also considered a nonlinear filter, has a harder time processing images with noisy data. When using a median filter, the noise variance of a picture with a normal distribution and zero average noise is calculated as in Eq. (1).

\[
\sigma_{\text{med}}^2 = \frac{1}{M^2} \sum_{M} \left( \frac{x - \mu}{\sigma} \right)^2
\]

In Eq. (1), \( M \times M \) depicts size of median filter mask, \( \sigma_{\text{med}}^2 \) depicts input noise power difference, \( F(M) \) and depicts density function of sound. The noise variability is represented in Eq. (2) when using the average Filter.

\[
\sigma_{\text{av}}^2 = \frac{1}{M^2} \sum (x - \mu)^2
\]

When comparing Eq. (1) with Eq. (2), the benefits of median filtering rely on distributions and mask size. Additionally, the outcomes of the median filtering provide low random noise, and it is higher than the average performance of the Filter. In the impulse noise, narrow pulses occur because the pulse is lower than the efficient median filter. Whenever the median and average filtering algorithms are combined, the

---

### Table 1

Reviews on conventional COVID-19 diagnosis utilizing chest X-ray images: Features and Challenges.

| Author          | Adopted scheme     | Features                                                                 | Challenges                                                                 |
|-----------------|--------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Wang et al. [32]| NRAPM Method       | * Better Precision                                                       | * Need concentration on the web applications.                             |
|                 |                    | * High Accuracy                                                          |                                                                           |
|                 |                    | * Increased F1-score                                                     |                                                                           |
| Shankar et al.  | FM-HCF-DLF model   | * Maximum sensitivity                                                    | * The FM-HCFDLF model needs to be enhanced using other classifiers rather |
|                 |                    | * Better specificity                                                     |                                                                           |
|                 |                    | * High precision                                                        |                                                                           |
|                 |                    | * Increased accuracy                                                     |                                                                           |
|                 |                    | * Higher F score                                                        |                                                                           |
|                 |                    | * Better kappa value                                                    |                                                                           |
|                 |                    | * We need to develop optimization algorithm                              |                                                                           |
| Wang et al. [34]| PSCNN              | * Better precision                                                      |                                                                           |
|                 |                    | * High Recall                                                            |                                                                           |
|                 |                    | * Maximum F1 score                                                      |                                                                           |
|                 |                    | * Larger accuracy                                                       |                                                                           |
| Bejoy et al. [35]| Multi-CNN model    | * Better AUC                                                            | * All possible combinations of multi CNNs are not investigated.            |
|                 |                    | * Higher accuracy                                                       |                                                                           |
| Govardhan et al.| Transfer learning  | * Increased accuracy                                                    | A network design was not enhanced for detect COVID-19                     |
|                 | with the CNN model | * Recall is greater                                                     |                                                                           |
| Suat et al. [36]| Novel ANN (Convolutional CapsNet) model                                 | * Increased accuracy                                                    |                                                                           |
|                 |                    | * Recall is greater                                                     |                                                                           |
|                 |                    | * Increased specificity                                                 |                                                                           |
|                 |                    | * High precision                                                        |                                                                           |
|                 |                    | * Better F1-score                                                      |                                                                           |
| Asif et al. [39]| CoroNet (Deep CNN)  | * High precision                                                        | * The proposed model does not use dissimilar data types like CT images    |
|                 |                    | * Better recall rate                                                   | for diagnosing COVID-19.                                                  |
|                 |                    | * Maximum accuracy                                                      |                                                                           |
|                 |                    | * Higher specificity                                                   |                                                                           |
|                 |                    | * Increased F-measure                                                  |                                                                           |
|                 |                    | * We need to overcome the limitations in-depth analysis.                |                                                                           |
average filtering performance is raised to resize the mask based on the noise density.

4.2. Feature Extraction: Extraction of deep and texture features

The feature extraction phase includes three processes as follows.

✓ Proposed LVP
✓ LBP
✓ Deep features by Inception V3

(i) Proposed LVP: Through computing the values between the adjacent pixels from various directions with varied distances and the reference pixel, it is established that the LVP [41] symbolizes the structure information and one-dimensional direction of the local text in the existing LVP the local features are extracted by using the higher order derivatives. The mathematical representation of the LVP is determined in Eq. (3), Eq. (4), Eq. (5), and Eq. (6).

\[
H_{ik}(k, \gamma) = \sum_{j \in \Omega} 2^{j-k} g_{ik}^{D}(p, k, \gamma)
\]  

\[
g_{ik}^{D}(p, k, \gamma) = \begin{cases} 
1 & \text{if } CS_{ik}^{D} \neq 0 \\
0 & \text{otherwise} 
\end{cases}
\]  

\[
CS_{ik}^{D}(p) = V_{ik}^{D}(p) - \left[ \frac{V_{ik}^{D}(k)}{V_{ik}^{D}(k)} \times V_{ik}^{D}(k) \right]
\]

\[
V_{ik}^{D}(p) = J(p, D) - J(k)
\]

In Eq. (3), \(H_{ik}(k, \gamma)\) indicates the LVP at a nearby distance \(D\), and \(\gamma\) refers to the index angle. As per the proposed LVP, the mutual information \(\mu I\) is also computed as in Eq. (7).

\[
\mu I = \sum_{i \in \Omega, j \in \Omega} P(s_1, s_2) \log \frac{P(s_1, s_2)}{P(s_1)P(s_2)}
\]  

\[
P(s_1) = \sum_{s_2} w_{12}
\]

\[
P(s_1, s_2) = \sum_{s_{12}} w_{11} w_{22}
\]

In Eq. (7), \((s_1, r_{12})\) indicates the pairs of elements of LVP. \(w_{11}, w_{22}\) denotes the weights, \(t_1, t_2\) refers to the tags and \(T_i\) indicates the set of resources tagged with it.

Color transformation: The distance between 2 features is calculated using cosine distance in Eq. (10).

\[
D_{is}(G, G') = \frac{1}{n} \sum_{q=1}^{n} \left\| \frac{G - G'}{\|G\| \|G'\|} \right\|
\]  

In Eq. (10), \(G\) indicates the query image feature, \(\cdot\) denotes the dot product, \(G\) specifies the feature library, and \(n\) depicts the total image in the dataset. The extracted proposed LVP feature is indicated as \(FE_{LVP}\).

(ii) LBP: The LBP [32] has computational simplicity and higher discriminative power. Additionally, the LBP operator labels each image pixel with decimal integers. The labeling process calculates every image pixel with its surrounding pixels by subtracting the mid-pixel value. Additionally, it encodes the negative values produced as 0, and the positive and 0 values as 1. Additionally, all the basic codes are spliced starting from the top-left and moving clockwise to create the binary number. This is known as the LBP codes. Those many local descriptions that make up the global description are constructed using the texture descriptor. Additionally, based on the capacity to distinguish between these texture items, the characteristics are retrieved. In Eq. (11), \(SH_{pl}\) and \(SH_{di}\) indicates the mid pixel intensities and the imagemid pixel from relativepl, correspondingly. The pixel’s LBP descriptor is displayed as \(LBP(\bullet)\) and \(NE_{pl}\) indicates the neighbor’s count. \(FE_{LBP, pl, di}\) - LBP
descriptor function is expressed in Eq. (12).

\[ LBP(S_{H_i}) = \sum_{p \in \Omega} FE_{LBP(p,i)}2^{o-1} \]  

(11)

\[ FE_{LBP(p,i)} = \begin{cases} \frac{1}{\gamma}, & \text{if } SH_{\gamma_i} - SH_{\lambda_i} \geq 0 \\ 0, & \text{otherwise} \end{cases} \]  

(12)

(iii) Deep features by Inception v3: Five layers make up a CNN: convolutional, source layer, FC, output, and pooling. Additionally, the GoogLeNet network is set up in Google and is meant to function like CNN. Furthermore, the inception network model [33] is applied to improve a system’s depth and limit the number of network attributes. As a result, it is frequently used in image classifications.

Convolution layer: It differs from NN in that bias and weight is unrelated to all of the subsequent layer’s pixels. Nevertheless, the entire image is categorized into small areas following it using bias and weights. Moreover, the bias and weights are called kernels or filters and are convoluted with all the little areas of the input image, which provides a feature map. These filters are the easy ‘features’ discovered from this layer’s input image. The size of the local region, count of filters, padding, and stride are known as the hyperparameters of this layer. The hyperparameters endure the tuning for optimal outcomes based on the genre and size of an input image.

Pooling layer: This layer minimizes the parameter count and the image’s physical dimensions. Additionally, a set function to an intake with no parameters was accomplished via the pooling layer. The max pooling, average pooling, and stochastic pooling are further forms of pooling layers.

FC layer: Additionally, the FC layer receives the compressed results of the last pooling layer as input. Moreover, it acts as a CNN where each neuron of the conventional layer is associated with the presentation layer. Consequently, the parameter count present in the convolution layer is higher.

Activation function: It is used in the dissimilar CNN framework. The nonlinear activation function provides an optimal result than the previous tangent or sigmoid functions. The nonlinear functions are meant to improve the training speed.

CNN learning model relies on the chain rule vector and calculus. Consider \( y \in RH \) as a scalar and \( x \in RG \) as a vector. While the partial derivative of \( y \) in terms of \( x \) is a vector, \( y \) is a function of \( x \) and it is given in Eq. (13).

\[ \frac{dy}{dx} = \begin{bmatrix} \frac{dy}{dx_1} \\ \vdots \\ \frac{dy}{dx_n} \end{bmatrix} \]  

(13)

In Eq. (13), \( \frac{dy}{dx} \) denotes the vector of the same size as \( x \), and the \( ith \) element is represented as \( \frac{dy}{dx_i} \). Moreover, it is obvious that \( \frac{dy}{dx^T} = \frac{dy}{dx}^T \). Suppose \( x \) is a function of \( z \), and \( z \in RH^p \) is the other vector. The partial derivative of \( X \) based on \( z \) is specified in Eq. (14).

\[ \frac{dy}{dx} = \frac{dy}{dx^T} = \frac{dy}{dx^T} \]  

(14)

Moreover, the \( L \times B \) matrix present in the fractional derivative and access at the juncture of \( oth \) column and \( ith \) row i.e., \( \frac{dy}{dx} \). Further, \( y \) is a function of \( z \) in a chain-like argument. In addition, a function maps \( z \) to \( x \), and other function maps \( x \) to \( y \). Eq. (15) determines the chain rule.

\[ \frac{dy}{dx} = \frac{dy}{dx^T} = \frac{dy}{dx^T} \]  

(15)

The loss or cost function is used to calculate the difference between the target \( ts \) and the prediction of a CNN, \( z^{th} \rightarrow w_{g_1}, z^{th} \rightarrow, \ldots, z^{th} \rightarrow w_{g^2} = qd \), it is used a single loss function \(qd = \|ts - z^{th}\|^2 \).

The predictive outcome is specified as \( argmax \|z^{th}\|^2 \). Further, the convolution model is defined in Eq. (16).

\[ x_i^{l+1}, o_i^{l+1}, dp = \sum_{o=0}^{L} \sum_{d_0=0}^{d_1 \times d_2 \times d_3} \hat{a}_{l_1, o_1, dp} \times \frac{\partial LBP}{\partial x_i^{l+1+i_1, o_1, dp}} \]  

(16)

The Filter \( fh \) with size \( \frac{L \times B \times DL}{k} \), and convolutional layer contains \( \frac{L^k - L + 1}{k} \times \frac{B^k - B + 1}{k} \) spatial size with DL slices that imply \( x^{(i+1)} \) in \( RH^{K-k} \times DP^{k+1} \times DP^{k+1} \), \( L^k - L + 1 \), \( B^k - B + 1 \) and \( DL^{k+1} = DL \).

The possibility of all labels \( k \in \{1, \ldots, K\} \) is used for training the instance that is computed through the \( PT(k) = \frac{exp(y_k)}{\sum_{k=1}^{K} exp(y_k)} \), in which \( y \) indicates the non-normalized log possibility. Further, the ground truth shares over labels \( qv(k) \) are normalized as \( \sum_{k=1}^{K} qv(k) / K \). The loss \( lc \) is provided via the cross entropy and it is given in Eq. (17).

\[ lc = \sum_{k=1}^{K} \log(PT(k))qv(k) \]  

(17)

Moreover, the loss of cross-entropy is a differential value with respect to the \( \log y_k \) and used as deep systems during gradient training as \( \frac{d}{d\theta} = \frac{d}{d\theta} \). Thus, the label-smoothing regularization is the same as the loss of a single cross-entropy during execution, \( L(qv, PT) \) and a pair of losses \( L(qv, PT) \) and \( DL(\theta, PT) \). Further, the 2nd loss punishes the difference of the forecast label shared PT from prior \( \theta \) with \( \frac{\partial L}{\partial \theta} \) relative weight.

GoogLeNet’s primary goal is to operate similarly to the Inception framework, which is why the GoogLeNet architecture is also known as the Inception network. This is because it consists of more GoogLeNet versions than any other, including Inception v1, v2, v3, v4, and Inception-ResNet. Therefore, the inception consists of 3 different sizes of maximum pooling and convolution. The network outcome in the preceding layer is known as the channels that are gathered after nonlinear fusion is performed once the convolution task is completed. Inception v3 preferred a network model organized via Keras that is advanced trained in Image Net. The Inception v3 network framework determined a convolution kernel splitting method to categorize the integrals of massive volume into lower convolutions than the Inceptions v1 and v2. The extracted deep features by Inception v3 were represented as \( SE_{CA} \).

Finally, the overall extracted features \( FE \) are given in Eq. (20).

\[ FE = FE_{LBP} + FE_{I} + FE_{CA} \]  

(20)

5. Optimization assisted Ensemble classification for COVID-19 detection model

5.1. SVM

\( FE \) was the extracted feature which is provided as a source of an SVM. To perform a simple regression model, SVM is defined [43]. To find the separating hyperplane of the SVM model, the saddle purpose, which effectively reduces the quadratic programming problems, is used to conduct the double issues of the Lagrange job. The SVM classifier assigns
the decision boundary among all potential hyperplanes using the maximum margin. Further, \(|\|LP\|\) reduces the objective for maximizing the \(M\), which was determined in Eq. (21). Where denotes the margin separating the hyperplane and the closest data point for both classes.

\[
\min \frac{||LP||^2}{2} \quad \text{subject to} \ n \in \{LP, R_o \} + b || \geq 1 \\
(21)
\]

In Eq. (21), \(b\) depicts scalar threshold number, \(O\) depicts source data count in SVM, \(R_o\) depicts source data points, and \(LP\) vector expresses showing boundary. \(FK(R)\) portrays the attained optimal hyperplane of SVM as given in Eq. (22).

\[
FK(R) = \sum_{i=0}^{n} x_{i}A_{S_{O}} < R_o, R > + bs
\]

(22)

Here, \(R_o\) is also said to be a support vector as a source data point \(R_o\) with a non ‘0’ Lagrange multiplier (\(A_{S_{O}}\)). It was not that important to calculate \(FK(R)\) by outer data points support vectors. Additionally, the SVM classifier’s default settings offer the best possible data grouping. They brighten up area is crucial for analyzing the accuracy of the SVM. They provide the nonlinear characteristics of multi-layer networks, is how the nonlinear characteristics is more crucial. The RF algorithm has a few main predictors and each tree depending on the value of a randomly selected feature on the map \(\{\mathbf{V}^P_{E}, \mathbf{V}^D_{E}\}\). And then, the \(\mathbf{V}^D_{E}\) is displayed as \(\hat{y}_s\) and \(\mathbf{V}^P_{E}\) is also said to be a support vector as a source data point \(R_o\) with a non ‘0’ Lagrange multiplier (\(A_{S_{O}}\)). The activation function, which forecasts the nonlinear characteristics of multi-layer networks, is how the nonlinearity is obtained. Keep in mind activation value \(\{\mathbf{A}_r, \mathbf{A}_l\}\) and nonlinear activation function as \(\mathbf{AF}(\cdot)\) was displayed in Eq. (26). Furthermore, the granularity of feature maps as shown in Eq. (27) is minimized in order to estimate the switch inside the pooling layer. Each feature map’s pooling function is described as \(pool()\) and the neighborhood surrounding each feature on the map \(\{\mathbf{A}_r, \mathbf{A}_l\}\) at position of \(e, x_l\) is displayed as \(\mathbf{H}_{e,x_l}\).

\[
SP_{e,x_l} = \mathbf{WE}^E_{r} g^E_{r} + \mathbf{BF}^E_{r}
\]

(25)

\[
AF_{e,x_l} = \mathbf{A}\left(SP_{e,x_l}\right)
\]

(26)

\[
OF_{e,x_l} = \mathbf{pool\left(AF_{e,x_l}\right)}, \mathbf{Y}(cs, cr) \in JH_{e,x_l}
\]

(27)

The loss function in CNN is established in Eq. (28). The condition of CNN are coupled to wanted IO source-outcome relation, and this was displayed as \(\{\mathbf{V}^P_{E}, \mathbf{V}^D_{E}\}\). And then, CNN output, \(\mathbf{P}_r\) source data, and connected target score are displayed as \(\text{OUT}^D_{E, \mathbf{V}^D_{E}} , \mathbf{V}^D_{E}\), and \(\text{OUT}^P_{E, \mathbf{V}^P_{E}}\), accordingly.

\[
\text{Loss} = \frac{1}{\text{Num}} \sum_{p=1}^{\text{Num}} \mathbf{P}_r \mathbf{V}^D_{E}, \mathbf{OUT}^D_{E, \mathbf{V}^D_{E}}
\]

(28)

The RF classifier’s categorized results are indicated as \(\text{CLS}_{\text{CNN}}\).

5.4. Optimized NN

"NN is a series of algorithms that endeavours for recognizing the underlying relationships in a set of data through a process that mimics the way the human brain operates." The extracted features \(FE\) are subjected to the NN as its input [46] and are given in Eq. (29). Here, \(\mathbf{E}\) extracted features are total.

\[
\mathbf{FE} = \{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_{n}\}
\]

(29)

The input, hidden, and output layers make up the NN framework. And hidden layer outcome \(\mathbf{Q}^D_P\) was displayed in Eq. (30). Here, \(\mathbf{E}\) and \(\mathbf{E}\) depicts neurons in the source layer and hidden layer, \(\mathbf{X}\) display activation function, \(\mathbf{W}^P_E\) display bias weight, and \(\hat{y}_s\) hidden neuron, \(\tilde{y}_s\) related to neuron count, and \(\mathbf{W}^P_E\) depicts weight among \(\mathbf{E}\) input neurons to \(\tilde{y}_s\) hidden neuron. The suggested SAKHO technique is then used to update the weights. And network result \(\mathbf{P}_{\mathbf{r}}\) was displayed in Eq. (31), here, \(\tilde{O}\) which denotes result neurons, \(\mathbf{W}^P_E\) display the output bias weight of \(\tilde{y}_s\) output layer, \(\mathbf{N}\) shows hidden neurons total, and \(\mathbf{W}^P_E\) display weight on \(\tilde{y}_s\) hidden layers to \(\tilde{y}_s\) output layer. Finally, error (\(\mathbf{E}\)) gathered in actual and guessed values must less which was depicted in Eq. (32). Here, \(\tilde{y}_s\) displays output neuron total, \(\mathbf{P}_{\mathbf{r}}\) displayed actual and guessed output.

\[
\mathbf{Q}^D_P = \mathbf{X} \left( \mathbf{W}^P_E + \sum_{\gamma=1}^{\mathbf{N}} \mathbf{W}^P_E \right)
\]

(30)

\[
\mathbf{P}_{\mathbf{r}} = \mathbf{X} \left( \mathbf{W}^P_E + \sum_{\gamma=1}^{\mathbf{N}} \mathbf{W}^P_E \right) \mathbf{Q}^D_P
\]

(31)

\[
\mathbf{E} = \mathbf{argmin} \left\{ \sum_{\gamma=1}^{\mathbf{N}} \mathbf{P}_{\mathbf{r}} - \mathbf{P}_{\mathbf{r}} \right\}
\]

(32)
The optimized NN classifier’s classified outcome is displayed as $\text{CLS}_{\text{NN}}$. Thus, the overall classified output $\text{CLS}$ is the average of classified outcomes such as SVM, RF, optimized NN, and CNN, which is given in Eq. (33).

$$\text{CLS} = \text{avg}(\text{CLS}_{\text{SVM}}, \text{CLS}_{\text{RF}}, \text{CLS}_{\text{CNN}}, \text{CLS}_{\text{NN}})$$  \hspace{1cm} (33)

Ensemble learning creates a new classifier that outperforms all of its constituent classifiers from various base classifiers. The algorithm employed, the hyper-parameters, the representation, or the training data may vary amongst these base classifiers. The classified outcome of each classifier is either 0 or 1. COVID positive is denoted as 0 and negative as 1. The output from all the four classifiers, such as SVM, RF, CNN, and NN, is averaged to get the final classification score. If a final score obtained is the decimal point, it is rounded off to the nearest integer i.e. 0.5 and above is taken as 1, and below 0.5 is taken as 0. Based on this final score, the image is predicted as COVID Positive or Negative.

6. Weight optimization of Neural network via self-improved krill swarm optimization algorithm

6.1. Objective function and solution Encoding

As previously stated, the suggested SAKHO approach allows for the best tuning of the weights of NN. The input solution for the suggested SAKHO model is shown in Fig. 2. In this case, the overall amount of weights in NN is shown as in Eq. (34), and the objective function is identified.

$$\text{Obj} = \min \left( \sum \frac{1}{2} \right)$$  \hspace{1cm} (34)

6.2. Proposed SAKHO algorithm

Although the traditional KHO [47] provides a more robust solution for solving complex optimization problems, the fundamental motions in the KH model are not improved and an optimal strategy of the early krill distribution is not selected. To tackle this disadvantage of existing KHO, the algorithm was improved, called the SAKHO approach. Generally, self-improvement is proved to be promising in solving many complex optimization issues [29 30 31 22 23]. In the existing KHO no essential strategy is chosen at the initial stage. The random process is not considered for the diffusion. For determining the parameters efficient strategy is not considered. To overcome all the drawbacks SAKHO algorithm are used. The convergence is improved in the proposed model.

Further, the krill individual is based on the herding behavior of the KH algorithm. The distances reduced from the largest density of the herd and foods in each krill are regarded as the movement of krill.

**Herding behavior of krill swarms:** The best-studied species related to the marine animal is krill. The krill herd is also collected without parallel alignment while still living over distances of 10 to 100 m and during periods ranging from hours to days.

**Lagrangian method of the krill herding:** Moreover, the Predation eliminates the individuals. The average krill density is minimized as the krill swarm moves away from the food source. In the KH algorithm, it is assigned as the initialization process. The three main actions in a 2D surface dictate the time-dependent positioning of each individual krill. Movement simulated by another individual krill.

✓ Foraging activity

![Fig. 2. Solution Encoding.](image)

Moreover, the KH approach has the searching ability in spaces with its arbitrary dimensionality. Consequently, the Lagrangian scheme is utilized in the $m$ dimensional decision space.

$$\frac{dz_b}{dt} = A_b + R_b + S_b$$  \hspace{1cm} (35)

In Eq. (35), $R_b$ denotes the foraging motion, $S_b$ specifies the physical diffusion of the $b^{th}$ individual krill, and $A_b$ indicates the motion induced via other krill individuals.

(i) Movement simulated by another individual krill: Further, the single krill tried to continue a larger density and moved owing to the mutual effects depending on the studied argument. The induced direction of movement was displayed as $\text{CLS}_b$, and it is determined from a repulsive swarm density, the target swarm density, and local swarm density. The individual krill movement is determined in Eq. (36).

$$A_{b}^\text{new} = A_{b}^\text{max} x_a + \theta_b A_{b}^\text{old}$$  \hspace{1cm} (36)

Where,

$$\zeta_b = x_{b}^\text{local} + x_{b}^\text{target}$$  \hspace{1cm} (37)

In Eq. (16), $A_{b}^\text{max}$ indicates the highest induced speed, $\theta_b$ denotes the inertia weight of the motion ranges [0,1], $A_{b}^\text{old}$ refer to the previous motion stimulated. In Eq. (37), $x_{b}^\text{local}$ indicates the local effect given through the neighbors and $x_{b}^\text{target}$ denotes the target direction effect offered through the most excellent krill individual. The measured value of the highest induced speed is 0.01 (ms$^{-1}$).

In a krill movement individual, the effect of the nearby individual is specified in Eq. (38), Eq. (39), and Eq. (40).

$$x_{b}^\text{local} = \frac{NF}{\sum_{i=1}^{NF} \zeta_i} Z_i$$  \hspace{1cm} (38)

$$\zeta_i = \frac{Z_i - Z_0}{\|Z_i - Z_0\| + \beta}$$  \hspace{1cm} (39)

$$\zeta_b = \frac{O_b - O_s}{O_{\text{best}} - O_{\text{worst}}} - O_{\text{avg}}$$  \hspace{1cm} (40)

Where, $O_{\text{best}}$ and $O_{\text{worst}}$ refers to the single krill with best and the worst fitness values, $O_s$ indicates the fitness value or objective function of the $b^{th}$ krill individual; $O_b$ represents the fitness of $a^\text{th}$ ($a = 1, 2, ..., \text{NF}$) neighbor; $Z$ indicates the neighbor location, and NF represents the neighbor total.

The sensing distance ($d_s$) is used to determine the krill individual and find the neighbors. Moreover, various heuristic models compute the sensing distance for every krill individual as per Eq. (41).

$$d_{b} = \frac{1}{\sqrt{N_D}} \sum_{s=1}^{N_D} \|Z_b - Z_s\|$$  \hspace{1cm} (41)

In Eq. (41), $N_D$ indicates the count of the individual krill, and $d_b$ denotes the sensing distance of the $b^{th}$ each krill’s.

The impact of each krill having the best fitness on $b^{th}$ individual krill was determined in Eq. (42).

$$x_{b}^\text{target} = C_{\text{best}} \frac{O_b}{O_{\text{best}}} \zeta_b$$  \hspace{1cm} (42)

In Eq. (42), $C_{\text{best}}$ indicates the efficient coefficient of individual krill’s to $b^{th}$ individual krill with greatest fitness. Moreover, the $C_{\text{best}}$ value is given in Eq. (43).

$$C_{\text{best}} = 2 \left( \frac{\text{rand} + \frac{\text{IM}}{\text{IM}_{\text{max}}}}{1} \right)$$  \hspace{1cm} (43)

Eq. (43) shows random values between 0 and 1, intended to enhance
explore, $IM_{\text{max}}$ specifies the maximum count of iterations, and $IM$ denotes the actual iteration number.

(ii) Foraging motion: It is examined fortwo major effective parameters, such as earlier knowledge regarding the food location and the food location. This motion of the $i^{th}$ krill individual is expressed in Eq. (44).

$$R_i = \text{VG}_i a_0 + \text{Z}_i R_i^\text{old}$$

(44)

Where,

$$a_i = a_i^\text{food} + a_i^\text{best}$$

(45)

Here, $a_i^\text{food}$ portrays the food attractive, $\text{VG}_i$ denotes the foraging speed, $Z_i$ specifies the inertia weight of foraging movement among [0,1], and $a_i^\text{best}$ indicates the effect of the most excellent fitness of the $b^{th}$ krill. The foraging speed of the measured value is 0.02 (m/s$^{-1}$).

The center of the food is identified for each iteration using Eq. (46).

$$Z_{\text{food}} = \frac{\sum_{b=1}^{\text{UB}} Z_{b}}{\sum_{b=1}^{\text{UB}}}$$

(46)

Consequently, for the $b^{th}$ krill individual, the food attraction is given in Eq. (47).

$$a_i^\text{food} = C_i^\text{food} \frac{\text{Z}}{b_{\text{food}} \cdot b_{\text{food}}}$$

(47)

In Eq. (47), $C_i^\text{food}$ represents the food coefficient and it is specified in Eq. (48).

$$C_i^\text{food} = 2 \left(1 - \frac{IM}{IM_{\text{max}}} \right)$$

(48)

The best fitness effect of the $b^{th}$ individual krill is determined in Eq. (49).

$$a_i^\text{best} = \frac{\text{O}_{\text{best}}}{b_{\text{best}} \cdot b_{\text{best}}}$$

(49)

In Eq. (49), $O_{\text{best}}$ indicates the best earlier appointment position of the $b^{th}$ krill individual.

(iii) Physical diffusion: It is measured as a random process. This motion is determined with respect to the random directional vector and highest diffusion speed, and it is given in Eq. (50).

$$S_b = S_{\text{max}} \theta$$

(50)

In Eq. (50), $\theta$ denotes the random directional vector among −1 and 1, and $S_{\text{max}}$ indicates the maximum diffusion speed. Moreover, depending on the time, it linearly decreased the random pace. and worked as the geometrical annealing schedule as per Eq. (51)

$$S_b = S_{\text{max}} \left(1 - \frac{IM}{IM_{\text{max}}} \right) \theta$$

(51)

**Motion Process of the KH model:** The position vector of an individual krill by various parameters efficacy of movement throughout $t$ tot + $\Delta t$, and it is determined in Eq. (52).

$$Z_b(t + \Delta t) = Z_b(t) + \Delta t p + \frac{dZ_b}{dt}$$

(52)

Moreover, $\Delta p$ depends completely on the search space and it is expressed in Eq. (53).

$$\Delta p = C_P \sum_{n=1}^{NV} \left(UB_b - LB_b\right)$$

(53)

In Eq. (53) $NV$ indicates the total number of variables, $C_P$ is a constant number between [0,2], $UB_b$ and $LB_b$ refers to the upper and lower bounds of the $b^{th}$ variables ($b = 1, 2, ..., NV$), correspondingly.

As per the proposed SAKHO method, the global search is performed using step size after exploration stage and it as given in Eq. (54).

$$\text{stepsizes} = \begin{cases} \exp(-nd), & \text{if } nd > 10 \\ \exp(-2), & \text{otherwise} \end{cases}$$

(54)

In Eq. (54), $nd$ denotes the dimension. Subsequently determines a probability $\text{value prob}$, and if the probability is less than 0.5, then the update is done by Eq. (55) or Eq. (56).

$$Z_b^{\text{new}} = Z_b^{\text{old}} + 0.001 \cdot \frac{z_{\text{best}} - z_{\text{random}}}{2} \oplus \text{Levy}(\delta)$$

(55)

$$Z_b^{\text{new}} = Z_b^{\text{old}} + \text{step}$$

(56)

Then,

$$\text{step} = (Z_b^{\text{old}} - z_{\text{random}}) \oplus \text{random}[−1, 1] + \text{stepsizes}$$

$$\oplus \frac{Z_b^{\text{max}} - Z_b^{\text{min}}}{2} \oplus \text{random}[−1, 1]$$

The pseudo-code for the proposed SAKHO methodology is given in Algorithm 1.

**ALGORITHM 1: Proposed SAKHO method**

| Determine the algorithm parameters, the simple bounds, etc. |
| Initialization | Determine a determination of each krill’s worth based on its location. |
| Movement calculation | ✓ Motion induced by the other individuals |
| ✓ Foraging motion | ✓ Physical diffusion |

The individual krill position is updated in the search space. If $\text{prob} < 0.5$

The proposed update $Z_b^{\text{new}}$ is given in Eq. (55).

else

The proposed update $Z_b^{\text{new}}$ is given in Eq. (56).

They are repeated until the stop criteria are reached.

End

7. Results and discussion

7.1. Simulation procedure

The suggested COVID-19 diagnosis using chest X-ray pictures and the SAKHO system was implemented in MATLAB, and the outcomes were confirmed. The performance of the adopted SAKHO model with COVID-19 detection was computed over the conventional schemes such as FM-HCF-DLF [33], DBN [48], FCM [42], LSTM [45], Ensemble + KHO [47], Ensemble + EHO [49], Ensemble + WOA [50], and Ensemble + BOA [51], correspondingly. The sample images of the proposed work are illustrated in Fig. 3. Additionally, different training data, such as 50, 60, 70, 80, and 90 for various performance indicators, were used to calculate the performance. Contains accuracy, sensitivity, specificity, precision, FMS, FDR, FNR, FPP, FOR, NPV, and MCC, correspondingly.

7.2. Dataset description

Datasets were gathered from [36]. The dataset consists of COVID-19 X-ray samples and non-COVID samples. A dataset of 5,000 images with binary labels was created for the detection of COVID-19 in chest X-ray samples and non-COVID samples. A dataset of 5,000 images with normal chest X-ray pictures. This dataset may be used by the scholarly community as a standard. The COVID-19 class labels are designated by a radiologist with pictures. This dataset may be used by the scholarly community as a standard.
The proposed Ensemble + SAKHO scheme is contrasted to the existing schemes, such as FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, and Ensemble + BOA, by certain metrics. The Ensemble + SAKHO model also performs better in terms of positive metrics like sensitivity, accuracy, precision, and specificity than other current models for detecting COVID-19. Additionally, as shown in Fig. 4, the reliability of the chosen Ensemble + SAKHO approach for training examples 90 is 2.69 percent better than the reliability of training data 50. (a). For a model to operate effectively, texturing and deep features are used. There is no need to split the image independently when utilizing deep learning. The image will be automatically segmented using efficient features. The NN weights are fine-tuned to improve prediction accuracy using the SAKHO approach. The observed improvement provides the impact of the proposed SAKHO model on training the classifier in a superior manner.

The chosen ensemble + SAKHO model’s negative metrics, such as FPR, FNR, FDR, and FOR, in comparison to other traditional schemes are shown in Fig. 5. Further, a FOR of chosen Ensemble + SAKHO approach achieves superior outcomes that is 93.01 %, 91.04 %, 90.49 %, 88.86 %, 80.20 %, 77.93 %, 66.15 %, and 55.74 % than other existing models like FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, and Ensemble + BOA, respectively.
Fig. 6 represents the other measures analysis, such as NPV, MCC, and FMS of the adopted Ensemble + SAKHO model and existing models. The graph demonstrates that for training data 80, the NPV of the adopted Ensemble + SAKHO scheme holds a higher value (89.2145). In contrast, the compared existing models, such as FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, and Ensemble + BOA, each achieve lower values, as shown in Fig. 6. As a result, the performance of the given Ensemble + SAKHO system has improved over other current models.

7.4. Overall performance analysis

The overall performance analysis of the presented Ensemble + SAKHO model over other existing schemes for different training data to accuracy measures are represented in Table 2 shows that the chosen Ensemble + SAKHO system has demonstrated superior detection performance for any training data compared to other traditional models. Which are FM-HCF-DLF, DBN, FCM, LSTM, ensemble + KHO, ensemble + EHO, ensemble + WOA, and ensemble + BOA, respectively. Similarly, the proposed Ensemble + SAKHO system gets a better accuracy score than other standard systems like FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, Ensemble + BOA, SVM, RF, CNN and NN without optimization correspondingly for training data 60. Additionally, compared to the training data 50, the suggested ensemble + SAKHO system achieves the highest accuracy score (92.1375) for training data 90. An adopted ensemble + SAKHO model holds the highest value for training data 80 than other traditional models like FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, Ensemble + BOA, SVM, RF, CNN and NN without optimization correspondingly in Table 2. Offers an ensemble classifier for the final prediction of COVID-19 chest pictures that combines SVM, RF, CNN, and optimal NN. The outcomes have summarized that the adopted Ensemble + SAKHO model performance is improved over conventional systems.

7.5. Statistical analysis

Table 3 presents the statistical comparison of the chosen Ensemble + SAKHO system to the conventional system based on accuracy measures. “The meta-heuristic algorithms are stochastic in nature; therefore, the algorithms are executed for several times to determine the objective of the work by evaluating different case scenarios.” Moreover, the adopted Ensemble + under ideal circumstances, the SAKHO prototype has demonstrated greater performance. that was 10.48 %, 14.13 %, 14.858 %, 12.76 %, 12.88 %, 7.48 %, 20.74 % and 22.54 % better than standard systems like FM-HCF-DLF, DBN, FCM, LSTM. As a result, the suggested ensemble + SAKHO framework has been improved and successfully validated.

7.6. Convergence analysis

Fig. 7 illustrates the convergence study of the proposed Ensemble + SAKHO method to various traditional models for different iterations ranging from 0 to 25. Additionally, the graph illustrates how the cost function gradually decreases with more repetitions. For example, in Fig. 4, the performance of the proposed Ensemble + SAKHO model at the 25th iteration shows the best cost function value (~10.2637) than the traditional models such as FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, and Ensemble + BOA. The graph shows that “at first the chosen ensemble + SAKHO model obtains the worst value and at the final stage it obtains better solution converging to precise detection of COVID-19. Still, the adopted Ensemble + SAKHO method remains enhanced over other traditional
The analysis of the ensemble + proposed feature set at certain iterations.

7.7. Analysis based on feature set: Existing vs Proposed

The analysis of the ensemble + proposed feature set over Ensemble + existing feature set for a certain measure is represented in Table 4. Compared to other Ensemble + existing feature sets, the ensemble + proposed feature set has demonstrated a better ability to recognize objects, according to the table. The ensemble + suggested feature set achieves an accuracy value of 6.61 percent higher than the ensemble + existing feature set. Additionally, when contrasted to the Ensemble + current feature set, the ensemble + suggested feature set achieves the highest MCC values (0.92365). Compared to the Ensemble + current feature set in Table 4, the ensemble + suggested feature set has a lower FPR value and performs better. The outcomes have summarized that the ensemble + proposed feature set provided improved performance over the Ensemble + existing feature set.

8. Conclusion

This study introduces a COVID-19 detection model with three key phases: (i) preprocessing (ii) Feature extraction (iii) Classification. The input image was initially used for the preprocessing step, and the deep and texture features were extracted from the preprocessed image. Particularly, it extracts the deep features by inceptionv3. Then, features like the proposed LVP and LBP were extracted from the preprocessed image. Moreover, the extracted features were subjected to the proposed ensemble model based classification phase, including SVM, CNN, Optimized NN, and RF. A new SAKHO method was used to optimize the weight of the NN to improve classification accuracy and precision. Finally, the effectiveness of the proposed scheme was compared to the performance of the conventional ways using a variety of measures.
including recall, FNR, MCC, FDR, Thread score, FPR, precision, FOR, accuracy, specificity, NPV, FMS, and sensitivity, etcetera. From graph, a chosen ensemble + SAKHO model attains higher sensitivity (~91.45) in COVID-19 detected results than other existing schemes like FM-HCF-DLF (~84.39), DBN (~84.52), FCM (~86.35), LSTM (~88.75), ensemble + KHO (~86.69), ensemble + EHO (~85.63), ensemble + WOA (~84.69), and ensemble + BOA (~85.68), respectively for training data 70. Additionally, the suggested ensemble + SAKHO model’s FOR yields superior results that was 93.01 %, 91.04 %, 90.49 %, 88.86 %, 80.20 %, 77.93 %, 66.15 %, and 55.74 % than other existing models like FM-HCF-DLF, DBN, FCM, LSTM, ensemble + KHO, ensemble + EHO, ensemble + WOA, and ensemble + BOA, respectively for training data 70. Moreover, the performance of the proposed Ensemble + SAKHO model at the 25th iteration provides a better cost function value (~10.2637) than the traditional models such as FM-HCF-DLF, DBN, FCM, LSTM, Ensemble + KHO, Ensemble + EHO, Ensemble + WOA, and Ensemble + BOA, respectively. The advantage of the proposed model is the efficiency is high. The limitation of the proposed method it does not concentrate on the multi-class classification. In the future, multi-class classification can be considered where the image can be pneumonia or lung cancer. Moreover, segmentation can be performed.

9. Author Statement

Balasubramaniam S has made a substantial contribution to the concept or design of the article. Satheesh Kumar K drafted the article for important intellectual content and approved the version to be published. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

CRediT authorship contribution statement

Balasubramaniam S: Conceptualization, Methodology, Software,
References

[1] S. M. Swamy, B. R. Rajakumar and I. R. Valaromathi, “Design of Hybrid Wind and Photovoltaic Power System using Opposition-based Genetic Algorithm with Cauchy Mutation”, IET Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems (SEICON 2013), Chennai, India, Dec. 2013, DOI: 10.1049/ic.2013.0361.

[2] S. H. Wang, Z. Zhu, Y.-D. Zhang, PSCNN: PatchShuffle convolutional neural network for COVID-19 explainable diagnosis, Front. Public Health 9 (2021) .

[3] A. S. M. Shafique, M. S. Khan, and A. A. Ashiri, “A novel framework for COVID-19 detection and classification”, Multi-Con. 2 (2020) pp. 617-622.

[4] A. S. M. Shafique, M. S. Khan, and A. A. Ashiri, “A novel framework for COVID-19 detection and classification”, Multi-Con. 2 (2020) pp. 617-622.

[5] A. S. M. Shafique, M. S. Khan, and A. A. Ashiri, “A novel framework for COVID-19 detection and classification”, Multi-Con. 2 (2020) pp. 617-622.

[6] A. S. M. Shafique, M. S. Khan, and A. A. Ashiri, “A novel framework for COVID-19 detection and classification”, Multi-Con. 2 (2020) pp. 617-622.
[48] H.Z. Wang, G.B. Wang, G.Q. Li, J.C. Peng, Y.T. Liu, Deep belief network based deterministic and probabilistic wind speed forecasting approach, Appl. Energy 182 (2016) 80–95.

[49] Mostafa A. Elhosseini, Ragab A. El Sehiemy, Yasser I. Rashwan, X. Z. Gao, “On the performance improvement of elephant herding optimization algorithm” Knowledge-Based Systems, Volume 166, February 15 2019, Pages 58-70.

[50] S. Mirjalili, A. Lewis, The Whale Optimization Algorithm, Adv. Eng. Softw. 95 (May 2016) 51–67.

[51] Asma CHAKRI, Rabia KHELIF, Mohamed BENOUARET, Xin-She YANG, “New directional bat algorithm for continuous optimization problems”, Expert Systems with Applications, vol. 69, pp.159-175, 2017.

[52] A.K. Das, et al., Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network, Pattern Anal. Appl. 24 (3) (2021) 1111–1124.

[53] A.K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, A. Chakrabarti, Automatic COVID-19 detection from X-ray images using Ensemble learning with convolutional neural network, Pattern Anal. Appl. 24 (3) (2021 Aug) 1111–1124.

[54] S. Mohana, S.A. Sahaaya, A. Mary, A comparative framework for feature selection in privacy preserving data mining techniques using pso and k-anonumization, Iioab Journal 7 (9) (2016) 804-811.