

COVID-19 Diagnosis Using Chaotic Logistic Map Based Modified Whale Optimization: A Robust Feature and Parameter Selection Approach



Suganthi Nachimuthu*^{ORCID}, Sarojini Kaliyamoorthi^{ORCID}

Department of Computer Science, L.R.G Govt Arts College for Women, Tirupur 641604, Tamilnadu, India

Corresponding Author Email: sugancaphd@gmail.com

<https://doi.org/10.18280/ria.370508>

ABSTRACT

Received: 6 May 2023

Revised: 28 July 2023

Accepted: 5 August 2023

Available online: 31 October 2023

Keywords:

chaotic map, Coronavirus disease 2019, Particle Swarm Optimization, Whale Optimization Algorithm, Support Vector Machine

The ongoing Coronavirus (COVID-19) pandemic poses a significant global health crisis due to its rapid transmission among humans and animals. Projections suggest that 90% of the world's population could potentially be affected in the years to come, underlining the critical need for early and accurate detection to mitigate the mortality rate. Previous models developed for COVID-19 prediction have primarily relied on manually-extracted features, a process that is time-intensive and prone to human error. In response to this challenge, this study introduces a novel diagnostic tool for COVID-19, leveraging machine learning approaches for efficient feature extraction and optimal parameter selection. Initially, features are extracted from collected Computed Tomography (CT) images using both the Gray Level Co-Occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRM) techniques. Subsequently, a Chaotic Logistic Map-based Modified Whale Optimization (CLM-MWO) algorithm is applied to select the optimal hyperparameters for a Neural Network (NN) classifier and to perform Feature Selection (FS) from the extracted features. The CLM-MWO is inspired by the prey-searching behaviours of whales and incorporates elements of chaos theory and logistic map to enhance exploration and exploitation capabilities. This approach improves the stability and convergence speed of the algorithm, enabling it to identify optimal features and fine-tune the parameters of the NN for superior classification performance. The features extracted through the CLM-MWO are then input into an improved Neural Network (INN) classifier, which facilitates the classification and prediction of COVID-19 from CT images. The proposed CLM-MWO-INN technique is validated through comparison with other classifiers frequently utilized in recent researches. Performance measurements indicate that the proposed method achieves an accuracy of 91.13% and 93.11% on two different CT image datasets. This accuracy surpasses that of other classifiers, demonstrating the potential of the proposed method for effective early detection of COVID-19.

1. INTRODUCTION

The infectious disease known as COVID-19 is attributed to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). First identified in Wuhan, the capital of China's Hubei province, in 2019, this novel disease rapidly propagated across the globe, culminating in the unprecedented COVID-19 pandemic of 2019-2020. The global dissemination of this virus has precipitated severe health threats and economic disruption worldwide. Common symptoms of the disease include fever, cough, shortness of breath, muscle aches, diarrhea, and sore throat [1].

Medical imaging techniques such as computed tomography (CT) and chest x-ray (CXR) are commonly used to detect lung disorders. Chest CT scans are superior to CXR because they are less influenced by other chest tissues and are better at identifying lung problems [2]. CT imaging also reveals the fundamental anatomical characteristics of COVID-19. Radiologists recommend chest CT scans as the primary lung clinical diagnosis technique since COVID-19 development may be effectively quantified using CT scans [3].

The aspects of CT scan digital images with the development of computer-assisted diagnostic (CAD) technologies have

benefited physicians in making rapid diagnosis. Depending on the features displayed in the CAD techniques and medical imaging may be used to provide a diagnosis. To sort Infected and not Infected CT scanning images, these frameworks typically use feature extraction, selection, and classification. When the dataset dimensions of predictive modelling continue to increase, feature selection (FS) process is adopted [4]. Various feature selection models have been adopted for COVID-19 prediction tasks. But still, the existing models struggle to provide high results on datasets that include complex feature relationships and high redundancy.

Many researchers have been inspired to examine the application of Artificial Intelligence (AI) models for identifying COVID-19 from different medical images and then categorize the patients [5]. AI is categorized into Machine Learning (ML) and Deep Learning (DL). ML techniques which have been widely utilized for many applications for medical imaging CAD attempts for rapid identification of COVID-19 at earlier stage.

For instances, an automated ML algorithm is developed for the detection of COVID-19 using CT and X-Ray images [6]. Initially, these images were pre-processed by normalization to enhance the quality of the image and removing of noise.

Secondly, the images were segmented using fuzzy c-means clustering. Then, essential features were selected from the feature vector using principal component analysis (PCA). Finally, the ML algorithms like k-Nearest Neighbour (k-NN) Sparse Representative Classifier (SRC), Artificial Neural Network (ANN) and Support Vector Machine (SVM) were employed to make decisions for normal, pneumonia, COVID-19 positive patients. On the other hand, classifier's performances are decided based on the predefined parameter alone. The regularization of hyper-parameters is very important for improving a network's generalization while handling different sample size datasets. In order to select the appropriate features and hyper-parameter for COVID-19 prediction, a Chaotic Logistic Map based Modified Whale Optimization with Improved Neural Network (CLM-MWO-INN) is presented in this paper to enhance the classifier's performances.

1.1 Contribution of this research

The major objective of this framework is to select the relevant features and hyperparameters for improving the classifier performance for COVID-19 detection. At first, the features are extracted from the collected CT images using GLCM and GLRM techniques. Then, CLM-MWO is applied to select the suitable features from the extracted features and hyper parameter tuning of NN for the best possible results. In this method, hyper parameter tuning and FS are executed concurrently, since the selection of a model's features might impact the hyperparameters' performance. Finally, the INN model is constructed to classify and detect the COVID-19 risk levels. As the COVID-19 cases increases, the proposed model automatically categorizes the diseases patients from CT scans potentially saving time for doctors and physicians.

2. LITERATURE SURVEY

Li et al. [7] constructed a multi-threshold optimization model for contour feature extraction of medical images. Initially, the gray theory is utilized to examine the degree of correlation between each pixel in accordance with the varying degrees of noise present in various medical images. Image contours were extracted after a series of thresholds which was established to develop a narrow band restriction. Once a number of constraint thresholds for contour features in the medical images have been determined, the key segmentation threshold for the image was using the Otsu approach. Genetic algorithms (GA) speed up the process of finding these thresholds. Additionally, random population generators employ processes like selection, crossover, mutation, etc. to formulate populations that are better able to adapt to their new environment. The golden section method was employed to solve various histograms in medical images, while the greedy approach was adopted to reach the ultimate threshold of contour feature extraction. However, the complete model was manually designed and results in high computational complexity.

Sharath Chander et al. [8] proposed a novel technique for detecting and classifying brain tumours. This model considers the Magnetic Resonance (MR) Images as input and the regions were segmented using adaptive K-means clustering algorithm. The Imadjust function was constructed to adjust the image intensities and Imfill operation was employed on the adjusted

images to fill the image regions and holes. The irrelevant pixels were removed by fixing the appropriate threshold ranges. The left-over region in the image would be the tumour region which was mapped on the actual MR image. However, this model results in slower convergence rate.

Zargari Khuzani et al. [9] developed an automated ML model called COVID-Classifier to diagnose the COVID-19 using chest x-ray images. In this model, a dimensionality reduction method was used to generate a set of optimal image features to construct an efficient ML classifier to distinguish COVID-19 cases from non-COVID-19 cases. Then, spatial and frequency-domain features were employed for features extraction by utilizing the texture, GLDM, GLCM, Fast Fourier Transform (FFT), and Wavelet Transforms (WT). Finally, the ML model was used to detect the COVID-19 disease. But, the quantity of learning data was limited.

Du et al. [10] suggested a ML application for the prediction of SARS-CoV-2 infection using blood tests and X-Ray images. Initially, a statistical comparison of blood tests in patients was performance with different aetiologies of pneumonia and COVID-19 infection. Then, the trained and validated ML models with basic blood tests comparison in comparison to RT-PCR testing was used to predict COVID-19 infection status, and explore different scenarios with adjunction of chest radiographs. However, the accurate hyperplane between the dataset ranges was highly complex to locate and detect the COVID-19.

Nageswaran et al. [11] presented a ML model for lung cancer classification and prediction. In this method, the lung CT images were collected and pre-processed geometric mean filter. Then, the K-means was utilized to segment the pre-processed images by effectively the facilitating the Region of Interests (ROIs). Finally, the ML models Random Forest (RF) and Artificial Neural Network (ANN) for the prediction of lung cancer. However, this model was easy prone to overfitting issues and the performance was ineffective.

3. MATERIALS AND METHODS

3.1 Feature extraction

The technique of assessing picture texture involves feature extraction. The findings provide a clearer picture of how texture and object shape are determined. When the method has a larger input data set, it should be reduced in size to make it easier to handle. Feature extraction is the process of converting an input image into a standard set of features. The clustered pixels were turned into quantitative data by using the feature extraction method on CT images of the healthy and COVID-19 types. The features used in this study are primarily the GLCM for extracting statistical features and GLRM for extracting run length features [12], as well as their procedures.

3.1.1 GLCM technique

The GLCM is described as a statistical paradigm that considers the spatial relationship between pixels [13]. It is a 2D histogram that calculates the relevance of each pixel in an image divided by a certain distance D . Assume that $J(a, b)$ is an image of size $U \times V$ with l gray levels, and that, $J(a_1, b_1)$ and $J(a_2, b_2)$ are two pixels with gray level intensities x_1 and x_2 respectively. Considering, $\Delta_a = a_2 - a_1$ in the direction of a and $\Delta_b = b_2 - b_1$ in the direction of b , the intersecting line has the identical interval θ as

$\arctan\left(\frac{\Delta_b}{\Delta_a}\right)$. The modified co-incident matrix $CO_{\theta,D}$ is depicted in Eq. (1):

$$CO_{\theta,D}(i_1, i_2) = \left(Num\left\{\left((a, b_1), (a_2, b_2)\right) \in (U \times V) * \right\} \setminus M\right) \quad (1)$$

In above Eq. (1), $U \times V$ defines the numerical values in the co-incident matrix whereas M represents the total number of pixels. This model defines Z as a resulted condition given in Eq. (2):

$$D\sin(\theta), \Delta_b = D\cos(\theta) \quad (2)$$

In general, θ equals to 0.459013 degrees and D ranges from 1 to 5. In the study [14], the five distinct texture features like power, variance, contrast, and autocorrelation, are concisely presented based on the applicability of co-incident matrices.

3.1.2 GLRM technique

GLRM is referred to as a mathematical element network. It returns a percentage of the pixels' intensity along the supplied guidance, referred to as Run length. The GLRM estimates the dimension of homogenous cycles for every level of gray. It is calculated for four different orientations, each of which has eleven texture indices derived from the matrices. The (a, b) component of GLRLM connects homogeneous segments of b pixels with intensity a from an image which is termed as GLRLM. Considering, the above-mentioned GLRM matrix with 11 features is achieved. 44 features (11 features 4 directions) have been completed for all images [15]. Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-uniformity (GLN), Run Length Non-uniformity (RLN), Run Percentage (RP), Low Gray Level Run Emphasis (LGRE), and High Gray Level Run Emphasis (HGRE) are utilized in this method to determine the categorization performance [14].

These features use pixel grey levels in sequence to distinguish textures with similar SRE and LRE values but different grey level distributions. After that, using Histogram features, GLCM, GRLM, and their combinations, create feature sets. Totally For each CT scan normal and COVID-19 picture independently, twenty-nine characteristics were retrieved. The data set obtained by both techniques of feature extraction procedure is too vast and hence, it is difficult to implement as input to the classifier. As a result, the FS technique, which is detailed in the next section, is used to reduce dimensionality.

3.2 Feature and parameter optimization

To achieve adequate classification accuracy, a FS strategy is used to reduce the number of retrieved characteristics by excluding irrelevant and redundant features [16]. In categorization problems, FS is a common and crucial strategy for gaining more precision and a quicker convergence. This section controls the selection of the best attributes from a dataset using optimization agents. To study and evaluate the effectiveness of the proposed optimizer, the Genetic Alogrithm (GA), Particle Swarm Optimization (PSO), and WOA-based FS algorithms are presented. This component handles the optimizer's agents in order to efficiently choose features. Here, the historical context of GA, PSO, and WOA is discussed.

3.2.1 GA

The normal selection and natural selection theories communicated in genetic engineering begin with the Genetic Algorithm. The selection cycle is an interaction in which the number of components is reduced to a small subset, resulting in a high precision order. The genetic algorithm [17] is outlined in the steps below.

Algorithm: GA (n, χ_f, χ_p, μ), where n is the number of individuals in the population; χ_f is the total number of chosen features from each chromosome; χ_p is the cumulative number of accepted parameters from each chromosome; and μ is the mutation rate.

//Initialize generation 0: $k := 0$; $P_k :=$ a population of n individuals produced at random;

//Determine P_k :

Configure

$$Fitness F(i) = \frac{\max_{\vartheta} (accuracy(\theta_{ki}) - \alpha * \frac{selected\ features(\theta_{ki})}{Total\ number\ of\ features(n)})}{\vartheta} \quad (3)$$

// $F(i)$ = fitness function; ϑ is the initial population, θ_{ki} represents the i^{th} chromosome in population, $accuracy(\theta_{ki})$ depicts the classification performance achieved by using chromosome θ_{ki} , α is a weighting parameter

Do

{

/Create generation $k + 1$;

1 (i) **Copy**: Select $(1 - \chi_f) \times n$ members and replace them in $P_k + 1$;

(ii) **Copy**: Choose $(1 - \chi_p) \times n$ members and transfer them into $P_k + 1$;

2 (i) **Crossover**: Identify $\chi_f \times n$ members of P_k , couple them, generate offspring, and add the offspring into $P_k + 1$;

(ii) **Crossover**: Acquire $\chi_p \times n$ members of P_k ; couple them; produce offspring; and integrate the offspring into $P_k + 1$.

3 (i) **Reconstruct**: Evaluate $\mu_f \times n$ members of $P_k + 1$; shift an arbitrary bit in every feature. //Compute $P_k + 1$;

(ii) **Reconstruct**: Determine $\mu_p \times n$ members of $P_k + 1$; reverse a randomized bit in every parameter //Compute $P_k + 1$;

Calculate $fitness(i)$ for every $i \in P_k$;

Increment: $k := k + 1$;

}

While the fitness of the fittest individual in P_k is insufficient;
Return the fittest member of P_k ;

3.2.2 PSO

PSO is not a well-established improvement approach based on herds of birds or, more likely, schools of fish. The flock of birds (swarm) has developed a cooperative strategy for obtaining food, with every bird in the flock changing their chase strategy in accordance to updated information [18]. The concept of PSO calculation is linked to developmental calculations and a plethora of fictitious life frameworks. The Birds fly through the multidimensional chase space invisibly. Throughout the journey, each particle moves at its own unique speed in relation to space. The entire population is renewed by refreshing each particle. The multiplicity game plan propels itself forward, aiming for the highest objective capacity esteem point, where the particles eventually congregate. The following are the methods for improving molecular swarms:

Step 1: Initialization X_f and X_p using swarm particles $\setminus X_f$ selected feature by particle, X_p selected parameter by particle.

Step 2: Calculate fitness of each particle $F(i)$ using Eq. (3).

Step 3: Evaluate the particle fitness for selected feature.

If Fitness of $X_{fi} > pbest_i \setminus X_{fi}$ is the current position of particle i for feature f .

$$pbest_i = X_{fi}.$$

If fitness of $pbest_i > gbest_i$.

$$gbest_i = pbest_i.$$

Step 4: Evaluate the particle fitness for selected parameter.

If Fitness of $X_{pi} > pbest_i \setminus X_{pi}$ is the current parameter of particle i for parameter p .

$$pbest_i = X_{pi}.$$

If fitness of $pbest_i > gbest_i$.

$$gbest_i = pbest_i.$$

Step 5: Update the velocity of particle i for feature and parameter-At each cycle the rates of the multitude particles are determined by Eq. (4) and Eq. (5):

$$F(V_{id}^{t+1}) = W * V_{id}^{t+1} + c_1 * r_{1i} * (p_{fi} - x_{fi}^t) + c_2 * r_{2i} * (g_{fi} - x_{fi}^t) \quad (4)$$

$$P(V_{id}^{t+1}) = W * V_{id}^{t+1} + c_1 * r_{1i} * (p_{pi} - x_{pi}^t) + c_2 * r_{2i} * (g_{pi} - x_{pi}^t) \quad (5)$$

where, t denotes the replication in the process;

d denotes the dimension in the exploration space.

W is inertia weight;

c_1 and c_2 are acceleration constants.

r_1 and r_2 are arbitrary range consistently dispersed in $[0, 1]$.

$pbest_i$ defines the best position of the particle.

$gbest_i$ defines the best area recalled by the particle.

p_{fi} and p_{pi} represent the best position of particle in p_{best} for feature and parameter selection.

g_{fi} and g_{pi} represent the best area reassigned by particle in g_{best} for feature and parameter selection.

The PSO is created by taking into account the components of each particle. The weighting perspectives c_1 and c_2 protect the particles from colliding (people). After refreshing particle I , the speed (v) and random number (r) are verified and ensured in a way shown, to avoid a crash.

Step 6: Update the position of particle i for feature and parameter updating using:

$$x_{fi}^{t+1} = x_{fi}^t + V_{fi}^{t+1} \quad (6)$$

$$x_{pi}^{t+1} = x_{pi}^t + V_{pi}^{t+1} \quad (7)$$

Step 7: Step 4. If stopping criterion is not met, continue Steps 2 and 3.

Step 8: Return g_{best} and its fitness values.

This is a considerable computational benefit over GA when the population is considerably large. The activity of reducing actual numbers is used to calculate movement and position. The PSO alters its speed (accelerating) toward p_{best} (Personal Best) and g_{best} (Global Best) regions (neighbourhood interpretation of PSO) and accelerates toward p_{best} and g_{best} areas.

3.2.3 WOA

Recently, there has been a lot of buzz about WOA, which

was proposed in the study [19]. WOA is a population-based meta-heuristic technique that simulates humpback whale hunting behavior. Humpback whales will dive twelve meters underwater and lead a spiral of bubbles raising to the surface after they have identified their meal. Encircling prey, spiral bubble-net feeding maneuver, and hunt for prey are the three primary stages of position information [19].

Exploitation phase

Throughout the exploitation process, humpback whales assess the available best location as the target prey (global optimum) and modify their locations toward the global optimum based on two systems like shrinking encircling (circling prey) and circular elevating location in spiral bubble-net feeding function. The module of the shrinking encircling operation for FS and parameter selection is represented as follows:

$$D_f = |C \bullet X_f^*(t) - X_f(t)| \quad (8)$$

$$X_f(t+1) = X_f^*(t) - A \bullet D_f \quad (9)$$

$$D_p = |C \bullet X_p^*(t) - X_p(t)| \quad (10)$$

$$X_p(t+1) = X_p^*(t) - A \bullet D_p \quad (11)$$

From Eq. (8) to Eq. (11), X_f is the position vector for selecting features and parameters, X_f^* and X_p^* indicate the optimal elucidation acquired for identifying features and parameters, which will be changed in every repetition if a finest solution is determined, t denotes the present iteration, D denotes distance between X position of i^{th} whale and X^* best solution whale. $|\cdot|$ is the relative valuation, and \bullet is an element-to-element multiplication. Here, A and C are two constants that are computed using the formula:

$$A = 2a \bullet r - a \quad (12)$$

$$C = 2 \bullet r \quad (13)$$

where, r is an arbitrary integer between 0 and 1. Throughout the repetitions in both the exploration and exploitation stages, it drops-out linearly from 2 to 0 so that the shrinking encircling behavior may be realized. The statistical expression of the spiral upgrading position for feature and parameter is a spiral Eq. (14)-Eq. (17):

$$F(D_f') = |X_f^*| (t) - X_f(t) - X_f(t) \quad (14)$$

$$X_f(t+1) = D_f' \bullet e^{bl} \bullet \cos(2\cos\pi l) + X_f^*(t) \quad (15)$$

$$P(D_p') = |X_p^*| (t) - X_p(t) \quad (16)$$

$$X_p(t+1) = D_p' \bullet e^{bl} \bullet \cos(2\cos\pi l) + X_p^*(t) \quad (17)$$

where, D' indicates the distance between the i^{th} whale and the best solution obtained so far, l is a random value in the interval $[-1, 1]$, and b is a constant that defines the form of a logarithmic spiral. It is important to note that when whales grab prey, both the spiral-shaped course and the diminishing encirclement occur concurrently. To simulate this behavior, each mechanism is constructed with a 50% chance for each characteristic and parameter.

$$X_f(t+1) = \begin{cases} X_f^*(t) - A \bullet D_f \text{ if } r < 0.5 \\ D_f' \bullet e^{bl} \bullet \cos(2\cos\pi l) + X_f^*(t) \text{ if } p < 0.5 \end{cases} \quad (18)$$

$$X_p(t+1) = \begin{cases} X_p^*(t) - A \bullet D_p \text{ if } r < 0.5 \\ D_p' \bullet e^{bl} \bullet \cos(2\cos\pi l) + X_p^*(t) \text{ if } r < 0.5 \end{cases} \quad (19)$$

where, r is a random integer between 0 and 1.

Exploration phase

Here, a global query is constructed to facilitate the investigation of feature and parameter data. Its statistical structure is comparable to Eq. (20) to (23), with the distinction that an arbitrary search agent, rather than the best agent, is assigned to direct the search. To upgrade the location, the stochastic parameter $|A|$ with a range greater than 1 or less than 1 is used to identify between exploration (search for prey) or exploitation (shrinking encircling mechanism).

$$D_f = |C \bullet X_{rand} - X_f| \quad (20)$$

$$X_f(t+1) = X_{rand} - A \bullet D_f \quad (21)$$

$$D_p = |C \bullet X_{rand} - X_p| \quad (22)$$

$$X_p(t+1) = X_{rand} - A \bullet D_p \quad (23)$$

where, X_{rand} is a location vector selected randomly from the present generation. The WOA pseudo code is described in Algorithm. Note that WOA may seamlessly transition among the stages of exploration and exploitation based on a single parameter. Although WOA is comparable with certain conventional MAs, it still has disadvantages such as poor precision and abrupt convergence. Therefore, modified WOA is used in this work.

Algorithm for WOA

- Step 1: Begin
- Step 2: Determine the whale population for feature X_f ($i = 1, 2, 3 \dots, n$) and parameter X_p ($i = 1, 2, 3 \dots, n$).
- Step 3: Compute the fitness of each search agent $F(i)$ using Eq. (3).
- Step 4: Determine best search agent for feature X_f^* and parameter X_p^* .
- Step 5: While ($t < \text{maximum number of iterations}$).
- Step 6: For each exploration agent, update a, A, C, l and r .
- Step 7: if 1 ($r < 0.5$).
- Step 8: if 2 ($|A| < 1$).
- Step 9: Improve the ranking of present search agent for a feature using Eq. (9).
- Step 10: Improve the present search agent's position for parameter using Eq. (11).
- Step 11: else if 2 ($|A| \geq 1$).
- Step 12: Choose a random search agent (X_{rand}).
- Step 13: Enhance the location of current search agent for feature using Eq. (21).
- Step 14: Determine the current search agent position for parameter using Eq. (23).
- Step 15: end if 2.
- Step 16: else if 1 ($r \geq 0.5$).
- Step 17: Explore the location of current search agent for feature using Eq. (15).

Step 18: Upgrade the location using current search agent for parameter using Eq. (17).

Step 19: end if 1.

Step 20: end for.

Step 21: Inspect and fix redundant genes to ensure the validity of each search agent.

Step 22: Place X_f^* and X_p^* if there is a preferable solution is determined.

Step 23: $t = t + 1$.

Step 24: end while }.

Step 25: return to step 4 and step 5.

For the classifier to categorize the numerous feature categories, FS will be converted into a more consistent and relevant sort of feedback. This research introduces the chaotic-based whale optimized characteristics for machine approaches to categorize COVID-19 and non-COVID 19 characteristics in CT images. Using a ML algorithm, the accuracy, precision, and recall of the suggested method are compared to those of the GA, PSO, and whale methods.

4. PROPOSED METHODOLOGY

The Figure 1 devise the framework of the proposed experiment.

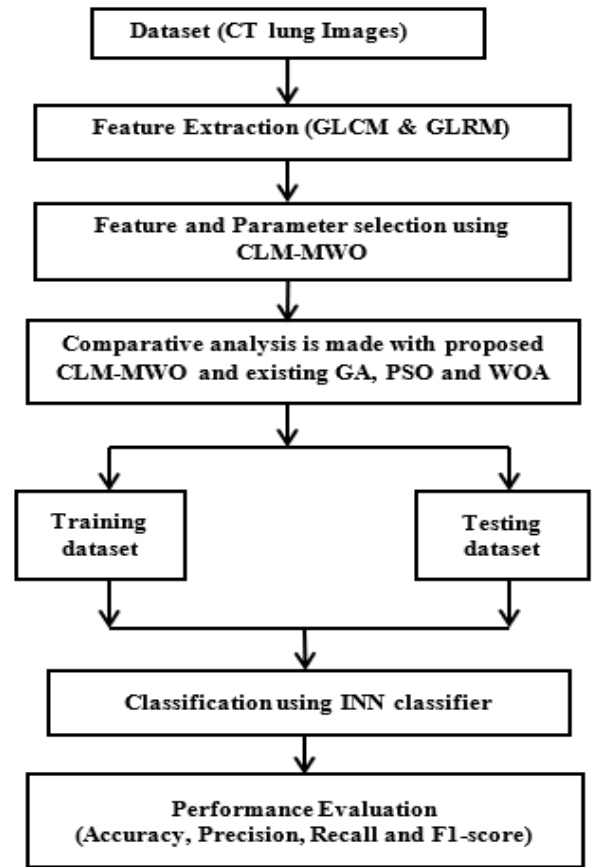


Figure 1. Block structure of the proposed model

In this paper, the proposed CLM-MWO-INN model is briefly illustrated. The features of CT images are extracted using a GLCM and GLRM techniques. The GLCM uses the second-order statistics of images to efficiently obtain the textural relationship among pixel areas whereas GLRM is a texture representation model which employs high-order

statistics to determine the spatial plane properties of each pixel. These two approaches for feature extraction boost pixel intensity, making it simpler to locate areas of varying tissue quality. In order to select the relevant features from the complete extracted features, CLM-MWO algorithm is used. For feature selection, CLM-MWO selects the optimal feature randomly initialized search agents with a set of solutions which are feature subsets. At each iteration, each search agent updates its position based on a predefined fitness function. The classification performance is used as the fitness function. The best feature subset is one which maximizes the classification accuracy and minimize the selected features.

For the parameter optimization, this model optimizes the weights initialization and regularization hyper-parameters concurrently with the Multilayer Perceptron (MLP) topology and learning hyper-parameters. In addition, the correlation among these additional hyperparameters and classification efficiency will be examined in order to determine the implications of these hyperparameters on categorization performance. It will enable the identification of hyperparameters space locations with the greatest identification performance. It will then be able to minimize the query space and construct a CLM-MWO. Moreover, hyper parameter tuning and FS are also addressed concurrently in this phase, since the identification of a features in the model's structure might determine which hyperparameters functions well. Finally, the selected features are given as input to the INN model for COVID-19 detection and classification.

4.1 CLM-MWO algorithm

The MWO algorithm is an improved version of the WOA algorithm that enhances the hunting behavior of humpback whales. By modifying the control factor and incorporating other exploring techniques, the WOA is adjusted. The key concern of MWO algorithm is the purposeful combination speed. As a result, this study incorporates chaos technique into the MWO algorithm process in order to accelerate total convergence speed and get superior results. In the proposed method for managing the essential factor of WOA, several chaotic map functions are assessed to aid in governing study and utilisation. On numerous well-known test maps, the proposed approach is predefined. The findings demonstrate that chaotic maps, particularly logistic maps, can improve WOA presentation.

i. Chaotic systems

Chaos is an unbalanced active behaviour that includes infinite unbalanced periodic waves in nonlinear ways, as well as difficult requirements on the primary conditions. It has a logical core structure and three key active features. Initially, the Ergodic characters is the largest, and it might utilize absolute ranging expressions to explore all nodes in the system's query plane [20]. These ergodic characters will provide a statistical representation of the chaotic dynamical system by attaching relevant probability measures which improves the convergence speed of the classifier. The Chaos technique is presented to enhance the value of exploring the global optimum in numerous optimization applications by avoiding optimal problems in local optima. As a result, the searching range [0, 1] can be used for a FS system that uses chaos theory to optimise the features. This n-dimensional map is a time-isolated nonlinear method described by the preceding equation:

$$cu_i^{(k+1)} = f(cu_i^{(k)}), i = 1, 2, 3, \dots, n \quad (24)$$

A chaotic order may be assessed by executing the network function and defining the early state of $cu_i^{(0)}$ $cu_i(0)$; chaotic orders can then be described in the form of $cu_i^{(k)}$, where $k = 0, 1, 2$, and so on. Calculating a chaotic vector value, which is determined using a chaotic factor from a chaotic system with the Ergotis condition, is used to create the search function of a chaotic evolution system. By using a variety of chaotic functions, the proposed system can evolve its versions in a variety of ways. To analyse the FS/optimization performances of the proposed system, this research presents three types of chaotic map systems: logistic p, tent, and Gaussian.

ii. Chaotic maps

In this work, a chaotic strategy is used to drive the WOA display in order to avoid a stuck issue at the neighbourhood optima while also speeding it up. It's also used to deal with WOA's random factor values [20]. A chaotic map is an evolution operation that depicts the chaotic behavior. A discrete or continuous-time parameter can be used to parameterize the maps. Discrete maps are often iterated functions. These maps depict the complex and dynamic behaviours that occur in nonlinear systems as well as the determination of system states. Chaotic maps were widely employed to produce random numbers and to replace the pseudo-random numbers of gaussian distributions. The maps might be exploited to remove the uncertainty of pseudo randomness, and hence improved stability could be desired for the classifier. Three different chaotic maps are used in this study to improve FS performance, which are discussed below.

(A) Logistic map: This type of map is a two-degree polynomial mapping that may be created from a chaotic situation utilising a basic nonlinear system. The mathematical expression is demonstrated in the following equation:

$$x_n = \mu * x_{n-1} * (1 - x_{n-1}) \quad (25)$$

This is a factor that adopts the system's performance; it's usually set to (0, 4). For values of $\mu > 4$, this map does not travel into interval [0, 1], but if $\mu = 4$, the map system emerges from chaotic map.

(B) Tent map: In this system, the parameter μ is classified (0, 2). If the value of $\mu = 2$, the tent map proves a range of dynamical performances from likely to chaotic. The chaotic map defined by the below equation:

$$x_n = \begin{cases} \mu * x_{n-1} & \text{if } 0 \leq x_n < 0.5, \\ \mu * (1 - x_{n-1}) & \text{if } 0.5 \leq x_n \leq 1. \end{cases} \quad (26)$$

(C) Gaussian map: It is the non-linear map with underlying function is a Gaussian, as described in Eq. (27):

$$x_n = \exp(-\alpha * x_{n-1}^2) + \beta \quad (27)$$

The above mentioned chaotic maps are determined to enhance the WOA function by preventing it from becoming trapped at the local optimum and increases the convergence rate. These chaotic maps are used to change the values of WOA's arbitrary parameters. The initial point of all chaotic maps is adjusted to 0.7 while the rest of the parameters are configured since the initial values of chaotic map might leads to the considerable influence on the fluctuation pattern. Based on trial and error, these settings have been determined to be

the best. The parameters a, A, C, l and r are thought to be the most important elements influencing WOA's convergence behavior. a, A, C and r influences the bubble-net attacking strategy (exploitation phase) with a and A influencing the shrinking encircling mechanism and l influencing the spiral model. The parameter X_p indicates the probability of to select the spiral-model or the shrinking encircling mechanism to update the whales' locations during the optimization. The l strikes modified positions in the exploration phase. The effectiveness of each of these parameters alone and in combination with various chaotic maps over WOA.

The proposed CLM-MWO is employed as a feature selection algorithm in a wrapper mode to choose the best possible collection of features. The CLM-MWO based feature selection method initiates with randomly initialized search agents which are considered to be the feature subsets. A separate combination of features of varying sizes is used in each of the feature subsets. Each search agent constantly adjusts its position to maximize some fitness function which is depicted in Eq. (3). Finally, the maximum classification accuracy with minimum feature selection is determined which is adopted as the ideal feature subset to increase the detection performances.

4.2 Classification model

For the classification, this work adopts for INN model to address the arrangement as a standard characterization problem involving logical elements. The controlled learning calculations (INN) were employed to characterize the COVID-19 disease in this study. The controlled learning strategies were based on the indicated examples in the preparation informative index, which aids the learning models in effectively preparing. As a result, excellent characterization precision can be achieved.

The hyper parameter optimized NN is known as INN model which composed of layers and the connections made up between neurons from their adjacent layers. The input feature vector from CLM-MWO are fed as input to the first layer and moving onto layers to layers which is transformed to a high level features vectors. All the layers in INN classifier constitutes a set of trainable layers, and it determines the dot production among the input of the layer and the trainable layers to produce an activation map. The filters are known as kernels which enable the detection of the same features from various locations. The number of neurons in the output layer that eventually contribute to classification which is equal to the number of classifying classes i.e., COVID-19 prediction.

The pseudocode for the suggested method is shown below, based on the information presented above.

Pseudo-code for proposed algorithm

Step 1: Initialize the generation counter g and randomly initialize the whale population for feature X_f ($i = 1, 2, 3 \dots, n$) and parameter X_p ($i = 1, 2, 3 \dots, n$).

Step 2: Utilize an Eq. (3) to examine the performance of every search agent in order to discover the optimal search agent.

Step 3: Initialize the value of the chaotic map x_0 randomly.

Step 4: while ($t <$ maximum number of iterations).

Step 5: Update the chaotic map using the respective chaotic map Eq. (26) and Eq. (27).

Step 6: For each search agent, upgrade $a, A, C, l, 1 p$ and $2 p$.

Step 7: if1 ($1 p < 0.5$).

Step 8: if2 ($|A| < 1$).

Step 9: Modify the location of the present search agent for feature and parameter by using Eq. (9).

Step 10: else if 2 ($|A| \geq 1$).

Step 11: Determine a random search agent (X_{rand}).

Step 12: Using the Eq. (21) and Eq. (23), adjust the location of the present search agent for feature and parameter.

Step 13: end if 2.

Step 14: else if 1 ($1 p \geq 0.5$).

Step 15: if 3 ($2 p < 0.6$).

Step 16: Adjust the feature and parameter positions of the present search agent using Eq. (15) and Eq. (17).

Step 17: else if 3 ($2 p \geq 0.6$).

Step 18: The Eq. (9) and Eq. (10) are used to modify the location of the current search agent for feature and parameter.

Step 19: end if 3.

Step 20: end for.

Step 21: Identify and replace redundant genes to ensure the validity of each search agent.

Step 22: Calculate the fitness of each search agent.

Step 23: X_f^* and X_p^* should be updated if a better solution is initialized.

Step 24: $t = t + 1$.

Step 25: end while }

Step 26: return to step 2.

5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Dataset description

In order to prove the efficiency of the proposed model, two different CT images dataset have been collected which is illustrated below.

Dataset 1: SARS-CoV-2 CT scan dataset [21] have been collected from real patients in hospitals from Sao Paulo, Brazil. The aim of this dataset is to encourage the research and development of artificial intelligent methods which are able to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans. The total CT images obtained from the given dataset for the performance evaluation are listed in Table 1.

Table 1. Observed CT images for COVID-19

Lung Disease Categorizes	Number of Images Considered
Covid	1252
Normal	1230
Total	2482

Table 2. Observed CT images for various lung diseases categories

Lung Disease Categorizes	Number of Images Considered
Covid	1002
Normal	984
Pneumonia	1762
Atelectasis	310
Infiltrate	260
Total	4318

Dataset 2: In this dataset, The CT images also opted for five diseases categories like atelectasis, infiltration, pneumonia along with COVID-19 and non-Covid. In this study, different

open public portals are analyzed to collect the CT data for the experiment purposes. The lung atelectasis images are taken from [22], The Covid and Non-Covid (Normal) images are gathered from [23], the viral pneumonia is taken from the study [24] and infiltration is obtained from [25]. The total CT images obtained from the given dataset for the performance evaluation are listed in Table 2.

5.2 Performance measures parameters

The attributes collected from the datasets are used in the training and testing procedures. In this assessment, 60% of the data was utilized for training and the remaining 40% for testing. The performance evaluation metric evaluated for the various methods is discussed below.

- TP (True Positive)-The number of normal images is identified as normal.
- TN (True Negative)-The number of disease images identified as disease.
- FP (False Positive)-The number of normal images is identified as disease images.
- FN (False Negative)-The number of diseased images is identified as normal.

Accuracy: The proportion of correct predictions. It can be calculated using:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{28}$$

Precision: The proportion of anticipated positive models which truly are positive.

$$Precision = \frac{TP}{TP+FP} \tag{29}$$

Recall: Recall (hit rate/sensitivity) is a metric that indicates how well a classifier recognizes positive examples.

$$Recall = \frac{TP}{TP+FN} \tag{30}$$

F1 Score: It is the ‘Harmonic Mean’ of recall with accuracy.

$$F1_Score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{31}$$

In this work, the confusion matrix was employed to characterize the efficiency of the proposed algorithm on given dataset 1 and dataset 2. It facilitates for the modeling of algorithm execution. It also makes detecting discrepancies across classes much easier. The purpose of this technique is to evaluate the efficiency of CLM-MWO using the INN algorithm. The suggested model is contrast with MLP, Random Forest (RF) and Extreme Learning Machine (ELM). Different dataset was used to evaluate it, as stated in the Table 3 and Table 4.

The Figure 2 shows the parameters like Accuracy, Precision, Recall and F1-Score have been determined and that the suggested feature selection approach outperforms previous algorithms for dataset-I. For instance, the accuracy of CLM-MWO-INN is 13.71%, 10.11%, 6.61% and 3.08% higher than the existing SVM, MLP, RF and ELM classifiers. The suggested approach improves INN classification accuracy by incorporating feature optimization techniques.

The Figure 3 shows comparison values of accuracy, precision, recall and F1-Score for proposed and existing

models using the dataset-II. In dataset-II, different categories of lung diseases have been considered for this research work to prove the proposed model’s efficiency. From the Figure 3, it has been observed that the accuracy of CLM-MWO-INN is 10%, 6.64%, 4.36% and 2.75% higher than the SVM, MLP, RF and ELM models correspondingly.

Table 3. Performance evaluation for dataset-I

Performance Metrics	SVM	MLP	RF	ELM	CLM-MWO-INN
Accuracy	80.14	82.76	85.48	88.41	91.13
Precision	78.42	81.52	83.66	87.84	90.91
Recall	82.72	84.65	87.76	89.42	91.63
F1-Score	80.51	83.05	85.66	88.63	91.27

Table 4. Performance evaluation for dataset-II

Performance Metrics	SVM	MLP	RF	ELM	CLM-MWO-INN
Accuracy	85.34	87.31	89.22	90.61	93.11
Precision	69.04	72.52	76.27	80.33	85.30
Recall	73.11	76.85	79.95	81.47	85.92
F1-Score	71.02	74.62	78.07	80.90	85.61

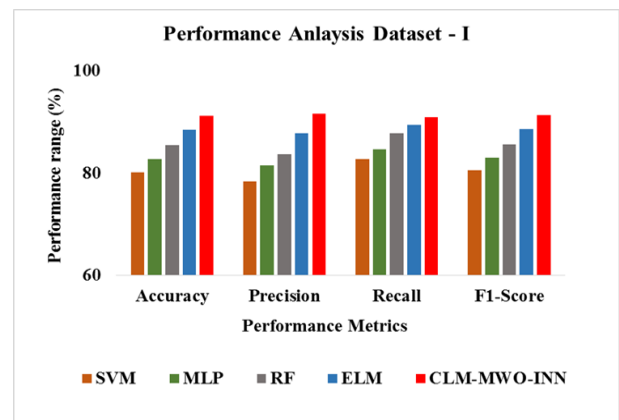


Figure 2. Performance analysis for dataset-I

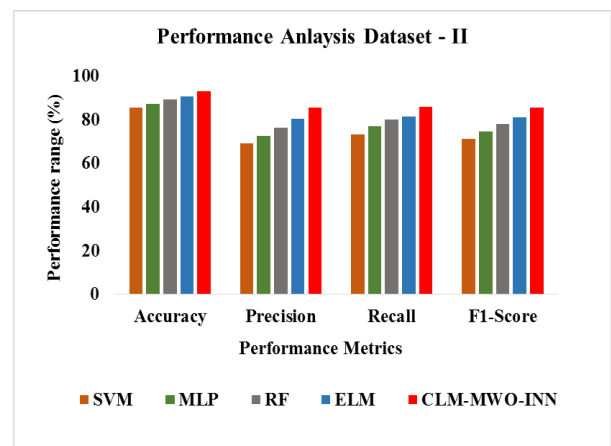


Figure 3. Performance analysis for dataset-II

6. CONCLUSION

In this paper, CLM-MWO-INN is developed to detect and classify the COVID-19 dataset as normal or abnormal. The feature extraction models like GLCM and GLRM techniques

are used to extract the features from the collected CT images. Then, the CLM-MWO is utilized for the relevant selection of features and appropriate hyperparameters to improve the model's performances. Finally, the INN model is employed for the prediction and classification of COVID-19 diseases. The experimental finding revealed that the proposed CLM-MWO-INN achieves the accuracy of 91.13% and 93.11% on CT dataset I and dataset II respectively, outperforms than other existing models. In hospitals, the proposed model automatically detects the diseased patients from the CT images which lessens the manual inceptions and time of doctor and physicians. In future, DL model will be adopted with CLM-MWO to enhance the classification outcomes for COVID-19 detection.

REFERENCES

- [1] Elmuogy, S., Hikal, N.A., Hassan, E. (2021). An efficient technique for CT scan images classification of COVID-19. *Journal of Intelligent & Fuzzy Systems*, 40(3): 5225-5238. <https://doi.org/10.3233/JIFS-201985>
- [2] Kassania, S.H., Kassanib, P.H., Wesolowski, M.J., Schneidera, K.A., Detersa, R. (2021). Automatic detection of coronavirus disease (COVID-19) in x-ray and CT images: A machine learning based approach. *Biocybernetics and Biomedical Engineering*, 41(3): 867-879. <https://doi.org/10.1016/j.bbe.2021.05.013>
- [3] Vliegenthart, R., Fouras, A., Jacobs, C., Papanikolaou, N. (2022). Innovations in thoracic imaging: CT, radiomics, AI and x-ray velocimetry. *Respirology*, 27(10): 818-833. <https://doi.org/10.1111/resp.14344>
- [4] Jain, D., Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, 19(3): 179-189. <https://doi.org/10.1016/j.eij.2018.03.002>
- [5] Soomro, T.A., Zheng, L.B., Afifi, A.J., Ali, A., Yin, M., Gao, J.B. (2022). Artificial intelligence (AI) for medical imaging to combat coronavirus disease (COVID-19): a detailed review with direction for future research. *Artificial Intelligence Review*, 55: 1409-1439. <https://doi.org/10.1007/s10462-021-09985-z>
- [6] Bhargava, A., Bansal, A., Goyal, V. (2022). Machine learning-based automatic detection of novel coronavirus (COVID-19) disease. *Multimedia Tools and Applications*, 81(10): 13731-13750. <https://doi.org/10.1007/s11042-022-12508-9>
- [7] Li, W., Huang, Q., Srivastava, G. (2021). Contour feature extraction of medical image based on multi-threshold optimization. *Mobile Networks and Applications*, 26: 381-389. <https://doi.org/10.1007/s11036-020-01674-5>
- [8] Sharath Chander, P., Soundarya, J., Priyadharsini, R. (2020). Brain tumour detection and classification using k-means clustering and SVM classifier. In *RITA 2018: Proceedings of the 6th International Conference on Robot Intelligence Technology and Applications*, Springer Singapore, 49-63. https://doi.org/10.1007/978-981-13-8323-6_5
- [9] Zargari Khuzani, A., Heidari, M., Shariati, S.A. (2021). COVID-classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. *Scientific Reports*, 11(1): 9887. <https://doi.org/10.1038/s41598-021-88807-2>
- [10] Du, R., Tsougenis, E.D., Ho, J.W.K., Chan, J.K.Y., Chiu, K.W.H., Fang, B.X.H., Ng, M.Y., Leung, S., Lo, C.S.Y., Wong, H.F., Lam, H.S., Chiu, L.J., So, T.Y., Wong, K.T., Wong, Y.C.I., Yu, K., Yeung, Y., Chik, T., Pang, J.W.K., Wai, A.K., Kuo, M.D., Lam, T.P.W., Khong, P., Cheung, N., Vardhanabhuti, V. (2021). Machine learning application for the prediction of SARS-CoV-2 infection using blood tests and chest radiograph. *Scientific Reports*, 11(1): 14250. <https://doi.org/10.1038/s41598-021-93719-2>
- [11] Nageswaran, S., Arunkumar, G., Bisht, A.K., Mewada, S., Kumar, J.N.V.R.S., Jawarneh, M., Asenso, E. (2022). Lung cancer classification and prediction using machine learning and image processing. *BioMed Research International*, 2022. <https://doi.org/10.1155/2022/1755460>
- [12] Tadist, K., Najah, S., Nikolov, N.S., Mrabti, F., Zahi, A. (2019). Feature selection methods and genomic big data: A systematic review. *Journal of Big Data*, 6(1): 1-24. <https://doi.org/10.1186/s40537-019-0241-0>
- [13] Raj, R.J.S., Shobana, S.J., Pustokhina, I.V., Pustokhin, D.A., Gupta, D., Shankar, K. (2020). Optimal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access*, 8: 58006-58017. <https://doi.org/10.1109/ACCESS.2020.2981337>
- [14] Shankar, K., Perumal, E., Tiwari, P., Shorfuzzaman, M., Gupta, D. (2022). Deep learning and evolutionary intelligence with fusion-based feature extraction for detection of COVID-19 from chest x-ray images. *Multimedia Systems*, 28(4): 1175-1187. <https://doi.org/10.1007/s00530-021-00800-x>
- [15] Khojastehnazhand, M., Ramezani, H. (2020). Machine vision system for classification of bulk raisins using texture features. *Journal of Food Engineering*, 271: 109864. <https://doi.org/10.1016/j.jfoodeng.2019.109864>
- [16] Shuaib, M., Abdulhamid, S.I.M., Adebayo, O.S., Osho, O., Idris, I., Alhassan, J.K., Rana, N. (2019). Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification. *SN Applied Sciences*, 1: 1-17. <https://doi.org/10.1007/s42452-019-0394-7>
- [17] Wang, H.Z., He, C.Q., Li, Z.P. (2020). A new ensemble feature selection approach based on genetic algorithm. *Soft Computing*, 24: 15811-15820. <https://doi.org/10.1007/s00500-020-04911-x>
- [18] Tran, B., Xue, B., Zhang, M.J. (2014). Overview of particle swarm optimisation for feature selection in classification. In *Simulated Evolution and Learning: 10th International Conference*, Springer International Publishing, 605-617. https://doi.org/10.1007/978-3-319-13563-2_51
- [19] Mirjalili, S., Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95: 51-67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
- [20] Demir, F.B., Tuncer, T., Kocamaz, A.F. (2020). A chaotic optimization method based on logistic-sine map for numerical function optimization. *Neural Computing and Applications*, 32: 14227-14239. <https://doi.org/10.1007/s00521-020-04815-9>
- [21] <https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset>
- [22] <https://radiopaedia.org/articles/lung-atelectasis?lang=us>
- [23] <https://www.kaggle.com/datasets/mehradaria/covid19-lung-ct-scans>

[24] <https://radiopaedia.org/articles/viral-respiratory-tract-infection>

[25] <https://radiopaedia.org/playlists/41156?lang=us>