healthcare

MDPI

Article

Enhanced Gravitational Search Optimization with Hybrid Deep Learning Model for COVID-19 Diagnosis on Epidemiology Data

Mahmoud Ragab [1,2,3,*], Hani Choudhry [2,4], Amer H. Asseri [2,4], Sami Saeed Binyamin [5] and Mohammed W. Al-Rabia [6,7]

[1] Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2] Centre for Artificial Intelligence in Precision Medicines, King Abdulaziz University, Jeddah 21589, Saudi Arabia; hchoudhry@kau.edu.sa (H.C.); ahasseri@kau.edu.sa (A.H.A.)
[3] Mathematics Department, Faculty of Science, Al-Azhar University, Nasr City, Cairo 11884, Egypt
[4] Biochemistry Department, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[5] Computer and Information Technology Department, The Applied College, King Abdulaziz University, Jeddah 21589, Saudi Arabia; ssbinyamin@kau.edu.sa
[6] Department of Medical Microbiology and Parasitolog, Faculty of Medicine, King Abdulaziz University, Jeddah 21589, Saudi Arabia; mwalrabia@kau.edu.sa
[7] Health Promotion Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia
* Correspondence: mragab@kau.edu.sa

check for updates

Citation: Ragab, M.; Choudhry, H.; H. Asseri, A.; Binyamin, S.S.; Al-Rabia, M.W. Enhanced Gravitational Search Optimization with Hybrid Deep Learning Model for COVID-19 Diagnosis on Epidemiology Data. *Healthcare* **2022**, *10*, 1339. https://doi.org/10.3390/healthcare10071339

Academic Editor: Rocío De Andrés Calle

Received: 25 May 2022
Accepted: 14 July 2022
Published: 19 July 2022

Abstract: Effective screening provides efficient and quick diagnoses of COVID-19 and could alleviate related problems in the health care system. A prediction model that combines multiple features to assess contamination risks was established in the hope of supporting healthcare workers worldwide in triaging patients, particularly in situations with limited health care resources. Furthermore, a lack of diagnosis kits and asymptomatic cases can lead to missed or delayed diagnoses, exposing visitors, medical staff, and patients to 2019-nCoV contamination. Non-clinical techniques including data mining, expert systems, machine learning, and other artificial intelligence technologies have a crucial role to play in containment and diagnosis in the COVID-19 outbreak. This study developed Enhanced Gravitational Search Optimization with a Hybrid Deep Learning Model (EGSO-HDLM) for COVID-19 diagnoses using epidemiology data. The major aim of designing the EGSO-HDLM model was the identification and classification of COVID-19 using epidemiology data. In order to examine the epidemiology data, the EGSO-HDLM model employed a hybrid convolutional neural network with a gated recurrent unit based fusion (HCNN-GRUF) model. In addition, the hyperparameter optimization of the HCNN-GRUF model was improved by the use of the EGSO algorithm, which was derived by including the concepts of cat map and the traditional GSO algorithm. The design of the EGSO algorithm helps in reducing the ergodic problem, avoiding premature convergence, and enhancing algorithm efficiency. To demonstrate the better performance of the EGSO-HDLM model, experimental validation on a benchmark dataset was performed. The simulation results ensured the enhanced performance of the EGSO-HDLM model over recent approaches.

Keywords: COVID-19; disease detection; health promotion; fusion model; hybrid deep learning; epidemiology data; parameter optimization

1. Introduction

COVID-19 or Coronavirus is a recent disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. In February 2021, over 106 billion SARS-CoV-2 infections and a mortality rate of above 2.3 billion were recorded globally, in the worst epidemic to have troubled humans since the Spanish flu occurred in 1918, and has consequently overwhelmed global healthcare systems resulting in serious economic disturbance [2]. In order to combat this infectious disease, doctors and researchers have tirelessly attempted to gain further knowledge and develop technical devices for the diagnosis of this disease. Attempts include the advancement of drugs and vaccines [3], the formation of epidemiologic methods for forecasting the spread of the disease among people [4], the advancement of mobile device apps to track patients who are infected, the enhancement of policies, and the implementation of novel technologies for managing the capacities and resources in clinics [5]. Owing to the increasing prevalence of coronavirus infection worldwide, many studies analyzed various prospects of this infection. Research includes recognizing the methods of virus detection, the origin of coronavirus along with examining its gene series, early cases in the countries infected, forecasting COVID-19 cases in the epidemiologic pandemic, and patient information [6].

Machine learning (ML) procedures first accumulate data individually, i.e., from various sources [7], followed by fixing the preprocessed data to solve data-oriented problems and minimize space size by erasing incorrect data and selecting relevant data. However, in some cases, the dataset value may differ for the system in making decisions, thus, ML methods were devised with the help of other ideas such as probability statistics and theory control for examining data and deriving useful and original information, concealed information, or patterns from past knowledge [8]. Lastly, the performance assessments of the methods is conducted and model optimization concludes the process, which enhances the method with the help of novel rules and a dataset. ML approaches are utilized in various sectors such as traffic management, medicine, education, manufacturing and production, robotics, engineering, and so on [9–11]. Currently, ML methods are utilized in the examination of high dimensional bio-medical structured and unstructured datasets. Malaria Diagnosis, diabetes risk assessment, typhoid, and vascular diseases classification, and genomic and genetic data studies are a few instances of the bio-medical use of ML approaches.

The authors in [12] developed a novel image processing-based approach for the health care system called "C19D-Net", for the detection of "COVID-19" in CXR images to assist radiologist in enhancing COVID-19 classification performance. It uses the InceptionV4 model with a multi-SVM classifier for COVID-19 diagnosis in distinct class labels. In [13], the researchers present a solution to analyze the COVID-19 pandemic dataset. Specifically, the solution focuses on analyzing useful knowledge and valuable information (for example, frequency, distribution, and pattern) of healthcare statuses and characteristics in populations. The availability of knowledge and information assists users (for example, civilians, and researchers) in better understanding the disease, and playing an active role in combating, fighting, or controlling the disease. Li et al. [14] performed a qualitative and quantitative assessment of Chinese social networking media posts originating in Wuhan City on the Chinese microblogging platform Weibo during the earlier stage of the COVID-19 epidemic.

The researchers in [15] reviewed the available approaches while anticipating the challenges and difficulties in the improvement of a data-driven strategy to combat the COVID-19 outbreak. A 3M-analysis was presented, which included making, monitoring, and exhibiting decisions. The emphasis was placed on the potential of a familiar data-driven scheme to tackle various problems which increased following the outbreak; firstly, by forecasting and monitoring the spread of the epidemic; secondly, by assessing the efficiency of government decisions; and lastly, by making appropriate decisions. Every step of the roadmap may be exhaustive for an analysis of consolidated theoretic outcomes and the potential applications in the COVID-19 context. Supervised ML methods for COVID-19 infection were designed in [16] with learning algorithms including DT, LR, NB, ANN, and SVM using epidemiology labeled data for negative and positive COVID-19 cases. A correlation coefficient (CC) analysis between independent features and dependent features was performed to determine stronger relationships among every dependent and independent feature of the prior dataset to develop the model.

The researchers in [17] processed available information on US states to create an incorporated dataset on potential factors which cause the epidemic to spread. Then, the researchers made use of the supervised ML approach to reach a consensus and rank the crucial factor. The author performed a regression analysis to identify the crucial pre-lockdown factors that affected post-lockdown contamination and death to inform lockdown-related policy making. Kim et al. [18] aimed to examine the relationship between obesity, as determined by the body mass index (BMI), with mortality and morbidity owing to COVID-19. Information from 5628 approved COVID-19 patients were gathered by the Center for Disease Control and Prevention of Korea. The odds ratio (OR) of diabetes and morbidity in the BMI groups was examined using LR models attuned for similar covariates. Hawkins et al. [19] designed a nationwide COVID-19 dataset at the county level that was coupled with the Distressed Communities Index (DCI) and their constituent metrics of socio-economic status.

This study developed an Enhanced Gravitational Search Optimization with Hybrid Deep Learning Model (EGSO-HDLM) for COVID-19 diagnosis using epidemiology data. The proposed EGSO-HDLM model employs a hybrid convolutional neural network with gated recurrent unit based fusion (HCNN-GRUF) model. In addition, the hyperparameter optimization of the HCNN-GRUF model was improved by the use of the EGSO algorithm, which was derived by including the concepts of cat map and the traditional GSO algorithm. To assure the enhanced performance of the EGSO-HDLM model, an extensive set of experimental validation processes on a benchmark dataset was performed. In short, the paper's contribution is summarized as follows.

- An intelligent EGSO-HDLM model which includes pre-processing, HCNN-GRUF classification, and EGSO-based parameter optimization is presented. To the best of our knowledge, the EGSO-HDLM model has not previously been presented in the literature.
- Hyperparameter optimization of the HCNN-GRU model with the EGSO algorithm using cross-validation helps to boost the predictive outcome of the EGSO-HDLM model for unseen data.
- Validating the performance of the proposed EGSO-HDLM model on a benchmark dataset from the Kaggle repository.

2. The Proposed Model

In this study, a new EGSO-HDLM model was designed for the identification and classification of COVID-19 using epidemiology data. Primarily, the presented EGSO-HDLM model applied data normalization to pre-process the actual data. Then, the HCNN-GRUF model was applied for the classification process. Moreover, the hyperparameter optimization of the HCNN-GRUF model was improved by the use of the EGSO algorithm. Figure 1 depicts the overall process of the EGSO-HDLM algorithm.

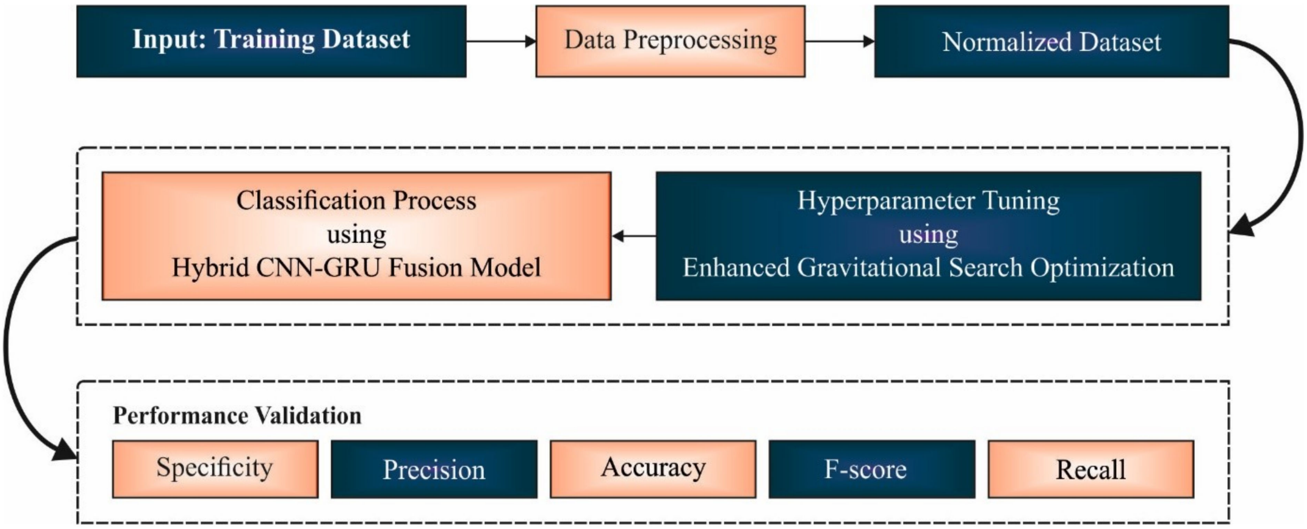
Input: Training Dataset
Data Preprocessing
Normalized Dataset
Classification Process using Hybrid CNN-GRU Fusion Model
Hyperparameter Tuning using Enhanced Gravitational Search Optimization
Performance Validation
Specificity
Precision
Accuracy
F-score
Recall

Figure 1. Overall process of EGSO-HDLM technique.

2.1. Data Pre-processing

When the numerical processing was executed, a huge variance in magnitude amongst the values became apparent. This condition can cause problems such as minimal convergence of networks and saturation of neuron outcomes [20]; therefore, normalized original data were required. The following Min-Max normalized approach (Equation (1)) can be used for normalizing the data to intervals of zero and one:

$$x^* = \frac{x - x_{\text{ min}}}{x_{\text{ max}} \ - x_{\text{ min}}}. \tag{1}$$

where x^* signifies the normalized data, x represents the original data, x_{min} implies the minimal data values from the present element, and x_{max} denotes the maximal data value from the present elements.

2.2. Design of HCNN-GRUF Based Classification

CNN [21] is a hierarchical FFNN approach that is dependent upon convolutional as well as pooling functions. Convolution was utilized to extract local features for further processing with subsequent layers. Primarily, the resultant input layers are convoluted with several $n \times h$ filters (convolutional kernel). Next, the resultant convolutional layer was created by linking the outcomes of the convolutional function that differs with filter size h.

$$m_i = f(w \cdot x_{i+h-1} + b) \tag{2}$$

$$M = [m_1,\ m_2,\ m_3,\ \ldots,\ m_{l-h+1}] \tag{3}$$

In Equation (2), m_i refers the ith feature extracted with convolutional function, f denotes the non-linear function, w signifies the weighting of filters, h denotes the filtering window size, and b implies the bias. A combination of GRU and LSTM was applied for a gated recurrent neural network (RNN) which remembers a long order of data and thus decreases data loss. Related to LSTM, GRU decreases the amount of gated units that reduces the processing time but preserves the accuracy. GRU has two gated sub-units, namely reset as well as update gates. In every moment, the GRU receives the existing as well as hidden states in the preceding moment with their update gate, and defines the activation state of their individual neurons. Additionally, the reset gate receives both states and defines if any input data are forgotten. Figure 2 depicts the framework of the GRU technique. Input at the present moment is then integrated with the weighted and resultant reset gate to obtain the memory contents at the present moment with activation functions. Next, the upgrade gate obtains the memory content at the present moment and hidden state in the preceding moment for determining the output and hidden state at the present moment. The function of GRU is summarized in the following formulas.

$$z_t = \sigma\left(W^{(z)} \cdot x_t + U^{(z)} \cdot h_{t-1}\right) \tag{4}$$

$$r_t = \sigma\left(W^{(z)} \cdot x_t + U^{(r)} \cdot h_{t-1}\right) \tag{5}$$

$$h'_t = tanh(W \cdot x_t + r_t \cdot U \cdot h_{t-1}) \tag{6}$$

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot h'_t \tag{7}$$

In Equation (4), z_t signifies the upgrade gate, $W^{(z)}$ and $U^{(z)}$ imply the weighting of z_t, σ stands for the activation function, x_t indicates the input at present moment, and h_{t-1} represents the hidden output in the preceding moment. r_t stands for the reset gate, and $W^{(r)}$ and $U^{(r)}$ refer the weighting of r_t. h'_t defines the memory satisfied at the present moment, tanh implies the activation functions, and W and U represent the weighting at the present moment. h_t defines the outcome at the present moment. Pooling is the procedure of extracting data for the purpose of reducing the size of the resultant GRU layer. The maximal pooling approach was utilized for extracting the maximal vector value of all the inputs as the result of pooling layers.

$$\hat{m} = \ \max\{M\} \tag{8}$$

$$z = [\hat{m}_1,\ \hat{m}_2,\ \hat{m}_3,\ \ldots\ldots,\ \hat{m}_k] \tag{9}$$

where M denotes the typical vector in the GRU to pooling layers, $\hat{m}$. implies the maximal of M, z signifies the output resultant in the pooling layer, and k denotes the amount of features input to pooling layers. The fusion layer was planned to merge more than 2 layers of tensors; during this case, it integrates several tensors in the polling layer as opposed to one tensor. It can be executed by the "Concatenate" approach that receives the final bit as axis, and splices all the resultant polling layers for creating the result of this layer.

$$b = [a_1,\ a_2] \tag{10}$$

In Equation (10), b refers the output of this layer, aa denotes the resultant in the polling layers. The softmax classification is usually employed for several classifier problems. The fully connected (FC) layer was utilized to connect an input in the fusion layer and the value of all the neurons was computed utilizing the softmax function. Additionally, the maximal value of every neuron output was obtained as the classifier outcome.

$$\hat{p} = soft\max\,(W \cdot x + b) \tag{11}$$

$$y = argmax(\hat{p}_y) \tag{12}$$

In Equation (11), $\hat{p}_y$ signifies the probability of y, forecast labels. W and b denote the weights as well as biases. In Equation (12), y denotes the last forecast label.

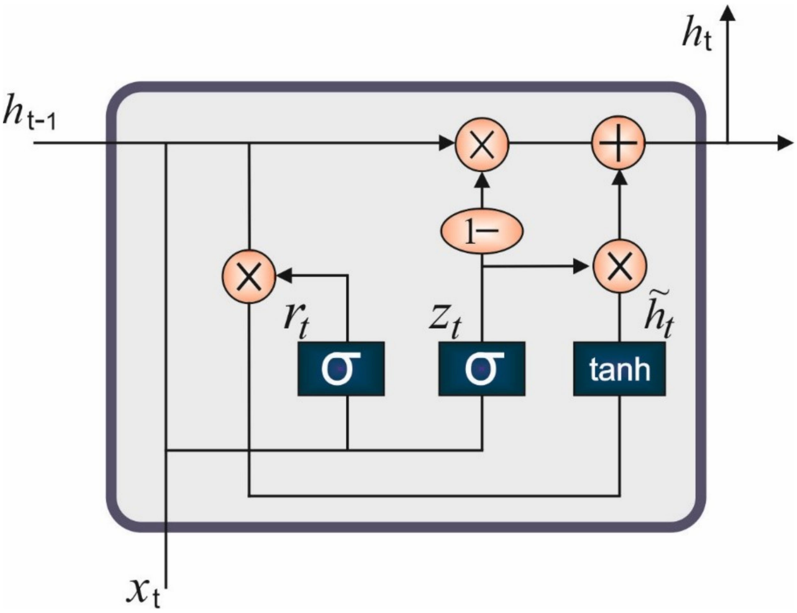

Figure 2. Structure of GRU Model.

2.3. Hyperparameter Tuning

At this stage, the hyperparameter tuning of the HCNN-GRUF model can be tuned by the design of the EGSO algorithm. Rashedi [22] developed a GSO algorithm, which is an optimization algorithm that gained popularity in the past few decades. It depends on the second law of motion, as illustrated in Equation (15), and the law of gravity, as demonstrated in Equation (13). Additionally, it depends on the physical concepts of passive gravitational mass, inertial mass, and active gravitational mass. The second law of motion states that once a force (F) is applied to all the particles, the (a) acceleration is defined by the mass (M) and force. The law of gravity says that all the particles attract all the other particles with a gravitational force (F). The gravitational force (F) between these particles is directly proportionate to the product of (M_1 and M_2) masses and inversely proportionate to the square of $\left(R^2\right)$ distance.

$$F = G\frac{M_1 M_2}{R^2} \tag{13}$$

G indicates the gravitational constant that reduces with increasing time, and it is evaluated by the following expression:

$$G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta} \text{ , } \beta < 1 \tag{14}$$

$$a = \frac{F}{M} \tag{15}$$

The GSA algorithm is briefly described in the following steps:

- Step one: Initialization

An isolated scheme with N number of particles is assumed, where the location of *i-th* particles is represented as follows:

$$P_i = \left(p_i^1,\ p_i^2,\ \ldots,\ p_i^d,\ \ldots,\ p_i^N\right) \ for\ i = 1, 2, 3, \ldots, N \tag{16}$$

In Equation (16), p_i^d indicates the location of *i-th* particles in dimension d.

- Step two: Fitness assessment of particle

Here, the best and worst fitness values are evaluated using Equations (17) and (18) correspondingly. Minimization and maximization problems are evaluated by Equations (19) and (20).

$$worst(t) = \max_{j,\{1,\ldots,N\}} fitness_j(t) \tag{17}$$

$$best\ (t) = \min_{j,\{1,\ldots,N\}} fitness_j(t) \tag{18}$$

$$worst(t) = \min_{j,\{1,\ldots,N\}} fitness_j(t) \tag{19}$$

$$best\ (t) = \max_{j,\{1,\ldots,N\}} fitness_j(t) \tag{20}$$

In the equation, $fitness_j(t)$ refers to the fitness of *j-th* particles at t time.

- Step three: Evaluate the $G(t)$ gravitational constant

Here, the $G(t)$ gravitational constant at t time can be evaluated using Equation (21):

$$G(t) = G_0 \times \left(1 - \frac{t}{t_{\max}}\right) \tag{21}$$

where G_0 is the primary value of the gravitational constant initialized at a random fashion, t indicates the present time, and $t_{\max}$ denotes the overall time.

- Step four: Upgrade the inertial and gravitational mass

Here, the gravitational and inertia mass are upgraded with the help of the fitness function. Equality of gravitational and inertia masses is assumed, and the masses value is evaluated using the following expression:

$$M_{ii} = M_{pi} = M_{ai} = M_i \ for\ i = 1, 2, 3, \ldots N \tag{22}$$

$$m_i(t) = \frac{fitness_i(t) - worst(t)}{best(t) - worst(t)} \tag{23}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \tag{24}$$

In the expression, $fitness_i(t)$ represent the fitness of *i-th* particles at time and t, $M_i(t)$ denotes the mass of *i-th* particles at t time.

- Step five: Calculate the overall force

Here, the complete force $F_i^d(t)$ that is exerted on the *i-th* particles in a *d-th* dimension at t time is evaluated by the following equation:

$$F_i^d(t) = \sum_{j=1, j \neq i}^{\mathrm{k}best} rand_j F_{ij}^d(t) \tag{25}$$

In Equation (25), $rand_j$ represent a random integer $\in [0,\ 1]$, and k*best* is the set of initial particles with the biggest masses and the best fitness value. The force applied from mass '*j*' on mass '*i*' at '*t*' time can be expressed as $F_{ij}^d(t)$ and is evaluated using Equation (26):

$$F_{ij}^d(t) = G(t)\frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \tau}\left(p_j^d(t) - p_i^d(t)\right) \tag{26}$$

Consider that M_{pi} refers to the passive gravitational mass correlated to *i-th* particles, Ma_j denotes the active gravitational mass interrelated to *j-th* particles. τ indicates a smaller positive constant to avoid division by 0, and $R(t)$ indicates the Euclidian distance between *i-th* and *j-th* particles:

$$R_{ij}(t) = \|p_i(t), p_j(t)\|_2 \tag{27}$$

- Step six: Calculate the acceleration and velocity

Here, $F_i^d(t)$, the acceleration of *i-th* particles; $a_i^d(t)$ at t time in the d direction; and the velocity of *i-th* particles in the *d-th* direction, $u_i^d(t+1)$, are evaluated by:

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \tag{28}$$

$$u_i^d(t+1) = rand_i \times u_i^d(t) + a_i^d(t) \tag{29}$$

From the expression, $M_{ii}(t)$ refers to the inertial mass of *i-th* particles, and $rand_i$ indicates an arbitrary integer $\in [0,\ 1]$.

- Step seven: Upgrade particle location

Here, the location of *i-th* particles in the *d-th* direction, $p_i^d(t+1)$, is evaluated by the following expression:

$$p_i^d(t+1) = p_i^d(t) + u_i^d(t+1) \tag{30}$$

- Step eight: Repeat steps two to seven until the end condition is met

In the EGSO algorithm, the GSO algorithm is integrated into the cat map concept to reduce the ergodic problem, avoid premature convergence and enhance the algorithm's efficiency. In optimization algorithms, the grouping of SI optimization algorithms and chaotic mapping improves the outcomes of SI optimization. This concept was confirmed by several instances and depends on Chaos features including sensitivity, randomness, and ergodicity to the initial condition. Hence, an attempt was made to integrate the Chaos concept with GSO to attain good accuracy and efficiency [23]. Until now, the SI optimization approach was generally integrated into tent or logistic mapping models to accomplish a certain amount of progress. However, the chaotic sequence produced from the logistic mapping is of a non-uniform distribution. This follows Chebyshev's Distribution, where density is lower at the center and is higher at both ends. These features have some level of impact on the ergodicity of optimization solution space. The chaotic sequence made from tent mapping models follows uniform distribution; however, the values rapidly fall into the smaller cycle due to the impact of word length. It therefore lacks better randomness. This method could generate a uniformly distributed chaotic sequence, to enhance GSO single distribution and might improve the model searching efficacy. A chaotic algorithm, Cat Map, is generally utilized in image encryption.

$$\begin{aligned} x_{n+1} &= (x_n + y_n) \\ y_{n+1} &= (x_n + 2y_n)^{mod1} \end{aligned} \tag{31}$$

Here, the chaotic sequence produced by the higher dimensional Cat Map at the range of [0, 1] is fixed as an arbitrary value rather than the rand function to determine individual distribution. The mapping value lies within the range of [0, 1] when building the multi-dimension chaotic sequence Cat(i,j). The sequential dataset in the dimension is independent of others and has stronger ergodicity. These features decrease the ergodic problem carried by random distribution and assures the ergodicity of individuals flying in the solution space. This technique might improve the algorithm performance and prevent premature convergence.

3. Results and Discussion

This section describes the experimental validation of the EGSO-HDLM technique, which was carried out with the help of the benchmark epidemiology dataset from Kaggle repository [24]. In this article, a set of 5000 samples in positive class and 5000 samples in negative classes were considered. The proposed model was simulated using the Python 3.6.5 tool on PC i5-8600k, GeForce 1050Ti 4 GB, 16 GB RAM, 250 GB SSD, and 1 TB HDD. The parameter settings were as follows: dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU.

The confusion matrices generated by the EGSO-HDLM model on the test data are demonstrated in Figure 3. For 70% of the training (TR) data, the EGSO-HDLM model identified 3330 samples in the positive class and 3465 samples in the negative class. Additionally, for 30% of the testing (TS) data, the EGSO-HDLM system identified 1425 samples in the positive class and 1496 samples in the negative class. In addition, for 80% of the TR data, the EGSO-HDLM method identified 3885 samples as positive class and 3896 samples as negative class. Finally, on 20% of TS data, the EGSO-HDLM system identified 977 samples in the positive class and 974 samples in the negative class.

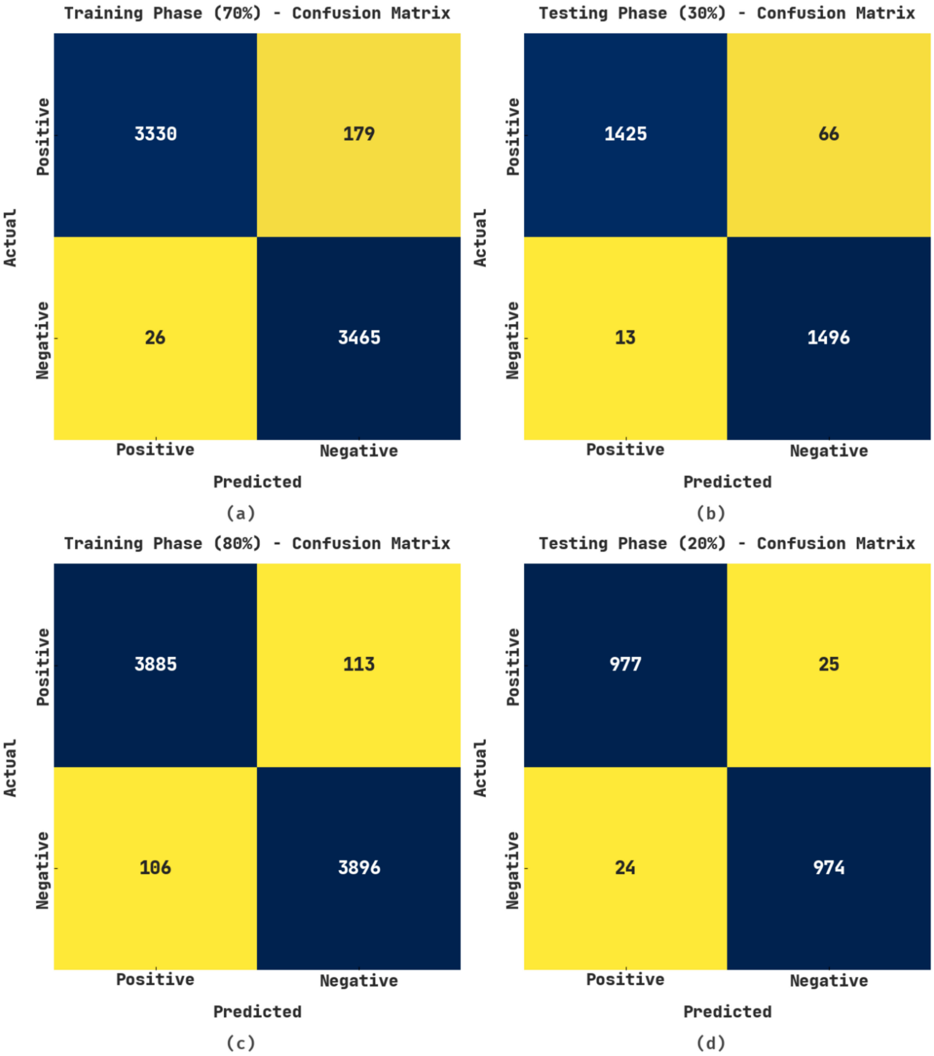

Figure 3. Confusion matrices of EGSO-HDLM algorithm (**a**) 70% of TR; (**b**) 30% of TS; (**c**) 80% of TR; (**d**) 20% of TS data.

Table 1 offers detailed COVID-19 recognition outcomes of the EGSO-HDLM model for 70% of the TR and 30% of the TS data. A brief result analysis of the EGSO-HDLM model in the identification of COVID-19 for 70% of the TR data is depicted in Figure 4. The results implied that the EGSO-HDLM model classified all the samples into corresponding positive and negative classes. For instance, the EGSO-HDLM model identified positive samples with $accu_y$ of 97.07%, $prec_n$ of 99.23%, $reca_l$ of 94.90%, $spec_y$ of 99.26%, and F_{score} of 97.01%. Additionally, the EGSO-HDLM technique identified negative samples with $accu_y$ of 97.07%, $prec_n$ of 95.09%, $reca_l$ of 99.26%, $spec_y$ of 94.90%, and F_{score} of 97.13%.

Table 1. Result analysis of EGSO-HDLM technique with various measures for 70% of TR and 30% of TS data.

Class Labels	Accuracy	Precision	Recall	Specificity	F-Score
		Training Phase (70%)			
Positive	97.07	99.23	94.90	99.26	97.01
Negative	97.07	95.09	99.26	94.90	97.13
Average	97.07	97.16	97.08	97.08	97.07
		Testing Phase (30%)			
Positive	97.37	99.10	95.57	99.14	97.30
Negative	97.37	95.77	99.14	95.57	97.43
Average	97.37	97.44	97.36	97.36	97.37

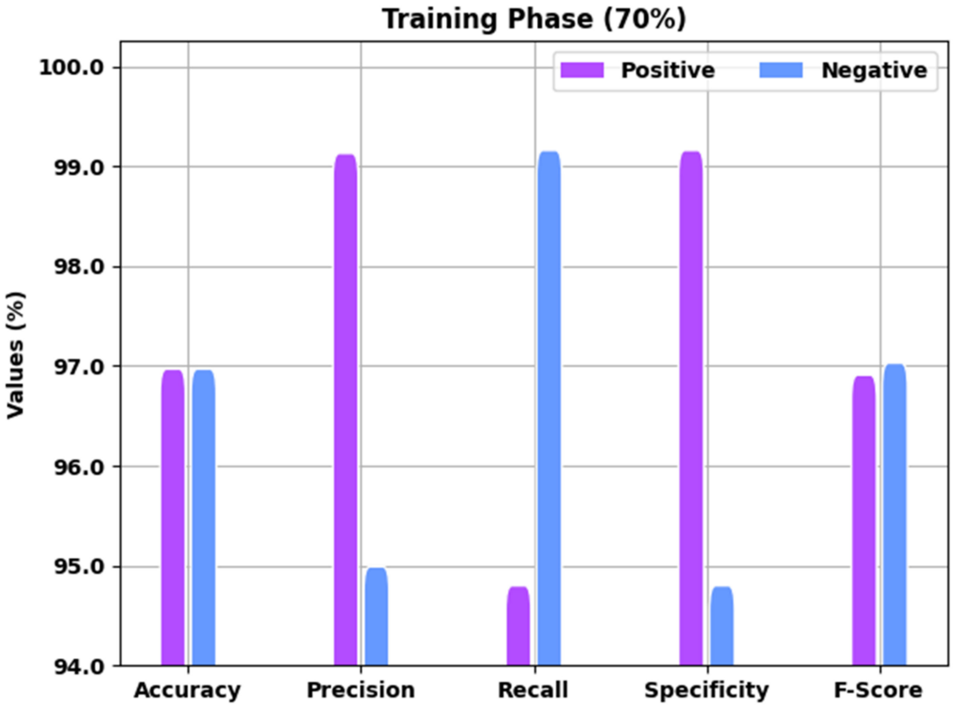

Figure 4. Result analysis of EGSO-HDLM approach under 70% of TR data.

A brief overview of the EGSO-HDLM technique in the identification of COVID-19 under 30% the TS data is shown in Figure 5. The results implied that the EGSO-HDLM methodology recognized all the samples as corresponding positive and negative classes. For example, the EGSO-HDLM system identified positive samples with $accu_y$ of 97.37%, $prec_n$ of 99.10%, $reca_l$ of 95.57%, $spec_y$ of 99.14%, and F_{score} of 97.30%. Moreover, the EGSO-HDLM approach identified negative samples with $accu_y$ of 97.37%, $prec_n$ of 95.77%, $reca_l$ of 99.14%, $spec_y$ of 95.57%, and F_{score} of 97.43%.

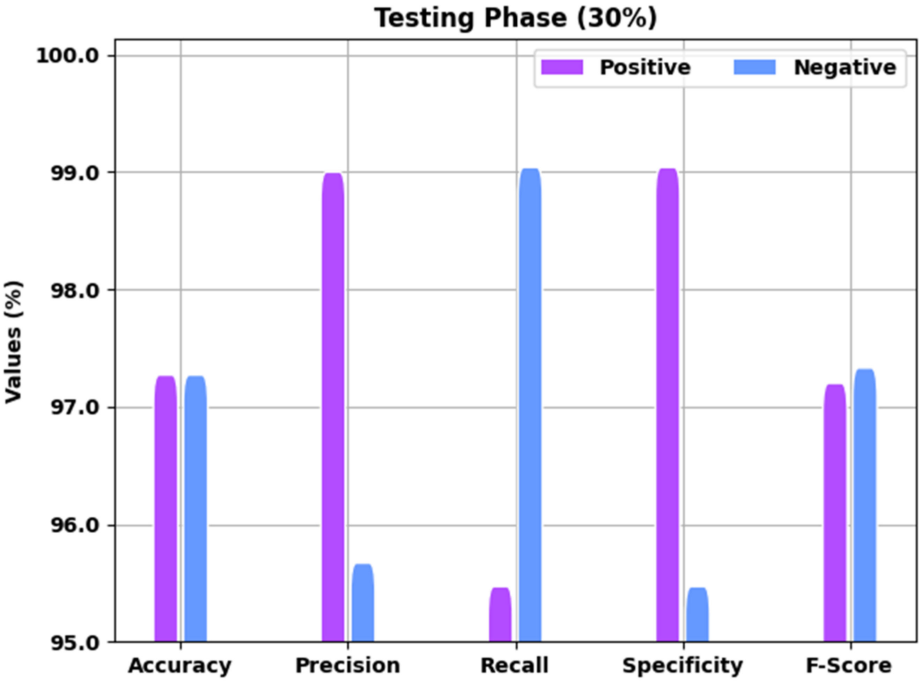

Figure 5. Result analysis of EGSO-HDLM approach under 30% of TS data.

Table 2 presents detailed COVID-19 identification outcomes of the EGSO-HDLM system for 80% of the TR and 20% of the TS data. A brief analysis of the EGSO-HDLM approach in the identification of COVID-19 for 80% of the TR data is portrayed in Figure 6. The results implied that the EGSO-HDLM model identified all the samples into corresponding positive and negative classes. For example, the EGSO-HDLM methodology identified positive samples with $accu_y$ of 97.26%, $prec_n$ of 97.34%, $reca_l$ of 97.17%, $spec_y$ of 97.35%, and F_{score} of 97.26%. Additionally, the EGSO-HDLM model identified negative samples with $accu_y$ of 97.26%, $prec_n$ of 97.18%, $reca_l$ of 97.35%, $spec_y$ of 97.17%, and F_{score} of 97.27%.

A brief overview of the EGSO-HDLM approach in the identification of COVID-19 for 20% of the TS data is depicted in Figure 7. The results implied that the EGSO-HDLM technique identified all the samples as corresponding positive and negative classes. For example, the EGSO-HDLM method identified positive samples with $accu_y$ of 97.55%, $prec_n$ of 97.60%, $reca_l$ of 97.50%, $spec_y$ of 97.60%, and F_{score} of 97.55%. Additionally, the EGSO-HDLM system identified negative samples with $accu_y$ of 97.55%, $prec_n$ of 97.50%, $reca_l$ of 97.60%, $spec_y$ of 97.50%, and F_{score} of 97.55%.

Table 2. Result analysis of EGSO-HDLM technique with various measures for 80% of TR and 20% of TS data.

Class Labels	Accuracy	Precision	Recall	Specificity	F-Score
Training Phase (80%)					
Positive	97.26	97.34	97.17	97.35	97.26
Negative	97.26	97.18	97.35	97.17	97.27
Average	97.26	97.26	97.26	97.26	97.26
Testing Phase (20%)					
Positive	97.55	97.60	97.50	97.60	97.55
Negative	97.55	97.50	97.60	97.50	97.55
Average	97.55	97.55	97.55	97.55	97.55

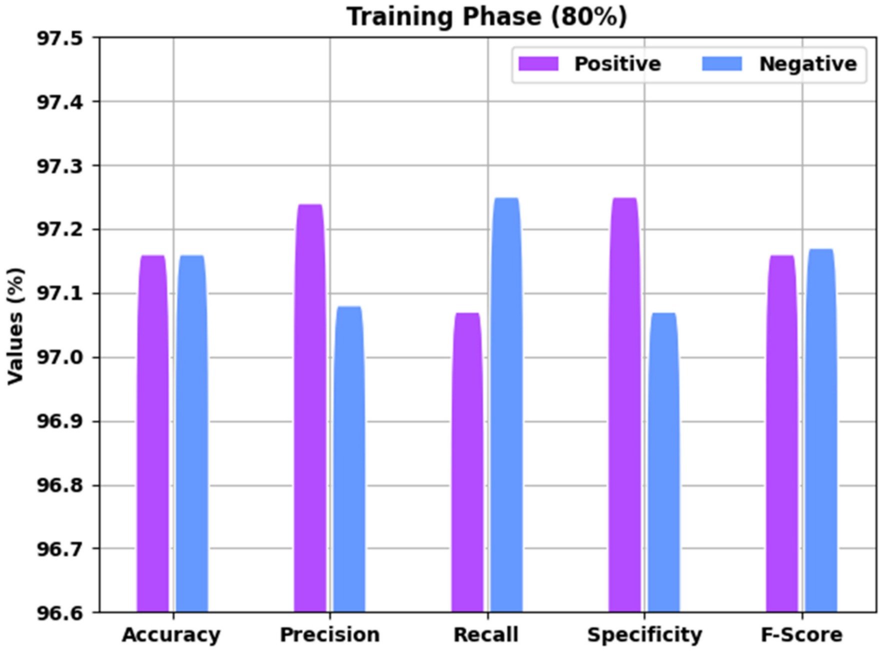

Figure 6. Result analysis of EGSO-HDLM approach under 80% of TR data.

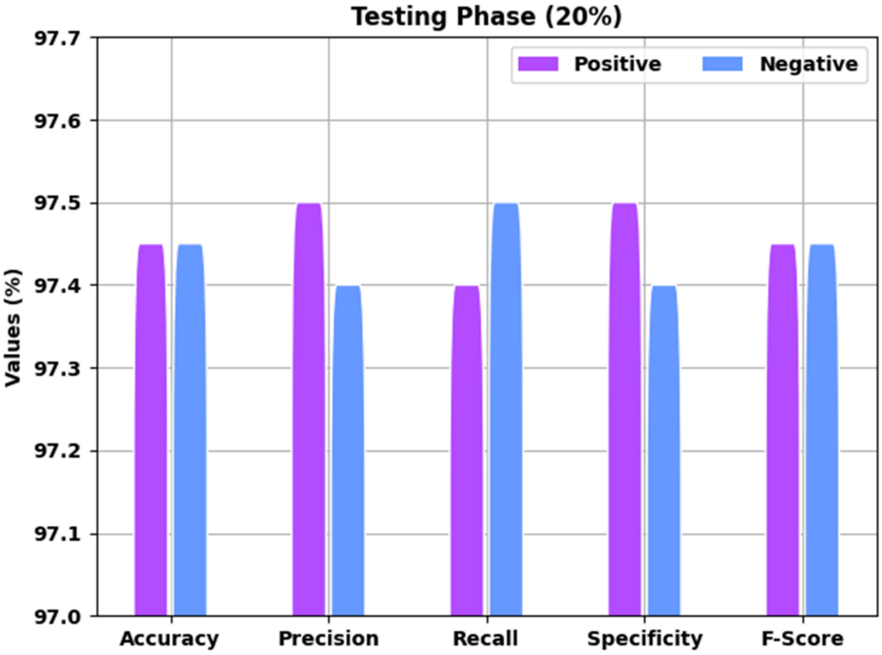

Figure 7. Result analysis of EGSO-HDLM approach under 20% of TS data.

The training accuracy (TA) and validation accuracy (VA) attained by the EGSO-HDLM method on test dataset are illustrated in Figure 8. The experimental outcome implied that the EGSO-HDLM system gained maximum values of TA and VA. Particularly, the VA seemed to be higher than TA.

The training loss (TL) and validation loss (VL) achieved by the EGSO-HDLM system on test dataset are established in Figure 9. The experimental outcome implied that the EGSO-HDLM methodology had the lowest values of TL and VL. In particular, the VL was lower than TL.

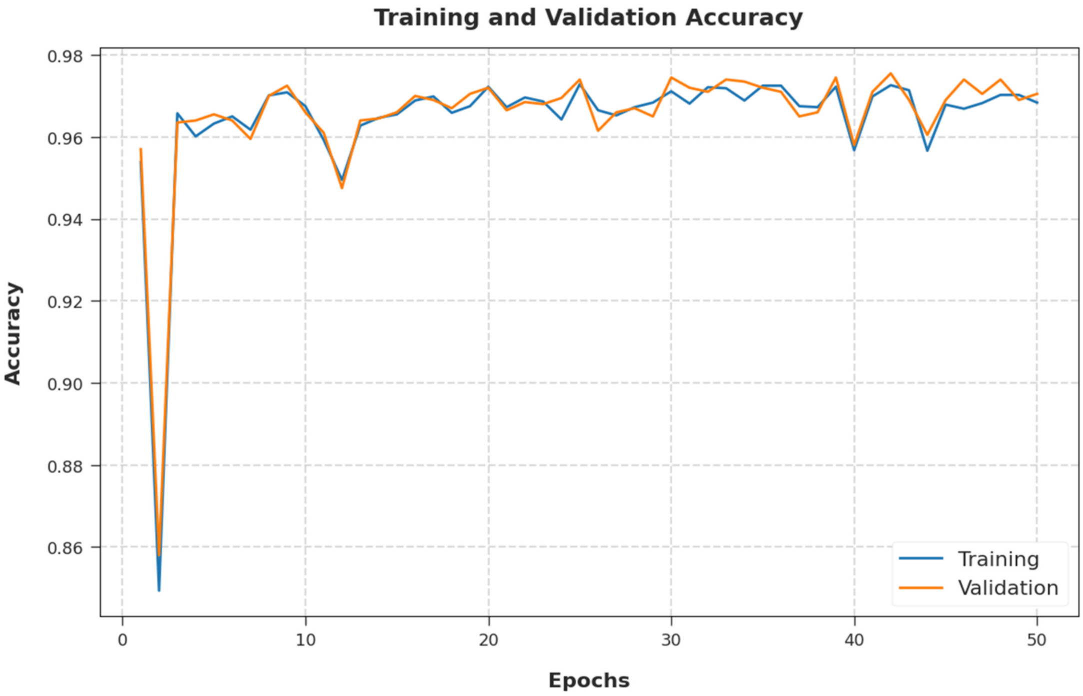

Figure 8. TA and VA analysis of EGSO-HDLM methodology.

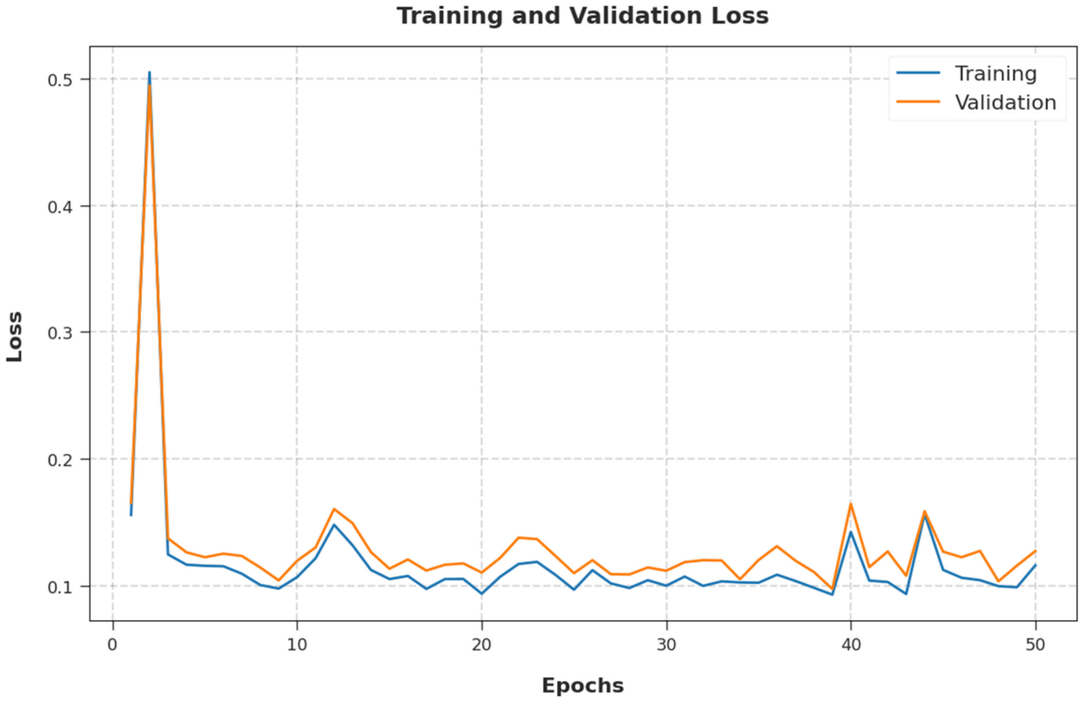

Figure 9. TL and VL analysis of EGSO-HDLM methodology.

A brief precision-recall analysis of the EGSO-HDLM methodology on test dataset is portrayed in Figure 10. By observing the figure, it can be seen that the EGSO-HDLM technique accomplished maximum precision-recall performance under all classes.

A detailed ROC inquiry of the EGSO-HDLM algorithm on the test dataset is portrayed in Figure 11. The results implied that the EGSO-HDLM technique exhibited its ability in categorizing two distinct classes on the test dataset.

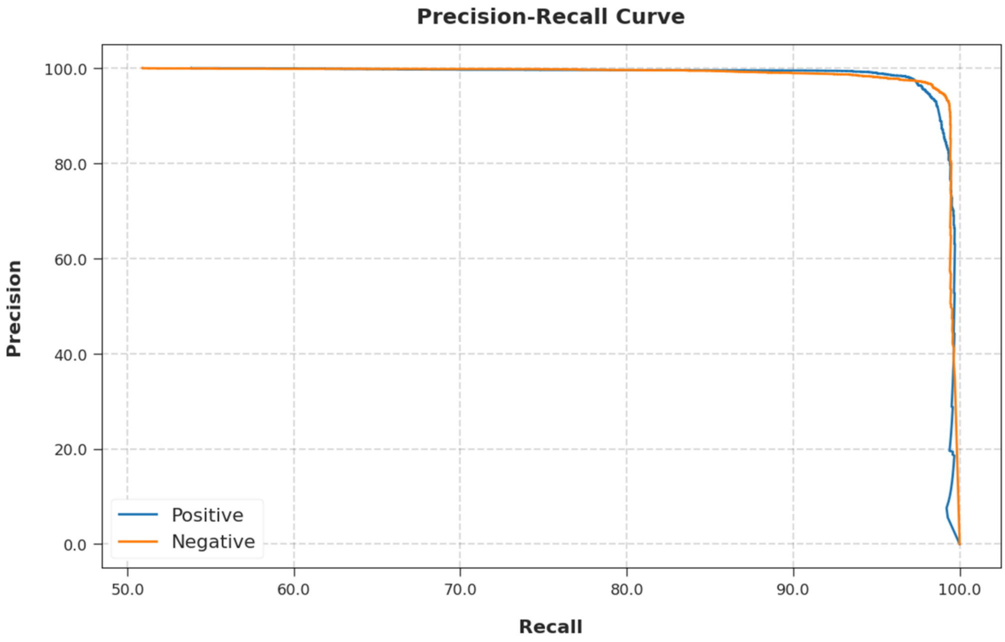

Figure 10. Precision-recall curve analysis of EGSO-HDLM methodology.

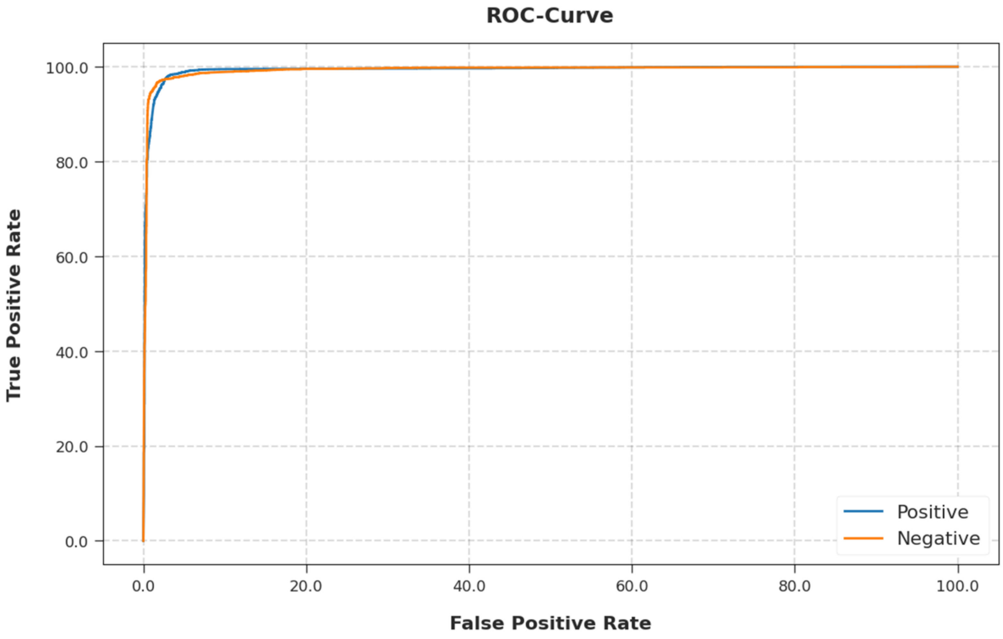

Figure 11. ROC curve analysis of EGSO-HDLM methodology.

Table 3 portrays a detailed comparison study of the EGSO-HDLM model with recent models under several measures [16]. Figure 12 illustrates an extensive $accu_y$ in the comparison of the EGSO-HDLM model with existing models. The figure indicates that the SGD and ACO models displayed poor performance with minimal $accu_y$ values of 90.01% and 90.67%, respectively. Additionally, the ELM model obtained a slightly enhanced $accu_y$ value of 91.46%. Next, the MLP and LSTM models demonstrated a reasonably closer $accu_y$ of 94.27% and 94.64%, respectively. However, the EGSO-HDLM model offered a maximum $accu_y$ of 97.55%.

Figure 13 demonstrates an extensive $prec_n$ analysis of the EGSO-HDLM technique with existing models. The figure indicates that the SGD and ACO algorithms demonstrated poor performance with minimal $prec_n$ values of 90.08% and 94.40%, respectively. The ELM approach obtained a slightly enhanced $accu_y$ value of 91.18%. Next, the MLP and LSTM models demonstrated a reasonably closer $prec_n$ of 94.58% and 93.18% correspondingly. However, the EGSO-HDLM method provided the maximum $prec_n$ of 97.55%.

Table 3. Comparative analysis of EGSO-HDLM technique with existing approaches [14].

Methods	Accuracy	Precision	Recall
EGSO-HDLM	97.55	97.55	97.55
Extreme Learning Machine	91.46	91.18	92.59
Multilayer Perceptron	94.27	94.58	93.66
SGD Model	90.01	90.08	92.37
LSTM Model	94.64	93.18	94.96
ACO Model	90.67	94.40	93.36

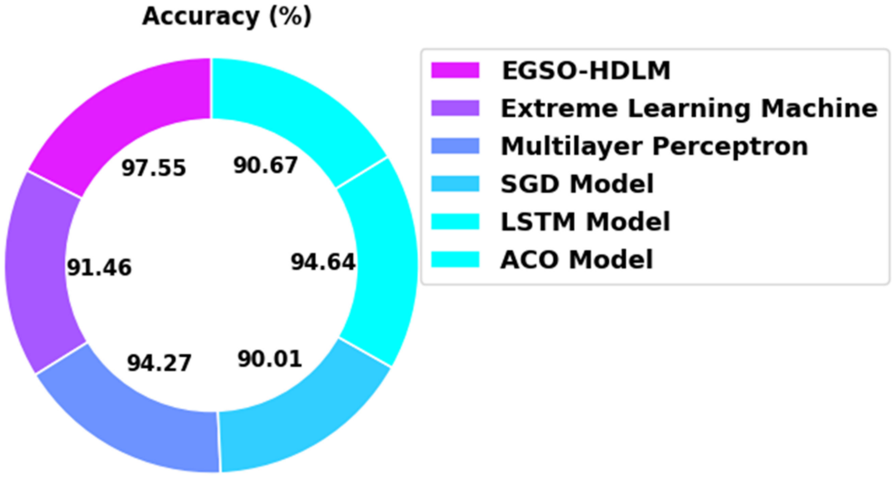

Figure 12. $Accu_y$ analysis of EGSO-HDLM algorithm with existing methodologies.

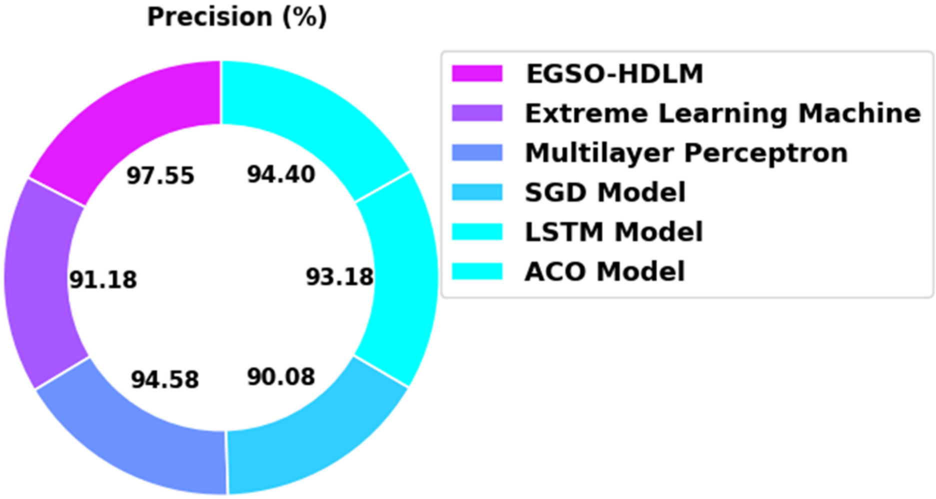

Figure 13. $Prec_n$ analysis of EGSO-HDLM algorithm with existing methodologies.

Figure 14 illustrates an extensive $reca_l$ inspection of the EGSO-HDLM algorithm in comparison with existing models. The figure indicates that the SGD and ACO systems demonstrated poor performance with minimal $reca_l$ values of 92.37% and 93.36%, respectively. Additionally, the ELM approach gained a slightly enhanced $accu_y$ value of 92.59%. Next, the MLP and LSTM models displayed a reasonably closer $reca_l$ of 93.66% and 94.96%, respectively. However, the EGSO-HDLM system offered the maximum $reca_l$ of 97.55%.

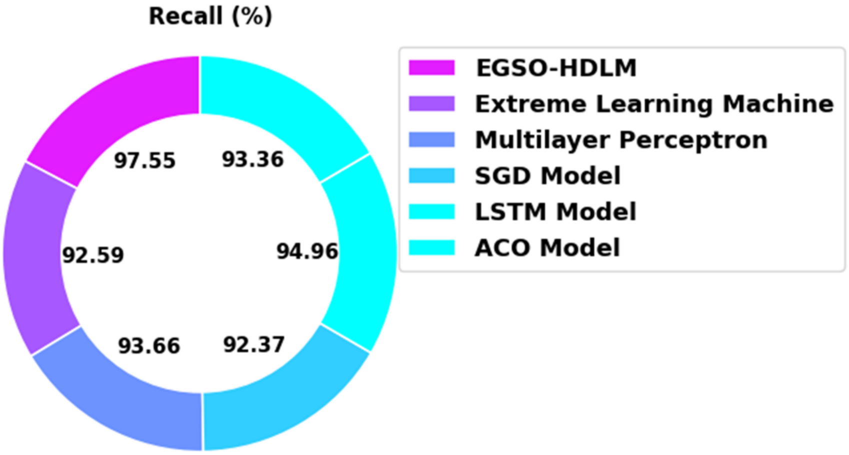

Figure 14. $Reca_l$ analysis of EGSO-HDLM algorithm with existing methodologies.

Therefore, the experimental results indicated that the EGSO-HDLM model has the ability to classify COVID-19 in an epidemiology dataset. The enhanced performance of the proposed model is due to the fusion process and EGSO-based hyperparameter tuning process.

4. Conclusions

In this study, a new EGSO-HDLM model was designed for the identification and classification of COVID-19 using epidemiology data. Primarily, the presented EGSO-HDLM model applied data normalization to pre-process the actual data. Then, the HCNN-GRUF model was applied for the classification process. Moreover, the hyperparameter optimization of the HCNN-GRUF model was improved by the use of the EGSO algorithm, which was derived by including the concepts of cat map and the traditional GSO algorithm. The design of the EGSO algorithm helps in reducing the ergodic problem, avoiding premature convergence, and enhancing algorithm efficiency. To assure the better performance of the EGSO-HDLM model, an extensive set of experimental validations on a benchmark dataset was performed. The simulation results ensured the enhanced performance of the EGSO-HDLM model over recent approaches. In future, the classification performance of the EGSO-HDLM model can be improved by the use of feature selection and outlier removal approaches. Moreover, the proposed model can be extended to the detection of other diseases.

Author Contributions: Conceptualization, M.R. and H.C.; methodology, M.W.A.-R. and A.H.A.; software, S.S.B. and M.R.; validation, H.C. and A.H.A.; formal analysis, S.S.B. and M.W.A.-R.; investigation, M.R. and H.C.; resources, M.W.A.-R.; data curation, S.S.B.; writing—original draft preparation, H.C.; writing—review and editing, A.H.A.; visualization, S.S.B.; supervision, M.W.A.-R.; project administration, M.R.; funding acquisition, M.W.A.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (G:813-140-1441).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated during the current study.

Acknowledgments: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah. The authors, therefore, gratefully acknowledge with thanks DSR for technical and financial support.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Shuja, J.; Alanazi, E.; Alasmary, W.; Alashaikh, A. COVID-19 open source data sets: A comprehensive survey. *Appl. Intell.* **2021**, *51*, 1296–1325. [CrossRef] [PubMed]
2. Yuki, K.; Fujiogi, M.; Koutsogiannaki, S. COVID-19 pathophysiology: A review. *Clin. Immunol.* **2020**, *215*, 108427. [CrossRef] [PubMed]
3. Rockett, R.J.; Arnott, A.; Lam, C.; Sadsad, R.; Timms, V.; Gray, K.A.; Eden, J.S.; Chang, S.; Gall, M.; Draper, J.; et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **2020**, *26*, 1398–1404. [CrossRef] [PubMed]
4. Böhmer, M.M.; Buchholz, U.; Corman, V.M.; Hoch, M.; Katz, K.; Marosevic, D.V.; Böhm, S.; Woudenberg, T.; Ackermann, N.; Konrad, R.; et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series. *Lancet Infect. Dis.* **2020**, *20*, 920–928. [CrossRef]
5. Pan, A.; Liu, L.; Wang, C.; Guo, H.; Hao, X.; Wang, Q.; Huang, J.; He, N.; Yu, H.; Lin, X.; et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* **2020**, *323*, 1915–1923. [CrossRef]
6. Xia, X.Y.; Wu, J.; Liu, H.L.; Xia, H.; Jia, B.; Huang, W.X. Epidemiological and initial clinical characteristics of patients with family aggregation of COVID-19. *J. Clin. Virol.* **2020**, *127*, 104360. [CrossRef]
7. Wu, Q.; Xing, Y.; Shi, L.; Li, W.; Gao, Y.; Pan, S.; Wang, Y.; Wang, W.; Xing, Q. Coinfection and other clinical characteristics of COVID-19 in children. *Pediatrics* **2020**, *146*, e20200961. [CrossRef]
8. Ibrahim, N.K. Epidemiologic surveillance for controlling Covid-19 pandemic: Types, challenges and implications. *J. Infect. Public Health* **2020**, *13*, 1630–1638. [CrossRef]
9. Chauhan, H.; Gupta, D.; Gupta, S.; Singh, A.; Aljahdali, H.M.; Goyal, N.; Noya, I.D.; Kadry, S. Blockchain Enabled Transparent and Anti-Counterfeiting Supply of COVID-19 Vaccine Vials. *Vaccines* **2021**, *9*, 1239. [CrossRef]
10. Mansour, R.F.; Escorcia-Gutierrez, J.; Gamarra, M.; Gupta, D.; Castillo, O.; Kumar, S. Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. *Pattern Recognit. Lett.* **2021**, *151*, 267–274. [CrossRef]
11. Ragab, M.; Alshehri, S.; Alhakamy, N.A.; Mansour, R.F.; Koundal, D. Multiclass Classification of Chest X-Ray Images for the Prediction of COVID-19 Using Capsule Network. *Comput. Intell. Neurosci.* **2022**, *2022*, 6185013. [CrossRef]
12. Kaur, P.; Harnal, S.; Tiwari, R.; Alharithi, F.S.; Almulihi, A.H.; Noya, I.D.; Goyal, N. A hybrid convolutional neural network model for diagnosis of COVID-19 using chest X-ray images. *Int. J. Environ. Res. Public Health* **2021**, *18*, 12191. [CrossRef]
13. Zhao, C.; Leung, C.K.; Pazdor, A.G.; Wen, Q. October. Analyzing COVID-19 epidemiological data. In Proceedings of the 2021 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Edmonton, AB, Canada, 25–28 October 2021; pp. 985–990.
14. Li, J.; Xu, Q.; Cuomo, R.; Purushothaman, V.; Mackey, T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: Retrospective observational infoveillance study. *JMIR Public Health Surveill.* **2020**, *6*, e18700. [CrossRef]
15. Alamo, T.; Reina, D.G.; Millán, P. Data-driven methods to monitor, model, forecast and control COVID-19 pandemic: Leveraging data science, epidemiology and control theory. *arXiv* **2020**, arXiv:2006.01731.
16. Muhammad, L.J.; Algehyne, E.A.; Usman, S.S.; Ahmad, A.; Chakraborty, C.; Mohammed, I.A. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN Comput. Sci.* **2021**, *2*, 11. [CrossRef]
17. Roy, S.; Ghosh, P. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. *PLoS ONE* **2020**, *15*, e0241165. [CrossRef]
18. Kim, S.Y.; Yoo, D.M.; Min, C.; Wee, J.H.; Kim, J.H.; Choi, H.G. Analysis of mortality and morbidity in COVID-19 patients with obesity using clinical epidemiological data from the Korean Center for Disease Control & Prevention. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9336.
19. Hawkins, R.B.; Charles, E.J.; Mehaffey, J.H. Socio-economic status and COVID-19–related cases and fatalities. *Public Health* **2020**, *189*, 129–134. [CrossRef]
20. Man, D.; Zeng, F.; Yang, W.; Yu, M.; Lv, J.; Wang, Y. Intelligent intrusion detection based on federated learning for edge-assisted internet of things. *Secur. Commun. Netw.* **2021**, *2021*, 9361348. [CrossRef]
21. Ding, L.; Fang, W.; Luo, H.; Love, P.E.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [CrossRef]
22. Rashedi, E.; Nezamabadi-Pour, H.; Saryazdi, S. GSA: A gravitational search algorithm. *Inf. Sci.* **2009**, *179*, 2232–2248. [CrossRef]
23. Gao, S.; Vairappan, C.; Wang, Y.; Cao, Q.; Tang, Z. Gravitational search algorithm combined with chaos for unconstrained numerical optimization. *Appl. Math. Comput.* **2014**, *231*, 48–62. [CrossRef]
24. Available online: https://www.kaggle.com/marianarfranklin/mexico-covid19-clinical-data/metadata (accessed on 24 May 2022).