

An Efficient Covid-19 Prediction Using Dynamic Parameter Based Artificial Fish Swarm Algorithm with Ensemble Learning

1. T. Sundaravadivel,

Research Scholar, Dept. of Computer Science and Engineering, FEAT, Annamalai University, Tamilnadu, India.

2.Dr. V. Mahalakshmi,

Assistant Professor, Dept. of Computer Science and Engineering, FEAT, Annamalai University, Tamilnadu, India

ABSTRACT

Corona virus infection spreads quickly. COVID-19 is a serious pneumonia virus that is projected to have a major influence on the healthcare industry. Prenatal recognition is essential for proper behavior, which relieves burden on the health-care system. Recently, the various Machines Learning (ML) techniques are used to predict the cases of daily increase Covid-19. The previous system designed an Improved Coefficient based Chicken Swarm Optimization (ICCSO) with Exponential Distribution based Long Short-Term Memory (EDLSTM) approach for covid-19 prediction. However, the ensemble models produce better prediction accuracy compared to single model. For successful COVID -19 prediction, the suggested system constructed a Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) with Ensemble Learning (EL). The COVID-19 case dataset is used as a starting point, and it is standardised using Z-score normalisation. Then Independent Component Analysis(ICA) is used to diminish the dimensions. The Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) is used to find the best attributes to increase classification accuracy. Ensemble Learning (EL) with Exponential Distribution based Long Short-Term Memory (EDLSTM), Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS), and Support Vector Machine (SVM) classifier is used to predict COVID-19 based on the specified attributes. An actual data set would be used to carry out the experiment. The suggested system outperforms the existing system in terms of accuracy, precision, recall, and f-measure, according to the results.

Keywords: COVID-19, Ensemble Learning (EL), Independent Component Analysis (ICA), Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) and SVM classifier

1. INTRODUCTION

The ongoing outbreak, termed COVID-19, is a disease that is characterized by the virus severe acute respiratory syndrome coronavirus2 (SARS-CoV2). The disease COVID-19 was initially diagnosed in late December 2019 in Wuhan, the capital of Hubei Province, China, creating the initial epidemic of the century [1-2]. COVID-19, an infectious illness that has been spreading internationally since the beginning of the year 2020, has resulted in about three million optimistic cases to date (There are approximately one million closed cases with a 20% fatality rate through April 27, 2020). COVID-19 has now become a contagion, as we all know, for a variety of whys and wherefores. Some of these are the lack of an appropriate vaccine and specific treatment, (ii) a high illness recurrence rate, and (iii) the virus SARS-specific CoV-2's nature being undetermined. COVID-19 has affected as many as 210 nations and territories throughout the world today.

There is apparently no human vaccination authorised to fight it [3]. When people are close together, the covid-19 spreads quickly. Travel limitations, facial covering with a mask or cloth, social isolation, maintaining a lockdown, and periodic hand hygiene are all suggested to avoid viral infections. Meanwhile, the most frequent illness signs are fever and cough [4]. Other signs and symptoms include chest pain, sputum production, and a sore throat. COVID19 has the potential to proceed to viral pneumonia, which has a 5.8% fatality rate.COVID-19 has a mortality rate that is 5% that of the 1918 Spanish flu pandemic. The World Health Organization (WHO) has issued preventative measures to avoid infection with the COVID-19 virus. Covering one's face with a mask or cloth, crouching with namaste instead of shaking hands, social isolation, and establishing a lockdown are among them.

This outbreak continues to pose a number of challenges to pharmaceutical systems all over the world, including increased demand for inpatient care and critical shortages of diagnostic instruments, as well as the infection of many healthcare personnel. As a consequence, the capacity to make timely clinical decisions and use healthcare resources efficiently is crucial. RT-PCR (Reverse Transcriptase Polymerase Chain Reaction), the most extensively used COVID-19 diagnostic test, has long been in short supply in disadvantaged countries [5]. As a result, infection rates rise, and critical prophylactic measures are postponed. COVID-19 may be diagnosed quickly and efficiently with effective screening, reducing the load on healthcare schemes. Forecast models that incorporate many variables to assess the probability of contagion have been created in the hopes of supporting health professionals throughout the world in triaging patients, particularly in areas where healthcare resources are few.

Numerous big data based on a meta have thoroughly characterised and summarised the clinical aspects of COVID-19 patients, allowing early identification and monitoring of illness progression. In clinical characteristics-based medical diagnosis, Artificial intelligence (AI) techniques such as Fuzzy Logic (FL), Genetic algorithms (GA), Decision Trees (DT), SVM, Artificial Neural Networks (ANNs), and Deep learning (DL) have gained popularity in the field of health care for detection, identification, and estimation of various medical problems. Omar and Gupta [6] use the Long Short-Term Memory (LSTM) model to forecast the quantity of instances and to investigate the impact of social isolation and lock-down. [7] To display the trend of the covid impacted instances, multiple regressions such as Linear and Multiple Linear Regression approaches are employed to the data set. However, these methods do not provide the satisfactory accuracy results. The main goal of the proposed exertion is to develop an efficient collaborative learning prototypical that could foresee whether a patient is suffering from COVID-19 and analyze the severity.

This is how the rest of the paper is organised. Previous covid-19 prediction techniques are given in Section 2. The sources of data and proposed approach are described in depth in section 3. Section 4 contains the experiments and analysis of the data. Finally, the primary findings of this study are described in Section 5.

2. LITERATURE REVIEW

For COVID-19 prediction, Zheng et al (2020) suggested a mixed Artificial Intelligence (AI) model. Initially, like traditional modelling tools, which assume that all coronavirus-infected people have the identical contagion rate, the Improved Susceptible-Infected (ISI) model is designed to examine the diversity of contagion rates in order to study distribution laws and trends. Second, the ISI model is combined with the Natural Language Processing (NLP) module and the LSTM network to create a hybrid AI model for COVID-19 forecast that takes into consideration the impacts of adopting sustainability as well as improved public awareness. Thus, according to experimental research relying on outbreak data from various typical areas and cities in China, entities diagnosed with corona virus had a higher contagion rate during the third to eighth days following infection, which is much more in line with the epidemic's genuine distribution laws. Furthermore, the built hybrid AI model can considerably diminish prediction errors, with MAPEs of 0.52 percent, 0.38 percent, 0.05 percent, and 0.86 percent for the next six days in Wuhan, Beijing, Shanghai, and the entire country [8].

Peng and Nagata (2020) used Support Vector Regression (SVR) to forecast the quantity of positive cases in the 12 most pretentious regions, using Kernel functions to examine for various configurations of nonlinearity and 3-D approximated textures to analyse the responsiveness of the forecasting model' generalization accuracy to distinct hyper parameter setups. In this investigation, the approach with the largest nonlinearity (Gaussian Kernel) had the best in-sample efficacy but the poorest out-of-sample predictions, which is a wonderful illustration of standardisation in an instructional approach. The quadratic Kernel function, on either hand, outperformed in-sample but provided the greatest out-of-sample estimates. The outcomes of this study offer an analytical review of key ideas in data collecting and highlight the need for precaution when using machine learning models to aid critical decision-making, particularly in the context of the pandemic problems [9].

Deep learning models for forecasting the number of new corona virus positive confirmed cases for 32 Indian states and union territories were published by Parul Arora et al (2020). Using an Indian sample, the range of potential cases is forecasted using RNN-based LSTM variants such as Deep LSTM, Convolutional LSTM, and Bi-directional LSTM. Considering daily and weekly situations, an LSTM model with the lowest error is used. The developed technique produces great exactness for temporary forecasts, with error rates of a smaller amount than 3% for day-to-day projections and less than 8% for daily estimates. Indian provinces are classified into sectors depending on the distribution of optimistic cases and day-to-day progress rate for comfortable discovery of small coronavirus outbreaks. Precautionary procedures to minimise disease transmission in certain zones are also advocated [10].

Guhathakurata et al. (2021) proposed a new method for predicting infected rate based on SVM. As a starting point, a symptom-based COVID-19 dataset is used. Using our suggested method, we can determine whether or not someone is infected based on their symptoms. After that, the dataset is sent to the SVM classifier. Not infected, slightly infected, and severely infected are the three classifications for the outcome, i.e. infected status. Based on the results of the experiments, it can be determined that the constructed system predicts situations with an accuracy of 87 percent [11].

Zheng et al (2020) established a hybrid AI model for corona prognosis. To begin, although typical pandemic models assume that all corona virus-infected people have the same infection rate, an ISI model is presented to evaluate infection rate variability in order to examine distribution rules and trend. Second, the NLP unit and the LSTM structure are integrated into the ISI framework to create a hybrid AI Optimization approach for prognosis, which takes into consideration the impacts of adopting sustainability as well as improved public awareness. Extensive experimental results demonstrate hybrid AI model can suggestively diminish generalization error, with mean unconditional percentage errors (MAPEs) of 0.52 percent, 0.38 percent, 0.05 percent, and 0.86 percent for another six days in Wuhan, Beijing, Shanghai, as well as the entire country, respectively [12].

Niazkar (2020) utilised ANN to foresee the pandemic. ANN-based strategies were used to predict the number of cases reported in China, Japan, Singapore, Iran, Italy, South Africa, and the United States of America. These models rely on historical information of proven instances, with the key variation being the number of days they believe will influence the estimating process. Training and testing were separated from the COVID-19 data. The first one was used to create the ANN models, whereas the second was utilised to assess the objectives. The study of the data indicates not just to substantial daily validated case fluctuations, but also discrete varieties of aggregate established cases reported across the time period examined [13].

3. PROPOSED METHODOLOGY

In this proposed work, Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) with Ensemble Learning (EL) is utilized for covid 19 prediction. It is divided into steps that include input, preprocessing, dimension reduction, feature selection, and classification. Figure 1 depicts the process flow for the integrating technology.

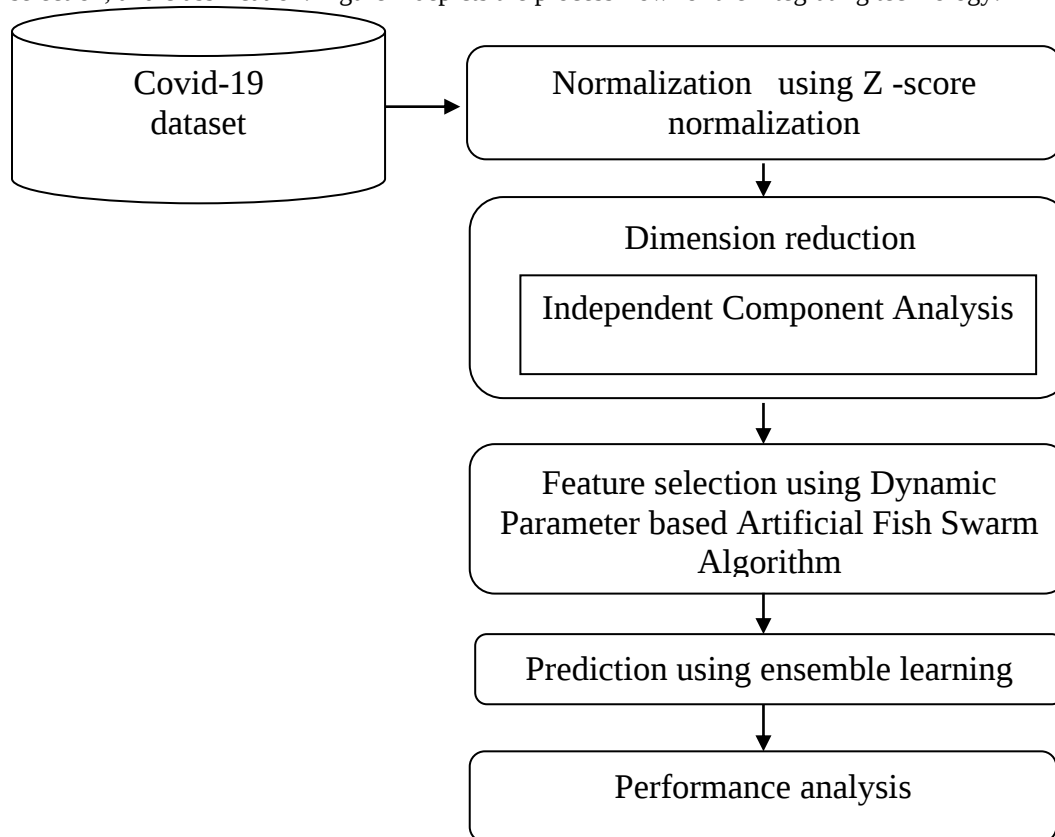


Figure 1: Flow diagram of the planned work

3.1 Input

The information was gathered since <https://www.kaggle.com/iamhungundji/covid19-symptoms-checker>. The dataset comprises seven important factors that will influence whether or not anyone gets corona virus illness; the following is an explanation of each parameter:

Country: A list of the nations that the individual has visited.

Age: According to WHO, each person's age group is classified, category by Age Standard

Symptoms: Fever, tiredness, breathing difficulty, dry cough, and sore throat are the five prominent indications of COVID-19, according to the WHO.

Understanding any other indications: Pains, Nasal Congestion, Runny Nose, Diarrhea and Other.

Severity: Mild, Moderate, and Severe are the three levels of severity.

Contact: Is this individual in contact with another affected patient?

3.2 Preprocessing using normalization

Scaling the numeric value to be between 0 and 1, and then scaling all of the values prior to actually building the Machine Learning models. The participation dataset is normalised via z-score normalisation in this suggested study. Data

normalisation is a type of data analysis in which characteristic data or characteristics are mounted to fall inside a certain variety, such as -1.0 to 1.0 or 0.0 to 1.0. The most frequent technique of normalising is Z-score normalisation. Zero mean normalisation is also known as Z score normalisation. The data is standardised using the mean and standard deviation in this case. The formula is then,

$$d' = \frac{d - \text{mean}(p)}{\text{std}(p)} \quad (1)$$

Where,

Mean(p) -Sum of the all-attribute values of p

Std (p)-Standard deviation

3.3 Dimension reduction

ICA is used to reduce the number of dimensions. The ICA is a helpful PCA modification that was created to help blindly separate distinct sources from their linear mixes. This type of blind separation technique has been used in a range of applications, including acoustic signal isolation, health signal processing, and so on. Instead of requiring the parameters of the data vectors' inception modules to be statistically independent as in PCA, ICA demands that they be reciprocally independent. This suggests that in direction to determine the ICA expansion, higher-order statistics are required.

Bearing in mind an $n \times p$ data matrix X (input dataset), whose rows $\mathbf{r}_i$ ($i=1, \dots, n$) resemble to databased variables and whose columns $\mathbf{c}_j$ ($j=1, \dots, p$) are the entities of the matching variable quantity, the ICA model of X can be inscribed as

$$X = AS \quad (2)$$

A is a $n \times n$ mixing matrix, and S is a $n \times p$ source matrix, with both the rows of S being as linearly separable as feasible. The empirical variables are linear mixes of different components, while the 'independent components' are the particular alterations in the rows of S . Joint distribution can be used to quantify statistical independence between variables. $I = \sum_k H(s_k) - H(S)$ the marginal entropy of the variable s_k and $H(S)$, while the joint entropy is $H(S)$. We may predict the independent components by determining the appropriate parameterization of the estimated parameters meanwhile we can reverse the mixing matrix.

$$U = S A^{-1} X = W X \quad (3)$$

The ICA approach, which has been created to handle features extraction challenges, is used in this study. A 'contrast function' is used to represent mutual information in this approach:

$$J(s_k) = (E\{G(s_k)\} - E(G(v)))^2 \quad (4)$$

Where, G is an subjective non-quadratic function and v is a typically circulated variable.

ICA, like PCA, may eliminate all linear relationships. It also uses a non-orthogonal basis to account for relatively high links in the data. PCA, which is only vulnerable to second-order data correlations, is better than ICA. By tradition, the ICA model allows for considerable scaling and ranking freedom; although the separate mechanisms are usually clambered to unit deviation, their signs and ordering can be changed at any time.

3.4 Feature selection

In this proposed research work, Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) is utilized for feature selection. Artificial fish (AF) is a made-up creature that looks like a real fish. Fish, in instance, travel to a site with more consistent food by engaging in social search activities [14-15]. Prey behaviour, follow behaviour, swarm activity, and jump behaviour are four of AF's social behaviours. These actions are used to undertake the problem analysis and explanation.

In this proposed work, features in the dataset $X = (x_1, x_2, \dots, x_n)$ is taken as an input, where x_i is number of features in the dataset. The fitness function $Y = f(X)$ may be used to indicate the food attentiveness in the existing location of the artificial fish. The suggested study considers categorization accuracy to be a fitness function. The variety in which AFs may exploration is called visual, and the greatest length that an AF can travel is called step. The distance between two AFs X_i and X_j may be calculated using the Euclidean distance equation:

$$D_{i,j} = |X_i - X_j| \quad (5)$$

The crowd factor δ ($0 < \delta < 1$), is used to manage the crowding of AF's around a location, and the best position with the most features will be inserted into the bulletin. The behaviours of AF's will be detailed in depth in the next sections.

(1) *Prey Behavior*. Prey behaviour is a basic genetic activity used by fish to novelty food in the wild. The point X_j in the pictorial purview of the AF i can be determined at random. Y is the food saturation, which can be represented as the unbiased function $Y = f(X)$. X_i is the recent location of feature i , X_j is a arbitrary state of its visual distance. The following equation can be used to compute the location X_j :

$$X_j = X_i + \text{Visual} \times \text{rand}(0,1) \quad (6)$$

Y_j and Y_i determine the food attentiveness of X_i and X_j ; if $Y_i < Y_j$, AF interchanges forward a phase from its existing location to X_j , which is ended by

$$X_i(t+1) = X_i(t) + \frac{X_j(t) - X_i(t)}{\|X_j(t) - X_i(t)\|} \times \text{step} \times \text{rand}(0,1) \quad (7)$$

If $Y_i > Y_j$, we pick another state X_j at random and see if its food consistency meets with a forward condition. When the feature i is not fulfilled with the onward disorder after trying for number of times, the related AF executes jump behaviour.

(2) *Swarm Behavior*. AFs seek to travel towards the centre location in every iteration to maintain swarm generality. The following equation may be used to calculate the central position:

$$X_c = \frac{1}{N} \sum_{i=1}^N X_i \quad (8)$$

The arithmetic average of entirely characteristics is X_c . And N denotes the population size. The quantity of AF swarms in the visual possibility of X_c is denoted by n_f . If $\frac{n_f}{N} < \delta$ and $Y_c > Y_i$, which indicate that the swarm's centre location has greater food consistency and the inhabitants is not a crowd, feature i advances one step to the companion center is given as

$$X_i(t+1) = X_i(t) + \frac{X_c - X_i(t)}{\|X_c - X_i(t)\|} \times \text{step} \times \text{rand}(0,1) \quad (9)$$

Otherwise, AF accomplishes prey performance.

(3) *Follow Behavior*. When a single fish or a group of fish finds food during the AF's movement, the neighbouring fishes will follow and arrive fast. Assume that the current positioning of the feature i is X_i , and that position X_j is its perceptual scope neighbour. The frequency of AF swarms in the visual scope of X_c is denoted by n_f ; if $Y_i < Y_j$ and $\frac{n_f}{N} < \delta$, feature i advances a step to the neighbour X_j . The following is how the statement is ascertained:

$$X_i(t+1) = X_i(t) + \frac{X_j(t) - X_i(t)}{\|X_j(t) - X_i(t)\|} \times \text{step} \times \text{rand}(0,1) \quad (10)$$

Feature i implements the prey behaviour if there are no neighbours surrounding X_i or if all of them are unsatisfied with the situation.

(4) *Leap Behavior*. Leap behaviour is the basic activity of searching for food or partners over wide distances, essentially preventing local optimum feature. The leap behaviour is performed, and the characteristic is changed to jump out of the present location. To avoid the local outliers, it picks a visual state and goes towards it. Hence,

$$X_i(t+1) = X_i(t) + \text{visual} \times \text{rand}(0,1) \quad (11)$$

Dynamic parameter based AFSA

The visual in traditional AFSA is fixed. The search scope, convergence rate, and accuracy of conventional AFSA are determined by the visual. The strong visual values cause the AFs to get nearer to the best solution more rapidly, increasing the convergence speed at the start of the typical AFSA optimization procedure. When the AFs come close to the target point, however, the huge visual values cause them to overlook the location with superior food density. However, if the graphical values are too little, the local search capability increases but the optimization capability decreases. As a consequence, AFs may miss the worldwide optimum and instead slip into the local optimum, lowering the output's dependability. To meet the needs of convergence rate and stability of worldwide and resident search, a continuously constraint is identified to enhance the visual of AF. The variable's value is computed as trails:

$$\alpha = k e^{-\beta d_1} + (1-k) e^{-\left(\frac{t}{T_{max}}\right)^s} \quad (12)$$

$$d_1 = c_1 - \frac{t}{T_{max}} \quad (13)$$

Where, $k \in [0,1]$, c_1 is constants. The present amount of repetitions is t , while the algorithm's maximum number of repetitions is T_{max} . The parameter s has a range of values from 1 to 20. The d_1 is a set of tailored control settings that ensures a high level of local search capability and improves convergence accuracy. The visual expressiveness is calculated as follows:

$$\text{Visual}_{new} = \text{Visual} \times \alpha \times \text{Visual}_{min} \quad (14)$$

Where Visual_{min} is the bare minimum given by the issue space's requirements. In equation, the Visual_{new} equation is modified.

Algorithm 1:

Input: Number of features in the dataset

Output: Selected features

Step 1: Initialize the features in the dataset $X = (x_1, x_2, \dots, x_n)$, where $x_i (i = 1, 2, \dots, n)$

Step 2: Set up the fake fish's settings, such as step, visual, try number, and maximum iterations.

Step 3: Decide on the best value for each feature and write it down in the statement.

Step 4: Incorporate prey, swarm, track, and jump behaviour into your training.

Step 5: The improved individual optimum updates the ideal value in the bulletin board.

Step 6: If the expiry state is met, the best characteristics is produced

Step 7: Or else, coming back to Step 2.

3.5 Classification

An Ensemble Learning (EL) classifier, which includes EDLSTM, IFGF-ANFIS, and SVM classifier, is used to classify people with infection possibility.

Exponential Distribution based Long Short-Term Memory (EDLSTM)

The EDLSTM classifier is used in this proposed study to categories persons with infection potential into severity levels based on the specified characteristics. The LSTM is a RNN topology that was created to more precisely describe temporal patterns and their long-range relationships than ordinary RNNs. As depicted in Figure 2, a characteristic LSTM cell contains of input, forget, and output gates, as well as a cell activation component. These devices receive activation signals from many sources and use the predefined multipliers to regulate the cell's activation. For several time steps, the LSTM gates can prevent the remainder of the system from changing the values of the memory cells. Signals are preserved and mistakes are propagated for far longer in LSTM networks than in RNNs.

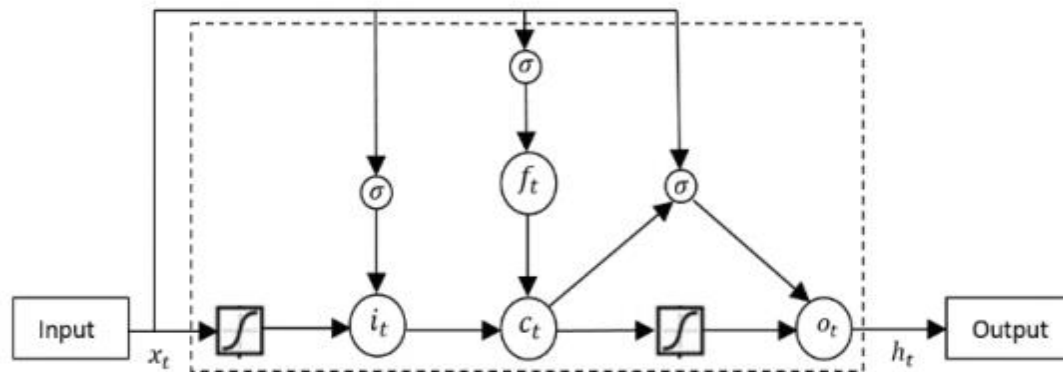


Figure 2. Long Short-Term Memory Cell

The input gate of LSTM is defined as

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (15)$$

The forget gate is defined as

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (16)$$

The cell gate is defined as

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (17)$$

The output gate is defined as

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (18)$$

Finally, the hidden state is computed as

$$h_t = o_t \tanh(c_t) \quad (19)$$

tanh - hyperbolic tangent initiation purpose

x_t - input at time t

W and b - network parameters (Weights and Biases)

The input gate, forget gate, output gate, and cell state are all represented by the logistic sigmoid function, where i, f, o and c are the input gate, forget gate, output gate, and cell state, respectively. Weight matrices for peephole associates are designated by the letters W_{ci} , W_{cf} and W_{co} . The information flow in an LSTM is controlled by three gates. The input gate determines the input ratio. This ratio has an impact on the equation when computing the cell state (17). The forget gate regulates whether the prior memory $ht1$ is passed or not. In equation (16), the prior memory ratio is computed and employed in the calculation (17). The output gate decides whether the memory cell's output is passed or not. This is depicted in equation (19). Due to the three gates, we can tackle the disappearing and explosion gradient problems with LSTM. In the LSTM-RNN architecture, the recurrent hidden layer is substituted by an LSTM cell.

In this proposed research work, optimal value of the w are updated based on the exponential distribution. For a given features a_i ($i=1, \dots, n$) the associated weight values are $(w_1, w_2, \dots, w_n)$. Here weight values are computed based on the exponential distribution.

$$w_i = \frac{e^{-\frac{(a_i - \mu)^2}{2\sigma^2}}}{\sum_{i=1}^n e^{-\frac{(a_i - \mu)^2}{2\sigma^2}}}, i=1, 2, \dots, n \quad (20)$$

Where,

$$\mu = \frac{1}{n} \sum_{i=1}^n a_i,$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mu - a_i)^2}$$

Where, μ is the mean of the collection features, and σ is the standard deviation of features.

Based on the weight matrices of the input features the classification is performed. The output of the Exponential Distribution based Long Short-Term Memory (EDLSTM) is covid -19 probability prediction with stages.

A. Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS)

A IFGF-ANFIS classifier is used to categories patients who are at risk of infection by analysing symptoms such as fever, fatigue, dry cough, sore throat, and breathing problems. This is a collection of nodes that is built on the base of a neuro fuzzy network. Because it's a research tool that allows, some of its nodes are also adaptable, which means their outputs are dictated by the parameters they take. The architecture of the ANFIS model is seen in Figure 3.

The output membership function for various rules is different, and these numbers must be the same. Two fuzzy IF-THEN rules based on the first order Sugeno model are studied in order to implement this architecture.

Rule (1): IF x is A_1 AND y is B_1 , THEN

$$f_1 = p_1x + q_1y + r_1$$

Rule (2): IF x is A_2 AND y is B_2 , THEN

$$f_2 = p_2x + q_2y + r_2$$

Where,

The participations are x and y

Fuzzy sets are A_i and B_i

f_i - productions within the fuzzy section quantified by the fuzzy rule

p_i , q_i , and r_i , are the strategy restrictions which are strong minded at the keeping fit process.

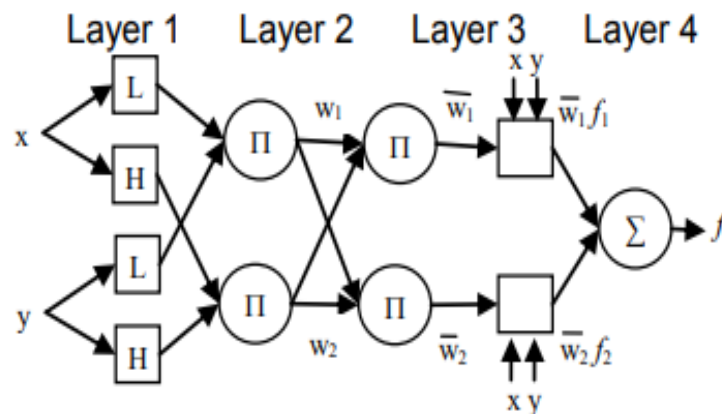


Figure 3: Architecture of ANFIS

Layer 1: Adaptive nodes make up the first layer. The selected characteristics are used as input and passed to the first layer. The above mathematical representation denotes the outputs of Layer 1, fuzzy involvement grade input:

$$O_{1,i} = \mu_{A_i}(x) v_A(x), i=1, 2, \quad (21)$$

$$O_{1,i} = \mu_{B_{i-2}}(y) v_B(y), i=3, 4, \quad (22)$$

The inputs for node i are represented by x and y , whereas A_i and B_i are the linguistic labels (high, low, etc.) related with the node purpose. All involvement functions may be adapted using $\mu_{A_i}(x)$ and $\mu_{B_{i-2}}(y)$. The Intuitionistic Fuzzy Gaussian

Function (IFGF) is constructed for member function computation in this suggested study. There are two parameters that define the IFGF. The Gaussian function has a centre value of m and a breadth of $k > 0$. The curve narrows as the k value lowers. These are the fuzzy Gaussian functions:

$$\mu_A(x) = \exp\left(-\frac{(x-m)^2}{2(k)^2}\right) - \epsilon \quad (23)$$

$$v_A(x) = 1 - (\exp\left(-\frac{(x-m)^2}{2(k)^2}\right)) \quad (24)$$

Here,

$\mu_A(x)$ is the membership function

$v_A(x)$ is the non membership function

Figure 4 depicts a graphical representation depiction of an intuitionistic fuzzy Gaussian function.

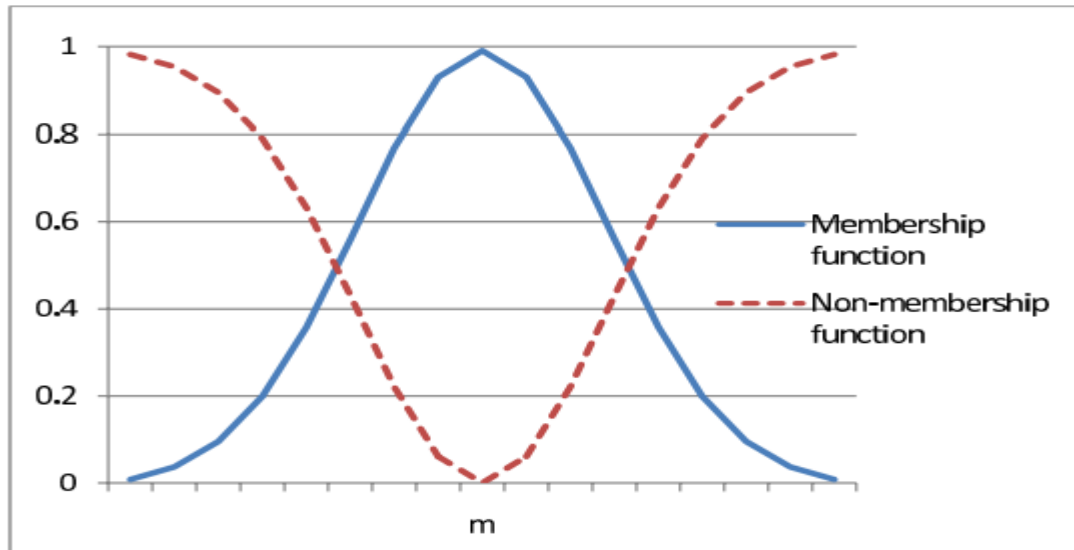


Figure 4: Intuitionistic fuzzy Gaussian function.

Layer 2: The nodes in Layer 2 are fixed. It fuzzifies the inputs using fuzzy operators and the AND operator. This is represented as $\mu_{A_i}(x)v_{A_i}(x)$, which is a simple multiplier. The following is the result of this layer:

$$O_{2,i} = w_i = (\mu_{A_i}(x)v_{A_i}(x))(\mu_{B_i}(y)v_{B_i}(y)), i = 1, 2 \quad (25)$$

Layer 3: N nodes, which are fixed nodes, contribute up the third layer. The nodes are in charge of standardizing the shooting strengths obtained from the second level. The following is an example of the finished product:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2 \quad (26)$$

The outputs are referred to as normalised firing assets (weight). The weight values of the characteristics are w_1 and w_2 .

Layer 4: This layer's nodes are extremely adaptable. The combination of the normalised weight and a first order polynomial represents the output shaped by each node of the fourth layer:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i), i = 1, 2, (27)$$

Where, w_i stands for the output of Layer 3, and p_i , q_i , and r_i represent the subsequent parameters.

Layer 5: The fifth layer is made up of a single fixed node named P that sums up all incoming signals. The following is an example of Layer 5 output:

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (28)$$

The outcome of layer 5 is the covid -19 likelihood estimation.

B. Support Vector Machine (SVM)

The SVMs is a collection of machine learning algorithms that can be used to recognize patterns in given features. In this work, initially the selected features are taken as input and utilized for covid-19 classification. In classification, the training features belonging to two separate classes, $(x_1, y_1), \dots, (x_n, y_n)$, where, $x_i \in R^n$ which represents the i -th training features and $y_i \in \{-1, +1\}$ which represents the corresponding class label. In order to perform separation, one wants to find a hyperplane $\omega \cdot x + b = 0$. In this weight vector ω is a normal vector of hyperplane, b is deviation, and x is training features. The category formula is:

$$(w \cdot x) + b \geq \text{if } y_i = +1 \quad (29)$$

$$(w \cdot x) + b \leq -1 \text{ if } y_i = -1 \quad (30)$$

To find the optimum hyper plane, $\|w\|^2$ must be minimized under the constraint $y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \dots, n$. Therefore, it is required to solve the optimization problem given by

$$\min \frac{1}{2} \|w\|^2 \quad (31)$$

Subject to $y_i(x_i \cdot w + b) \geq 1$

The positive slack variables ξ_i are now included to replace the optimization issue and enable the approach to be extended to a nonlinear decision surface. The new optimization problem is formulated as follows:

$$\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (32)$$

s.t $y_i(\omega \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i=1, 2, \dots, n$

Where C is a penalty parameter that handles the tradeoff amongst maximization of margins and minimization of errors.

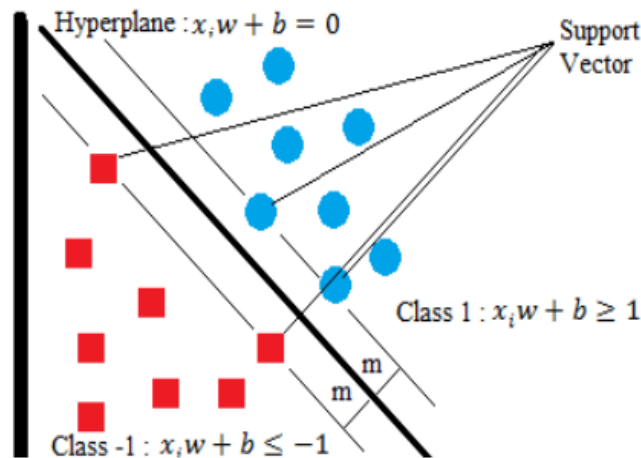


Figure 5: Illustration of classification in SVM

The ensemble learning such as EDLSTM, IFGF-ANFIS and SVM) approaches are mentioned in figure 5.

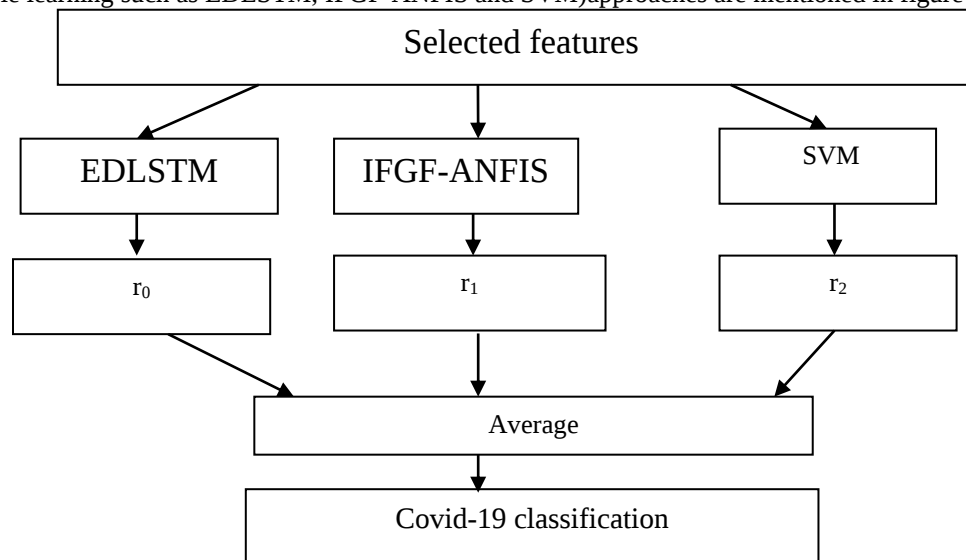


Figure 6. Ensemble learning

For a given input, the output probabilities from EDLSTM, IFGF-ANFIS and SVM are averaged before making a covid-19 classification. For output i, the average output S_i is given by:

$$S_i = \frac{1}{n} \sum_{j=1}^n r_j(i) \quad (33)$$

Where, $r_j(i)$ is the output i of network j for a given input pattern. According to the average output of the network, the maximum classification accuracy is selected to predict the covid-19.

4. EXPERIMENTAL RESULTS

The experiments are evaluated by java. The dataset is collected from <https://www.kaggle.com/iamhungundji/covid19-symptoms-checker>. The performance of the proposed Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) with Ensemble Learning (EL) is compared with existing XGBoost, Support Vector Machine (SVM) and Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) and Improved Coefficient based Chicken Swarm Optimization (ICCSO) with Exponential Distribution Based Long Short-Term Memory (EDLSTM) approaches in terms of accuracy, precision, recall and f-measure. Table 1 represents the overall performance comparison.

Table 1: Performance comparison

Metrics	Methods				
	XGBoost	SVM	IFGF-ANFIS	ICCSO - EDLSTM	DPAFSA-EL
Accuracy	87	91	93	94.5	95.94
Precision	92.94	95.51	96.7	97	98.14
Recall	91.86	94.44	95.65	96.8	97.60
F-measure	92.4	94.97	96.17	97.5	97.87

1. Accuracy

Accuracy is the most extreme intrinsic capacity estimation that accurately orders the instance of occurrence and it is basically a proportion of anticipated right observation to the total observation.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (34)$$

here,

TP- True Positive

FN - False Negative

FP - False Positive

TN- True Negative

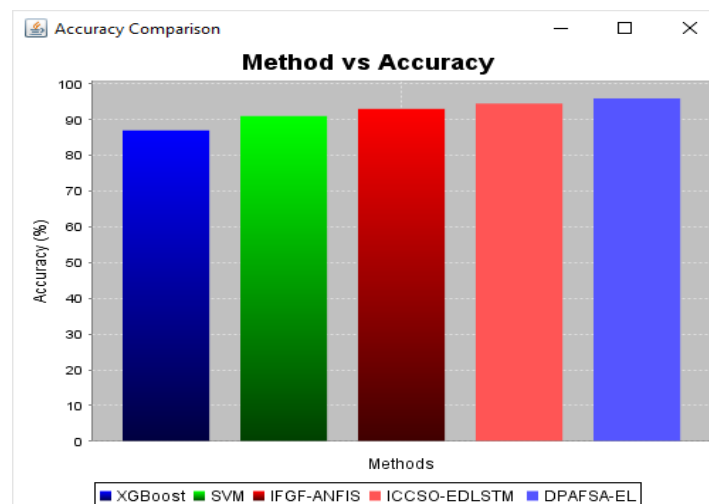


Figure 7: Accuracy comparison

Accuracy of the proposed DPAFSA with EL is compared with existing XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM approaches. In x-axis approaches are taken and accuracy is taken as y-axis. In this proposed research work, Dynamic Parameter based Artificial Fish Swarm Algorithm (DPAFSA) is exploited for optimal feature selection. It improves the accuracy rate of the classifiers. From the investigational outcomes, it can be decided that the proposed system attains 95.94% of accuracy whereas other methods such as XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM achieves 87%, 91%, 93% and 94.5% correspondingly.

2. Precision

Precision defines the relevance of the results and is given by the ratio positive observations predicted correctly to the positive observations predicted in total.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (35)$$

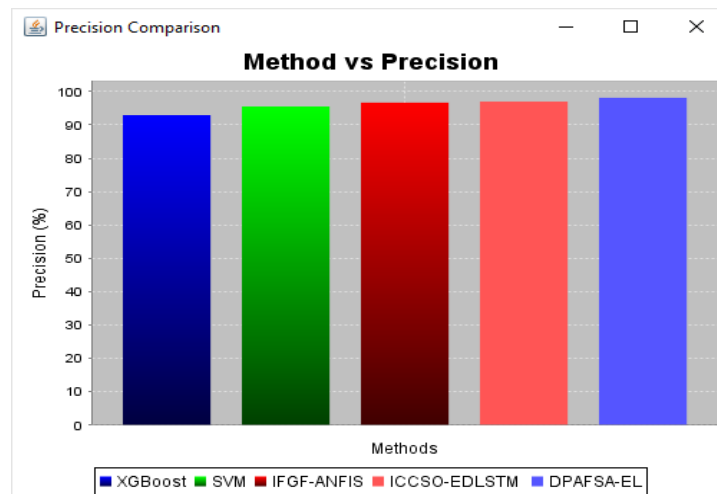


Figure8 : Precision comparison

The proposed DPAFSA with EL is compared with existing XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM approaches in terms of precision. In x-axis methods are taken and precision is taken as y-axis. The investigational grades shows that the proposed system attains 98.14% of precision whereas other methods such as XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM provides 92.94%, 95.51%, 96.7% and 97% correspondingly.

3. Recall

Recall, the ratio of right estimation of positive observation to all observations in actual class - yes.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (36)$$

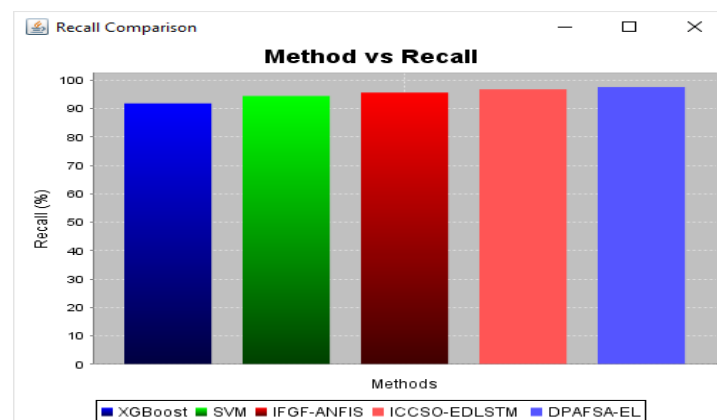


Figure9 : Recall comparison

Recall of the proposed DPAFSA with EL is compared with existing XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM approaches. In x-axis, methods are taken and recall is measured as y-axis. In this proposed work, ensemble learning is exploited for classification which expands the true positive rate. From figure 9, it can be decided

that the proposed system attains 97.60% of recall when other methods such as XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM achieves 91.86%, 94.44%, 95.65% and 96.8% correspondingly.

4. F-measure

F1 Score is also an accuracy estimation of experiment as well as is represented by the weighted mean of Precision and Recall. Hence, this score reflects both FPs and FNs.

$$\text{F-measure} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (37)$$

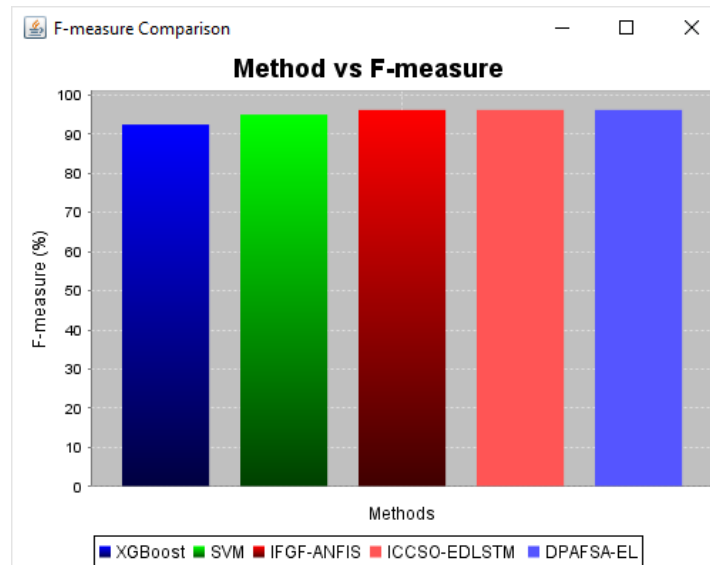


Figure10 :F-measure comparison

F-measure of the proposed DPAFSA with EL is compared with existing XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM approaches. In x-axis, methods are taken and f-measure is measured as y-axis. From the investigational outcomes, it can be concluded that the designed system reaches 97.87% of f-measure whereas other approaches such as XGBoost, SVM and IFGF-ANFIS and ICCSO with EDLSTM achieves 92.4%, 94.97%, 96.17% and 97.5% correspondingly.

5. CONCLUSION

COVID-19's fast meal over the world and the rising amount of deaths require immediate response from all segments. In this proposal, DPAFSA with EL is used to improve the exactness of covid -19 prediction. The input dataset is first normalised by Z-score normalisation. The ICA technique is used to accomplish dimension reduction. DPAFSA is used to pick the best characteristics to increase classification performance. For prediction, EL is added based on the specified characteristics. In terms of accuracy, precision, recall, and f-measure, the investigational findings suggest that the new system outperforms the prior system.

References

- [1] Santosh, K. C. (2020). COVID-19 prediction models and unexploited data. *Journal of medical systems*, 44(9), pp.1-4.
- [2] Mandal, M., Jana, S., Nandi, S. K., Khatua, A., Adak, S., & Kar, T. K. (2020). A model based study on the dynamics of COVID-19: Prediction and control. *Chaos, Solitons & Fractals*, 136, 109889.
- [3] Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., & Alhyari, S. (2020). COVID-19 prediction and detection using deep learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 12, pp. 168-181.
- [4] Yadaw, A. S., Li, Y. C., Bose, S., Iyengar, R., Bunyavanich, S., & Pandey, G. (2020). Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digital Health*, 2(10), e516-e525.
- [5] Zoabi, Y., Deri-Rozov, S and Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj digital medicine*, 4(1), pp. 1-5.
- [6] Tomar, A., & Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Science of The Total Environment*, 728, 138762.

- [7] Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), pp.1467-1474.
- [8] Zheng, N., Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., ... & Xin, J. (2020). Predicting COVID-19 in China using hybrid AI model. *IEEE transactions on cybernetics*, 50(7), pp.2891-2904.
- [9] Peng, Y., & Nagata, M. H. (2020). An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos, Solitons & Fractals*, 139, 110055.
- [10] Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, 139, 110017.
- [11] Guhathakurata, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2021). A novel approach to predict COVID-19 using support vector machine. In *Data Science for COVID-19* (pp. 351-364). Academic Press.
- [12] Zheng, N., Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., ... & Xin, J. (2020). Predicting COVID-19 in China using hybrid AI model. *IEEE transactions on cybernetics*, 50(7), pp.2891-2904.
- [13] Niazkar, H. R., & Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. *Global Health Research and Policy*, 5(1), pp.1-11.
- [14] Yao, X. G., Zhou, Y. Q., & Li, Y. M. (2010). Hybrid algorithm with artificial fish swarm algorithm and PSO. *Application Research of Computers*, 27(6), pp.2084-2086.
- [15] Zainal, N., Zain, A. M., & Sharif, S. (2015). Overview of artificial fish swarm algorithm and its applications in industrial problems. In *Applied Mechanics and Materials* (Vol. 815, pp. 253-257). Trans Tech Publications Ltd.