Materials Today: Proceedings 56 (2022) 3317–3324

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Weighted butterfly optimization algorithm with intuitionistic fuzzy Gaussian function based adaptive-neuro fuzzy inference system for covid-19 prediction

T. Sundaravadivel [*,1], V. Mahalakshmi [2]

Department of Computer Science and Engineering, FEAT, Annamalai University, Tamil Nadu, India

ARTICLE INFO

Article history:
Available online 25 October 2021

Keywords:
Covid-19 prediction
Weighted Butterfly Optimization Algorithm (WBOA)
Intuitionistic fuzzy Gaussian function
Adaptive-Neuro Fuzzy Inference System (IFGF-(ANFIS)
Classification accuracy

ABSTRACT

Covid-19 cases are increasing each day, however none of the countries successfully came up with a proper approved vaccine. Studies suggest that the virus enters the body causing a respiratory infection post contact with a disease. Measures like screening and early diagnosis contribute towards the management of COVID- 19 thereby reducing the load of health care systems. Recent studies have provided promising methods that will be applicable for the current pandemic situation. The previous system designed a various Machine Learning (ML) algorithms such as Decision Tree (DT), Random Forest (RF), XGBoost, Gradient Boosting Machine (GBM) and Support Vector Machine (SVM) for predicting COVID-19 disease with symptoms. However, it does not produce satisfactory results in terms of true positive rate. And also, better optimization methods are required to enhance the precision rate with minimum execution time. To solve this problem the proposed system designed a Weighted Butterfly Optimization Algorithm (WBOA) with Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier for predicting the magnitude of COVID- 19 disease. The principle aim of this method is to design an algorithm that could predict and assess the COVID-19 parameters. Initially, the dataset regarding COVID-19 is taken as an input and preprocessed. The parameters included are age, sex, history of fever, travel history, presence of cough and lung infection. Then the optimal features are selected by using Weighted Butterfly Optimization Algorithm (WBOA) to improve the classification accuracy. Based on the selected features, an Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is utilized for classifying the people having infection possibility. The studies conducted on this proposed system indicates that it is capable of producing better results than the other systems especially in terms of accuracy, precision, recall and f-measure.

Selection and peer-review under responsibility of the scientific committee of the First International Conference on Design and Materials (ICDM)-2021

1. Introduction

Wuhan, China reported the World Health Organization (WHO) with prevalence and threat of a novel virus in December 2019. The virus outbreak was then termed as COVID- 19 by WHO on 11 February 2020 [1,2]. COVID-19 virus is a strain of SARS and ADRS. This outbreak was announced as a community health emergency by WHO and stated that the mode of transmission of the virus is through the respiratory tract, however there are also other ways of transmission which are yet unknown [3]. Any person infected with the disease will manifest symptoms between 2 and 14 days, which also depends on incubation period of Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). Commonly seen signs and symptoms are cough, fever, general malaise, shortness of breath. People with co-morbidities like long term respiratory infections, Type I and Type II diabetes and cardiac ailments are at a higher risk to get infected by COVID-19 virus leading to severe illness.

Fig. 1 illustrates the data on corona virus worldwide. As no available treatment is capable of treating COVID-19, several pharmaceutical companies have initiated developing a vaccine for this

* Corresponding author.
[1] ORCID Id: 0000-0002-2705-3626.
[2] ORCID Id: 0000-0003-1071-9265.

https://doi.org/10.1016/j.matpr.2021.10.153
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the First International Conference on Design and Materials (ICDM)-2021

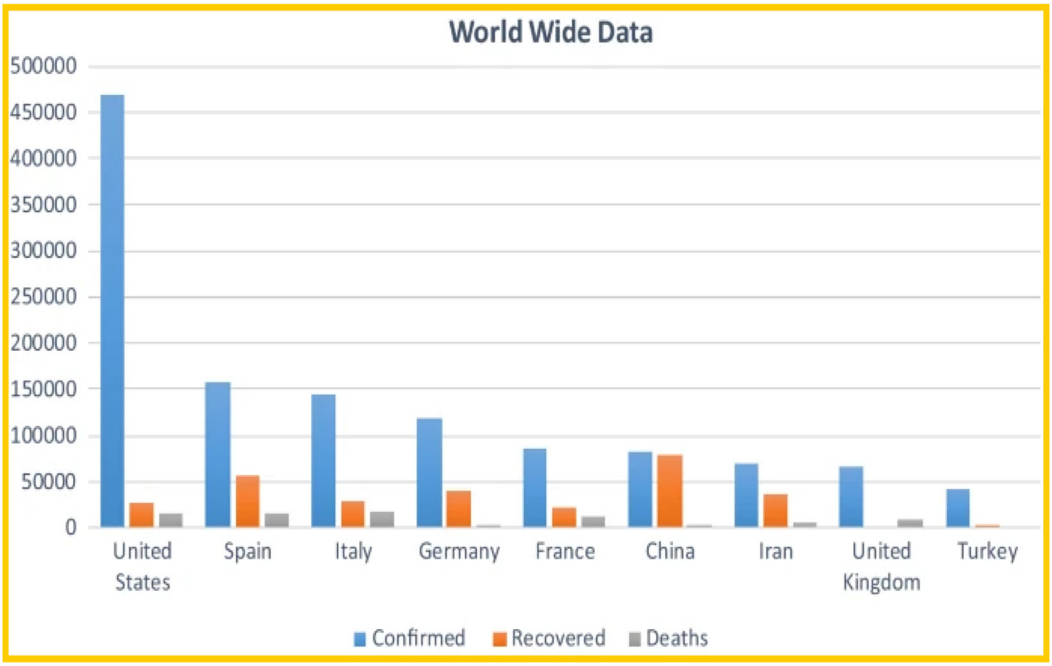

Fig. 1. Worldwide corona virus as of 10th April 2020.

virus. Lack of clinical trials and medical resources have led to the increased number of cases during the global pandemic [3].

Various data mining methods can be applied for obtaining the automatic symptom based data classification system. Major challenge faced while developing this technique is analyzing data with higher accuracy. Previously conducted studies have prediction methods to trace the incidence, magnitude, and the changing trends of the disease [4]. Markov chain models [5], Bayesian networks are examples of some of these machine learning models.

In Liu et al. [6] the Artificial Intelligence users used Machine learning to process data from sources like internet, news and health care reports to assess the magnitude of the disease in China. In Sujatha et al. [7], the researchers have suggested a method that could help predict the extent of COVID-2019, with the help of linear regression, and the Multilayer perception and Vector auto regression algorithm in order to produce data on the COVID-19 data. The Kaggle data will help in assessing the epidemiological factors related to the disease and foretell the number of infected individuals in our country, India. Nevertheless, application of machine learning will require a large amount data for classifying or predicting the disease magnitude.

This research paper is arranged according to the following sections. Section 2: Overview of literatures related to this paper. The Weighted Butterfly Optimization Algorithm (WBOA) with Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier has been explained in the 3rd section. Section 4 illustrates a comprehensive detailed performance analysis and comparisons. Conclusion of the paper is mentioned in Section 5 along with some future research proposals.

2. Literature survey

Pirouz et al (2020) has analyzed the numbers of individuals infected by the COVID disease using binary classification and Artificial Intelligence and regression analysis. According to this study, Hubei province of China was selected to design the model. Input dataset included parameters like daily temperature records, (minimum, average and maximum) humidity, wind speed and the output data set was the total COVID-19 infected cases for a period of 30 days. This model is capable of providing higher performance capacity for prediction of confirmed cases. Additionally, regression analysis is also performed by comparing the confirmed cases to the fluctuations of daily weather parameters. Results obtained indicate that factors such as humidity and maximum temperature highly affect the number of infected cases. The relative humidity effects 77.9% on average, positive cases, where as the maximum temperature, with 15.4 °C on an average negatively affect the confirmed cases [8].

Khanday et al (2020) developed machine learning method using clinical data set for determining the number of individuals infected with COVID-19. According to this study, methods like inverse document frequency or term frequency (IDF/TF) are used. Bag of words such as unigrams, bigrams were also included and were obtained. 40 parameters were identified. The parameters classifies the clinical reports as four classifications with the help of classical and ensemble machine learning algorithms. The outcomes indicated that the Logistic regression and Multinomial Naïve Bayes have better testing accuracy with 96.2% compared to other ML algorithms [9].

Celestine Iwendi et al (2020) designed a Random forest methodology for assessing and anticipating the COVID-19 parameters. This algorithm mainly includes two sections; AdaBoost and Random Forest Classifier method comprising several decision trees. The model uses information such as geographical locations, travel history, and demographic data for assuming the magnitude of disease, the outcomes, recovery from the disease and the number of deaths. This method is 94% accurate and also has a F1 Score of 0.86 on the dataset used. Statistics indicate positive association between the gender of the infected patients' and the deaths, and age. Most of the infected patients are aged between 20 and 70 years of age [10].

Ahamad et al (2020) introduced a model on the basis of supervised machine learning algorithms for determining the parameters used for predicting the magnitude of COVID-19 outbreak. The parameters included were age, sex, and recent history of fever, history of travel, and symptoms like cough and presence of lung infection. Various machine learning methods were used to collect information. The outcomes indicated that the XGBoost algorithm was the most accurate model (>85%).It was effective in predicting and selecting features that precisely point to COVID-19 status. Statistics suggested that the most important parameters were fever (41.1%), followed by cough (30.3%), lung infection (13.1%) and rhino rhea (running nose) (8.43%). However 54.4% of the total people who were assessed were non-symptomatic. This model has greatly contributed towards improving the prediction of COVID-19 prevalence, along with the initial phases of infection [11].

Jia et al. (2020) designed a trend analysis based on the data of cumulative confirmed positive cases, deaths, and treated cases on the basis of Wuhan statistics belonging to the Hubei Province of China from January 23, 2020 to April 6, 2020. The data collection was done with the help of Elman neural network, Long Short-Term Memory (LSTM), and Support Vector Machine (SVM). In order to assess the rise in the newly diagnosed COVID-19 cases, deaths and treated cases, SVM and fuzzy granulation is used. Studies conducted have proven that Elman neural network and SVM can be used to predict the growing trends infected cases, deceased and treated numbers. However, LSTM is more appropriate for predicting cumulative confirmed cases. Hence, to confirm the status of positive cases, deaths and treated cases, SVM and fuzzy granulation can be used, despite the fact that the average predicted values are large [12].

3. Proposed methodology

In this proposed research work, Weighted Butterfly Optimization Algorithm (WBOA) with Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is designed to assess covid-19 based on the symptoms. It consists of four phases such as input, preprocessing, feature selection and classification. The flow chart of the proposed method is illustrated in Fig. 2.

3.1. Input

The dataset is collected from https://www.kaggle.com/iamhungundji/covid19-symptoms-checker. The parameters that will influence the presence of COVID-19 disease in an individual are as follows:

- Country: Countries visited by the individual recently.
- Age: Classification of the individuals according to WHO age group criteria.
- Clinical manifestations: The 5 major signs and symptoms of COVID-19 according to WHO, that are Fever, General malaise, Shortness of breath, Dry cough, and throat infection.
- Other signs and symptoms: Chest pain, Nasal congestion, diarrhea.
- Severity: Severity of the disease is categorized as Mild, Moderate, Severe
- Contact: Past contact with a COVID-19 positive patient.

3.2. Data normalization

In this proposed work, input dataset is normalized by using z-score normalization. Data normalization technique is a sub division of data analysis, where the attribute data or features are scaled in order to fit in a specific range like −1.0 to 1.0, or 0.0 to 1.0. Z -score normalization is the most commonly used normalization methods. It is also known as Zero mean normalization. On the basis of mean and standard deviation, the required data is normalized. The formula is as follows,

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

Where,

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

3.3. Feature selection using Weighted Butterfly Optimization Algorithm (WBOA)

In this proposed research work, Weighted Butterfly Optimization Algorithm (WBOA) is utilized for feature selection. Butterfly Optimization Algorithm (BOA) is a novel nature based meta-heuristics similar to the mating process of butterflies. Framework for BOA is designed on the basis of the fragrance released by the butterflies, which in turn guides the other butterflies to look for food and mating partner. Butterflies act as the search agents for BOA and carry out optimization in a particular method and accomplish the task of obtaining optimal solution. Butterflies make use of the sense receptors to find their food and identify the fragrance over the body parts of butterfly such as palp, radio wires and legs. The receptors or the nerve cells on the body of butterfly is known as chemo receptors [13–15].

According to BOA, butterfly creates a fragrance that's related to its body fitness. As the butterfly migrates from one place to other, its fitness is changed as required. Other butterflies are capable of identifying it and by this process, the butterflies transmits its data to the other butterflies to form an aggregate social learning system. The stage, at which one butterfly identifies the fragrance of the other and moves towards it, is called a global search. However, in a circumstance where a butterfly is unable to recognize the fragrance, it starts moving randomly, known as local search. In BOA, every fragrance possesses a specific intriguing fragrance and individual touch. This specific feature acts as a guideline trademark that distinguishes BOA from other meta-heuristics. To understand the fragrance adjustment in BOA, it is essential to study the calculation of sound, light and temperature.

The idea of distinguishing the entire process is dependent on three basic parameters: stimulus intensity (I), sensory modality (c), and power exponent (a). According to the sensory modality method, tactile sense improves the power of the procedure and the algorithm mentions necessary information which is used by the sensors. At present the different parameters of BOA are temperature, sound and light. Modality is denoted as the fragrance emitted by the butterfly and I denote the intensity of stimulus, which is also related to the fitness of the solution/butterfly. In the proposed research work, the attributes in the dataset is considered as butterflies in the population. The classification accuracy is considered as objective function. When a butterfly releases Av, which is a strong recognizable fragrance, butterflies present in the surroundings can identify it and get attracted. An elevation in the intensity of the fragrance is called the energy or power of butterfly/ solution and a is the parameter that considers normal expression. Several research studies were conducted on bugs and other creatures to estimate the stimulus. It was assessed that as the stimulus is more grounded, insects become less impulsive towards changes in the environment. With respect to the above mentioned concepts, fragrance in BOA is calculated by:

$$\mathrm{f} = \mathrm{c} \cdot I^{a} \quad (2)$$

Here, f_i is denoted as the intensity of the emitted fragrance, i.e., grounded fragrance of the smell by i^{th} butterfly, sensory modality is represented by c, I is the objective function (classification accuracy) and power exponent reliant on modality is a, that monitors the varying levels of absorption. The two major search algorithms are: global and local search stage. Accordingly, butterflies produce

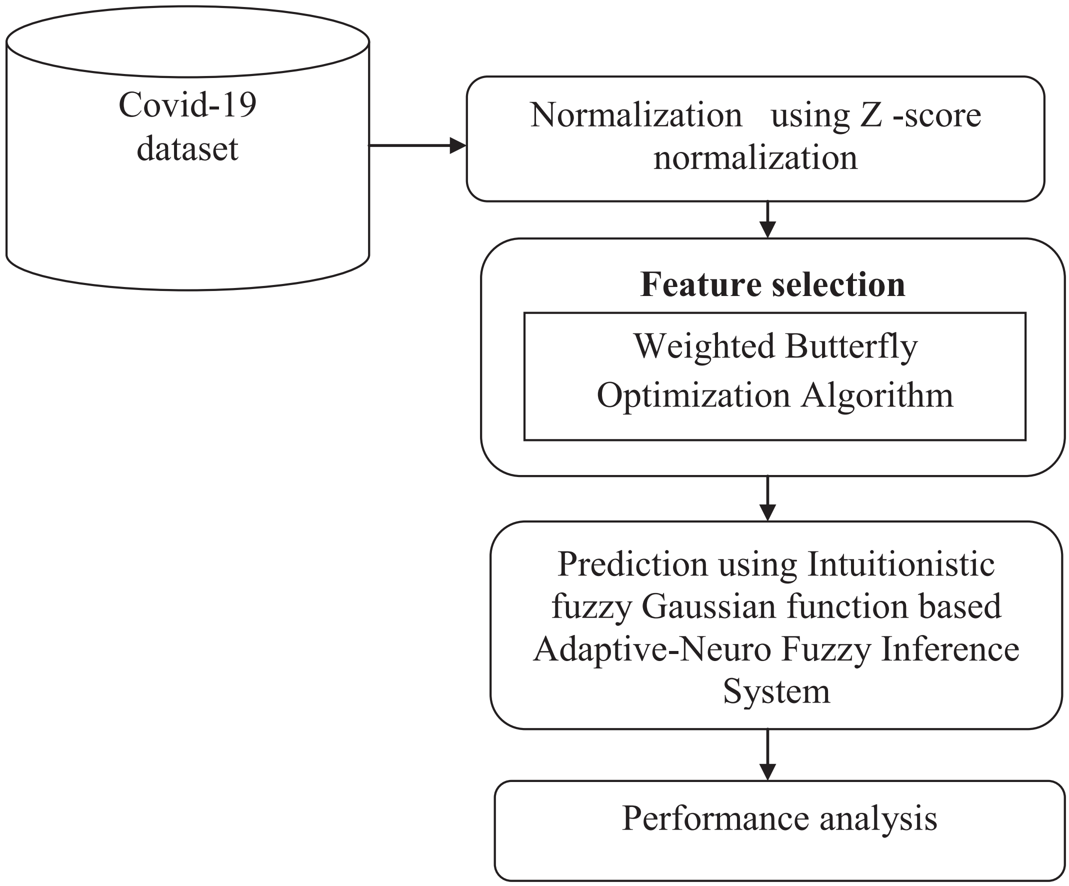

Fig. 2. Flowchart describing the proposed work.

fragrance that is identified from all surrounding places. At the initial (global) search stage, features favors the most appropriate feature g^* that is denoted by:

$$x_i^{t+1} = x_i^t + \left(r^2 \times g * - x_i^t\right) \times f_i \quad (3)$$

Where, x_i^t is the solution vector x_i for i^{th} attribute in iteration number t. $g*$ represents the most favorable solution as compared to the rest of the solutions in the current stage. f_i is fragrance emitted by the i^{th} butterfly and r stands for the random number in [0, 1]. The mathematical representation of Neighborhood (local) search stage is:

$$x_i^{t+1} = x_i^t + (r^2 \times x_j^t - x_k^t) \times f_i \quad (4)$$

Where, x_j^t and x_k^t are j^{th} and k^{th} attribute from the search space. In case when x_j^t and x_k^t has a place with a similar swarm and r is a random number in [0,1] then Eq. (4) turns into a neighborhood random walk. The need and search for nourishment and mating are possible in both neighborhood and worldwide scale. Hence a switch p is used to change from the normal worldwide quest and work on the local search.

3.4. Weighted Butterfly Optimization (WBO) Algorithm

Providing balance between exploration and exploitation process is mainly done by Inertia weight. Inertia weight is directly proportional to the global exploration ability of the algorithm, enabling it to assess new unknown areas. In this proposed research work, inertia weight and global exploration of BOA is combined to enhance the exploration ability. α stands for the adjusting factor that improves the efficiency of the algorithm. For the iteration, update Natural Exponent Inertia Weight which is as follows:

$$\mathrm{w(t)} = w_{min} + (w_{max} - w_{min}).e^{-\left(\frac{\left[\frac{t}{MAXITER}\right]}{10}\right)} \quad (5)$$

Where,

w_{max} - maximum weight value
w_{min} - minimum weight value
t-number of the tth iteration
N –maximum number of iterations

In favor of global exploration, mathematical equation (3) is rewritten as follows:

$$x_i^{t+1} = \mathrm{w(t)} * \alpha * x_i^t + (r^2 \times g^* - x_i^t) \times f_i \quad (6)$$

$$\alpha = \frac{S_{max}}{t^2} \quad (7)$$

Where,

S_{max} - max walk step

Algorithm 1: Weighted Butterfly Optimization (WBO) Algorithm

1. Objective function (Classification accuracy) f(x)
2. Initialize the features in the dataset x_i = (i = 1, 2… n)
3. Compute classification accuracy
4. Define sensor modality c, power exponent a and switch probability p
5. while stopping criteria not met do
6. for each features in the dataset do
7. Calculate fragrance for *bf*

end for
Find the best *bf*
for each feature in the dataset
Generate a random number *r* from [0, 1]
if *r* < *p* then
Move towards best solution using Eq. (6)
else
Move randomly using Eq. (4)
end if
end for
Update the value of *c*
end while
Output the best optimal features

3.5. Classification using Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier

Analyzing symptoms like fever, tiredness, dry-cough, sore throat and breathing problem a Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is utilized for classifying the people having infection possibility. The ANFIS is a neural network that functions on the basis of neuro fuzzy network. Because the ANFIS is an adaptive network, element of its nodes are adaptive, which means with the purpose of their outputs based on the parameters fit in to these nodes [16,17]. Fig. 3 illustrates the architecture of ANFIS model.

Various rules have different output membership function and these numbers must be the same. In order to carry out the ANFIS architecture, two fuzzy IF-THEN rules on the basis of first order Sugeno model are considered:

$Rule(1) : IFxisA_1ANDyisB_1, THEN$

$f_1 = p_1x + q_1y + r_1$

$Rule(2) : IFxisA_2ANDyisB_2, THEN$

$f_2 = p_2x + q_2y + r_2$

Where,

The inputs are x and y
Fuzzy sets are A_i and B_i
f_i - outputs within the fuzzy region specified by the fuzzy rule
p_i , q_i , and r_i , are the design parameters which are determined at the training process.

Layer 1: First layer consists of adaptive nodes. Here, the selected features are taken as an input and it is given to first layer. The outputs of Layer 1, fuzzy membership grade input is denoted by the given mathematical representation:

$$O_{1,i} = \mu_{Ai}(\mathrm{x})\ \upsilon_A(\mathrm{x})-,\ \mathrm{i} = 1, 2, \tag{8}$$

$$O_{1,i} = \mu_{Bi-2}\ (\mathrm{y})\ \upsilon_A(\mathrm{y}),\ \mathrm{i} = 3, 4, \tag{9}$$

Here, x and y represent the inputs for node i, where as A_i and B_i represent the linguistic labels (high, low, etc.), that are related to the node function. $\mu_{Ai}(x)$ and $\mu_{Bi-2}(y)$ are adaptable to all membership functions. In this proposed work, Intuitionistic Fuzzy Gaussian Function (IFGF) is designed for member function computation. The IFGF is specified by two parameters. The Gaussian function is represented by a central value m and width k > 0. As the k value decreases, the curve gets narrower. Intuitionistic fuzzy Gaussian membership and non-membership functions are defined as

$$\mu_A(x) = exp\left(\frac{(x-m)^2}{2(k)^2}\right) - \epsilon \tag{10}$$

$$\upsilon_A(x) = 1 - \left(exp\left(-\frac{(x-m)^2}{2(k)^2}\right)\right) \tag{11}$$

Here,

μ_A (x) is the membership function
υ_A (x) is the non membership function

The diagrammatic representation of intuitionistic fuzzy Gaussian function is shown in Fig. 4.Fig. 5.Fig. 6.

Layer 2: Layer 2 has fixed nodes. It uses the fuzzy operators along with AND operator to fuzzify the inputs. This is denoted by $\prod$, performing a simple multiplier. Result obtained from this layer is denoted by:

$$O_{2,i} = w_i = \left(\mu_{A_i}(x)\upsilon_A(x)\right)(\mu_{B_i}(y)\upsilon_B(y)),\quad i = 1, 2 \tag{12}$$

These are the firing strengths of the rules.

Layer 3: The third layer consists of fixed nodes termed as N. The nodes carry out normalization to the firing strengths obtained from the second layer. The result obtained is represented by:

$$\mathrm{O}_{3,i} = \bar{\mathrm{w}}_i = \frac{\mathrm{w}_i}{\mathrm{w}_1 + \mathrm{w}_2}, \mathrm{i} = 1, 2 \tag{13}$$

Outputs produced in layer 3 are known as normalized firing strengths (weight). The w_1 and w_2 are weight value of the attributes.

Layer 4: Nodes of this layer are very adaptive. Output produced by every node of the fourth layer is obtained by the product of the normalized weight and a first order polynomial (for a first order Sugeno model) which is represented by:

$$\mathrm{O}_{4,i} = \bar{\mathrm{w}}_i f_i = \bar{\mathrm{w}}_i(p_i x + q_i y + r_i), i = 1, 2, \tag{14}$$

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

Layer 5:5th layer comprises of a single fixed node that is labeled P, that summarizes all incoming signals. Output generated by the Layer 5 is represented by:

$$\mathrm{O}_{5,i} = \sum_i \bar{\mathrm{w}}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{15}$$

The covid -19 probability prediction are the output of the layer 5.

4. Experimental results

The performance of the proposed work is evaluated using java. The dataset is collected from https://www.kaggle.com/iamhungundji/covid19-symptoms-checker. The performance of the proposed Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is compared with the previous XGBoost and Support Vector Machine (SVM) methods interms of accuracy, precision, recall and f-measure. Table 1 represents the overall performance comparison.

4.1. Accuracy

Accuracy is the most intuitive performance measure that correctly classifies the instance of occurrence and it is simply a ratio of predicted correct observation to the total observations.

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{16}$$

Where,

TP - True Positive

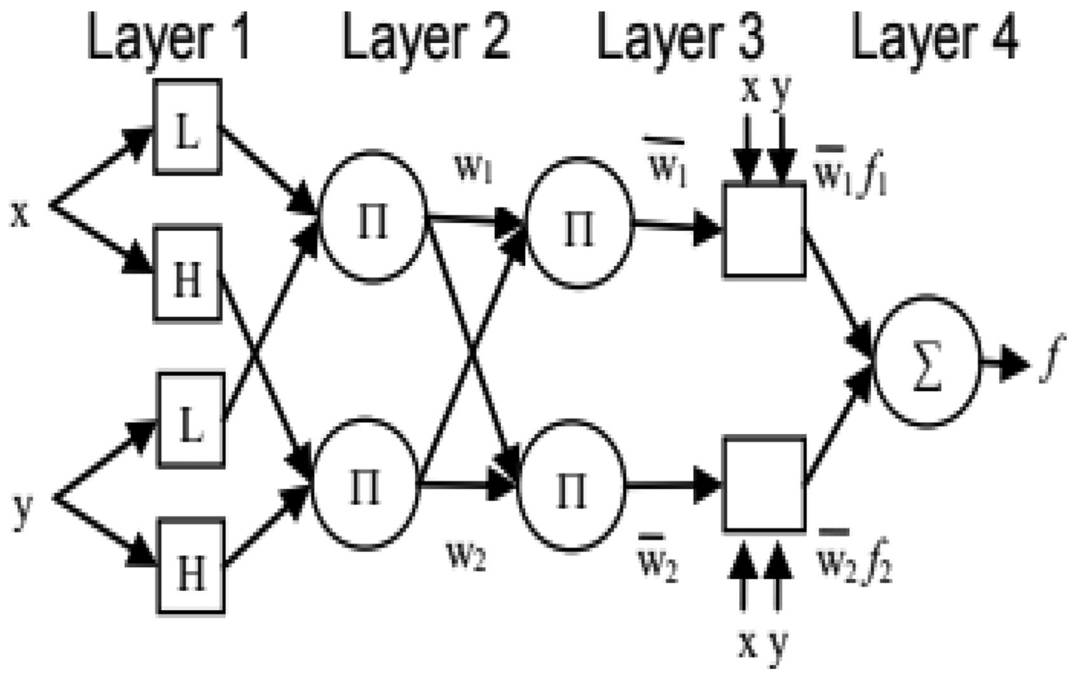

Fig. 3. Architecture of Adaptive Neuro Fuzzy Inference System (ANFIS).

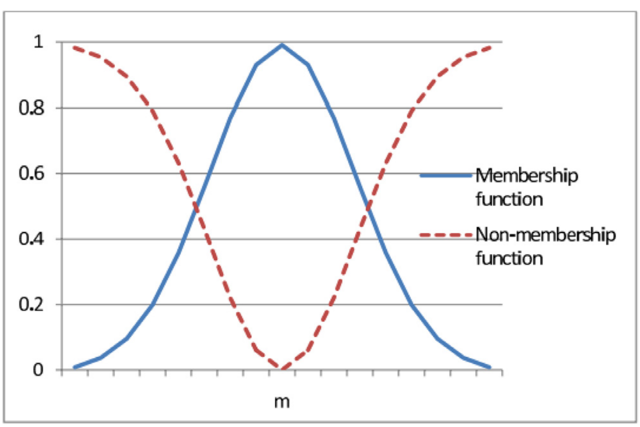

Fig. 4. Intuitionistic fuzzy Gaussian function.

FN - False Negative
FP - False Positive
TN- True Negative

The accuracy of the proposed Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is compared with the previous XGBoost and Support Vector Machine (SVM) methods. In x-axis methods are taken and accuracy is taken as y-axis. In this proposed work, optimal features are selected by using Weighted Butterfly Optimization Algorithm (WBOA). It improves the accuracy rate. The experimental results show that the proposed system attains 93% of accuracy whereas other method such as XGBoost and Support Vector Machine (SVM) achieves 87% and 91% respectively.

4.2. 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.

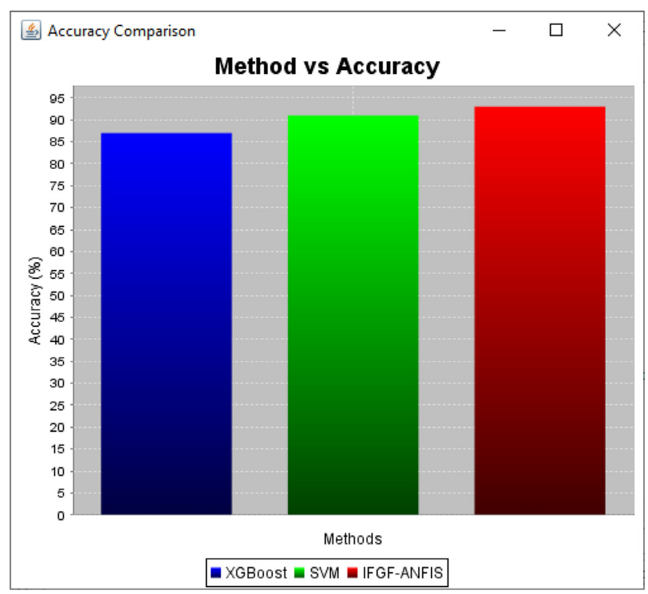

Fig. 5. Accuracy comparison.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

The performance of the proposed Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is compared with the previous XGBoost and Support Vector Machine (SVM) methods in terms of precision. In x-axis methods are taken and precision is taken as y-axis. In this proposed research work, classification is done by using Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier. In IFGF-ANFIS, Intuitionistic Fuzzy Gaussian Function (IFGF) is designed for member function computation. It improves the true positive rate. The experimental results show

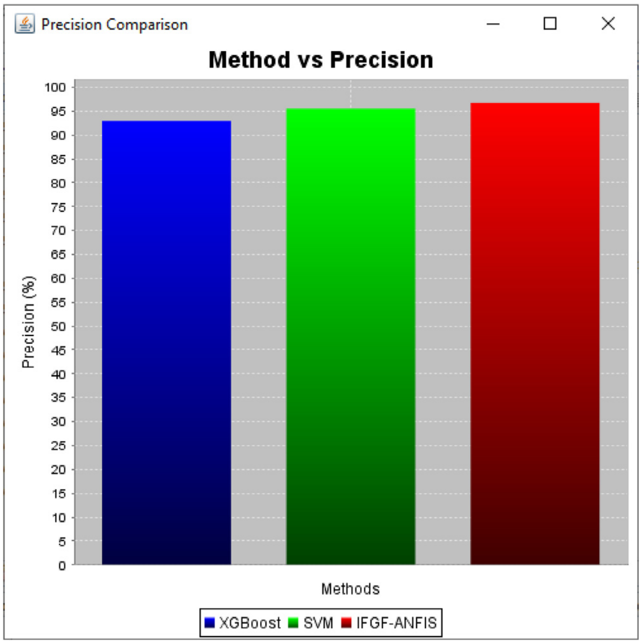

Fig. 6. Precision comparison.

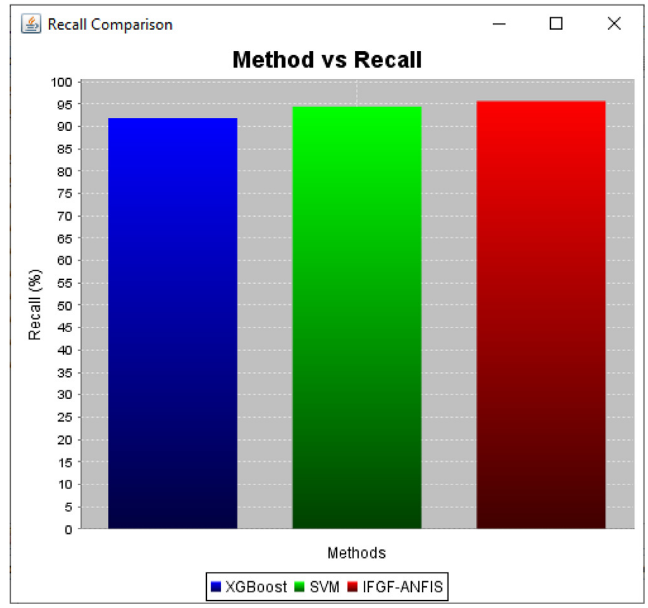

Fig. 7. Recall comparison.

Table 1
Performance comparison.

Metrics	Methods		
	XGBoost	SVM	IFGF-ANFIS
Accuracy	87	91	93
Precision	92.94	95.51	96.7
Recall	91.86	94.44	95.65
F-measure	92.4	94.97	96.17

that the proposed system attains 96.7% of precision whereas other methods such as XGBoost and Support Vector Machine (SVM) achieve 92.94% and 95.51 % respectively.

4.3. Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

Fig. 7 show the recall of the proposed Intuitionistic Fuzzy Gaussian Function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier and previous XGBoost and Support Vector Machine (SVM) methods. In x-axis methods are taken and precision is taken as y-axis. From the experimental results, it can be concluded that the proposed system attains 95.65% of recall where as other methods provides 91.86% and 94.44% respectively.

4.4. F-measure

F1 Score is also a measure of accuracy of experiment and is defined by the weighted mean of Precision and Recall. Therefore, this score takes account of both false positives and false negatives.

$$\text{F-measure} = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \tag{19}$$

Fig. 8 shows the f-measure of the proposed Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier and previous XGBoost and Support Vector Machine (SVM) methods. In x-axis methods are taken and f-measure is taken as y-axis. From the experimental results, it can be concluded that the proposed system attains 96.17% of f-measure where as other methods provides 92.4% and 94.97% respectively.

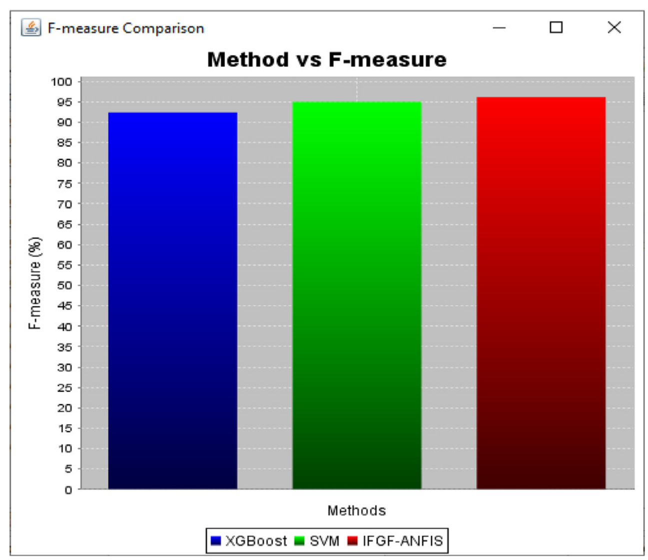

Fig. 8. F-measure comparison.

5. Conclusion

The proposed system is designed with a Weighted Butterfly Optimization Algorithm (WBOA) with Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference System (IFGF-ANFIS) classifier is designed for covid-19 prediction. In order to perform normalization, the Z score normalization method is utilized. In order to enhance the precision of prediction, Weighted Butterfly Optimization Algorithm (WBOA) is used for selecting optimal attributes. Based on the selected features, Intuitionistic fuzzy Gaussian function based Adaptive-Neuro Fuzzy Inference

System (IFGF-ANFIS) classifier is utilized for classifying the people having infection possibility. The results obtained by the experiment illustrate that this proposed system attains superior performance when compared to the conventional system in terms of accuracy, precision, recall and f-measure. In future, deep learning methods can be explored to achieve improved prediction accuracy.

CRediT authorship contribution statement

T. Sundaravadivel: Validation, Visualization, Writing - original draft. **V. Mahalakshmi:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.A. Achterberg, B. Prasse, L. Ma, S. Trajanovski, M. Kitsak, P. Van Mieghem, Comparing the accuracy of several network-based COVID-19 prediction algorithms, Int. J. Forecast. (2020), https://doi.org/10.1016/j.ijforecast.2020.10.001.

[2] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, S. Alhyari, COVID-19 prediction and detection using deep learning, Int. J. Comput. Inf. Syst. Ind. Manage. Appl. 12 (2020) 168–181.

[3] Saumendra Kumar Mohapatra, Mohan Debarchan Mohanty, Mihir Narayan Mohanty, Corona virus infection probability classification using support vector machine, Int. J. Adv. Sci. Technol. 29 (8s) (2020) 3093–3098.

[4] Y. Chen, C.W. Chu, M.I.C. Chen, A.R. Cook, The utility of LASSO-based models for real time forecasts of endemic infectious diseases: a cross country comparison, J. Biomed. Inform. 81 (2018) 16–30.

[5] C.P. Jewell, T. Kypraios, R.M. Christley, G.O. Roberts, A novel approach to real-time risk prediction for emerging infectious diseases: a case study in Avian Influenza H5N1, Prev. Veter. Med. 91 (1) (2009) 19–28.

[6] Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J.T., ... Santillana, M. (2020). A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. arXiv preprint arXiv:2004.04019.

[7] R. Sujath, J.M. Chatterjee, A.E. Hassanien, A machine learning forecasting model for COVID-19 pandemic in India, Stoch. Env. Res. Risk Assess. 34 (7) (2020) 959–972.

[8] B. Pirouz, S. Shaffiee Haghshenas, S. Shaffiee Haghshenas, P. Piro, Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (new type of corona virus) through a binary classification using artificial intelligence and regression analysis, Sustainability 12 (6) (2020) 2427.

[9] A.M.U.D. Khanday, S.T. Rabani, Q.R. Khan, N. Rouf, M. Mohi Ud Din, Machine learning based approaches for detecting COVID-19 using clinical text data, Int. J. Inf. Technol. 12 (3) (2020) 731–739.

[10] C. Iwendi, A.K. Bashir, A. Peshkar, R. Sujatha, J.M. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, O. Jo, COVID-19 patient health prediction using boosted random forest algorithm, Front. Public Health 8 (2020) 357.

[11] M.M. Ahamad, S. Aktar, M.d. Rashed-Al-Mahfuz, S. Uddin, P. Liò, H. Xu, M.A. Summers, J.M.W. Quinn, M.A. Moni, A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients, Expert Syst. Appl. 160 (2020) 113661.

[12] L. Jia, K. Li, Y. Jiang, X. Guo, Prediction and analysis of Coronavirus Disease 2019. arXiv preprint (2020).arXiv:2003.05447.

[13] S. Arora, S. Singh, Butterfly optimization algorithm: a novel approach for global optimization, Soft. Comput. 23 (3) (2019) 715–734.

[14] M. Ghetas, C.H. Yong, P. Sumari, (2015, November). Harmony-based monarch butterfly optimization algorithm, n: *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 156-161). IEEE.

[15] G.-G. Wang, S. Deb, Z. Cui, Monarch butterfly optimization, Neural Comput. Appl. 31 (7) (2019) 1995–2014.

[16] V.S. Ghomsheh, M.A. Shoorehdeli, M. Teshnehlab, (2007, June). Training ANFIS structure with modified PSO algorithm, in: *2007 Mediterranean Conference on Control & Automation* (pp. 1-6). IEEE.

[17] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey, Artif. Intell. Rev. 52 (4) (2019) 2263–2293.