

Mobile Application to Detect Covid-19 Pandemic by Using Classification Techniques: Proposed System

<https://doi.org/10.3991/ijim.v15i16.24195>

Azhar F. Al-zubidi¹(✉), Nadia F. AL-Bakri¹, Rajaa K. Hasoun²,
Soukaena Hassan Hashim³, Haider TH Salim Alrikabi⁴

¹AL Nahrain University, Baghdad, Iraq

²University of Information Technology and Communication, Baghdad, Iraq

³University of Technology, Baghdad, Iraq

⁴Wasit University, Wasit, Iraq

azhar.flaih@nahrainuniv.edu.iq

Abstract—Various mobile applications such as Mobile Health (mHealth) have been developed and spread across the world which has played an important role in mitigating the Coronavirus pandemic (COVID-19). As the COVID-19 pandemic spreads, several people have drawn parallels to influenza. While both viruses cause respiratory infections, they propagate in very different ways. This has a major impact on the public health measures that can be used to fight each virus. These viruses are pandemic-causing in the same way. That is, they both cause respiratory disease, and can present themselves in several ways, ranging from asymptomatic to severe and deadly. A proposal is presented in this paper that uses two algorithms to define and classify these pandemics, they are: The Back Propagation (BP) classification algorithm and the Fuzzy C-Mean (FCM) clustering algorithm. Two stages are implemented in the proposed system: in the first step, the FCM algorithm is used to find out the type of virus, and this algorithm is capable of handling ambiguous features of viruses. In the second step, a BP neural network is used as a classifier to detect the pandemic class. The proposed system was trained and tested using a well-known dataset (covid-19 vs influenza). Information Gain (IG) is used to optimize the related features that affect the classification process to improve speed and accuracy. The proposed mobile application is developed to support users easily detecting the COVID-19 infection by inputting the medical tests as significant features to the proposed system. The proposed system's accuracy is up to (89%), the framework was created using the Matlab programming environment and an Android Studio for Mobil application designing.

Keywords—COVID-19, Fuzzy C-Mean (FCM), propagation (BP) classification, information gain (IG), mobil application

1 Introduction

In different countries, every day, the number of cases of a new coronavirus (Covid-19) is increasing. COVID-19 is an infectious pandemic that has frightened the world and continues to put billions of people's lives in danger. The identification of coronavirus (COVID-19) has recently become a vital activity for medical practitioners. COVID-19, unfortunately, spreads. It has spread so quickly among people that it has now reached millions of people all over the world in just a few months. It is important to rapidly and reliably classify infected people to avoid the spread of the pandemic [1, 2]. To assess policies and strategies, the analysis of distributions of the virus's spread, and the study of the relationships between the features that have ratios. In this analysis, the distributions of Covid-19 or influenza virus spread were correlated and clustered using the fuzzy clustering technique. Despite the use of several diagnostic methods to identify individual injuries, the diagnosis reliability has yet to be achieved. In this post, the inputs are combined using the FCM clustering algorithm. Clustering's major part, according to news studies, is to split a dataset into clusters that represent similar objects in one cluster and dissimilar objects in another [3]. Back Propagation (BP) artificial neural networks can address several problems that other current intrusion detection techniques can't. There are three advantages of using a neural network to detect intrusions: In the intrusion detection process, neural networks provide elasticity by allowing the neural network to analyze data that is either entirely correct or partially correct. Similarly, a neural network that can perform nonlinear data analysis is useful. Neural networks have the potential to interpret data in a non-linear fashion from a variety of source [4–6]. Contact tracing was one of the first COVID-19 technologies to be industrialized and widely publicized. It was created to alert individuals whether they had crossed paths with another person infected with the coronavirus. In Singapore, the first national smartphone application for touch tracing was developed using Bluetooth technology. In response to COVID-19, mobile applications for symptom tracking have also appeared. These apps usually gather information about a user's health by presenting a list of questions about symptom detection, from which a distinct diagnosis is derived. Other sophisticated methods, such as automated assembly of the user's health data (such as temperature and pulse rates) from wearable like wristbands, have also been used. In the event of a suspected COVID-19 infection, the user is informed and directed to a nearby clinic for a checkup [7–11].

2 Related work

Now will present some of the related works we are standing on it to introduce the proposal of detecting COVID-19.

1. In [12], this research aims to create a fuzzy supported method to evaluate the safety of dental treatment concerning a variety of patient and environmental factors. For dental care providers, a fuzzy framework may be used as a diagnostic method to make diagnoses, identify and determine risks, and improve interventions. As a

result of this research, the fuzzy method will measure and classify the risk of dental treatment based on the patient's health status and dental hygiene conditions.

2. In [13], this work introduced an artificial intelligence paradigm that seeks to solve new cases by using COVID-19's vast database of cases. In this analysis, the researchers used an improved CBR model for a state-of-the-art reasoning challenge in the classification of alleged COVID-19 instances. The results showed that the suggested method in this work correctly categorized suspicious cases into their categories with a 94.54 percent accuracy.
3. In [14], based on previously reported COVID-19 cases identified in China, The number of confirmed COVID-19 cases over the next ten days is estimated and expected using new prediction models presented in this article. The suggested model is an enhanced flower pollination algorithm (FPA) and an advanced adaptive neuro-fuzzy inference approach (ANFIS) based on the salp swarm algorithm (SSA).
4. In [15], this article, the growing spectrum of confirmed COVID-19 cases was estimated using an SVM with fuzzy granulation. By showing that the Elman neural network and SVM used in this analysis can forecast the progression of combined reported cases, deaths, and healed cases, while the LSTM is best suited for cumulative confirmed case prediction. The SVM with fuzzy granulation will effectively forecast the growth continuum of recorded new cases and newly healed cases, despite the average expected values being marginally high.
5. In [16], this post would go into a Dual Diagnostic Strategy that has been suggested in detail (HDS). The primary objective of HDS is to classify COVID-19 incidents easily and reliably. Early diagnosis of COVID-19 cases allows for prompt care and separation of patients, which slows the transmission of the pandemic. HDS is given a training series consisting of laboratory findings from COVID-19 and non-COVID-19 individuals. Since the model has learned to recognize trends, new cases can be added to it (including laboratory findings). HDS determines whether or not the input case has COVID-19 infection.
6. In [17], a review was published concerning a systematic search strategy to identify the free mobile applications linked to the COVID-19 pandemic in the App and Google Play stores. Various applications have been created for various functions such as contact tracing, knowledge building, appointment booking, online consultation. a few applications have incorporated functions and features like self-assessment, consultation, assistance, and information access, designing and developing an integrated mobile health application that incorporates the majority of the features and functionalities.

In this work, a symptom monitoring application in the mobile phone as a web-based COVID-19 screen tool application is developed. A proposal system starts with the FCM algorithm, which is used to cluster diseases. Because of its intrinsic processing speed and ease of representation of nonlinear relationships between input and output, the BP algorithm is used to define the class of pandemic. A neural network will be capable of analyzing data from a network and providing the ability to distinguish and differentiate network behavior based on limited, imperfect.

3 Data mining techniques

Various data mining techniques, such as Clustering, Classification, and Information Gain used for feature selection. The two data mining techniques used in the proposed study are illustrated in this section.

3.1 Fuzzy C Mean (FCM) algorithm

Based on degree affiliation values, the fuzzy clustering algorithm splits data into separate clusters, each cluster contains data that is more similar to one another while still being different from data in other clusters. Soft clustering occurs when data points in the fuzzy c-means algorithm belong to more than one cluster with some degree of membership. The data is organized into clusters using a fuzzy participant ship function. [18 19] The fuzzy clustering algorithm estimates using equation (1) and is dependent on minimizing objective J . This is focused on the entity function minimization equation J (1).

$$J_m(U, C) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m d_{ij}^2(x_i, c_j) \quad (1)$$

Where m is the domain's actual number ($1 \leq m < \infty$). k is cluster number. n is the data sample number. u_i is the degree of membership that indicates the likelihood that the data sample will be used. x_i belongs to j^{th} Cluster. And c_j is the center of the cluster.

Using equations, repeatedly modify cluster center c_j and fuzzy membership u_{ij} to achieve fuzzy clustering (2) (3).

$$u_{ij} = \frac{1}{\sum_{j=1}^k (d_{ij}/d_{ik})^{2/m-1}}, \quad \forall i \quad (2)$$

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m}, \quad \forall i \quad (3)$$

Where u_{ij} denoted the degree of membership of data samples that belonged to a particular cluster and met the following criteria (see equations (4) and (5)) [20, 21]:

$$\sum_{j=1}^k u_{ij} = 1 \quad \forall k \quad (4)$$

$$\sum_{j=1}^k u_{ij} > 0 \quad \forall i \quad (5)$$

3.2 Back propagation NN

Backpropagation is a classification method in which the algorithm is told to validate the network using input and output tests. It's a multilayer feed-forward neural network

made up of layers of neurons with one output layer and one or more hidden layers. The forward step, in which information is applied and propagated to the output layer, and the information is extended and propagated to the input layer in the backward step. In the two stages of backpropagation learning, the error is measured and the weight is modified to reduce the error, allowing the ANN to learn the data [22, 23]. Weights between the input and hidden layers multiply the input of each input layer neuron in the input layer. Using equation (6), each hidden neuron (j) in the hidden layer obtained the value $Z_j(j)$.

$$z_j = \theta + \sum_{i=1}^n x_i w_{ij} \quad (6)$$

Equation (7) the hidden layer's output is processed using the binary sigmoid function activation:

$$f(x) = \frac{1}{1 + \exp^{-x}} \quad (7)$$

Equation (8) find the hidden layer output:

$$y_k = \theta + \sum_{j=1}^p z_j w_{jk} \quad (8)$$

$$y_k = f(y_k)$$

θ = bias.

The mean square error's worth E is a measure that can be used to evaluate how much you've learned. If the MSE is less than (0.1), the net has learned its training package. In the following equation (9) E was calculated as the mean square error value:

$$E = \frac{1}{2} \sum_p \sum_k (T_{pk} - Y_{pk})^2 \quad (9)$$

Where the p is Samples number, Y_{pk} is the actual output, and T_{pk} is target output.

If the network output in the backward step differed from the target output, the output error is measured, and the error is then sent to the input layer, changing the weights of the layers' neurons. Equation (10) can be used to calculate the error between the output and hidden layer:

$$\delta_{2k} = y_k(1 - y_k)(T_k - y_k) \quad (10)$$

Equation (11) can be used to calculate the error between the secret and input layers:

$$\delta_{1j} = z_j(1 - z_j) \sum_{k=1}^m \delta_{2k} w_{jk} \quad (11)$$

Using the equations (12) and (13) below, To reduce the error, the weights are adjusted:

$$w_{jk}(\text{new}) = \eta * \delta_{2k} * z_j + \infty * w_{jk}(\text{old}) \quad (12)$$

$$w_{ij}(\text{new}) = \eta * \delta_{1j} * x_i + \infty * w_{ij}(\text{old}) \quad (13)$$

Where w_{jk} = the weight differential between the hidden and output layers, w_{ij} = the weights differential between the hidden layers and the input layers, η = Learning rate, and ∞ = the MC [24, 25].

3.3 Information gain (IG)

The IG assesses characteristics by determining how much knowledge they receive concerning the class. Let C be a series of c data samples divided into m distinct groups. A sample of class I can be found in the training dataset C_i . The expected data used to identify a sample is determined [26–28].

$$I(c_1, c_2, c_m) = - \sum_{i=1}^m \frac{c_i}{c} \log_2 \left(\frac{c_i}{c} \right) \quad (14)$$

C_i is the set's subset that contains the value f_i for feature F . Let C_j contain C_{ij} class sample i . F is the entropy of the function.

$$E(F) = \sum_{j=1}^v \frac{C_{ij} + \dots + C_{mj}}{C} \times (C_{ij} + \dots + C_{mj}) \quad (15)$$

F 's information gain can be estimated as follows:

$$Gain(F) = I(C_1, \dots, C_m) - E(F) \quad (16)$$

4 COVID-19 pandemic detection

The coronavirus (COVID-19) is a highly infectious and disease-causing infection caused by the extremely essential breathing condition coronavirus 2 (SARS-CoV-2), which first appeared in Wuhan, China, and then spread across the world [29]. Figure (1) presents the structure and symptoms of this disease.

According to the World Health Organization, the COVID-19 diagnostic test is important for monitoring the infection, understanding the epidemiology, notifying case supervision, and stopping transmission. To classify the COVID-19 virus, diagnostic testing is carried out using a variety of in-house and profitable analyses. Tasters may be inhaled through the nose or the back of the mouth. A sample from the lower respiratory tract can provide the best results for hospitalized patients. Antigen testing reveals whether anyone is infected with Covid-19 and therefore can spread it to others. Antibody studies, on the other hand, use blood samples to determine the immunity induced by the previous infection. Antibody test kits look for proteins that the virus uses as “glue” to trap antibodies in the blood [30–32].

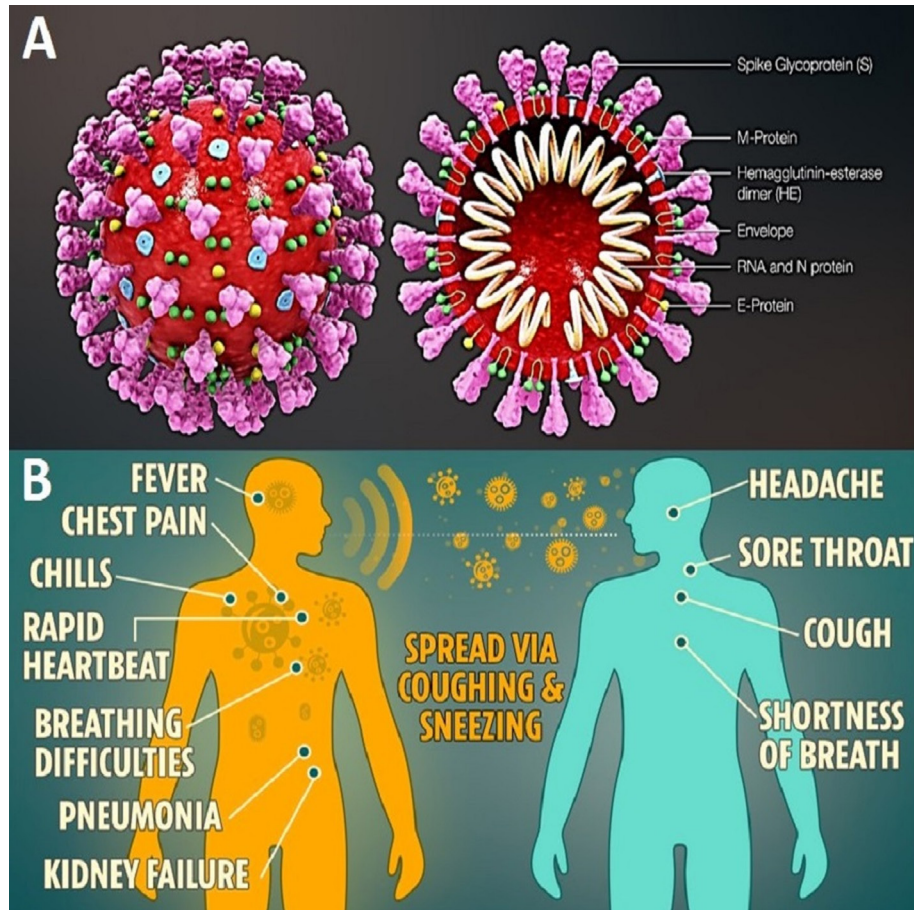


Fig. 1. (A) Corona Virus Structure. (B) COVID-19 Symptoms and its spreads

5 Mobile application covid-19 detection: proposed system

The following sections will present the details of the proposed system over four sections; general description, pre-processing the dataset, feature selection, Clustering using fuzzy c means, and classification using backpropagation neural networks are the primary approaches [33, 34]. Before describing the proposal of detection, the proposal structure collecting the symptoms is described as shown in Figure 2. In our model, there are eight sensor nodes on the human body. These sensor nodes all start with the same amount of resources, computation, and storage. The sink node is located in the network's middle to offset the network's energy usage [35].

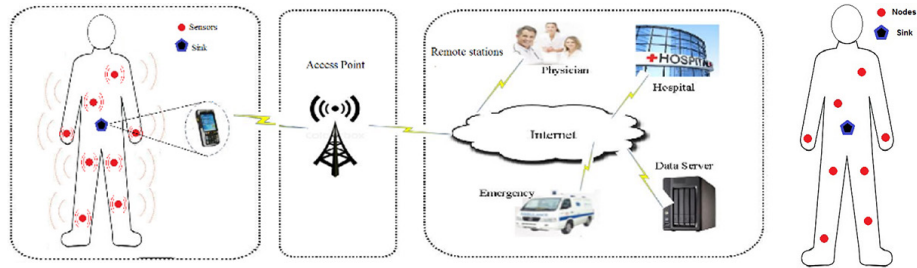


Fig. 2. Structure of mobile application

5.1 General description of the proposal

This paper aims to develop a Fuzzy C Means Clustering Neural Network (FCMNN) algorithm as pattern classifiers. The proposed system applied to cluster the symptoms features of patents like (Disease group, WBC, Neutrophil, Lymphocyte, Monocyte, C-reactive protein, Severity, Number of UMI, Number of Gene, Percentage of mitochondrial gene) to either influenza or covid-19 viruses from dataset to detect the class (type) of the virus. The (FCMNN) algorithm consists of four consequence steps explain as follows:

1. First step: Database building was a subset of data selected from the covid-19 dataset. After dataset selection Pre-processing operation conduct on the dataset to transform it into an appropriate form for further processing.
2. Second step: Information Gain (IG) algorithm used to find the relevant features form irrelevant feature, where coved19 dataset consists of 52 features that increase the computational cost and time of training and testing. This step reduces the computational cost and increase the speed of the proposed system make the proposed system highlight.
3. Third step: The Fuzzy C-Mean (FCM) algorithm was used to create two clusters based on the type of symptom (covid-19 or influenza).
4. Fourth step: in this final step backpropagation algorithm is used to build a classifier to classify the class of pandemic.

The following block diagram, see Figure 3, and present the details of the proposed system through four sections;

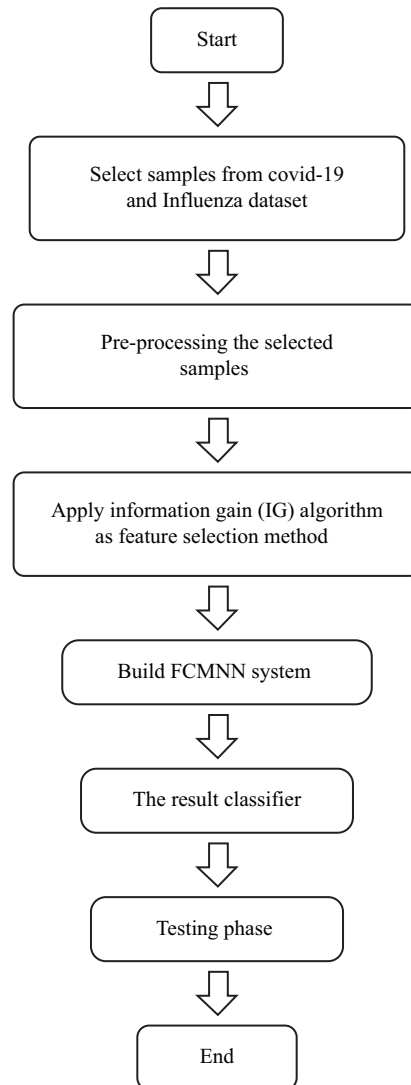


Fig. 3. The proposed system (FCMNN)

5.2 General description of the dataset

The dataset has been standardized for it to be suitable for use by the proposed algorithm. The dataset's features are divided into numeric and symbolic features, which are categorized into the following groups (Disease group, WBC, Neutrophil, Lymphocyte, Monocyte, C-reactive protein, Severity, Number of UMI, Number of Gene, Percentage of mitochondrial gene, and the class Disease condition). The phases of the standardization process are as follows:

A. Convert a symbolic feature’s value to a sequential integer value. Since the fuzzy c- mean algorithm and the backpropagation algorithm take numerical values into account. In the table, the symbolic function and its integer value are shown in Table 1.

Table 1. A symbolic feature of numeric value for the covid-19 dataset

Disease group	Feature value	Severity	Feature value	Pandemic condition	Feature value
Mild COVID-19	1	Mild	1	Asymptomatic	1
Mild COVID-19 (asymptomatic)	2	Mild (asymptomatic)	2	COVID-19 (Mild)	1
Severe COVID-19	3	Severe	3	COVID-19 (severe)	2
Severe influenza	4			flu	1
				normal	1

B. Data to prevent the bias that larger feature values can create, normalization is used. As a result, mining algorithms have become more efficient and accurate. When the data to be analyzed falls between [0, 1], these algorithms produce better results. A linear transformation called the Min-Max normalization method is used to scale data between [0, 1]. The new value was calculated using the following formula:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{17}$$

$$X_n = \frac{12900 - 2800}{22160 - 2800}$$

$$X_n = 0.5216$$

Figures 4 and 5 show the dataset before and after normalization.

1	Disease group	WBC/micro	Neutrophil/micro. (%)	Lymphocyte/micro. (%)	Monocyte/mic	C-reactiv	Severi	Number of UI	Number of Gr	Percentage of mi	Diseas
2	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	15679	3311	7.296383698	COVID
3	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	3630	1146	13.16804408	COVID
4	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	6796	1785	3.546203649	COVID
5	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	6803	2147	7.393796854	COVID
6	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	819	332	2.31990232	COVID
7	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	1188	488	5.303030303	COVID
8	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	5444	1958	5.180014695	COVID
9	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	11649	2946	9.228259936	COVID
10	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	4568	1654	11.2084063	COVID
11	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	6303	2078	6.139933365	COVID
12	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	5596	1932	9.810578985	COVID
13	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	7130	2178	7.13884993	COVID
14	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	8174	2472	7.022265721	COVID
15	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	4037	1391	10.10651474	COVID
16	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	1147	556	3.748910201	COVID
17	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	1231	697	11.61657189	COVID
18	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	4264	1589	6.167917448	COVID
19	severe COVID-19	21540	19235 (89.3)	1055 (4.9)	1271 (5.9)	7.58	severe	5319	1772	6.843391615	COVID

Fig. 4. Dataset before normalization

	B	C	D	E	F	G	H	I	J	K
1	Disease gro	WBC/micr	ltrophil/micr	ocyte/micr	onocyte/micr	e protein (mg/d	Severi	Number of UR	Number of Ger	ochondrial ges
2	0.75	0.97	0.92	0.24	0.72	0.24	1	0.1	0.39	0.49
3	0.75	0.97	0.92	0.24	0.72	0.24	1	0.02	0.12	0.88
4	0.75	0.97	0.92	0.24	0.72	0.24	1	0.04	0.2	0.24
5	0.75	0.97	0.92	0.24	0.72	0.24	1	0.04	0.24	0.49
6	0.75	0.97	0.92	0.24	0.72	0.24	1	0	0.02	0.15
7	0.75	0.97	0.92	0.24	0.72	0.24	1	0.01	0.04	0.35
8	0.75	0.97	0.92	0.24	0.72	0.24	1	0.03	0.22	0.35
9	0.75	0.97	0.92	0.24	0.72	0.24	1	0.08	0.34	0.62
10	0.75	0.97	0.92	0.24	0.72	0.24	1	0.03	0.18	0.75
11	0.75	0.97	0.92	0.24	0.72	0.24	1	0.04	0.23	0.41
12	0.75	0.97	0.92	0.24	0.72	0.24	1	0.04	0.22	0.65
13	0.75	0.97	0.92	0.24	0.72	0.24	1	0.05	0.25	0.48
14	0.75	0.97	0.92	0.24	0.72	0.24	1	0.05	0.28	0.47
15	0.75	0.97	0.92	0.24	0.72	0.24	1	0.03	0.15	0.67
16	0.75	0.97	0.92	0.24	0.72	0.24	1	0.01	0.04	0.25
17	0.75	0.97	0.92	0.24	0.72	0.24	1	0.01	0.06	0.77
18	0.75	0.97	0.92	0.24	0.72	0.24	1	0.03	0.17	0.41
19	0.75	0.97	0.92	0.24	0.72	0.24	1	0.03	0.2	0.46

Fig. 5. Dataset after normalization

5.3 Feature selection

The feature selection algorithm used in this work is Information Gain, which is a filter selection process. In the IG algorithm, features are weighted based on their material knowledge, with the target class's information being determined first, the entropy for each variable is then calculated. By subtracting the feature's entropy from the target feature's entropy, the information gain for each feature can be determined. Figure 6 shows the function selection criteria. See the algorithm for a detailed explanation of the IG equation (18, 19).

$$\text{Information class}(m) = - \sum_{i=1}^m \frac{\text{Distinct class}_i}{\text{Number of data sample}} \log_2 \left(\frac{\text{Distinct class}_i}{\text{Number of data samples}} \right) \tag{18}$$

m = number of distinct classes in the class.

$$\text{Entropy}(Feature) = \sum_{j=1}^v \frac{\text{Number of samples}_{ij} + \dots + \text{Number of samples}_{mj}}{\text{Number of data samples}} \times \text{Infomation}(C_{ij} + \dots + C_{mj}) \tag{19}$$

Then, using equation (20) to calculate the information gain for each function.

$$\text{Gain}(Feature) = \text{Information class}(C_1, \dots, C_m) - \text{Entropy}(Feature) \tag{20}$$

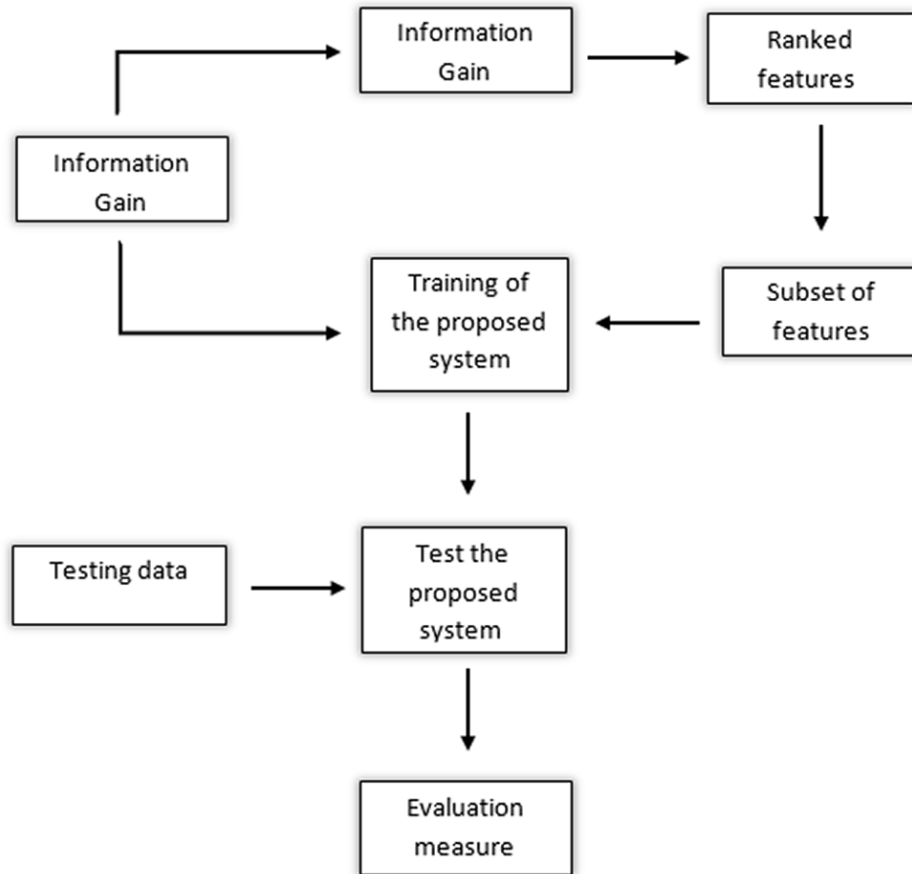


Fig. 6. The process of feature selection

5.4 Training and testing

The FCMNN proposed system consist of two phases:

A. Training phase

The proposed system's training process begins with a standardized training dataset as input data. The first stage was then used to detect covid19 from the influenza virus using the FCM algorithm. FCM is a clustering algorithm that uses membership to decide the degree to which each data point belongs to a cluster. Its central concept is to find the clustering center for each group by dividing n vector $x (i = 1, 2, n)$ into c fuzzy groups. Fuzzy C-means employs fuzzy partitioning, that is, membership is used to decide how often each data point belongs to a category, with each data point's value ranging from zero to one. The algorithm begins with a cluster center that has been initialized. The proposed system's second level is based on the BP algorithm. The BP algorithm starts with

random weights, input data specified by covid-19 dataset attributes, the number of hidden units, and two outputs, one for each covid-19 dataset class identified by (covid-19, influenza) class. Figure 7 shows that the steps of the proposed system.

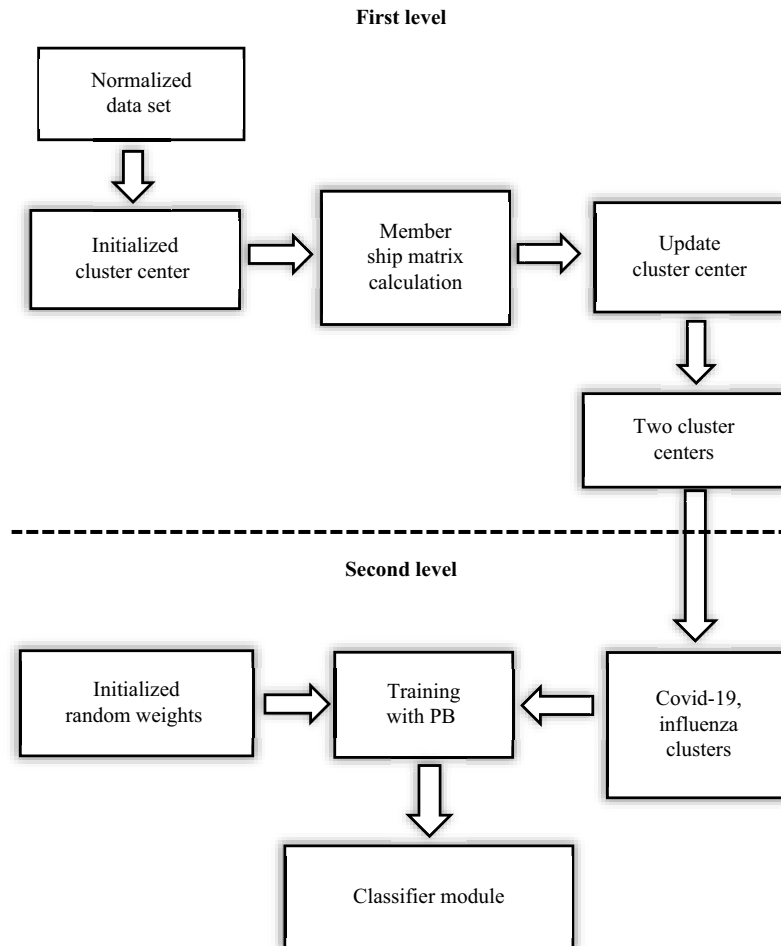


Fig. 7. The (FCMNN) algorithm’s training phase

B. Test phase

The proposed system’s testing process receives testing samples from the Covid-19 dataset, the optimal cluster center results from the first level of training, and the optimal vector of weights results from the second level of training. The membership matrix for each patient sample is calculated at the first step. After that, each sample is assigned to a cluster class. The covid-19 class then separates from the influenza classes. The second stage receives a vector of weights from the training phase and a pandemic cluster from the pandemic cluster phase, during which the BP algorithm forward phase begins to detect the pandemic type.

6 Discussion and experimental results

The proposed system’s first stage yields four potential outcomes and is referred to in the Table 2 confusion matrix.

Table 2. Confusion matrix

Actual Class	Predicted class	
	Negative Class (Influe)	Positive Class (COVID-19)
Influe	True Negative (TN)	False Positive (FP)
COVID-19	False Negative (FN)	True Positive (FP)

True negative (TN) indicates that normal behavior was correctly predicted, while true positive (TP) indicates that pandemic was correctly predicted, false positive (FP) indicates the incorrect prediction of influenza probability as covid-19, and false-negative (FN) indicates the incorrect prediction of influenza probability as covid-19. To assess the proposed system’s performance, two potential outcomes can be obtained, which are referred to as the pandemic confusion matrix (see Table 3).

Table 3. Covid-19 confusion matrix

Actual class	Predicted class	
	True prediction	False prediction
COVID-19	True Positive COVID-19 (TPC)	False Positive COVID-19 (FPC)

Accuracy (ACC) is a consistency metric that compares the rate of samples correctly identified as influenza or covid-19 to the total sample number. The equation is used to calculate it (22).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \dots \tag{21}$$

To evaluate the proposed system Table 4 describe the accuracy of the covid-19 dataset of (37557) patient sample with (10) features into methods FCM and BP.

Table 4. Dataset accuracy description

Method	Dataset class	No. of records	Covid-19	Influenza
Fuzzy C-means	Covid-19 Train dataset	26289	7207 89%	19082 75%
	Covid-19 Test dataset	11267	3088 87%	8178 78%
Back Propagation	Covid-19 Train dataset	26289	7207 87%	19082 71%
	Covid-19 Test Dataset	11267	3088 79%	8178 73%

There are two stages of the proposed system. The FCM algorithm was used in the first step to cluster covid-19 from influenza. FCM’s success is influenced by several factors. The following parameters are set experimentally:

- A. The fuzziness parameter's (m) value has an impact on FCM's efficiency. (m) Was measured with four different values during the training phase: 1.5, 1.9, 2, and 2.5. When the value of $m = 2$ is used, the best results are obtained.
- B. The cluster centers' initial values are set to random values.
- C. After much trial and error, the number of iterations used to find the FCM stopping criterion has been set to 9.

BP was used in the second step to detect the type of pandemic traffic. BP's success is influenced by several factors. Experiment with the following conditions:

- A. The basis value that has an impact on BP = 1's learning process.
- B. The BP was measured using four hidden units: 5, 10, 20, and 40. When the number of hidden units is equal to 20, the highest result is obtained.
- C. The number of input units equals 10, which is the number of features in the Covid-19 dataset that is presented.
- D. The output layer has a unit count of two, indicating that there are two classes of pandemics in the Covid-19 dataset.
- E. The learning rate and momentum coefficient have an effect on BP's output convergent with the goal output. Various values were tested during the training process. The best results were obtained when the learning rate was equal to one and the momentum coefficient was equal to one.
- F. The BP algorithms initialize weights. The value of weights set to random values during running time is examined in this work.
- G. The mean square error was used to find the BP stopping criterion. The mean square error in this work was 0.001.
- H. The highest number of iterations. The overall number of iterations in this work is 40.

7 Conclusions

This research has identified a mobile application that may be potentially useful to mitigate the COVID-19 pandemic. In this proposed work, a model is suggested to detect the type of the disease for which (covid-19 or influenza) with high accuracy and detection rate. Also, it was able to detect the class of the pandemic. It was shown from the experiments that, the sample of data that was incorrectly classified as covid-19 in the first level may be recognized as unknown classes in the second step of the model. In addition, this work is expanded to detect pandemic subtypes such as classified covid-19 to covid-19(mild) and covid-19(severe) sup classes, and classified influenza class to flu and asymptomatic sup classes. Furthermore, smartwatches and smart bands are used as symptom monitoring because they became more popular and incorporated into people's everyday lives and potentially assisting in the critical monitoring of vulnerable populations' health statuses. In the future, automated and rapid detection of suspected infections will become more effective using a convolution neural network (CNN). Also, the diagnosis accuracy can be increased by improving the symptom management algorithm and adapting it to the pandemic. Mobile doctors would be the way to go in

situations like this where self-isolation is needed, however, as technology advances, “digital humans” could be a viable option for reducing the burden on healthcare workers in future pandemics.

8 References

- [1] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. Abo-ElSoud, “Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network,” *Applied soft computing*, vol. 99, p. 106906, 2021. <https://doi.org/10.1016/j.asoc.2020.106906>
- [2] N. F. AL-Bakri and S. Hashim, “Collaborative Filtering Recommendation Model Based on k-means Clustering,” *Al-Nahrain Journal of Science*, vol. 22, no. 1, pp. 74–79, 2019. <https://doi.org/10.22401/ANJS.22.1.10>
- [3] K. V. Rajkumar, A. Yesubabu, and K. Subrahmanyam, “Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset,” *International Journal of Electrical & Computer Engineering*, vol. 9, no. 4, 2019. <https://doi.org/10.11591/ijece.v9i4.pp2760-2770>
- [4] H. H. Örkücü and H. J. Bal, “Comparing performances of backpropagation and genetic algorithms in the data classification,” *Expert systems with applications*, vol. 38, no. 4, pp. 3703–3709, 2011. <https://doi.org/10.1016/j.eswa.2010.09.028>
- [5] A. F. Hassan, “Using Unlabeled Data Set for Mining Knowledge from DDB,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 7, no. 1, pp. 1–8, 2021. <https://doi.org/10.26555/jiteki.v7i1.20164>
- [6] N. A. Huessein. H. A. Naman, M. Al-dabag, H. Th.Salim, “Encryption System for Hiding Information Based on Internet of Things,” *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 2, 2021. <https://doi.org/10.3991/ijim.v15i02.19869>
- [7] H. J. L. Singh, D. Couch, and K. Yap, “Mobile health apps that help with COVID-19 management: scoping review,” *Jmir Nursing*, vol. 3, no. 1, p. e20596, 2020. <https://doi.org/10.2196/20596>
- [8] M. Friedl, M. Ebner, and M. Ebner, “Mobile learning applications for android und iOS for German language acquisition based on learning analytics measurements,” *Int. J. Learn. Anal. Artif. Intell. Educ.*, 2020. <https://doi.org/10.3991/ijai.v2i1.12317>
- [9] S. Papadakis, M. Kalogiannakis, “Mobile educational applications for children: what educators and parents need to know,” *Mobile Apps for Boarding House Recommendation, International Journal of Mobile Learning and Organisation*, vol. 11, no. 3, pp. 256–277, 2017. <https://doi.org/10.1504/IJMLO.2017.085338>
- [10] S. Papadakis, N. Zaranis, and M. Kalogiannakis, “Parental involvement and attitudes towards young Greek children’s mobile usage,” *International Journal of Child-Computer Interaction*, vol. 22, p. 100144, 2019. <https://doi.org/10.1016/j.ijcci.2019.100144>
- [11] B. H. Majeed. L. F. Jawad, and Haider TH. ALRikabi, “The Impact of CATs on Mathematical Thinking and Logical Thinking Among Fourth-Class Scientific Students,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 10, 2021. <https://doi.org/10.3991/ijet.v16i10.22515>
- [12] A. J. Adwibowo, “Fuzzy logic assisted COVID 19 safety assessment of dental care,” *medRxiv*, 2020. <https://doi.org/10.1101/2020.06.18.20134841>
- [13] O. N. Oyelade and A. Ezugwu, “A case-based reasoning framework for early detection and diagnosis of novel coronavirus,” *Informatics in Medicine Unlocked*, vol. 20, p. 100395, 2020. <https://doi.org/10.1016/j.imu.2020.100395>

- [14] M. A. Al-Qaness, A. A. Ewees, H. Fan, and M. Abd El Aziz, "Optimization method for forecasting confirmed cases of COVID-19 in China," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 674, 2020. <https://doi.org/10.3390/jcm9030674>
- [15] L. Jia, K. Li, Y. Jiang, and X. Guo, "Prediction and analysis of Coronavirus Disease 2019," *arXiv preprint arXiv*, 2020.
- [16] K. Dittrich and G. Duysters, "Networking as a means to strategy change: the case of open innovation in mobile telephony," *Journal of product innovation management*, vol. 24, no. 6, pp. 510–521, 2007. <https://doi.org/10.1111/j.1540-5885.2007.00268.x>
- [17] T. M. H. Alanzi, "A review of mobile applications available in the app and google play stores used during the COVID-19 outbreak," *Journal of Multidisciplinary Healthcare*, vol. 14, p. 45, 2021. <https://doi.org/10.2147/JMDH.S285014>
- [18] S. M. Mehibs, and S. H. Hashim, "Proposed network intrusion detection system based on fuzzy c mean algorithm in cloud computing environment," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 26, no. 2, pp. 27–35, 2018. <https://doi.org/10.29196/jub.v26i2.471>
- [19] A. S. Hussein, R. S. Khairy, S. M. M. Najeeb, and H. T. ALRikabi, "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 5, 2021. <https://doi.org/10.3991/ijim.v15i05.17173>
- [20] N. F. AL-Bakri and S. H. Hashim, "A Study on the Accuracy of Prediction in Recommendation System Based on Similarity Measures," vol. 16, no. 1 Supplement, *Baghdad Science Journal*, 2019. [https://doi.org/10.21123/bsj.2019.16.1\(Suppl.\).0263](https://doi.org/10.21123/bsj.2019.16.1(Suppl.).0263)
- [21] N. Fadhil. and S. Hassan, "A Proposed Model to Solve Cold Start Problem using Fuzzy User-Based Clustering," in *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, 2019, pp. 121–125: IEEE.
- [22] S. M. Mehibs, and S. Hashim, "Proposed network intrusion detection system in cloud environment based on backpropagation neural network," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 26, no. 1, pp. 29–40, 2018. <https://doi.org/10.29196/jub.v26i1.351>
- [23] A. Salah, R. S. Khairy, H. Salim ALRikabi, "The Detection of Counterfeit Banknotes Using Ensemble Learning Techniques of AdaBoost and Voting," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 326–339, 2021. <https://doi.org/10.22266/ijies2021.0228.31>
- [24] N. AL-Bakri and S. H. Hashim, "MA Modified Similarity Measure for Improving Accuracy of User-Based Collaborative Filtering," *Iraqi Journal of Science*, pp. 934–945, 2018.
- [25] N. F. Al-Bakri and S. H. J. A.-N. J. o. S. Hashim, "Reducing data sparsity in recommender systems," vol. 21, no. 2, pp. 138–147, 2018. <https://doi.org/10.22401/JNUS.21.2.20>
- [26] T. Khairi, A. F. Al-zubidi, and E. Qahtan, "Modified Multipath Routing Protocol Applied On Ns3 Dcell Network Simulation System," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 10, 2021. <https://doi.org/10.3991/ijim.v15i10.22703>
- [27] N. Fadhi., A. F. Al-zubidi, A. B. Alnajjar, and E. Qahtan, "Multi label restaurant classification using support vector machine," *Periodicals of Engineering and Natural Sciences*, vol. 9, no. 2, pp. 774–783, 2021.
- [28] B. K. Mohammed, M. B. Mortatha, A. S. Abdalrada, and H. ALRikabi, "A comprehensive system for detection of flammable and toxic gases using IoT," *Periodicals of Engineering and Natural Sciences*, vol. 9, no. 2, pp. 702–711, 2021.
- [29] I. Mustapha, N. T. Van, M. Shahverdi, M. I. Qureshi, and N. Khan, "Effectiveness of Digital Technology in Education During COVID-19 Pandemic. A Bibliometric Analysis," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 8, 2021. <https://doi.org/10.3991/ijim.v15i08.20415>

- [30] B. H. Majeed, L. F. Jawad, and H. Alrikabi, "Tactical Thinking and its Relationship with Solving Mathematical Problems Among Mathematics Department Students," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 9, 2021. <https://doi.org/10.3991/ijim.v15i08.20415>
- [31] S. Wahyuningsih, A. Qohar, D. Satyananda, and N. A. Atan, "The Effect of Online Project-Based Learning Application on Mathematics Students' Visual Thinking Continuum in Covid-19 Pandemic," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 8, 2021. <https://doi.org/10.3991/ijim.v15i08.21565>
- [32] J. Verma and A. Mishra, "COVID-19 infection: Disease detection and mobile technology," *PeerJ*, vol. 8, p. e10345, 2020. <https://doi.org/10.7717/peerj.10345>
- [33] N. A. Hussien, A. Al-Magsoosi, H. Salim Alrikabi, and F. Abed, "Monitoring the Consumption of Electrical Energy Based on the Internet of Things Applications," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 7, pp. 17–29, 2021. <https://doi.org/10.3991/ijim.v15i07.20183>
- [34] A. H. Alaidi, O. H. Yahya, and H. T. Salim, "Using modern education technique in Wasit university," *International Journal of Interactive Mobile Technologies*, Article vol. 14, no. 6, pp. 82–94, 2020. <https://doi.org/10.3991/ijim.v14i06.11539>
- [35] M. Al-dabag, H. S. ALRikabi, and R. Al-Nima, "Anticipating Atrial Fibrillation Signal Using Efficient Algorithm," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 2, pp. 106–120, 2021. <https://doi.org/10.3991/ijoe.v17i02.19183>

9 Authors

Azhar F. Al-zubidi, Department of Computer Science, AL Nahrain University, Baghdad, Iraq. azhar.flaih@nahrainuniv.edu.

Nadia F. AL-Bakri, Department of Computer Science, AL Nahrain University, Baghdad, Iraq.

Rajaa K. Hasoun, Department of the Information System Management University of Information Technology and Communication, Baghdad, Iraq.

Soukaena Hassan Hashim, Computer science department, University of Technology, Baghdad, Iraq.

Haider TH Salim Alrikabi, Electrical Engineering Department, College of Engineering, Wasit University, Wasit, Iraq.

Article submitted 2021-04-22. Resubmitted 2021-05-26. Final acceptance 2021-06-27. Final version published as submitted by the authors.