



Kyasanur Forest Disease Classification Framework Using Novel Extremal Optimization Tuned Neural Network in Fog Computing Environment

Abhishek Majumdar¹ · Tapas Debnath² · Sandeep K. Sood³ · Krishna Lal Baishnab¹

Received: 19 July 2018 / Accepted: 20 August 2018 / Published online: 1 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Kyasanur Forest Disease (KFD) is a life-threatening tick-borne viral infectious disease endemic to South Asia and has been taking so many lives every year in the past decade. But recently, this disease has been witnessed in other regions to a large extent and can become an epidemic very soon. In this paper, a new fog computing based e-Healthcare framework has been proposed to monitor the KFD infected patients in an early phase of infection and control the disease outbreak. For ensuring high prediction rate, a novel Extremal Optimization tuned Neural Network (EO-NN) classification algorithm has been developed using hybridization of the extremal optimization with the feed-forward neural network. Additionally, a location based alert system has also been suggested to provide the global positioning system (GPS)-based location information of each KFD infected user and the risk-prone zones as early as possible to prevent the outbreak. Furthermore, a comparative study of proposed EO-NN with state of art classification algorithms has been carried out and it can be concluded that EO-NN outperforms others with an average accuracy of 91.56%, a sensitivity of 91.53% and a specificity of 97.13% respectively in classification and accurate identification of risk-prone areas.

Keywords e-Healthcare · Extremal optimization · Fog computing · Neural network

Abbreviations

KFD	Kyasanur forest disease
KFDV	Kyasanur forest disease virus
KNN	K- nearest neighbour
EHR	Electronic health record
EO-NN	Extremal optimization tuned neural network
GPS	Global positioning system
LDA	Linear discriminant analysis
MDC	Micro data center
MLP	Multi-layer perceptron
QoS	Quality of service
WBAN	Wireless body area network

Nomenclatures

C_{p1} , C_{p2}	Transfer Coefficients
C_{p3}	
ERR	Error
ERR_RMS	Average Root Mean Square
gen_min_e	Generation Minimum Error
hide	Number of Hidden Neurons
hide_in	Net Input to Hidden Neuron
hide_o	Net Output from Hidden Neurons
in	Number of Input Neurons
in_o	Net Output from Input
len	Length of the String
Loc_table	List of Infected Locations
new_case	Newly Occurred KFD Class
out	Number of Output Neurons
out_in	Net Input to Output Neuron
out_o	Net Output from Output Neuron
pop	Population
popsize	Population Size
prev_case	Previously Recorded KFD Class
$S_{Tr(prio)}$	Instance with the Priority in Training Dataset
$S_{(Train)}$	Training Dataset
$S_{Tr(i)}^{i^{th}}$	Instance of the Training Dataset
$S_{(Test)}$	Testing Dataset

This article is part of the Topical Collection on *Mobile & Wireless Health*

✉ Abhishek Majumdar
abhishek.nits@ieee.org

¹ Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

² Department of Mechanical Engineering, National Institute of Technology Silchar, Silchar, India

³ Guru Nanak Dev University, Regional Campus, Gurdaspur, India

tf	Transfer Function
tr_case	Total Number of Instances in Training Dataset
URN _o	User Registration Number Provided by the System
V	Hidden-Output Layer Weight
W	Input-Hidden Layer Weight

Introduction

Ticks are responsible for transmission of various infectious pathogens those have caused life-threatening human and animal diseases worldwide [1]. Ticks are blood-sucking ectoparasites that swarm over mammals, reptiles, birds, and amphibians, etc. Tick-borne diseases are seen in those particular areas where favourable ecological conditions exist for individual tick species.

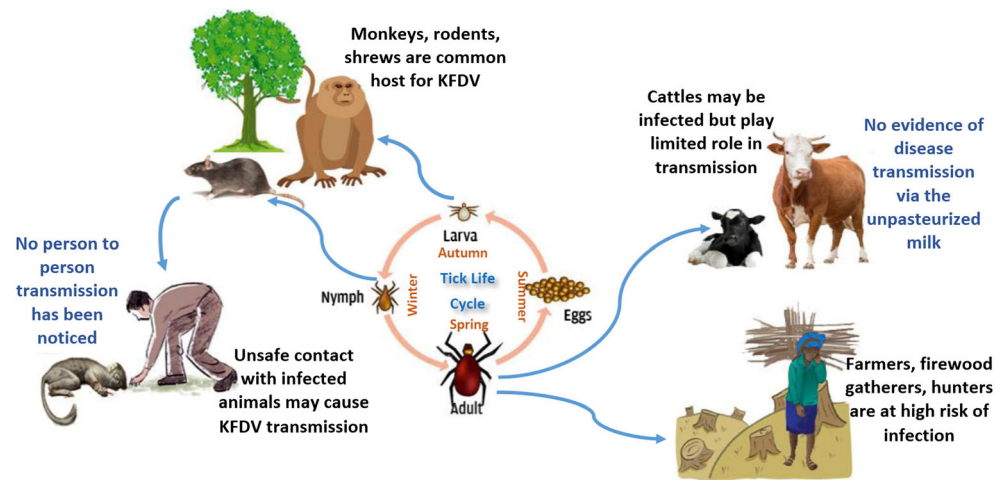
Globally, the incidence of tick-borne diseases is increasing gradually. A species of forest tick named *Haemaphysalis spinigera* is responsible for causing the deadly tick-borne viral disease Kyasanur forest disease (KFD). This highly infectious KFD and other closely related tick borne diseases have been causing periodic outbreaks in India and some other countries like China, France, Saudi Arabia, Egypt etc. which result in the death of thousands of lives over the decade.

KFD is an infectious disease in monkey and also termed as monkey fever in India. It is a tick-borne viral disease endemic in southern India that is caused by a highly pathogenic virus called KFD virus (KFDV) which belongs to the family of genus *flavi virus* and family *flavidae*. Ticks once infected, remain so for life and can pass KFDV to offspring by laying eggs. The larvae appear after the monsoon, and the nymphs start appearing in November or December. The nymphs feed on infected monkeys that serve as blood meal for these ticks. After that, these nymphs bite humans and thus picking up and passing on the virus. Cattle act as host to amplify and distribute the population of ticks but play no role in the KFD virus transmission. Figure 1 shows the KFD virus ecology to spread the disease [2]. In the enzootic state, KFDV spreads through small mammals such as rodents, shrews, ground birds, etc. [3]. When monkeys and the small mammals come in contact with the infected ticks, they also get infected and promulgate the infection. The people who visited forests to collect firewood, grass, and other forest products or pass through the forest are bitten by the infected nymphs of the tick (*Haemaphysalis spinigera*). Moreover, humans living near forest interfaces get infected on tick bite. This way the infection circulates among humans. Very recently thousands of ticks suspected to have KFDV related virus were collected from migratory birds captured in the Mediterranean basin. The results indicate that birds could also contribute to spreading the virus to new geographical areas. Due to the lack of post-exposure treatment

options for KFDV infections, the priority is given to the screening of replication inhibiting drugs against the virus [4].

Kyasanur Forest Disease virus (KFDV) was first identified in 1957, in the Kyasanur forest area of Shimoga district after few deaths of monkeys. It became a threat when humans started to die in the neighbourhood areas (Sagar Taluk, Shimoga District, Karnataka, India). Since then, 400 to 500 human cases per year have been reported [5]. During the initial outbreak, there were 466 human cases and 181 more in the following year. In 2003, KFD had affected more than 70 villages in four districts adjacent to Shimoga in western Karnataka. Until 2012 KFD had been endemic to five districts (Shimoga, Uttara Kannada, Dakshina Kannada, Chikkamagaluru and Udupi districts) of Karnataka state [5, 6]. In that period, a total of 3263 suspected KFD cases with 823 confirmed cases and 28 deaths were reported in those districts [7]. During 2012-2013, the disease was reported from new districts and new states in India. The KFD virus has spread into several regions within Karnataka state and surrounding Kerala, Tamil Nadu and Goa states [6, 8, 9]. Pieces of evidence of KFD virus or related virus were found in different parts of India (parts of the Saurashtra region in Gujarat, forest region west of Kolkata, West Bengal, and the Andaman Islands) after serological studies. It is found that 22.4% of people living in the Andaman and Nicobar islands are seropositive. KFD hit Wayanad, Kerala hard in 2015, infecting 102 people and killing 11, according to the district health department. The forest department, for its part, recorded the death of 400 monkeys. In the starting of 2017, a string of monkey fever cases was detected from neighbouring state Maharashtra after the disease left 3 dead and hundreds infected in Goa in 2016, according to the state health officials. According to Sindhudurg District Health officials, a total number of infected cases in the state has risen to 187 in 2017, from 128 in 2016 [10]. Maharashtra state health authorities have detected 45 cases of KFD in Sindhudurg district over the first 2 months of 2017 of which 3 persons have died and with this, the death toll has increased to 11 since 2016. Moreover, according to forest officials, more than 40 monkeys were found dead in Sindhudurg forests between January-May in 2017 [10]. Seventy-three cases of KFD have been detected, including one each at Patradevi in Pernem taluka and Savoi Verem in Ponda taluka in Goa since January 2017. With this, the KFD death toll in the taluk stood at 3. As per the state health officials of Goa, since January 2018, at least 35 people in Sattari taluka of North Goa have been tested positive with KFD. KFDV is common in young adults exposed during the dry season in the forest and has caused epidemic outbreaks of haemorrhagic fever affecting so many people per year since then, with a fatality rate that is estimated between 2 to 10% [7, 8]. The increasing number of new foci and cases indicates that eco-biological changes due to deforestation and use of new land for farming and cattle grazing could lead to spread of

Fig. 1 KFD virus ecology



KFD virus to newer geographical areas. Viruses related to KFDV have been identified in China [11] and Saudi Arabia [12–14] also. Alkhurma Hemorrhagic Fever Virus (AHFV) in Saudi Arabia and other regions of the Middle East and Nanjianyin Virus identified in Yunnan, China are found closely related to KFDV. Even in France, an average of 500–600 cases of KFD occur per year, with a 3.6% case fatality rate.

This virus has broken its endemic barrier and spread in new areas and will become a significant threat for an upcoming KFD epidemic in India and other countries very soon. So there is a serious need of an active KFD surveillance system that can find all the KFD related cases and take necessary actions upon it. Usually in many health monitoring systems, remote cloud servers have been used for storing and processing of the medical big data collected from a large pool of sensors and devices due to its capability of large storage volume, lack of unnecessary maintenance cost and inexpensive services. However, it can experience several issues regarding latency and data transmission which are the most crucial requirements for any healthcare system. Especially, in latency-sensitive cases such as medical emergency situations, a minor error in analysed data can cause wrong diagnosis which can risk a human life. In this respect the fog computing technology can be a proper solution to fulfil the requirement. The fog layer places itself between the conventional gateway and remote cloud server and offers pre-processing of data and other advanced services at the edge of the network. In addition, fog computing offers embedded data mining, low-latency real time response and notification service at the edge of network and reduces the burden of cloud. Considering these features, fog computing can be convenient for human healthcare systems.

This paper established a fog computing based KFD monitoring and prediction system model for controlling the disease outbreak through different ways such as early hospitalization and supportive therapy, Prioritize vaccinating areas, Tick control, etc. A novel Extremal Optimization tuned Neural Network (EO-NN) algorithm has been proposed here to

predict the KFD susceptibility status with higher accuracy. Moreover, a location based risk assessment and visualization system has been suggested to help the government authorities to take necessary actions as early as possible.

Background

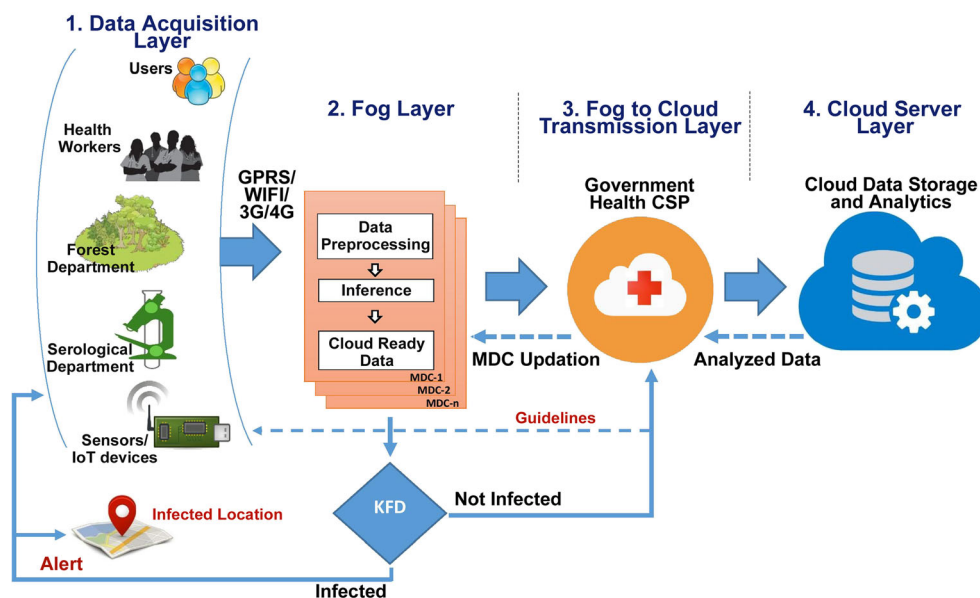
Neural network

A neural network (commonly referred as the artificial neural network) is an information processing paradigm inspired by the studies of the biological nervous system [15]. It mimics the processing features of human brain and attempts to extract underlying relationships in a set of data that are too complex to be noticed by other computing techniques. Neural networks can learn and adapt to changing input so the network generates the best possible result without needing to redesign the output criteria. Apart from this, a neural network has some other important characteristics such as self-organization, real-time operation, fault tolerance via redundant information coding etc. which make it an extensive research area to biomedical systems. Neural Networks are capable of detecting complex medical situations by fusing the data from the different biomedical sensors which makes it a flexible and powerful technique to be used in disease diagnosis.

Extremal optimization

Extremal Optimization (EO) is a stochastic search technique that has the properties of being a local and global search method. It is basically based on self-organized criticality concept. The main working principle of EO is to iteratively identify the most undesirable components of a given solution and successively replace them with newly generated random ones. It is motivated by its ancestors such as simulated annealing (SA) or genetic algorithms that use physical instinct to optimize [16].

Fig. 2 Proposed architecture of KFD prediction system



Literature survey

Literature Survey has been divided into two sections namely use of Cloud/Fog computing in the field of healthcare, and use of data mining techniques in disease prediction. The first section discusses the web-based analysis for disease monitoring in the field of healthcare. Secondly, a survey of different data mining techniques used for different kinds of disease prediction has been provided.

Web-based systems for healthcare monitoring

Computer Science and Information Technology have been extensively used for prevention and early prediction of deadly diseases and for analysing virus epidemiology. ANN was chosen as a classifier to diagnose the thyrotoxicosis disease and compared its performance with other classifiers [17]. Whereas another research work presented a healthcare information management system based on mobile application and cloud computing together for better sharing, storing, updating, and retrieval of electronic healthcare data. A prototype has been implemented using Android operating system, and it was evaluated using the Amazon S3 cloud service [18]. Some authors build a real-time predictive model for diagnosis of the dengue fever with high accuracy rate. They have developed a novel imputation procedure for identifying and imputing missing values in the dataset. Moreover, a wrapper-based feature selection strategy using genetic search was developed to extract the most likely symptoms of dengue fever [19]. A cloud-based intelligent system for diabetic patients was also established. In their methodology principal component analysis (PCA) was used for identifying minimum correlative variables from collected attribute variables. Naive Bayes and k-nearest neighbour classifiers were used to classify infected

and uninfected users [20]. Another strategy for diabetic patient classification was proposed by combining J48 decision tree, Bagging, and AdaBoost using J48 as a base learner. They performed the experimentation using Weka tool [21]. Some other researchers worked on a cloud-based health care system that aggregates the data from the Wireless Body Area Networks (WBAN) and later then the real-time analytics were done by combining the concept of STORM and the fuzzy logic [22]. A cloud-based monitoring model was also established to control the outbreak of Pandemic Influenza A(H1N1). Their methodology used a random decision tree to classify the infection in patient depending on H1N1 attributes. The system additionally introduced the concept of SNA graph and Outbreak Role Index to prevent the outbreak and to aware the users about how much they are probable to get infected respectively [23]. A web-based system was developed to predict dengue and identify the dengue hotspot areas, trends, outbreak behaviour, and case rate. For prediction, they used feed-forward ANN algorithm. For displaying the case rates, trends and prediction Google chart were used. Additionally, Google map and Google heat map was utilized to plot the dengue cases and the hotspot districts respectively [24]. A two layer strategy for diagnosing the condition of heart was proposed [25]. The work involved classifying the heart health attributes using KNN algorithm and predicting the rate of heart disease using a tree based ID3 algorithm. A cloud-based system was proposed to predict the airborne disease MERS-CoV, infected patients. For that, they used Bayesian belief network and additionally the system provides geographic-based risk assessment to control its outbreak [26]. A Fog-based zika virus outbreak monitoring system was also established where a fuzzy k-nearest neighbour is used to predict the possible infected users according to their clinical symptoms. Moreover, they implemented a Google

Table 1 Users personal attributes

Personal attribute	Description
URNo	Unique user registration number provided by the KFD System
Name	Name of user
Age	Age of user in years
Gender	Male/Female/Other
Address	Permanent address of user
Mob_no	Mobile number of user
Family_no	Mobile number of family member
GPS_loc	User’s current GPS location
Season	Current season
PIN	Postal code of user residence

map assisted representation of infected locations for risk assessment [27]. They also proposed another similar model to prevent zika virus by using a naive Bayesian network (NBN) as the classifier [28]. On the contrary, other researchers established an IoT and fog based model for controlling chikungunya virus attack. To diagnose the disease, they used fuzzy c-means as a classifier. Additionally, the real-time status

Table 2 Clinical attributes of users suffering from KFD

Tier-0 symptoms	Response	Tier-1 symptoms	Response
Sudden chills	Yes/No	Muscle Aches	Yes/No
Frontal headache	Yes/No	Joint Pain	Yes/No
Severe Myalgia	Yes/No	Low back & Extremities	Yes/No
Fever	High/Mild/No	Neck Pain	Yes/No
		Diarrhoea	Yes/No
		Vomiting	Yes/No
		Cough	Yes/No
		Photophobia	Yes/No
Tier-2 Symptoms	Response	Tier-3 Symptoms	Response
Bloody nose, gums	Yes/No	Neck Stiffness	Yes/No
Red eyes	Yes/No	Mental Disturbance	Yes/No
Blood in vomit	Yes/No	Giddiness	Yes/No
Blood in stool	Yes/No	Abnormality of Reflexes	Yes/No
Blood in Cough	Yes/No	Signs of Encephalitis	Yes/No
		BP	High/Normal/Low
		Platelet	High/Normal/Low
		RBC Count	High/Normal/Low
		WBC Count	High/Normal/Low
		Vision Deficits	Yes/No

of the outbreak was tracked continuously by using social network analysis (SNA) [29].

Data mining techniques in disease prediction

Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases [30]. It can be defined as a process of in-depth extraction of underlying information, previously obscure and conceivably valuable information from the database [31]. Data mining mainly uses two strategies: supervised learning and unsupervised learning. In supervised learning, a set of training examples is used to learn model parameters while in unsupervised learning no such training set is utilized [32]. The two most in-general objectives of data mining are classification and prediction. Classification models predict discrete and unordered functions whereas prediction model deals with the prediction of continuous-valued functions [33, 34]. For prediction of coronary disease, researchers proposed a method on a premise of coactive neuro-fuzzy inference system (CANFIS) by hybridizing the neural network with the fuzzy logic and genetic algorithm [35]. Another group of authors used the neural network with back propagation algorithm to predict heart disease, blood pressure and sugar in the human body [36]. After analysis, it was found that Neural Networks predict heart disease with the highest accuracy as compared to other data mining classifiers namely Decision Trees and Naive Bayes techniques [37]. Many studies was focused on hepatitis classification and feature reduction by using genetic algorithm [38], simulated annealing [39] and Linear Discriminant Analysis (LDA) [40]. For prediction, hepatitis information obtained from related patients is taken as instances to neural network (NN) algorithm, and fuzzy logic [40]. Whereas in another work, different types of neural network algorithms (Quick, Multiple, Dynamic and RBFN) with different factors (data size, learning cycle, and processing time to achieve the diagnostic accuracy and estimated error) are applied on the UCI dataset to predict hepatitis disease [41]. In hepatitis classification studies, other researchers achieved prediction diagnosis accuracies of 83.2%, 85.3%, and 86.4%, using Learning Vector Quantization (LVQ), K-Nearest Neighbor (KNN), and LDA techniques respectively [42]. Prediction of Parkinson’s disease has also been done using a radial basis function neural network based on particle swarm optimization [43]. An artificial neural network-based model has been developed for the diagnosis of coronary heart disease using a complex of traditional and genetic factors [44]. For training the network, they used a dataset that consists clinical, functional, laboratory, coronary angiographic, and genetic features acquired from several patients. By varying the number of input factors applied in the neural network, they created models which provide 64 to 94% accuracy.

Table 3 Data acquisition mediums for different types of attributes

Attribute type	Data acquisition
Personal attributes	Smart phone, monitor, GPS sensors etc.
Clinical attributes	Smart phone, smart wearable, body sensors, RFID tag, bio sensors etc.
Serological attributes	Smart phone, monitor
Forestry attributes	Smart phone, monitor, GPS sensors, humid sensor, barometric sensors etc.

Proposed system

Figure 2 represents the proposed architecture to monitor and detect KFD infected cases. The three main working scenarios of the proposed system have been illustrated in this section.

Scenario 1: After registering, the KFD system will provide a unique registration no “URNo” to the user, and after that, the user will be eligible to provide his/her symptoms details. To regularly update the health details, the system reminds every user after a given period. All the details will be submitted to the fog layer and based on that the fog layer will predict whether the user is suspected as KFD victim or not. If a user is found highly suspected, then immediately an alert will be sent to the user and his/her family as well and the nearby health worker group and the hospital to hospitalize the patient as soon as possible. Additionally, all the registered users within that suspected area will also be notified to take special care of themselves towards KFD and do a regular update of data and visit doctors. The system will also provide some guidelines to that user to follow strictly to avoid the KFD outbreak.

Scenario 2: If any monkey or bird is found infected or dead in the forest and after inspection, it is found to be KFD suspected then immediately that information will be uploaded to the system with the required attributes. Then, the system will transmit the data to the fog layer and update the `loc_table` with the infected location along with its 5 km vicinity. After cross-checking with the user updated symptoms table, if any user case is found suspected within the 5 km area from the forest department uploaded location then immediately it will be notified to the user as well as the health workers to hospitalize the user as soon as possible. Here, the attribute “GPS_loc” of the user attribute table act as the foreign key in the forest department table.

Scenario 3: If any patient comes to the hospital with complications similar to KFD symptoms, then the hospital authority will admit him to the hospital and do the needful serological tests to confirm the KFD. If the patient is found KFD infected, then the information will be updated instantly to the health attribute table with the patient “URNo” through the KFD system. Then, in the fog layer,

the given “URNo” will be checked with the user attribute table and the patient will be updated as KFD infected. Beside this, a notification will be sent to the other registered users within the 5 km area of the infected patient’s address (taken from the user attribute table) and an alert signal will be sent to the nearby hospitals and the health worker group so that they can take necessary actions immediately.

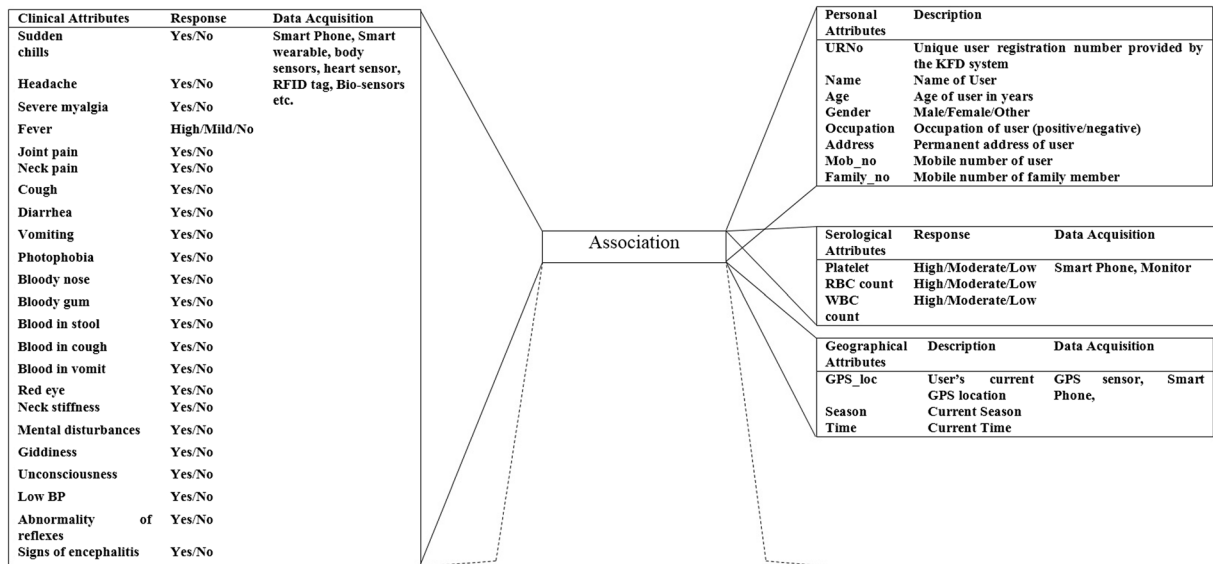
Data acquisition layer

Initially, each user has to be registered with the system by submitting their personal details through the mobile application installed on their mobile phone or devices. After getting registered, a unique user id ‘URNo’ will be provided to each user. After registration, the user can also upload the KFD related symptoms data to the cloud using the intended mobile application or web browsers.

This layer is responsible for collecting different kinds of data from various sources. It can be personal details of any user, data collected by the govt. health workers (ASHA in India), KFD related data from the forest department as well as the serological department, and different acquired data through different sensors. Moreover, the health workers can periodically scan the user’s health and check the KFD related symptoms, if any. Any recent instance related to KFD infected animals or birds and information about forest ticks can be uploaded from the forest department and serological department. Additionally, data can also be acquired and transmitted to the cloud from different sensors placed at different locations.

Fog layer

Fog computing brings cloud computing potentialities to the edge of the network so that, data can be processed more efficiently and if required it can be sent to the cloud for processing. Fog computing has the potential to offer data, compute, storage and application services with low latency. Moreover, it helps in improving QoS for streaming and real-time applications. Here, the efficiency of the system is enhanced by applying local data analysis at the edge. It can assist the system to detect and predict emergency situations. For instance, in highly susceptible cases of KFD, fog layer can locally offer the



Season	GPS location	Occupation	Sudden chills	Headache	Severe myalgia	Fever	Joint pain	Neck pain	Cough	Diarrhea	Vomiting	Photophobia	Bloody nose	Bloody gum	Blood in stool	Blood in cough	Blood in vomit	Red eye	Neck stiffness	Mental disturbances	Giddiness	Unconsciousness	Low KFD BP
P	P	P	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y HS
P	P	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N S
P	P	N	N	N	N	Y	Y	Y	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N PR
P	P	P	N	N	N	Y	N	N	N	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	N PR
N	N	P	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N U
P	N	P	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N S

Fig. 3 Association of different KFD related data sources

early hospitalization and related processing rather than sending parameters to the cloud and waiting for the responses.

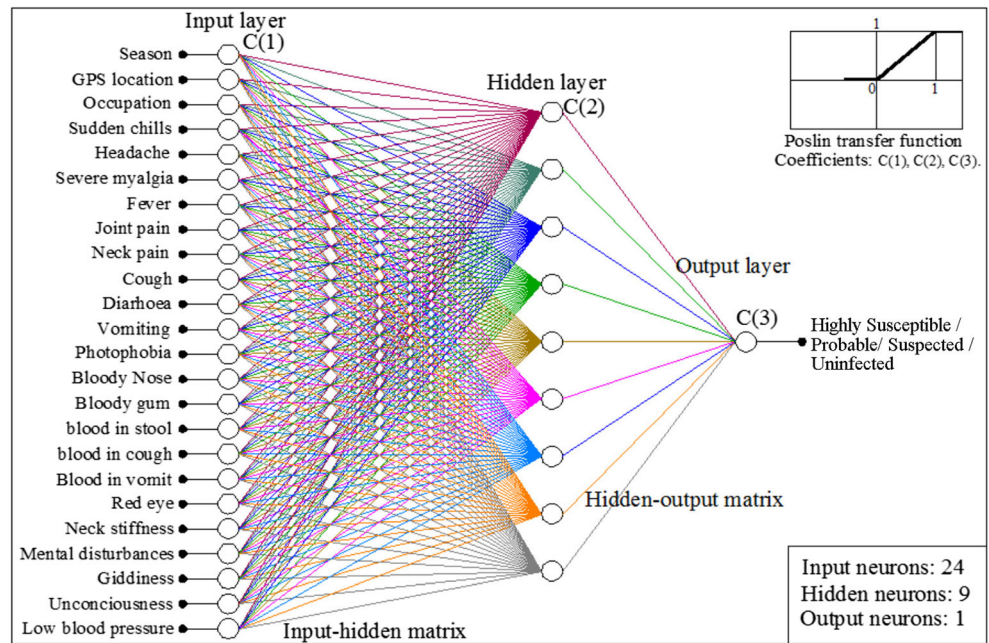
Here in the proposed architecture, the fog layer lies between the data acquisition layer and the cloud layer and comprises of a number of micro data centers (MDC) which are responsible for temporary and frequent data processing at the edge of the network. Each MDC has been associated with a significant number of users of a geographical area. Since data are collected from various sources, it can be of different types. Here in the proposed model, firstly the collected data has to be made ready for processing. Secondly, every MDC will be trained to determine the suspected cases of KFD. Once a suspected case is detected, it will immediately send an alert message to the user and notify all the nearby hospitals, health workers, forest department, etc. about it and at last send the processed data to the cloud server. If it fails to detect, then the processed data will be transmitted to the cloud server for further processing through a government cloud service provider. A set of fog servers must be distributed within predefined positions in each area under observation so that it can monitor risk-prone areas efficiently.

Data acquisition and pre-processing component

Since real-world data is often incomplete, inconsistent, lacks in certain behaviours or trends and is likely to contain many errors, data pre-processing is a proven method for resolving such issues. Data pre-processing prepare raw data for further processing. Every MDC is equipped with this component that

is configured with a data mining algorithm and a training data set to act upon the data collected from various sources. According to the nature of the sources and response, all the attributes acquired from various sources through numerous devices and sensors are divided into four groups namely, personal, clinical, serological and forestry. The lists of multiple personal and clinical attributes of the user have been shown in Tables 1 and 2 respectively whereas Table 3 lists different types of devices and sensors that can be used for data acquisition. The personal attributes are generally unchanged for most of the periods, whereas KFD attributes can change over the time. The location and other KFD related attributes collected from department of health, forest and serology are also stored separately. All the tables are interrelated so that every field related to KFD can monitor and predict the disease collectively. The values of all these parameters are stored in KFD database. In this work, all the clinical attributes are signs and symptoms of KFD though these are not exclusive for this disease. But some specific combination of these attributes are highly suggestive of KFD and can be graded as different KFD susceptibility classes. The initial prodromal stage of KFD in a patient is brought on by a sudden onset of a fever and severe headache along with a sore throat or neck pain, diarrhoea, vomiting, photophobia and severe pain in the lower and upper extremities. The inflammation of the conjunctiva (red eye) is also usually observed. The next stage is characterized by haemorrhagic complications such as low BP (hypotension), bloody nose (intermittent epistaxis), blood in vomit

Fig. 4 Schematic diagram of neural network for proposed system



(hematemesis), blood in stool (melena) and frank blood in the stool. At the severe stage, neurological manifestations such as mental disturbance, tremors, abnormal reflexes, bronchopneumonia or development of coma, giddiness, and vision deficits are generally observed. So all these clinical attributes along with some necessary serological attributes such as platelet count, RBC and WBC count have been included in the synthetic dataset. Additionally, the acquired GPS locations from multiple users have also been considered to confirm whether the locations lie in a risk-prone zone or not. An area where a

monkey death is reported is more likely to be a risky zone as KFD may be a cause of death of the monkey which can be transmitted to human. Similarly, the positive response of season has been reviewed in this component since season plays an important role in the transmission of KFD as more often cases are seen in dry season. The occupation attribute has also been considered because KFD is more often seen in hunters, foresters, wood-cutters etc. The KFD has been categorized depending upon these attributes. If some general prodromal symptoms are observed with the two most important attributes namely GPS location and occupation then the case is regarded as a suspected. In addition, if some haemorrhagic symptoms are observed on top of that then it is regarded as a probable case and if neurological symptoms are also positive then it is categorized as a confirmed or highly susceptible case. A set of these important attributes will be used for prediction at the MDC whereas the aggregated pre-processed data will be sent to the cloud server for further in-depth analysis. Figure 3 shows a graphical representation of sample KFD dataset generation with possible combinations of attributes.

Classification component

This component resides inside MDC and is responsible for the training of collected data using a predefined machine learning algorithm and predict the unspecified instances. In KFD monitoring system this component has been used to predict whether a certain instance from the collected real-time data is KFD infected or not. In this model, the proposed EO-NN has been used to train the dataset and predict the KFD status among users. This component will classify the registered user depending on KFD attribute data as uninfected (U), Suspected

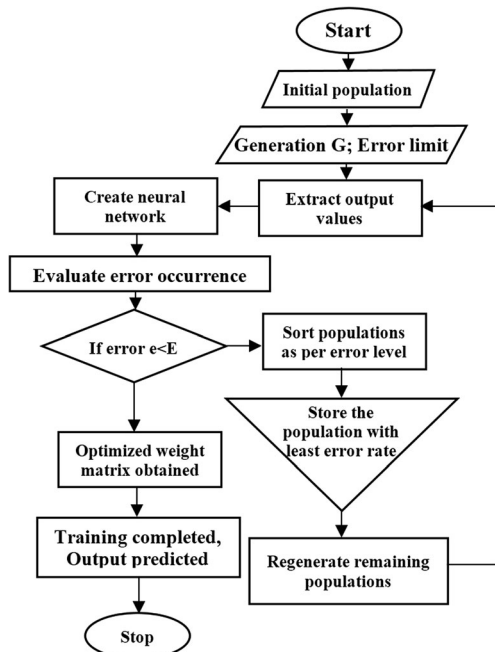


Fig. 5 Extremal optimization tuned neural network (EO-NN) flowchart

(S), Probable (PR), and Highly Susceptible (HS) category. Categorization has been done under the guidance of a group of doctors by following the guidelines of WHO. The ‘HS’ category encompasses the condition of patients which are almost confirmed to be infected to a high degree or level of KFD. ‘PR’ category defines the group of patients which are most likely to be infected with the KFD infection. ‘S’ category defines the group of patients whose health data is somehow suspicious to be KFD infected but their confident treatment can be made. The fourth category ‘U’ refers to all the persons who are found negative and safe from the KFD infection.

The proposed methodology involves the collection of experimental data, training the network, validating the network and forwarding it for the prediction of KFD conditions of the samples. For the classification of KFD, firstly a number of state of the art classification algorithms were used upon the dataset but yielded dissatisfactory outcomes. Therefore, in this paper, a real-coded hybrid neural network has been developed using feed-forward neural network along with Extremal optimization technique (EO-NN). The schematic diagram of the proposed neural network is depicted in Fig. 4.

In the training phase of the hybrid EO-NN model, initially weighted matrices along with the coefficients for the transfer functions are generated randomly many times to make a number of population. Using each population input-hidden matrices, hidden-output matrices, and coefficients of transfer function (for input layer, hidden layer and output layer) has been classified. In this neural network model, 24 symptoms have been considered as input information and 399 instances have been trained. Using multilayer feed-forward neural network the output has been calculated with the help of weight matrices and positive linear transfer function. Finally, the average RMS error has been calculated from the generated output and the actual output. Depending on the calculated error, sorting of the populations has been done. After that, the best population has been stored and all the remaining worst populations have been replaced with newly generated random solutions. Again from this newly generated populations output has been calculated by using the multi-layer feed-forward neural network. If any of these population gives less error than the previous best population; then it will become the new best one while all others will be regenerated. This process will continue until the desired accuracy level is achieved.

In the testing phase, the final input- hidden, hidden- output matrix, and the transfer function coefficients have been generated from the best population. Using a simple feed-forward neural network the output has been calculated, which gives the predicted result.

Figure 5 shows the flowchart of the proposed EO-NN algorithm and the internal functionality of the algorithm has been given in Algorithm 1 in detail. Using the obtained matrices, transfer functions, and applying simple feed-forward neural network the output has been classified.

Algorithm 1 Extremal Optimization Tuned Neural Network (EO-NN)

Input: Training Dataset [$S_{(Train)} = \{S_{Tr(1)}, S_{Tr(2)}, S_{Tr(3)}, S_{Tr(4)}, \dots, S_{Tr(m)}\}; 1 \leq m \leq \infty$];

Testing Dataset [$S_{(Test)} = \{S_{Te(1)}, S_{Te(2)}, S_{Te(3)}, S_{Te(4)}, \dots, S_{Te(n)}\}; 1 \leq n \leq \infty$]

Output: Trained Network /Classified Instance

For Training:

1. Define number of training cases (tr_case), input neurons (in), output neurons (out), hidden neuron (hide), transfer function (tf), and population (pop).
2. Define the string length, population size, and generation minimum error.
 - a. $len = (in \times hide) + (hide \times out) + tf$;
 - b. $popsize = rand(pop, len)$; // rand: randomly generated real values
 - c. $gen_min_e = 100$; //assigned a higher value
3. Call $S_{(Train)}$ and set priority to specific parameters and termed as $S_{Tr(prio)}$ for special conditions;
4. If $S_{Tr(prio)}$ satisfies the condition then, Predict output;

Else While $gen_min_e > 0.1$ // 10% error

limit allowed, can be changeable

- a. Store value of gen_min_e ;
- b. Generate input-hidden matrix (W_{pij}), hidden-output matrix (V_{pjk}), and transfer function (C_{p1}, C_{p2}, C_{p3}) ;
- c. Compute Net output from input layer

$$in_o(i, p, j) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq j \leq in \end{array} \right. = in_n(p, j) \left| \begin{array}{l} 1 \leq p \leq tr_case \\ 1 \leq j \leq in \end{array} \right. - C_{(i,1)} \left| \begin{array}{l} 1 \leq i \leq pop \end{array} \right.$$

- d. Compute Net input to hidden layer

$$hide_in(i, p, k) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq k \leq hide \end{array} \right. = V_{(i,j,k)} \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq j \leq in \end{array} \right. - in_o(i, p, j) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq j \leq in \end{array} \right.$$

- e. Normalize all the inputs for hidden layer and Compute Net output from the hidden layer

$$hide_o(i, p, k) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq k \leq hide \end{array} \right. = hide_in(i, p, k) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq k \leq hide \end{array} \right. - C_{(i,2)} \left| \begin{array}{l} 1 \leq i \leq pop \end{array} \right.$$

- f. Compute Net input to output layer

$$out_in(i, p, k) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq k \leq out \end{array} \right. = W_{(i,j,k)} \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq j \leq in \end{array} \right. - hide_o(i, p, j) \left| \begin{array}{l} 1 \leq i \leq pop \\ 1 \leq p \leq tr_case \\ 1 \leq j \leq in \end{array} \right.$$

- g. Normalize all the inputs for output layer and compute Net output from output layer

Table 4 Category wise patient monitoring interval

Patient category	Monitoring interval
Highly susceptible	12 h
Probable	24 h (1 day)
Suspected	72 h (3 days)
Uninfected	120 hour (5 days)

Continuous monitoring of KFDV infected users

To examine the appropriate infection status of each patient in the system, regular monitoring of treatment and symptoms of each user is highly desirable. According to infected categories of the patient as classified by EO-NN, monitoring will be done for a different period of times. Table 4 shows monitoring interval of infected patients. However, monitoring interval can also be changed by consulting a specialized doctor.

Cloud ready data component

All the pre-processed data have to be transmitted to the cloud server for further processing. The data mining at the cloud server side is responsible for extracting new features regarding the KFD and update the training dataset and send it to each fog layer to update their MDC accordingly.

Location based risk assessment and alert generation

This module is responsible for the location based risk assessment of the KFD outbreak. Once a tick gets on someone's body or cloths, they are likely to be there for many days if not checked properly. When they are in a desirable spot, they bite into the skin and begin to draw blood and remain attached to human body. After feeding, ticks may move to another body and bite or produce infected offspring by laying eggs. In this way ticks spread virus in urban areas also, even though they are far away from the forest areas. So, the humans or localities in vicinity of any infected person is also considered under the high risk prone zones which should be identified as soon as possible to control the spread of the virus. So instead of considering residential address only, the current geographic location of an infected user has been used here to distinguish and isolate the risk-prone areas. The objective of this module is to provide the real time information regarding the identification of infected population which can be a tough job for the government health authorities. After identifying the risk-prone zones, alert messages and infection control guidelines will be sent to the people close to those locations to control the spread of KFDV infection. The infected users will be diagnosed and monitored meticulously until their infection is fully recovered. In order to identify a newly infected user and risk-prone zone,

the fog server has been committed to acquire continuous data from the user. Google Maps Javascript API [45] has been used to locate the infectious regions according to their GPS locations. The Google map represents and updates the risk-prone zones automatically to help the uninfected users and government healthcare agencies in controlling the spread of infection. Algorithm 2 has been developed to precisely locate the infected users and the risk-prone zones in real time scenario. Figure 6 shows the Google map based visualization of the KFD infected user along with the risk-prone zones. The highly suspicious infected users have been denoted by a red coloured pin whereas a probable case has been represented by a yellow star. Moreover, the risk-prone zones have been represented with shaded circles. Here, the Cachar district, Assam, India and its surrounding regions have been taken as references to show the Google map based alert service. For example, user at the location of latitude 24.99017, and longitude 92.59541 has been found highly suspicious and user at the location of latitude 24.94613, and longitude 92.76288 has been found probable. In addition, an alert will be sent to all the registered users who fall within the risk-prone areas to avoid from the contact of KFDV infection. Furthermore, an alert message with a list of guidelines will also be sent to all the registered users within the risk-prone areas. This alert system warns the users to follow the guidelines to control the further spreading of the infection. The prevention suggestions have been structured as shown in Table 5.

Algorithm 2 Creation/Updating of Google map

Input: Location parameters of highly susceptible cases and probable cases.

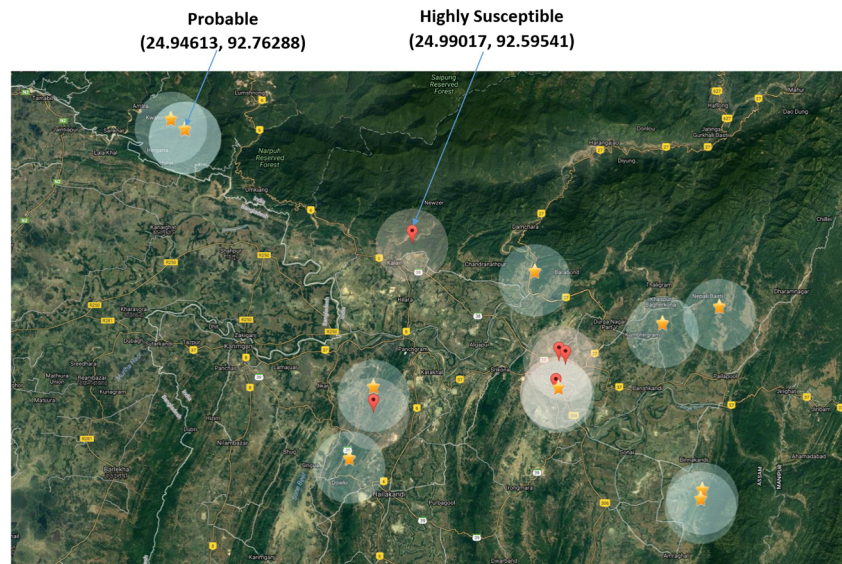
Output: Creation or updating of Google map.

1. Read the locations address and URNo of the patient;
2. If URNo already exists then
 - 2.1. If $prev_case \neq new_case$ then
 - a. Update the record with new location data with new case;
 - b. Compute the region within circular distance of 5 km radius from new location;
 - Else

No change
 - 2.2. End if
 - Else

Create a new record with URNo of the user;
3. End if
4. **Read the residential and current location of the user;**
5. Update the Google map pinpointing these locations as per found cases;
6. Highlight the circular boundary of each updated location;

Fig. 6 Google-map based visualization of KFD infected users and risk-prone areas (District: Cachar; State: Assam, India)



Health cloud service provider

The government can act as a cloud service provider, fully dedicated to the KFD big data transmission from each fog layer to the cloud server. It has the responsibility to maintain the security of confidential health related data or users personal details during transmission time or at storage.

KFD cloud server

The cloud server is responsible for secured storing of EHR as well as the in-depth analysis of the data received from the Fog layer. Once a highly suspicious or probable case is found, then the list of locations within the 5 Km radius from the infected location will be uploaded to the location table. The cloud server also analyses the inferred results coming from fog layer with their corresponding instances and do the needful for up gradation of the training data with new instances. From here the updated training dataset

and location table will be delivered to each MDC in the fog layer frequently.

Simulation and performance analysis

Creation of KFD dataset

Even after consulting with doctors and searching the internet very deeply, we have not found any KFD infection database based on the related symptoms of patients to test the proposed system. So, a synthetic database had been generated after consultation with a group of doctors. The synthetic database was developed systematically so that all possible cases and scenarios were included in the form of instances. Considering these, 500 cases were generated. To form the database, twenty-four different KFD attributes from the different attribute sets were associated together for the diagnosis of any user with KFDV infection. However, if required

Table 5 Guidelines to be followed for avoiding KFDV

S.No	Suggestions
1	Tick infected areas should be avoided strictly
2	Wear light coloured clothing
3	Wear Long boots
4	Tuck pants into boots
5	Regularly conduct tick checks after every outdoor activity
6	Apply DEET, DMP on exposed skin
7	Should be conscious about blood accepting/blood donations from those with a history of tick biting
8	Do regular health check-ups and inform health workers or hospitals if any complications faced which is similar to KFD symptoms.

these can be changed based on the situation. By increasing the values in the dataset, the number of records for training the EO-NN system can be increased and as a result the accuracy can be improved.

Training and testing of classifiers

The simulation of EO-NN algorithm along with other state of the art classifiers have been done in MATLAB R2017b platform for both training and testing. For evaluating the performance of the model in all the classifiers, four consecutive test sets have been made with 20, 40, 80, and 200 number of unspecified instances. Table 6 lists the comparative analysis of the proposed EO-NN classifier with four well known classification methods namely Naïve Bayes, Multilayer Perceptron neural network (MLP), LibSVM, Multi-Objective Evolutionary Fuzzy Classifier (MOEFC) on the KFD dataset for the four different test sets. The comparative analysis of various performance measures like accuracy, sensitivity, specificity, F-measure, Matthews’s Correlation Coefficient (MCC), Mean Absolute Error (EMA) and Root Mean Square Error (ERMS) have been conducted for all the classifiers. The F-Measure is basically the harmonic average of the precision and recall. It reaches its best value at 1 and worst at 0. Apart from the other measures, Matthews’s correlation coefficient is generally regarded as being one of the best as well as informative in evaluating classification problems.

The MCC has a range of -1 to 1 where -1 indicates a completely wrong classifier while 1 indicates a completely correct classifier. A value near to 1 represents better classification performance. These performance measures have been computed through Eqs. (1)-(6), where TN represents a number of True Negative samples, TP represents True Positive samples, FP is the number of False Positive samples, and FN is the False Negative samples.

$$Accuracy = \frac{\sum TP + \sum TN}{\sum Total\ Population} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

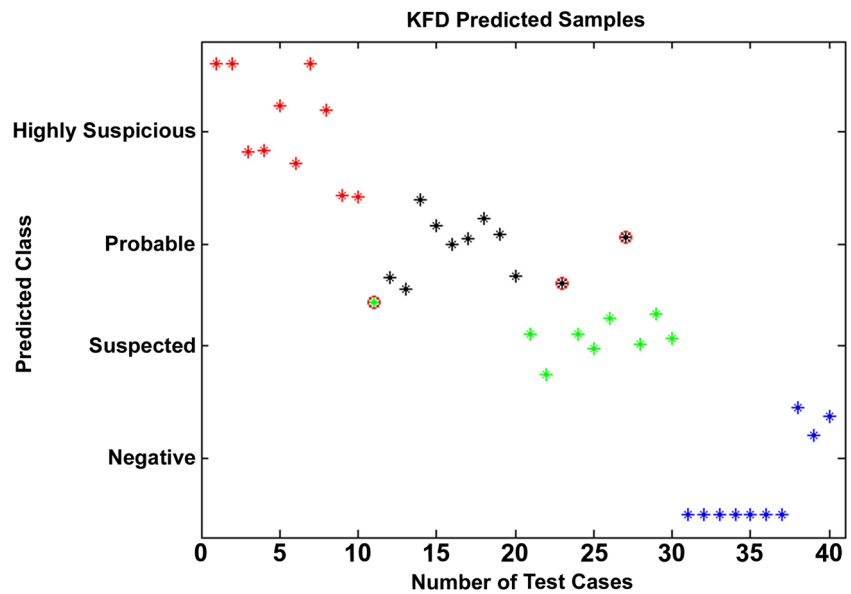
$$F-Measure = 2 \times \left(\frac{Precision \times Sensitivity}{Precision + Sensitivity} \right) \tag{5}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

Table 6 Performance analysis for different number of test cases

No. of test cases	Classifier	Correctly classified	Wrongly classified	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-Measure	MCC	E _{MA}	E _{RMS}
20	EO-NN	18	2	90	90.83	97.06	0.900	0.862	0.069	0.1963
	MLP	18	2	90	90.83	96.74	0.900	0.862	0.069	0.1963
	NaïveBayes	17	3	85	79.16	95.07	0.794	0.725	0.1473	0.2459
	MOEFC	10	10	50	47.92	55.91	0.505	0.407	0.2158	0.4646
	LibSVM	14	6	70	65	63.11	0.685	0.702	0.1585	0.3981
40	EO-NN	37	3	92.5	92.5	97.5	0.885	0.856	0.0753	0.2102
	MLP	35	5	87.5	84.24	95.12	0.860	0.821	0.0788	0.2315
	NaïveBayes	36	4	90	90	96.66	0.707	0.569	0.1552	0.2991
	MOEFC	24	16	60	47.85	84.72	0.514	0.378	0.2	0.4472
	LibSVM	32	8	80	71.43	91.3	0.731	0.668	0.1	0.3162
80	EO-NN	73	7	91.25	89.90	96.71	0.905	0.877	0.0593	0.1986
	MLP	71	9	88.75	90.2	96.69	0.875	0.838	0.0733	0.2135
	NaïveBayes	67	13	83.75	83.58	94.61	0.851	0.788	0.1199	0.2311
	MOEFC	44	36	55	48.89	82.96	0.451	0.296	0.2342	0.4839
	LibSVM	66	14	82.5	74.46	93.50	0.804	0.758	0.0823	0.2868
200	EO-NN	185	15	92.5	92.83	97.24	0.890	0.852	0.0741	0.2023
	MLP	179	21	89.5	86.57	96.16	0.890	0.853	0.0588	0.2027
	NaïveBayes	153	47	76.5	76.76	91.65	0.771	0.671	0.1393	0.2801
	MOEFC	111	89	55.5	41.16	82.62	0.468	0.291	0.2376	0.4875
	LibSVM	166	34	83	76.69	93.34	0.802	0.761	0.0842	0.2901

Fig. 7 Classified errors in EO-NN for testcase-2



It has been noticed that the EO-NN yielded a better result than the other methods in every respect and with a less error rate. Based on the comparison of all these measures, the performance of MOEFC was found dissatisfactory while that of the proposed EO-NN was found to be most satisfactory. The performance of the EO-NN based proposed model for the four test cases can be described as follows. For the 1st test case, the proposed model showed an accuracy of 90% by predicting 18 out of 20 instances correctly. For the 2nd test case, 37 out of 40 instances were predicted correctly, thereby giving an accuracy of 92.5%. For the 3rd and 4th test cases, the accuracy obtained was 91.25 and 92.5% respectively. The average accuracy for the proposed model is thus found out to be 91.56%. Figure 7 depicts the classification result for the second testset where the three wrongly predicted cases have been encircled with red

colour. Figure 8 illustrates the error occurrence at different iteration level for the proposed EO NN classifier. It can be seen that as the iteration level increases the error occurrence is either fixed or decreases.

It has been noticed that the EO-NN yielded a better result than the other methods in every respect and with a less error rate. Based on the comparison of all these measures, the performance of MOEFC was found dissatisfactory while that of the proposed EO-NN was found to be most satisfactory. The performance of the EO-NN based proposed model for the four test cases can be described as follows. For the 1st test case, the proposed model showed an accuracy of 90% by predicting 18 out of 20 instances correctly. For the 2nd test case, 37 out of 40 instances were predicted correctly, thereby giving an accuracy of 92.5%. For the 3rd and 4th test cases, the accuracy obtained was 91.25 and 92.5% respectively. The average accuracy for the proposed model is thus found out to be 91.56%. Figure 7 depicts the classification result for the second testset where the three wrongly predicted cases have been encircled with red colour. Figure 8 illustrates the error occurrence at different iteration level for the proposed EO NN classifier. It can be seen that as the iteration level increases the error occurrence is either fixed or decreases.

Figure 9 represents the comparative analysis performed between the abovementioned classifiers for different performance measures. From Fig. 9a it has been observed that the EO-NN yields the highest accuracy as compared to other classifiers in every test cases. Similarly from the Fig. 9b, it has been seen that the maximum sensitivity obtained for EO-NN is 92.83% whereas for the MLP, NB, LibSVM and MOEFC it is 90.83%, 90%, 48.89 and 76.69% respectively. Figure 9c illustrates the obtained specificity levels for all the classifiers in every testcases. It has been observed that the maximum specificity obtained by the EO-NN is 97.5% whereas for the

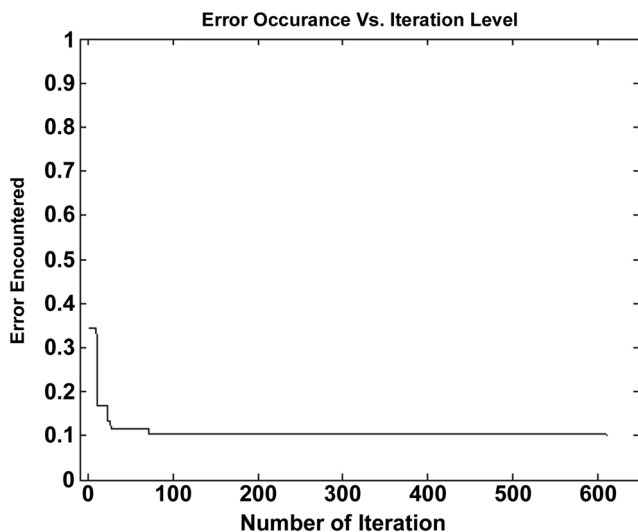


Fig. 8 Error occurrence during training at different iteration level

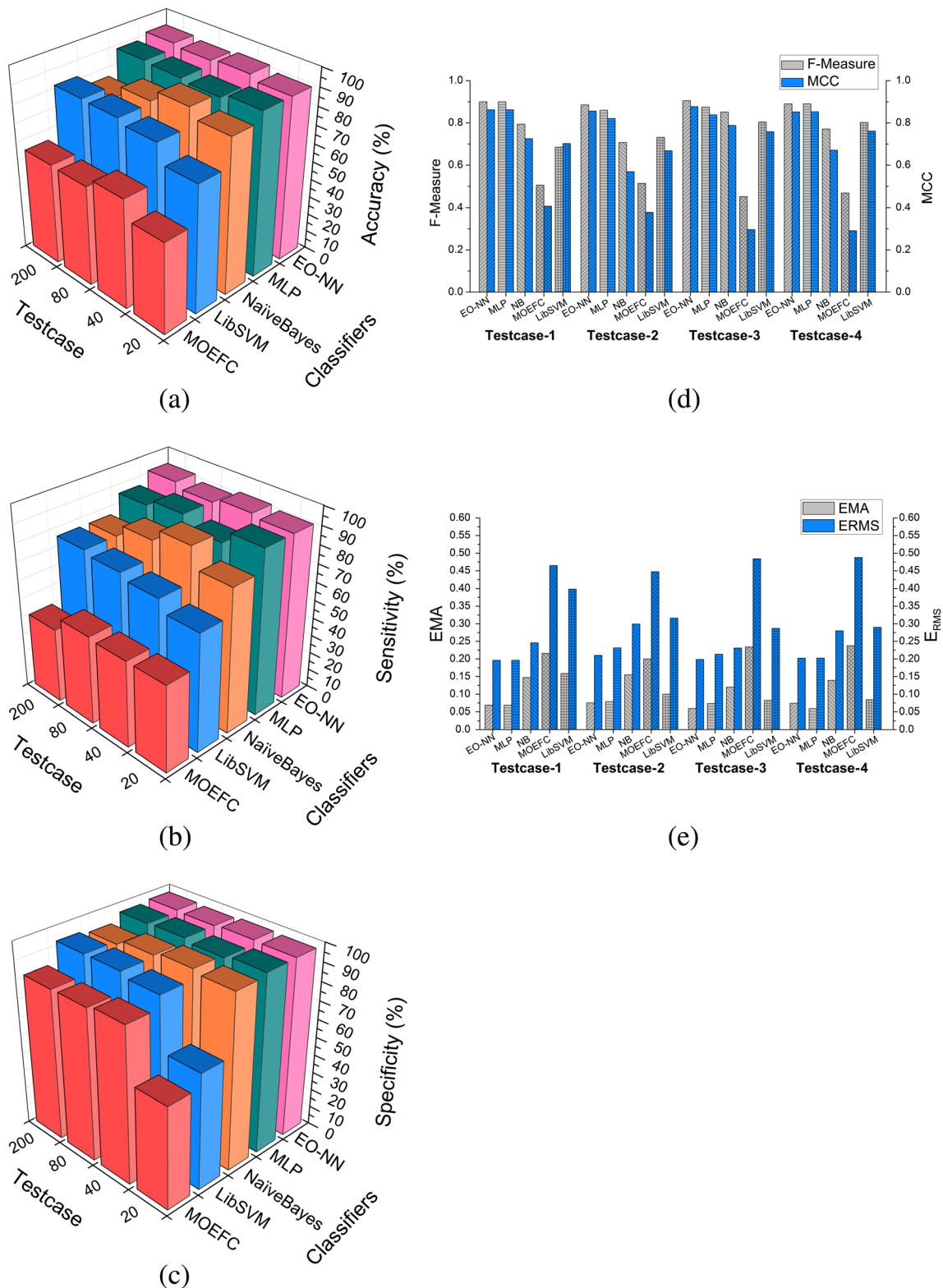


Fig. 9 Comparative analysis of the EO-NN with other classifiers. **a** Accuracy **b** Sensitivity **c** Specificity **d** F-Measure vs MCC **e** E_{MA} vs E_{RMS}

other classifiers it is 96.74%, 96.66%, 84.72 and 93.5% respectively. Figure 9d shows the observation of two performance measures together namely F-Measure and MCC. The

MCC defines how well the classification model is performing. From the graph, it has been observed that the best MCC obtained for the classifiers are 0.877, 0.862, 0.788, 0.407 and 0.761

respectively. Since the maximum MCC obtained for EO-NN is the nearest to 1, it has been treated as the most favourable result among all. Furthermore, the maximum F-Measure observed for the EO-NN is 0.905 which is nearest to 1 and thus considered as the best classification result. Even from the Fig. 9e it has been observed that the EO-NN outperforms other classifiers with least error rates (EMA and ERMS). Thus from the analysis, it can be clearly concluded that the EO-NN provides better performance in each scenario and is found to be suitable to configure in the proposed e-Healthcare system.

Conclusion

Due to the advancement in information technology, it has become much easier and effective to monitor, and diagnose many infectious diseases which help health sectors in taking necessary preventive measures to control the epidemic. In this paper, a fog-based intelligent system has been developed that can accurately diagnose and predict the KFDV infected users and also identify the risk-prone areas on the map. The proposed framework has been designed to monitor a large number of patients simultaneously by acquiring their clinical symptoms along with personal, and environmental data and transmitting the data to a dedicated Fog layer. Each Fog layer has been configured with a set of micro data centers which are distributed geographically and can handle the patient's big data in real time. From the data, the system will be able to perform inference locally that results in reducing the latency time and communication cost. The proposed EO-NN classifier has been developed for prediction of the KFD susceptibility to the user based on their symptoms, and cloud computing technology has been suggested for effective information analysis and sharing. The EO-NN provides good classification rate with an average accuracy of 91.56%. Moreover, the suggested location based alert services help to display KFDV-infected users and the risk-prone zones on Google map that will help the citizens to avoid regional exposure. Future work will include the prototype building of the framework with a suitable microcontroller and sensor devices and simulation in the real cloud scenario.

Acknowledgments Authors would like to thank Dr. Subhankar Debnath and Dr. Saswati Majumder, Government Medical Officer, Tripura Health Service, India for helping us in database generation and other support services.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interests to declare.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Shi, J., Hu, Z., Deng, F., and Shen, S., Tick-borne viruses. *Viol. Sin.* 1-23, 2018.
- Kyasanur Forest Disease (KFD) Virus Ecology. [Online]. Available: <https://www.cdc.gov/vhf/kyasanur/resources/virus-ecology.html>. [Accessed 02 June 2018].
- Pattnaik, P., Kyasanur forest disease: An epidemiological view in India. *Rev. Med. Virol.* 16(3):151–165, 2006.
- Cook, B. W., Nikiforuk, A. M., Cutts, T. A., Kobasa, D., and Theriault, S. S., Development of a subgenomic clone system for Kyasanur Forest disease virus. *Ticks. Tick. Borne. Dis.* 7(5):1047–1051, 2016.
- Sadanandane, C., Elango, A., Marja, N., Sasidharan, P. V., Raju, K. H. K., and Jambulingam, P., An outbreak of Kyasanur forest disease in the Wayanad and Malappuram districts of Kerala, India. *Ticks Tick Borne Dis.* 8(1):25–30, 2017.
- Mourya, D. T., and Yadav, P. D., Spread of Kyasanur Forest disease, Bandipur Tiger Reserve, India, 2012–2013. *Emerg. Infect. Dis.* 19(9):1540, 2013.
- Holbrook, M. R., Kyasanur forest disease. *Antivir. Res.* 96(3):353–362, 2012.
- Tandale, B. V., Balakrishnan, A., Yadav, P. D., Marja, N., and Mourya, D. T., New focus of Kyasanur Forest disease virus activity in a tribal area in Kerala, India, 2014. *Infect. Dis. Poverty* 4(1):12, 2015.
- Kyasanur Forest Disease - India (09): (Goa), ProMED-mail. [Online]. Available: <http://www.promedmail.org/direct.php?id=3306895>. [Accessed 04 April 2018].
- Kyasanur Forest Disease - India (10): (MH), ProMED-mail. [Online]. Available: <http://www.promedmail.org/post/20170517.5041642>. [Accessed 04 April 2018].
- Wang, J., Zhang, H., Fu, S., Wang, H., Ni, D., Nasci, R., Tang, Q., and Liang, G., Isolation of Kyasanur forest disease virus from febrile patient, Yunnan, China. *Emerg. Infect. Dis.* 15(2):326, 2009.
- Charrel, R. N., Zaki, A. M., Attoui, H., Fakeeh, M., Billoir, F., Yousef, A. I., de Chesse, R., De Micco, P., Gould, E. A., and de Lamballerie, X., Complete coding sequence of the Alkhurma virus, a tick-borne flavivirus causing severe hemorrhagic fever in humans in Saudi Arabia. *Biochem. Biophys. Res. Commun.* 287(2):455–461, 2001.
- Madani, T. A., Alkhurma virus infection, a new viral hemorrhagic fever in Saudi Arabia. *J. Infect.* 51(2):91–97, 2005.
- Musso, M., Galati, V., Stella, M. C., and Capone, A., A case of Alkhurma virus infection. *J. Clin. Virol.* 66:12–14, 2015.
- Sivanandam, S. N., and Deepa, S. N., *Introduction to neural networks using Matlab 6.0*. Tata McGraw-Hill Education, 2006.
- Boettcher, S., and Percus, A. G., Extremal optimization: an evolutionary local-search algorithm. In: *Computational Modeling and Problem Solving in the Networked World*. Boston: Springer, 21: 61-77.
- Shanker, M., Application of neural network classifiers to disease diagnosis. *The Journal of the Industrial Mathematics Society* 45(1): 51–62, 1995. Accessed 05 April 2018.
- Doukas, C., Pliakas, T., Maglogiannis, I., Mobile healthcare information management utilizing Cloud Computing and Android OS. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. p 1037–1040, 2010.
- Rao, V. S. H., and Kumar, M. N., A new intelligence-based approach for computer-aided diagnosis of dengue fever. *IEEE Trans. Inf. Technol. Biomed.* 16(1):112–118, 2012.
- Kaur, P. D., and Chana, I., Cloud based intelligent system for delivering health care as a service. *Comput. Methods Prog. Biomed.* 113(1):346–359, 2014.

21. Perveen, S., Shahbaz, M., Guergachi, A., and Keshavjee, K., Performance analysis of data mining classification techniques to predict diabetes. *Procedia Comp. Sci.* 82:115–121, 2016.
22. Sundharakumar, K. B., Dhivya, S., Mohanavalli, S., and Chander, R. V., Cloud based fuzzy healthcare system. *Procedia Comp. Sci.* 50:143–148, 2015.
23. Sandhu, R., Gill, H. K., and Sood, S. K., Smart monitoring and controlling of pandemic influenza a (H1N1) using social network analysis and cloud computing. *J. Comput. Sci.* 12:11–22, 2016.
24. Hwang, S., Clarite, D. S., Elijorde, F. I., Gerardo, B. D., and Byun, Y., A web-based analysis for dengue tracking and prediction using artificial neural network. *SERSC: ASTL* 122:160–164, 2016.
25. Thomas, J. and Princy, R.T., March. Human heart disease prediction system using data mining techniques. In *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on. IEEE.* pp. 1–5, 2016.
26. Sandhu, R., Sood, S. K., and Kaur, G., An intelligent system for predicting and preventing MERS-CoV infection outbreak. *J. Supercomput.* 72(8):3033–3056, 2016.
27. Sareen, S., Gupta, S. K., and Sood, S. K., An intelligent and secure system for predicting and preventing Zika virus outbreak using fog computing. *Enterprise Inf. Syst.* 11(9):1436–1456, 2017.
28. Sareen, S., Sood, S. K., and Gupta, S. K., Secure internet of things-based cloud framework to control zika virus outbreak. *Int. J. Technol. Assess. Health Care* 33(1):11–18, 2017.
29. Sood, S. K., and Mahajan, I., Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus. *Comput. Ind.* 91:33–44, 2017.
30. Thuraisingham, B., A primer for understanding and applying data mining. *IT Prof IEEE* 1(1):28–31, 2000.
31. Usama, F., Data mining and knowledge discovery in databases: implications for scientific databases. In *Proceedings of the 9th International Conference on Scientific and Statistical Database Management (SSDBM'97), Olympia, WA.* pp. 2-11, 1997.
32. Obenshain, M. K., Application of data mining techniques to healthcare data. *Infect. Control Hosp. Epidemiol.* 25(8):690–695, 2004.
33. Han, J., Pei, J., and Kamber, M., *Data mining: Concepts and techniques.* New York: Elsevier, 2011.
34. Kleissner, C., Data mining for the enterprise. In *Thirty-First Hawaii International Conference on System Sciences.* 7:295–304, 1998.
35. Parthiban, L., and Subramanian, R., Intelligent heart disease prediction system using CANFIS and genetic algorithm. *Int. J. Biol. Biomed. Med. Sci.* 3(3):157–160, 2008.
36. Guru, N., Dahiya, A., and Rajpal, N., Decision support system for heart disease diagnosis using neural network. *Delhi Bus. Rev.* 8(1): 99–101, 2007.
37. Dangare, C. S., and Apte, S. S., Improved study of heart disease prediction system using data mining classification techniques. *Int. J. Comput. Appl.* 47(10):44–48, 2012.
38. Huang, C. L., and Wang, C. J., A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* 31(2):231–240, 2006.
39. Sartakhti, J. S., Zangoeei, M. H., and Mozafari, K., Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Comput. Methods Prog. Biomed.* 108(2):570–579, 2012.
40. Dogantekin, E., Dogantekin, A., and Avci, D., Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system. *Expert Syst. Appl.* 36(8): 11282–11286, 2009.
41. Alshamrani, B. S., and Osman, A. H., Investigation of hepatitis disease diagnosis using different types of neural network algorithms. *International Journal of Computer Science and Network Security (IJCSNS).* 17(2):242, 2017.
42. Šter, B., and Dobnikar, A., Neural networks in medical diagnosis: Comparison with other methods. In *International Conference on Engineering Applications of Neural Networks.* pp. 427-30, 1996.
43. Wu, D., Warwick, K., Ma, Z., Burgess, J. G., Pan, S., and Aziz, T. Z., Prediction of Parkinson's disease tremor onset using radial basis function neural networks. *Expert Syst. Appl.* 37(4):2923–2928, 2010.
44. Atkov, O. Y., Gorokhova, S. G., Sboev, A. G., Generozov, E. V., Muraseyeva, E. V., Moroshkina, S. Y., and Cherniy, N. N., Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters. *J. Cardiol.* 59(2):190–194, 2012.
45. Google Maps. (2018). Google Maps. [Online]. Available: <https://www.google.co.in/maps/place/Cachar,+Assam/@24.7737965,92.6532688,75953m/data=!3m1!1e3!4m5!3m4!1s0x374e4a7895c9586b:0x921d35c0a735d9e8!8m2!3d24.7821253!4d92.8577105>. [Accessed 18 May 2018].