Journal of Scientific & Industrial Research
Vol. 82, February 2023, pp. 232-240
DOI: 10.56042/jsir.v82i2.70221

An Industry Framework for Remote Health Monitoring using Machine Learning Models to Predict a Disease

N Venkateswarulu[1]*, Chiranjeevi Manike[1] & O Obulesu[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad 500 075, Telangana, India
[2]Department of Computer Science and Engineering, G Narayanamma Institute of Technology & Science, Hyderabad 500 104, Telangana, India

Received 30 May 2022; revised 27 July 2022; accepted 07 October 2022

Remote health monitoring frameworks gained significant attention due to their real intervention and treatment standards. Most conventional works object to developing remote monitoring frameworks for identifying the disease at the earlier stages for an appropriate diagnosis. Still, it faced the problems with complexity in operations, increased cost of resources, misprediction results, which requires more time consumption for data gathering, and reduced convergence rate. Hence, the proposed work intends to design a machine learning based remote health monitoring framework for predicting heart disease and diabetes from the given medical datasets. In this framework, the Industry based smart devices are used to gather the health information of patients, and the obtained information is integrated together by using different nodes that includes the detecting node, visualization node, and prognostic node. Then, the medical dataset preprocessing is performed to normalize the attributes by identifying the missing values and eliminating the irrelevant qualities. Consequently, the Unified Levy Modeled Crow Search Optimization (U-CSO) algorithm is employed to select the optimal features based on the global fitness function, which helps increase the accuracy and reduce the training time of the classifier. Finally, the Most Probabilistic Guided Naïve Distribution (MP-ND) based classification model is utilized for predicting the label as to whether normal or disease affected. During an evaluation, two different datasets, such as PIMA and Hungarian, are used to validate and compare the results of the proposed model by using various performance measures. A Patients' health status can be monitored remotely for disease detection and proper diagnosis.

Keywords: Artificial intelligence, MP-ND, Smart devices, U-CSO

Introduction

In order to keep people safe from health threats, public health monitoring is one of the most important and fundamental concerns.[1] Heart disease and diabetes, which impact the nervous system and kidneys as well as the heart and kidneys, are now considered life-threatening disorders in the modern world. Most healthcare systems could benefit from using decision making tools for disease prediction and diagnosis because of poor quality treatments and inefficient clinical diagnosis.[2,3] Because early disease detection helps to prevent people from contracting serious illnesses. Health gadgets, smart watches, smart mobile phones, and other devices can be used to monitor patients' health status[4,5] via remote monitoring systems. There is a new remote health monitoring framework that uses advanced machine learning methods to predict different diseases. Patients' health status can be monitored remotely for disease detection and proper diagnosis in this paper.

*Author for Correspondence
E-mail: n.venkateswarulu@klh.edu.in

The major goal of the remote monitoring system is to ensure that patients receive the best possible care in the event of an emergency. In this study, an intelligent machine learning-based smart health framework is built, which helps to consolidate the patient data from smart devices for early detection of hazards and to keep healthcare practitioners updated with the current health information of the patient.[6,7] In this structure, the patient's health data is gathered through the detection, visualization, and prognostic nodes to provide an alarm for healthcare personnel. Data imputation, optimization, and categorization can all be used in this workflow. In this case, data imputation is used to scale the data from medical datasets, which aids in the improvement of pattern quality. Once the data has been reduced in dimensionality, the feature selection can be carried out utilizing the innovative optimization technique. Finally, an advanced AI-based machine learning technique can be used to accurately classify the given data.

To obtain the health information of patients from the smart devices, three different nodes such as detecting node, visualization node, and prognostic node have been utilized in this framework. To tune the given input datasets by eliminating the noise and normalizing the attributes, the dataset preprocessing is performed initially.

The study is different from the existing ones in the following manner:

- This system makes use of three separate nodes, such as a detection node, a visualization node, and a prognosis node, to gather patient health data from smart devices.
- In the beginning, dataset preprocessing is used to refine the input datasets by removing noise and standardizing their properties.
- Based on the global fitness, the features are selected using an algorithm called Unified Levy Modeled Crow Search Optimization (U-CSO).
- It is possible to classify the output as normal or disease-affected using the Probabilistic Guided Nave Distribution (MP-ND).
- Two prominent datasets, Hungarian and PIMA, are used to test the effectiveness of these techniques.

Related Works

In order to create a remote healthcare monitoring system for a variety of ailments, this part examines some of the usual research on data clustering, optimization, machine learning, and deep learning. According to their major features and functioning principles, the existing techniques are analyzed for their merits and drawbacks.

Using machine learning, Gondalia *et al.*[8] developed an IoT-based health monitoring system for soldiers in the military. The primary goal of this study was to locate soldiers who had been injured on the battlefield so that their health could be monitored. The GPS control and sensors like heart rate monitors were used to keep tabs on the soldiers' well-being (heartbeat and temperature). In addition, this study made use of the k-means clustering algorithm to examine the sensor data. Using sensor data, the clustering of features correctly predicted different types of movements like sitting, walking, or running in the event of wound or blast. For this reason, k-means clustering was not more suitable because of its lower efficiency, global clustering, and fluctuating starting partitions than any other method. It was Malasinghe *et al.*[9] who presented a detailed survey for identifying the best method to construct a remote patient health tracking system. E-health systems' security challenges were also addressed, along with the relevant solutions. A better machine learning technique was used[10] to design a new remote patient monitoring system. This system analyses the pilot data to see if wearable technology can be used to keep tabs on patients' health. A smart home health monitoring framework was created by Chatrati *et al.*[11] to monitor patients with diabetes and high blood pressure remotely.

In order to automate the transfer of patients' medical information to doctors, however, this project necessitates the creation of an appropriate graphical user interface. Based on data of heart rate and Electronic reports as well as chest sounds, Vitabile *et al.*[12] used a smart health monitoring system to analyze psychological circumstances and health status of patients. In this case, medical data has been kept private and secure using the block chain methodology. This system's key selling point was its faster processing of large-dimensional data. Patients' health state could be predicted remotely using a machine learning method developed by Nair *et al.*[13] In this case, the disease was predicted using the dataset's properties and a decision tree classification process. In addition, the division of feature models made it more capable of managing large datasets. It also has the ability to lessen generalization errors and over fitting errors. Still, this study relies on a series of computer procedures to forecast the classification labels.

Older adults can now be monitored from afar using a multi-stationary technique developed by Li and colleagues.[14] There are two different types of sensors employed to make the diagnosis in this framework: radar and sensors that can be worn by the patient. Filtering methods, wrapper methods, and embedded methods have all been used in this study to improve classification accuracy and speed up the process. In addition, the disease was predicted using KNN and SVM classification methodologies and the outcomes of these techniques were compared in terms of accuracy and processing time. It was determined through validation that the SVM model outperforms the KNN model in terms of performance. Patients' health status can be monitored remotely using the random forest classification technique developed by Kaur *et al.*[15] At the time of medical crisis, the IoT-based healthcare system was built to predict chronic diseases, such as diabetes. Classes such as KNN, linear SVM, decision tree and multilayer perception were evaluated and compared based on their performance in classifying different types of diseases using criteria such as accuracy, AUC, precision, and recall. According to the findings, the random forest method surpasses the other methods in terms of improved detection precision.[16]

An IoT-enabled architecture for remotely monitoring healthcare data via smart gateways was developed by Verma *et al.*[17] The temporal health index can be used to analyze the patient's health data using the event-triggering data transmission mechanism. Azimi *et al.*[18] developed a hierarchical healthcare monitoring architecture using an edge-based deep learning methodology. The primary goal of this research is to increase the availability and accuracy of classification while monitoring real-time events. The healthcare provider could access the cloud server's medical data through the sensors attached to the network under this framework.[20]

Proposed Methodology

An advanced optimization and classification model for a remote health monitoring system is discussed in this section. It is the primary goal of this research to design an electronic health monitoring framework that can be used to link medical data collected from patients via their personal smart devices. Using this medical data, diabetes and heart disease risk can be anticipated in advance for patients to receive medical updates.[21] In Fig. 1, we can see the suggested smart healthcare monitoring framework and its parts.

IoT technology implements a patch-based real-time healthcare monitoring system. As the primary objective, we sought out a highly reliable eHealth architecture that would save both time and money. Data transmission could be established with a lower bit error rate using low power radio frequency technology in this case.[22]

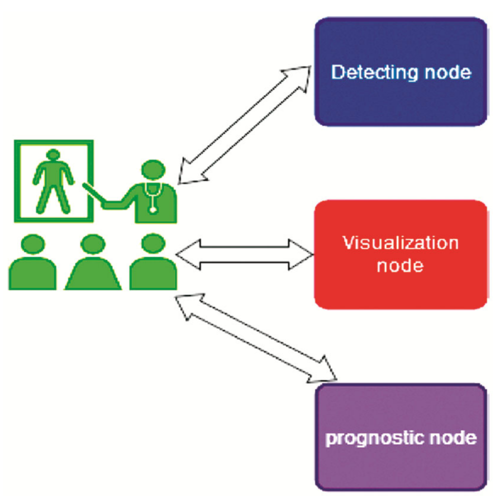

Fig. 1 — Smart remote healthcare framework

An analysis of the existing works shows that they are heavily focused on the development of remote healthcare systems for monitoring the health condition of patients using smart devices such as sensors, smart watches, and other intelligent gadgets. Nevertheless, the following drawbacks were encountered:

System performance is reduced due to inefficient data transfer between sensors and storage components, as well as a high bit error rate and complexity in classification.

The various sorts of nodes that can be found in this architecture are:

1 Detecting node - The primary role of this node is to collect patient data from smart devices. Medical data is collected from multiple sources and combined into a single web portal for easy access.
2 Visualization node - The machine learning classifier is used to predict the disease based on a subset of attributes derived from the patient data in the third node, the prognostic node.
3 Prognostic node - preprocessing and standardization of data Features selected using

- U-CSO (Unified Levy Modeled Crow Search Optimization).
- MP-ND (Probabilistic Guided Naive Distribution)

To begin, data pretreatment and normalization are carried out in order to enhance the dataset's quality by locating and removing any missing values and extraneous data. Because the outcomes of illness prediction are strongly dependent on dataset features, it is necessary to fine-tune the dataset to accurately detect and classify diseases. After that, a process known as feature scaling is used to uniformly scale the values of the dataset's properties. It is therefore necessary to use the U-CSO algorithm in order to reduce the number of features and improve classification accuracy. As a result, a classifier is trained with the specified attributes to better detect disease in patients with less computing complexity. Implementation: Integration is the key topic on this viewpoint. It presents the communication schemes, technologies, lifecycles to coordinate activities (usage viewpoint) and supportive of system capabilities (business viewpoint). i.e. technical representation. Standardization: a standard permit share of information across various enterprise levels – being that vertically or horizontally. Information share is possible between machines once there is a standard. Across employees the information standardization is important because every person cannot deliver the data as they want. The suggested methodology's flow

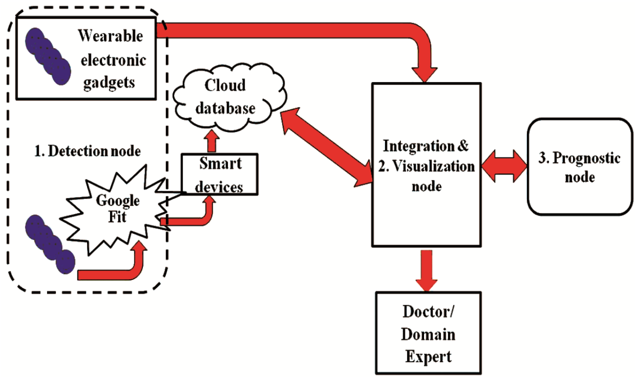

Fig. 2 — Industry Architecture model for healthcare

of work, its related architecture and components is portrayed in Fig. 2.

Dataset Normalization and preprocessing

To begin, missing values are identified and unnecessary attributes are removed using dataset preparation and normalization techniques. Preprocessing a dataset includes replacing missing data and removing background noise, both of which contribute to an enhanced detection rate. The input dataset is first analyzed for the amount of data points in order to detect the missing attributes using the median value. To arrive at the median value, the data are sorted in ascending order. The approximated median value replaces the irrelevant qualities and missing values. Once the data has been cleaned up, it is time to execute data normalization, which involves converting all of the numbers to a 0–1 scale. The regression model is used to estimate the standard deviation for data standardization as in Eqs (1 & 2):

$$P = \rho_0 + \rho_i D + \varepsilon_i \text{ for } i = 1, 2 \dots n \quad \dots (1)$$

$$P_i = \rho_0 + \rho_i D + \varepsilon_i^* \quad \dots (2)$$

By using these mathematical models, the residual value is computed. Where, P indicates the pair of data, ρ_0 and ρ_i are defined as the least square values, D is the input data, and ε_i denotes the error value. The average values are estimated for the sample data by using the standard deviation as in Eq. (3):

$$\mu = \frac{\sum_{i=1}^{n} T_i}{H_d} \quad \dots (3)$$

where, σ indicates the standard deviation, T_i represents the input data and H_d is the frequency of data. Consequently, the data normalization is performed as in Eqs (4 & 5):

$$M_D = \frac{\varepsilon_i^*}{\dot{\sigma_\iota}} \quad \dots (4)$$

$$M_D = \frac{T_i - \mu_i^*}{\dot{\sigma_\iota}} \quad \dots (5)$$

where, N_D indicates the normalized data, ε_i^* is the residual value, and $\dot{\sigma_\iota}$ defines the variance. Then, this preprocessed data can be used for the optimization and classification processes.

U-CSO based Feature Selection

The suggested U-CSO algorithm is used to pick the best number of features from the dataset after preprocessing, which reduces the dimensionality of features. The primary goal of employing the U-CSO technique is to maximize the number of features used in training the classifier's database. Predictive and classification systems typically use feature selection techniques to improve overall prediction performance. The Levy flight and CSO mechanisms are incorporated into the optimization process here. Meta-heuristic optimization techniques such as the proposed U-CSO can be used to discover the best fitness function for picking the most appropriate features.[22–26] The following are the primary advantages of utilizing this strategy:

- Convergence speed has been increased Reduced dimensionality of features Training and testing a classifier takes only a few minutes.
- Accuracy in classification Crows are often regarded as intelligent birds because of their propensity for food storage and concealment, as well as their large brains and long memories.
- CSO optimization is also widely used in a variety of systems to deal with the complexities of multi-objective optimization tasks. Reduced convergence rate of the CSO approach is a limitation of the technique due to its inadequate search capability and insufficient search space.
- For this reason, the levy flight technique is added into the CSO algorithm in order to improve the overall convergence speed and accuracy. Because of this, less iterations are required to find the optimal fitness value when using U-CSO. A random number of crows are chosen to populate the first population Fn. The maximum number of iterations is m k, and the position of the crow with the dimensional searching space d and the iteration t is obtained by applying Eq. (6):

$$Er_{c,t} = [Er_{c,t}^1, Er_{c,t}^2 \dots Er_{c,k}^n] \quad \dots (6)$$

where, c = 1, 2 … F_n and t = 1, 2 …m_t. In this model, each crow has the ability to evoke the location or position for hiding the food source, which accomplished before starting the next iteration as in Eq. (7):

$$K_{c,t} = [k_{c,t}^1, k_{c,t}^2 \dots k_{c,t}^n] \quad \dots (7)$$

Moreover, the crow can use the random path for safeguarding the food source from the other crow, which is represented the random path selection of optimization. In this integrated algorithm, the random path is identified by using the levy flight algorithm, where the searching probability is predicted for analyzing the behavior of crow. Let, consider that the random number W_c is uniformly distributed between the range of 0 to 1, which is represented as in Eq. (8):

$$x_c = Levy \sim x = a^{-\lambda}, \text{Where}(1 < \lambda \leq 3 \quad \dots (8)$$

This operation is explained by using the following model:

$$Dr_{c,t+1} = \begin{cases} Dr_{c,t} + W_c \cdot Am_{c,t} \cdot (Be_{v,t} - Dr_{v,t}) W_c \geq AP \\ Random\ placement\ Otherwise \end{cases} \quad \dots (9)$$

where, in Eq. (9), $Am_{c,t}$ defines the amplitude of crow c, $Be_{v,t}$ denotes the identified best possible solution of crow v, and W_c is the random number. Finally, the memory vector is updated by using Eq. (10):

$$Dr_{c,t+1} = \begin{cases} Dr_{c,t+1} & if\ O(Dr_{c,t+1})\ if\ better\ than\ (Be_{c,t}) \\ Be_{c,t} & Otherwise \end{cases} \quad \dots (10)$$

where, $O(.)$ indicates the objective function. By using this value, the best optimal function is identified for selecting the features to train the classifier.

Algorithm 1 — Unified Levy Modeled Crow Search Optimization (U-CSO)
Input: Preprocessed Dataset;
Output: Optimal selection of features;
Step 1: Initialize the set of populations F_n, Number of crows N, maximum count for iteration m_k, dimensional searching space d and iteration t;
Step 2: The population initialization with n number of crows with t iteration $Cr_{c,t}$ is represented in Eq. (1);
Step 3: Then, the location or position of hiding food source by the crow c is estimated by using the Eq. (7);
Step 4: Generate the random number between the range of [0, 1] by using the levy flight modeling as in Eq. (8);
Step 5: Then, its random placement $Cr_{c,t+1}$ is updated by using Eq. (9);
Step 6: Update the memory function according to the amplitude $Am_{c,t}$, best possible solution $Be_{v,t}$, and random number W_c.
Step 7: Based on the objective function $O(.)$, the best optimal solution is computed for optimally selecting the number of features;

MP-ND Classification

Using the ideal collection of characteristics, the classifier is trained to enhance its accuracy and reduce processing time. To detect the disease from the given datasets, the MP-ND classification process is developed utilizing the optimal amount of selected features. Probabilistic mixture model based on Gaussian distribution used for labeling datasets. Maximized likelihood functions are used in this approach to pick the model parameters. In order to better forecast the class, the sample mean and covariance values are calculated during this step. Using this method, the data overtraining problem is effectively avoided while increasing robustness. Large dimensional datasets can be handled more quickly and with less error with this approach. This is due to the fact that the hyper parameter tweaking is done using zero mean and unit variance measurements. An initial calculation is made by using T Gaussian distributions, as in Eq. (11):

$$p(opt_i | ch_i = t) = N(Me_t, \mathrm{Co}_t) \quad \dots (11)$$

where, opt_i is the optimized dataset, ch_i indicates the current operating condition, Me_t is the mean value, and Co_t defines the covariance. Also, there are T number of parameters used to define the feature space model of $\{(Me_1, \mathrm{Co}_1), (Me_2, \mathrm{Co}_2) \dots . (Me_t, \mathrm{Co}_t)\}$. Consequently, the statistical model of mixing properties is estimated by using Eq. (12):

$$\Theta = \{(Me_1, \mathrm{Co}_1, \Gamma_1), (Me_2, \mathrm{Co}_2, \Gamma_2) \dots (Me_t, \mathrm{Co}_t, \Gamma_t)\} \quad \dots (12)$$

where, Γ_t indicates the mixing proportions of class $t \in S$, $\Gamma = \{\Gamma_1, \Gamma_2 \dots \Gamma_t\}$. Then, the normal inverse wischart distribution NIW is computed based on the conjugate of distribution as shown in below Eq. (13):

$$p(Ne_t, Co_t) = NIW(nk_0, r_0, q_0, MV_0) \quad \dots (13)$$

where, nk_0 is the prior mean of Ne_t, MV_0 is the prior mean of Co_t, r_0 and q_0 indicates the strength of prior. Consequently, the identity matrix is constructed [ID × ID] with the d dimensional vector, and the distribution over the labelled space is denoted by using the model Eq. (14):

$$p(\Gamma) = Dir(\omega)\omega \prod_{t=1}^{T} \Gamma_t^{\omega_t - 1} \quad \dots (14)$$

where, the hyper parameters $\omega = \{\omega_1, \dots \omega_T\}$ that is used to integrate the posterior probability of each class. Then, its equally weighted factor $\omega_t = \frac{n}{T}, \forall t$ is computed for each class, and the generative statistical model $p(ch_i, opt_i, \Theta)$is also determined. Consequently, the labelled data LD_d is used to generate the initial number of classes T, where the model parameters are estimated by using the Bayesian function. Moreover, the posteriori probability is estimated and updated with the model parameters as shown in Eqs (15−19):

$$p(Ne_t, Co_t | ch_i = t, LD_d) = NIW(nk_n, r_n, q_n, MV_n) \quad \dots (15)$$

The parameters mk_0, r_0, q_0, MV_0 are computed as follows:

$$jk_n = \frac{r_0}{r_0 + n_t} jk_0 + \frac{n_t}{r_0 + r_t} \overline{opt}_t \quad \dots (16)$$

$$r_n = r_0 + n_{ch} \quad \dots (17)$$

$$q_n = q_0 + n_t \quad \dots (18)$$

$$JV_n = JV_0 + JV + r_0 jk_0 jk_0^S - r_n jk_n jk_n^S \quad \dots (19)$$

where, n_t indicates the number of observations in the labelled data LD_d, and $\overline{opt}_t$ defines the sample mean with the label *t*. Then, the sum of square matrix computed for each class *t* is shown in Eq. (20):

$$NV = \sum_{i=1}^{n_{ch}} opt_i opt_i^S \quad \dots (20)$$

The posterior probability is estimated based on the categorical distribution of the dirichlet function as shown in Eq. (21):

$$p(\Gamma | LD_d) \propto \prod_{ch=1}^{T} \Gamma_{ch}^{n_{ch} + \omega_{ch} - 1} \quad \dots (21)$$

According to the posteriori distribution function, the Bayes rule is applied to predict the classified label from the given data as in Eq. (22):

$$p\left(\widehat{ch}_i = t \middle| \widehat{opt}_i, LD_d\right) = \frac{p(\widehat{opt}_i | \widehat{ch}_i = t, LD_d) p(\widehat{ch}_i = t | LD_d)}{p(\widehat{opt}_i | LD_d)} \quad \dots (22)$$

By using this model, the classified label is predicted as whether normal or disease affected. The major benefits of using this mechanism are as follows:

- Requires minimum amount of time for training and testing the models.
- Minimized computational complexity.
- Increased detection efficiency and accuracy.
- Ensured reliability and scalability.

Results and Discussion

This section validates the results of both existing and proposed remote disease detection methodologies by using various evaluation metrics, where the MATLAB simulation tool is utilized to obtain the results. For this analysis, the diabetics and heart disease detection datasets are used, which includes Hungarian and PIMA (Indian Diabetes Database) (Indian Diabetes Database). The Hungarian dataset comprises 76 attributes, in which some of the essential attributes are listed in Table 1. The Fig. 3 shows the confusion matrix of the heart disease dataset with respect to the classes of 0 and 1. Based on the results, it is analyzed that the proposed mechanism accurately predicts the classes with increased TPR.

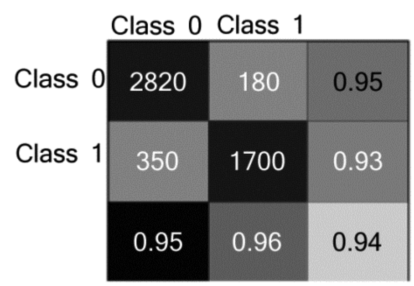

Fig. 3 — Confusion matrix for Hungarian dataset, Training and Testing

Table 1 — Hungarian dataset description

Attributes	Description
Age	Young, medium, old or very old
Gender	Male and female
Maximum Heart Rate (MHR)	Low, medium, high or very high
Chest Pain (CPT)	Type 1- typical Type 2 – Atypical Type 3 – Non-anginal Type 4 – Asymptomatic
Resting Blood Pressure (RBP)	Low, medium, high or very high
Blood sugar (FBS)	BS > 120 mg/dl
Serum Cholesterol (SCH)	Low, medium, high or very high
Major vessels (VCA)	Major vessels (0 to 3)
Thallium scan (TCA)	Normal, fixed defect, and reversible defect
Exercise Induced Angina (EIA)	Yes – 1 No – 0
Depression Induced by Exercise (OPK)	Low, Risk or Terrible

Table 2 — PIMA dataset description

Attributes	Description
No of pregnancies	Numerical attribute
Blood Pressure (BP)	mmHg
Age	Young, middle, old or very old (in terms of years)
BMI	Body mass index
Insulin	mU/mL
Glucose level	Estimated in terms of mg/dL
Skin type	Mm
Diabetes Pedigree	Yes – 1; No – 0

Table 3 — Accuracy, sensitivity, and specificity analysis using the Hungarian dataset

Methods	Precision (%)	Recall (%)	Accuracy (%)	F-Measure (%)
SVM	87.5	81.5	84.4	84.5
LR	89.2	95.2	92.2	92.2
MLP	93.3	85.3	89.3	89.3
RF	87.4	87.4	87.3	87.4
DT	84.6	77.7	77.6	77.6
NB	88.8	78.5	83.4	83.4
Ensemble DL	98.2	96.4	98.5	97.2
Proposed	98	98.5	98.5	98

Similarly, the PIMA dataset is obtained from the machine learning repository, which holds 768 number of samples with 9 different feature attributes. Also, it comprises the patient information aging from 21 to 81, and its attribute information are listed in Table 2.

The existing and proposed classification techniques are compared in Table 3 based on the measures of precision, recall, accuracy, and f-measure. These parameters are extensively used in many detection/classification application systems for analyzing the efficiency of methodologies. Also, the overall performance of classifier is highly depending on the improved values of these measures, which are calculated using Eqs (23−26):

$$Accuracy = \frac{PP+NN}{PP+NN+FF+NF} \quad \dots (23)$$

$$Precision = \frac{PP}{PP+FF} \quad \dots (24)$$

$$Recall\ or\ TPR = \frac{PP}{PP+NF} \quad \dots (25)$$

$$F-Measure = \frac{2PP}{2PP+FF+NF} \quad \dots (26)$$

where, PP – True Positives, NN – True Negatives, FF – False Positives, and NF – False Negatives. Based on the estimated results, it is analyzed that the proposed U-CSO-MP-ND outperforms the other approaches with increased precision, recall, accuracy, and f-measure values.

Table 4 — RMSE and MAE using the Hungarian dataset

Methods	RMSE	MAE
SVM	0.39	0.15
LR	0.22	0.11
MLP	0.27	0.10
RF	0.34	0.13
DT	0.39	0.31
NB	0.38	0.36
Ensemble DL	0.21	0.12
Proposed	0.13	0.9

Table 5 — Overall accuracy analysis using Hungarian dataset

Methods	Overall accuracy
GFLS	78.7
Ensemble classifier	85.4
Type 2 - FL	86
Hybrid ML	88.4
DL	89
ANN	91
Kernal RF	91
MLP	92
Fuzzy Diagnosis system	92.3
DT	92.8
ANFIS	94.1
Ensemble DL	98.5
Proposed	99

The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) value of both existing and proposed classification techniques using the Hungarian dataset are compared in Table 4. These are the error measures calculated as shown in Eqs (27 & 28):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \quad \dots (27)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \quad \dots (28)$$

For example, x i is the actual value, and predicted value x I is the total number of observations. In comparison to the other methods, the proposed U-CSO-MP-ND technique yields lower error outputs, as shown by the results. Therefore, by reducing error values through the use of normalization and feature selection processes, it is possible to train a classifier with the optimal features.[27]

As shown in Table 5 the overall accuracy of existing and proposed classification techniques can be evaluated using the dataset from Hungary. The detection mechanism's overall efficiency and improved performance are typically evaluated by estimating the accuracy of the detection mechanism as a whole. There appears to be a clear advantage to the proposed method over the others in terms of accuracy

Table 6 — Comparative analysis based on accuracy using PIMA dataset

Methods	Accuracy (%)
Firefly with cuckoo optimization based classification	81
Feed forward NN	82
NB	79.56
LDA with MWSVM	89.74
GA with NN	87.46
K-means with DT	90
K-means with PCA	72
U-CSO-MP-ND	98.5

in this evaluation. The training model of the classifier is improved as a result of the optimal feature selection, resulting in an increased accuracy value.

Using the PIMA dataset, Table 6 and compare the detection accuracy of the existing[28] and proposed classification techniques. As shown by these results, the proposed PSND technique has a higher level of accuracy than the other techniques.

On the basis of accuracy, precision, and recall, existing and proposed classification methods are compared in Table 6. With fewer iterations and higher convergence rates, a hybrid U-CSO algorithm is used in the proposed mechanism to find the best optimal solution and to select the most appropriate set of features. This information is used to train the classifier, which improves the system's accuracy, precision, and recall values, by a significant margin.

Conclusions

There is a new remote health monitoring framework that uses advanced machine learning methods to predict different diseases. Patients' health status can be monitored remotely for disease detection and proper diagnosis in this paper. For fusing medical data from sensors, a one-time authentication is used in this industry ready framework. As a result, the mobile application can directly access the cloud platforms to obtain the data it needs, thus eliminating the need for the user to enter their information. Patients' medical data obtained via mobile devices is stored in the Google Fit cloud for this purpose. In Google Fit, these different nodes are also used to incorporate medical information from various sources such as the visualization of node and the prognostication of node. Medical datasets are preprocessed and normalized at the beginning of disease prediction to improve the classifier's performance. In order to reduce the dimensionality and improve the classifier's accuracy, we then use the U-CSO algorithm. To detect the disease from the datasets, the MP-ND classification mechanism is implemented, which uses the optimal number of selected features, after optimization. Data sets such as PIMA and Hungarian are used to test the performance of the proposed model in order to validate results. As a result, the results are compared to the most recent state-of-the-art in terms of memory, accuracy, and precision; f-measure; error rate. Proposed outperforms other techniques in terms of performance, according to an evaluation.

References

1 Hong-Tan L, Cui-Hua K, Muthu B & Sivaparthipan C B, Aggression and Violent Behavior: Big data and ambient intelligence in IoT-based wireless student health monitoring system, *Aggression and Violent Behavior*, (2021). https://doi.org/ 10.1016/j.avb.2021.101601.

2 Albahri O S, Albahri A S, Mohammed K I , Zaidan A A, Zaidan B B, Hashim M & Salman O H, Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: Taxonomy, open challenges, motivation and recommendations, *J Med Syst*, **42** (2018) 1–27.

3 AlShorman O, AlShorman B, Alkhassaweneh M & Alkahtani F,A review of internet of medical things (IoMT)–based remote health monitoring through wearable sensors: A case study for diabetic patients, *Indones J Electr Eng Comput Sci*, **20** (2020) 414–422.

4 Divakaran J, Prashanth S K, Mohammad G B, Shitharth D, Mohanty S N, Arvind C, Srihari K, Abdullah R Y & Sundramurthy V P, Improved handover authentication in fifth-generation communication networks using fuzzy evolutionary optimisation with nanocore elements in mobile healthcare applications, *J Healthc Eng*, **2022** (2022) 1–8, https://doi.org/10.1155/2022/2500377.

5 Talal M, Zaidan A A, Zaidan B B, Albahri A S,Alamoodi A, Albahri O S, Alsalem M A, Lim C K, Tan K L, Shir W L & Mohammed K I, Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: Multi-driven systematic review, *J Med Syst*, **43** (2019) 1–34.

6 Hamim M, Paul s, Hoque S I, Rahman M N & Baqee I A, IoT based remote health monitoring system for patients and elderly people, in *Int Conf Robot Electr Signal Process Tech (ICREST)* (IEEE) (American International University, Bangladesh) 10–12 January 2019, 533–538.

7 Manoharan H, Haleem S L A, Shitharth S, Kshirsagar P R, Tirth V, Thangamani M & Chandan R R, A machine learning algorithm for classification of mental tasks, *Comput Electr Eng*, **99** (2022) 107785.

8 Gondalia A, Dixit D, Parashar S, Raghava V, Sengupta A & Sarobin V R, IoT-based healthcare monitoring system for war soldiers using machine learning, *Procedia Comput Sci*, **133** (2018) 1005–1013.

9 Malasinghe L P, Ramzan N & Dahal K, Remote patient monitoring: a comprehensive study, *J Ambient Intell Humaniz Comput*, **10** (2019) 57–76.

10 Ramkumar P N, Haeberle H S, Ramanathan D, Cantrell W A, Navarro S M, Mont M A, Bloomfield M & Patterson B M, Remote patient monitoring using mobile health for total knee arthroplasty: validation of a wearable and machine

learning–based surveillance platform, *J Arthroplasty*, **34** (2019) 2253–2259.

11 Chatrati S P, Hossain G, Goyal A, Bhan A, Bhattacharya S, Gaurav D & Tiwari S M, Smart home health monitoring system for predicting type 2 diabetes and hypertension, *J King Saud Univ Comput Inf Sci*, **34** (2022) 862–870.

12 Vitabile S, Marks M, Stojanovic S, Pllana S, Molina J M, Krzyszton M, Sikora A, Jarynowski A, Hosseinpour F, Jakobik A & Stojnev I, Medical data processing and analysis for remote health and activities monitoring, in *High-Performance Modelling and Simulation for Big Data Appli* edited by J Kolodziej & H González-Vélez, (Springer, Cham) 2019, 186–220.

13 Nair L R, Shetty S D & Shetty S D, Applying spark based machine learning model on streaming big data for health status prediction, *Comput Electr Eng*, **65** (2018) 393–399.

14 Li H, Shrestha A, Heidari H, Kernec J Le & Fioranelli F, A multisensory approach for remote health monitoring of older people, *IEEE J Electromagn RF Microw Med Biol*, **2** (2018) 102–108.

15 Seshadri Ramana K, Bala Chowdappa K., Obulesu O, Deena babu M & Suresh K, Deep convolution neural networks learned image classification for early cancer detection using lightweight, *Soft Comput*, **26** (2022) 5937–5943.

16 Padmaja M, Shitharth S, Prasuna K, Chaturvedi A, Kshirsagar P R & Vani A, Grow of artificial intelligence to challenge security in iot application, *Wirel Pers* Commun, **127(3)** (2021) 1–17.

17 Sathish K, Narayana Y V & Mekala M S, Efficient tumor volume measurement and segmentation approach for CT image based on twin support vector machines, *Neural Comput Appl*, **34** (2022) 7199–7207.

18 Azimi I, Takalo-Mattila J, Anzanpour A, Rahmani A M, Soininen J P & Liljeberg P, Empowering healthcare IoT systems with hierarchical edge-based deep learning, *Proc of the 2018 IEEE/ACM Int Conf Connected Health: Appl Syst Eng Technol* (IEEE) 2018, 63–68.

19 Mohsin A, Zaidan A A, Zaidan B B, Albahri A S, Albahri O S, Alsalem M & Mohammed K I, Real-time remote health monitoring systems using body sensor information and finger vein biometric verification: A multi-layer systematic review, *J Med Syst*, **42** (2018) 1–36.

20 Devarajan M, Subramaniyaswamy V, Vijayakumar V & Ravi L, Fog-assisted personalized healthcare-support system for remote patients with diabetes, *J Ambient Intell Humaniz Comput*, **10(10)** (2019) 3747–3760.

21 Hassan S R, Ahmad I, Ahmad S, Alfaify A & Shafiq M, Remote pain monitoring using fog computing for e-healthcare: An efficient architecture, *Sensors*, **20** (2020) 6574.

22 Hasan M, Shahjalal M, Chowdhury M Z & Jang Y M, Real-time healthcare data transmission for remote patient monitoring in patch-based hybrid OCC/BLE networks, *Sensors*, **19** (2019) 1208.

23 Mydukuri R V, Kallam S, Patan R, Al-Turjman F & Ramachandran M, Deming least square regressed feature selection and Gaussian neuro-fuzzy multi-layered data classifier for early COVID prediction, *Expert Syst*, **39(4)** (2022) e12694 1–14.

24 Mohan S, Thirumalai C & Srivastava G, Effective heart disease prediction using hybrid machine learning techniques, *IEEE Access*, **7** (2019) 81542–81554.

25 Nalluri M R & Roy D S, Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization, *J Healthc Eng*, **2017** (2017) 1–27, Article ID 5907264, https://doi.org/10.1155/2017/5907264.

26 Rizwan Patan, Suresh K, Amir H G, Thomas H & Ramachandran M, Gaussian relevance vector Map Reduce-based annealed Glowworm optimization for big medical data scheduling, *J Oper Res Soc*, **73(10)** (2022) 2204–2215,

27 Uyar K & İlhan A, Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks, *Procedia Comput Sci*, **120** (2017) 588–593.

28 Ahmed F, An Internet of Things (IoT) application for predicting the quantity of future heart attack patients, *Int J Comput Appl*, **164** (2017) 36–40.