

**DYNAMIC BAYESIAN NETWORK AND HIDDEN MARKOV MODEL OF PREDICTING IOT DATA FOR MACHINE LEARNING MODEL USING ENHANCED RECURSIVE FEATURE ELIMINATION***S. Noeiaghdam*<sup>1,2</sup>, *S. Balamuralitharan*<sup>3</sup>, *V. Govindan*<sup>4</sup><sup>1</sup>Irkutsk National Research Technical University, Irkutsk, Russian Federation<sup>2</sup>South Ural State University, Chelyabinsk, Russian Federation<sup>3</sup>Bharath Institute of Higher Education and Research, Chennai, India<sup>4</sup>DMI St John the Baptist University Central, Mangochi, Malawi

E-mail: snoei@istu.edu, balamuralitharan.maths@bharathuni.ac.in,

govindoviya@gmail.com

The research work develops a Context aware Data Fusion with Ensemblebased Machine Learning Model (CDF-EMLM) for improving the health data treatment. This research work focuses on developing the improved context aware data fusion and efficient feature selection algorithm for improving the classification process for predicting the health care data. Initially, the data from Internet of Things (IoT) devices are gathered and pre-processed to make it clear for the fusion processing. In this work, dual filtering method is introduced for data pre-processing which attempts to label the unlabeled attributes in the data that are gathered, so that data fusion can be done accurately. And then the Dynamic Bayesian Network (DBN) is a good trade-off for tractability becoming a tool for CDF operations. Here the inference problem is handled using the Hidden Markov Model (HMM) in the DBN model. After that the Principal Component Analysis (PCA) is used for feature extraction as well as dimension reduction. The feature selection process is performed by using Enhanced Recursive Feature Elimination (ERFE) method for eliminating the irrelevant data in dataset. Finally, this data are learnt using the Ensemble based Machine Learning Model (EMLM) for data fusion performance checking.

*Keywords:* dynamic bayesian network; hidden markov model; healthcare IoT data; machine learning; principal component analysis; enhanced recursive feature elimination.

**Introduction**

Health care is a key area where ubiquitous applications may be found. Pervasive computing refers to computing that takes place everywhere without participation of users [1, 2]. It alternates the conventional health-care system, which entails identifying symptoms, contacting a doctor, reporting symptoms, and receiving treatment [3]. Internet of Things (IoT) is a relatively novel method in ICT world which allows information to be sent and received using communication networks [4]. Internet of Things is a platform which is built on network of physical items, devices, vehicles, buildings, and so on, all of them are formed by electronic, software, and sensor systems [5, 6]. Kumar et al. [7] introduced a novel technique for improving lung cancer prediction health-care systems via addressing a designated gap by combining locally taught deep learning techniques with block chain technology. Gilula et al. [8] introduced an approach for estimating joint distribution of only the variables of interest directly. Uddin et al. [9] proposed using DRNN, a powerful DL method based on sequential information of a body sensor-based method for behavior detection. They combine data from a variety of body sensors, including electrocardiography (ECG), accelerometer, magnetometer, and others.

The technique proposed by Dautov et al. [10] was implemented using Complex Event Processing techniques supports hierarchical processing approach natively and concentrates on managing streaming data “on the fly”, which is major necessity for storage-constrained IoT devices and time-critical application areas. Begum et al. [11] used sensor signal fusion and case-based reasoning to categorize physiological sensor signals. The proposed method was tested using sensor data fusion to identify people as Stressed or Relaxed. During data collection phase, physiological sensor signals such as Heart Rate (HR), Finger Temperature (FT), Respiration Rate (RR), Carbon dioxide (CO<sub>2</sub>), and Oxygen Saturation (SpO<sub>2</sub>) are gathered. Sensor fusion is accomplished in 2 means:

- (1) decision-level fusion using features extracted by conventional methods,
- (2) data-level fusion using features extracted by employing Multivariate Multi-scale Entropy (MMSE).

The categorization of the signals is done using Case-Based Reasoning (CBR). In comparison to an expert in the subject, the developed approach can correctly diagnose Stressed or Relaxed individuals 87,5 % of the time. As a consequence, it showed potential in psychophysiological area, as well as it may be feasible to apply the technique to erstwhile important health care methods in future. In fog computing environment, Muzammal et al. [12] suggested data fusion supported Ensemble approach for working with clinical information gathered from BSNs. An comprehensive research study backs up the solution’s applicability, as well as the findings are encouraging, as they obtained 98 % accuracy whenever tree depth is equivalent to 15, the estimators number is 40, as well as the prediction work is based on 8 features. And the work [13] developed smart approaches for categorization of activities of daily living (ADL), which rely on data from inertial sensors fixed in user device. Sensitivity index for categories of falls and ADL studied in this article is 0,81, whereas specificity index is 0,98.

## 1. Proposed Model

Using IoT devices, healthcare systems may gather data from patients over a long phase of time. This research work focuses on developing the improved context aware data fusion and efficient feature selection algorithm. Finally this data are learnt using the Ensemble based Machine Learning Model (EMLM) for performance checking. Here the Enhanced Neural Network (ENN), Modified Extreme Gradient Boost Classifier (MXGB) and Logistic regression model (LR) are combined to construct a predictive model (ensemble model) for predicting the healthcare data. The detailed explanation of the proposed method is presented in next part. Fig. 1 shows overall process of the proposed methodology.

### 1.1. Algorithm

The Kalman filter is the statistical state estimation technique whereas the particle filter is a stochastic method to estimate moments.

- Kalman filter (KF)

KF is a widely used statistical state estimate technique for fusing dynamic signal level information. The system’s state estimations are calculated using a recursively implemented prediction and update method, which believes that the present state of a system is dependent on the preceding time interval’s state. KF can be used to identify postural sway throughout quiet standing (standing in one spot while doing no other activity or leaning on something).

- Particle Filtering (PF)

Depending on accelerometer and gyroscope information, PF may be used to estimate biomechanical condition.

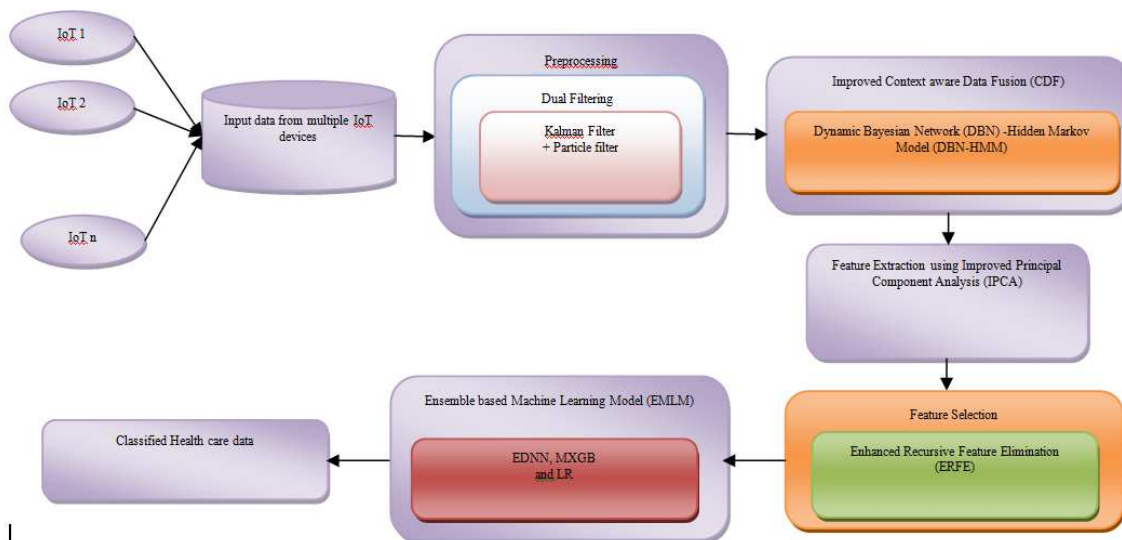


Fig. 1. The overall process of the proposed model

Let each data instance  $x_j$  be a multi-label set  $l_j = \{l_{jk}\}_{k=1}^R$ . For binary classifications, denote by  $C(+)$  and  $C(-)$  positive and negative labels in the set  $l_i$ , respectively. If  $pb(+)$  and  $pb(-)$  are positive and negative label probabilities in the set  $l_i$ , then

$$pr(+)=\frac{(C(+)+1)}{(C(+)+C(-)+2)}, \tag{1}$$

$$pr(-)=\frac{(C(-)+1)}{(C(+)+C(-)+2)} \tag{2}$$

for a Laplace correction that is applied. If  $C(+)$  is very near to  $C(-)$ , then the margin between classes  $|pr(+)-pr(-)|$  is small. Hence, for an instance  $x_i$ , if  $|pr(+)-pr(-)|$  is small, then inference algorithms may not integrate the instance and needs to be filtered. The proposed work uses Algorithm 1 which is listed below. Lines 1 to 7 execute preliminary filtering using  $|pr(+)-pr(-)|$ . The second level of filtering is in Line 8, while Lines 9 to 13 correct noisy labels and Lines 14 to 16 return noiseless data.

### 1.2. Context Aware Data Fusion (CADF)

DBN is used mainly to infer states of a known feature of interest and stands for the hidden variable  $v_t$ . Updates are performed based on sensory readings and their contexts.  $S_t = (S_t^1, \dots, S_t^n)$  is the set of sensory readings active in the time interval  $t$  and contexts set is represented by  $Cn_t = (Cn_t^1, \dots, Cn_t^n)$  based on the application’s environment. The probability distribution  $Pb(S_t|v_t)$  represents how sensor information is affected by system’s present state or sensor model while its state transition model is  $Pb(v_t|v_{t-1}, Cn_t)$ , indicates the probability that a state variable has a specific value, considering its prior value and current context. The DBN used is a first-order Markov model and a given system state in the time interval  $t$ , that is,  $v_t$  can be defined as

$$Bl(v_t)=Pb(v_t|S_{1:t}, Cn_{1:t}). \tag{3}$$

Bayes Filter analogous procedure is followed for a practical formulation of belief and via using Bayes rule it is likely to state equation (3) as

$$Bl(v_t)=Pb(v_t|S_{1:t}, Cn_{1:t})=Pb(v_t|S_{1:t-1}, S_t, Cn_{1:t})= \eta \cdot Pb(S_t|v_t, S_{1:t-1}, Cn_{1:t}) \cdot Pb(v_t|S_{1:t-1}, Cn_{1:t}), \tag{4}$$

**Algorithm 1.** Pre-processing Noisy Data set for Noise Reductions

**Input:**  $\hat{D} = \{(x_1, \hat{y}_i)\}_{i=1}^N$  is a training data set with integrated labels;  $\{l_i\}_{i=1}^N$  are the multiple label sets of  $\hat{D}$ ;  $\delta$  is a threshold

**Output:**  $\hat{D}$  is the corrected data set

1.  $A$  is an empty set
2. for  $i = 1$  to  $N$  do
3. Account the numbers of positive labels and negative labels in  $l_i$ , i.e.,  $C(+)$  and  $C(-)$ , respectively
4. Calculate  $pr(+)$  and  $pr(-)$
5. If  $|pr(+)-pr(-)| \leq \delta$
6. The instance  $i$  is added to the set  $A$
7. End for
8. A filter is applied to the set  $\hat{D}/A$  and all instances filtered out by the filter form a set  $B$
9.  $\hat{D}_c = \hat{D}/(A+B)$
10. Construct a classification model  $f$  on the set  $\hat{D}_c$
11. for  $i = 1$  to size of  $(A+B)$  do
12. Use the classifier  $f$  to relabel the instance  $i$  in the set  $A+B$
13. End for
14. Update the set  $A+B$  to  $\bar{A}+\bar{B}$  with corrected labels
15.  $\hat{D} = \hat{D}_c + \bar{A} + \bar{B}$
16. Return  $\hat{D}$  as the corrected data set

where  $\eta$  is a normalizing constant. In a Markov assumption, the sensor nodes in  $S_t$  do not rely on the context variables  $Cn_t$ , in  $S_t$ , a state variable and assuming sensor measurements are mutually independent, the parent node value of  $S_t$  can be expressed as

$$Pb(S_t|v_t, S_{1:t-1}, Cn_{1:t}) = Pb(S_t|v_t, Cn_{1:t}) = Pb(S_t|v_t) = \prod_{s_t^i} Pb(s_t^i|v_t), \tag{5}$$

where  $s_t^i$  is the specific value of the sensor  $i$  in the time interval  $t$ . Furthermore, the last term in equation (4) is also presented as

$$Pb(v_t|S_{1:t-1}, Cn_{1:t}) = \sum_{v_{t-1}} Pb(v_t, v_{t-1}|S_{1:t-1}, Cn_{1:t}) = \alpha \cdot \sum_{v_{t-1}} Pb(v_t|v_{t-1}, S_{1:t-1}, Cn_{1:t}) \cdot Pb(v_{t-1}|S_{1:t-1}, Cn_{1:t}), \tag{6}$$

where  $\alpha$  is a normalizing constant.  $Cn_t$  can be carefully neglected from the last term, as  $V_{t-1}$  do not rely on the next context  $Cn_t$  when the next state  $V_t$  is not considered. Therefore, using Markov assumptions, equation (6) can be expressed as

$$Pb(v_t|S_{1:t-1}, Cn_{1:t}) = \alpha \cdot \sum_{v_{t-1}} Pb(v_t|v_{t-1}, Cn_t) \cdot Pb(v_{t-1}|S_{1:t-1}, Cn_{1:t-1}) = \alpha \cdot \sum_{v_{t-1}} Pb(v_t|v_{t-1}, Cn_t) \cdot Bl(v_{t-1}). \tag{7}$$

By substituting equations (5) and (7) in (4), belief is described with the recursion

$$Bl(v_t) = \eta \cdot \prod_{s_t^i} Pb(s_t^i|v_t) \cdot \sum_{v_{t-1}} Pb(v_t|v_{t-1}, Cn_t) \cdot Bl(v_{t-1}), \tag{8}$$

where  $\alpha$  is integrated with the normalization constant  $\eta$ . Using equation (6), inference is executed via storing DBN two slices, where time as well as space updating network's

belief are not dependent on the length of sequence. Computational complexity involved in equation (8) is  $O(n + m)$ , where  $n$  is a number of sensors,  $m$  is a number of possible values of  $V_t$  and overall complexity of  $Bl(v_t)$  for all  $V_t$  is  $O(m^2 + m \cdot n)$ . However, the inference issue in DBNs is identical to inference trouble in BNs, in which the desired quantity is a posterior marginal distribution of a collection of hidden variables with an order of observations (up to date of belief):

$$P(X_h[t]|X_0[1], \dots, X_0[\tau]).$$

Here  $X[t] = \{X_h[t], X_0[t]\}$  is a collection of time evolution variables, where  $X_0[t]$  and  $X_h[t]$  specify observed and hidden variables, respectively. According to the time frame of observation used in calculations, time series inference is referred to as filtering ( $\tau = 1$ ), smoothing ( $\tau > 1$ ) and the forecast ( $\tau < 1$ ).

• Hidden Markov Model (HMM) consists of a finite number of states ( $N$ ), everyone with its own probability distribution. A collection of probabilities known as transition probabilities governs transitions between states following processes must be completed to construct a word recognition model that is formed on HMMs:

- 1) choose a number of states and observations,
- 2) select HMM topology,
- 3) choose training and samples,
- 4) train the system using training data,
- 5) perform testing using testing data.

Fig. 2 presents an example of a seven-state HMM that only allows transitions to same state, next state, as well as subsequent states. HMMs are often defined by the letter  $k$  and

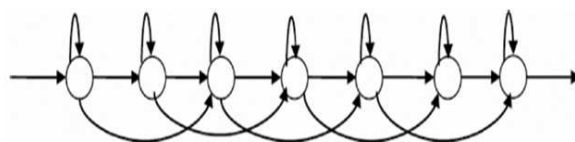


Fig. 2. 7-state Hidden Markov model (HMM)

are specified by a 3-parameter set  $k = (A, B, p)$ , where  $A, B$ , and  $p$  are parameters.

- $A$  is a transition probability matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad (9)$$

$$A = \{a_{ij} | a_{ij} = P(S_t = j) | S_{t-1} = i\}, \quad (10)$$

$$a_{mn} = P(S_n | S_m); \quad m, n = 1, 2. \quad (11)$$

Here  $a_{mn}$  represents a probability that the current state  $S_n$  is provided by the previous state  $S_m$ . The value of  $a_{mn}$  is computed as ratio of the expected number of transitions from  $S_m$  to  $S_n$  to the expected number of transitions out of the state  $S_m$ .

- $B$  is an emission probabilities matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}, \quad (12)$$

$$B = \{b_j(o_k) | b_j(o_k) = P(O_t = o_k | S_t = j)\}, \quad (13)$$

$$b_{np} = b_n(p) = P(O_p | S_n); \quad n = 1, 2; \quad p = 1, 2, 3. \quad (14)$$

Here  $b_n(p)$  is a probability that the present observation  $O_p$  is given by the present state  $S_n$ . The value of  $b_n(p)$  is computed as ratio of the expected number of times, where  $O_p$  is observed with  $S_n$ , to the expected number of times in the state  $S_n$ .

- $\pi$  represents the initial states probabilities

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix}, \tag{15}$$

$$\pi = \{\pi_i | \pi_i = P(S_1 = i)\}, \tag{16}$$

$$\pi_m = P(S_m); m = 1, 2. \tag{17}$$

Due to several advantages of the HMMs, the adjustment procedure for  $k = (A, B, p)$  is as follows.

1. Set  $\lambda = (A, B, \pi)$  arbitrarily:  $a_{mn}$  is  $1/N$ ,  $b_{mp}$  is  $1/M$ ,  $\pi_m$  is  $1/N$ .
2. Compute the parameters  $\alpha_t(m), \beta_t(m), \xi_t(m, n)$  and  $\gamma_t(m)$ .
3. Compute model's novel parameters  $\lambda^* = (A^*, B^*, \pi^*)$  using values computed at Step 2:

$$a_{mn}^* = \frac{\sum_{t=1}^T \xi_t(m, n)}{\sum_{t=1}^T \gamma_t(m)}, \quad b_n^*(p) = \frac{\sum_{t=1, o_t=O_p}^T \gamma_t(n)}{\sum_{t=1}^T \gamma_t(n)}, \quad \pi_m^* = \gamma_1(m).$$

4. Compute  $P(V|\lambda^*)$ . As the probability  $P(V|\lambda^*)$  is rising, repeat Steps 2 and 3.

The model parameters describe a model that best matches the training observation sequences once they converged to particular values. Owing to the synchronisation of both streams (DBN and HMM) at every time interval, all observation sequences in the proposed enhanced context aware data fusion architectures must perform better than the present fusion process.

### 1.3. Feature Extraction

This level is split into two halves. The first one performs measurements in the time or frequency domains. Signal itself, or preliminary information required to determine the features, can be used to make these measurements. Fig. 3 presents feature extraction procedure using the proposed method.

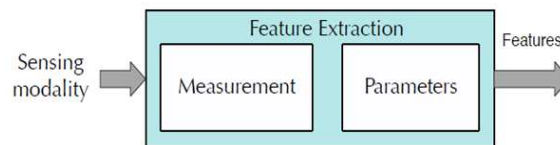


Fig. 3. Feature extraction process

- Improved Principal Component Analysis (IPCA). Whenever a multivariate database is shown as a collection of coordinates in a high-dimensional data space, PCA can provide the viewer with lower-dimensional image, or projection of object from its most relevant viewpoint.
- Constructing the Adaptive Gaussian Kernel Matrix. Consider a collection of  $s$  nodes  $V = \{v_i, 1 \leq i \leq s\}$ , each of which may interact with the central coordinator  $v_0$  in a distributed setup. A local data matrix  $P_i \in R^{n_i \times d}$  with  $n_i$  data points in the dimension  $d$ , where  $n_i > d$ , exists on every node  $v_i$ . A global data matrix  $P \in R^{n \times d}$  is formed by concatenating the local data matrix, that is  $P^T = [P_1^T, P_2^T, \dots, P_s^T]$  and  $n = \sum_{i=1}^s n_i$ . The  $i$ -th row of  $P$  is denoted by  $p_i$ . Assume that the data points are centred so that the mean is zero, i.e.,  $\sum_{i=1}^s p_i = 0$ .

Consider a nonlinear transformation  $\phi(x)$  from the original  $D$ -dimensional feature space to the  $M$ -dimensional feature space, where generally  $M \gg D$ . After that every data point  $x_i$ , a point  $\phi(x_i)$  is projected. Perform conventional PCA in a new feature space, however be aware that this might be both expensive and inefficient. The advantage of using kernel techniques is that these techniques do not manually calculate  $\phi(x_i)$  explicitly. Create the kernel matrix directly from the training data set  $\{x_i\}$ .

$$k(x, y) = (x^T y)^d \tag{18}$$

or

$$k(x, y) = (x^T y + c)^d, \tag{19}$$

where  $c > 0$  is a constant. The Gaussian kernel is

$$k(x, y) = \exp(-\|x - y\|^2/2\sigma^2), \tag{20}$$

where  $\sigma$  is an Adaptive gaussian kernel parameter. Whilst  $k = 1$  and the centre is an  $r$ -dimensional subspace, PCA is an unique case. Top  $r$  right singular vectors of  $P$ , defined as key components, span this optimum  $r$ -dimensional subspace, which may be determined via the Singular Value Decomposition (SVD).

#### 1.4. Feature Selection Using Enhanced Recursive Feature Elimination (ERFE)

**Algorithm 2.** Representation of ERFE method procedure for removing unnecessary information

<p><b>Input:</b> Dimensionality Reduced data  <b>Output:</b> Relevant features (discard the irrelevant data)</p> <ol style="list-style-type: none"> <li>1. Train the classification model by entire features with cross validation</li> <li>2. Compute model performance</li> <li>3. Compute feature importance or ranking</li> <li>4. For every subset <math>T_i</math>, <math>i = 0, 1, 2, 3, \dots, n</math>, do</li> <li>5. Maintain the most vital <math>T_i</math> features</li> <li>6. Update the adaptive learning function stage using (20)</li> <li>7. Re-compute the model's performance</li> <li>8. Re-compute the ranking importance of every feature</li> <li>9. end</li> <li>10. Determine optimal number of features</li> </ol>
---

So this work focuses on developing the new method (namely, ERFE), which redefines the criteria for deleting features from every state.

- Adaptive learning function based RFE. Here the adaptive learning function  $\phi$  is introduced. As a result, the suggested ERFE enhances generalization accuracy substantially, particularly for small numbers of features.

$$|w_j| = \left| \sum_{i=1}^n \alpha_i y_i x_{ij} \right|, \tag{21}$$

where  $x_{ij}$  is the  $j$ -th element of the  $i$ -th feature vector.

Following the calculation of (21) for each of  $P$  features, they may be ranked in the order of significance (high-value means more significant). An adaptive learning function  $\phi$  used to combine the weights was defined as:

$$\phi(w_j, r_i) = w_j \frac{1}{\text{rank}(r_i)}, \quad (22)$$

$$\text{rank}(r_i) = |r_l| |r_l \geq r_i|, \quad (23)$$

where  $w_i$  is RFE weight generated by equation (21),  $r_i$  is the outcome of attribute rank for the  $i$ -th feature in equation (22). Rather than using the original ranks provided by findings, the rank function was used to transform them into rank-based form (23). This implies that a feature with highest rank receives weight of 1, a feature with second highest rank receives weight of 2, a feature with third highest rank receives weight of 3 and so on, up to  $P$ .

### 1.5. Ensemble Based Machine Learning Model (EMLM)

Three machine learning-based classification algorithms are employed to classify data in this study. Here the Enhanced Neural Network (ENN), Modified Extreme Gradient Boost Classifier (MXGB) and Logistic regression model are combined to construct a predictive model (ensemble model).

## 2. Enhanced Deep Neural Network (EDNN)

Deep Learning was shown to be a successful approach for producing extremely accurate predictions from complicated data sets. A fuzzy neural network is a learning method that incorporates neural network methods to apply the attributes of fuzzy systems [14].

- Integration with Fuzzy Inference system. The benefit of structured rule-based algorithms is that they may be influenced by subjective data. This allows an analyst to give expert knowledge to the system, perhaps enhancing categorization findings or altering the system's behaviour [15].

Consider the rules that comprise three Takagi and Sugeno-style fuzzy if-then rules.

Rule 1: If  $a$  is  $X_1$ ,  $b$  is  $Y_1$ ,  $c$  is  $Z_1$ , then  $fn_1 = p_1a + q_1b + t_1c + r_1$ .

Rule 2: If  $a$  is  $X_2$ ,  $b$  is  $Y_2$ ,  $c$  is  $Z_2$ , then  $fn_2 = p_2a + q_2b + t_2c + r_2$ .

Rule 3: If  $a$  is  $X_3$ ,  $b$  is  $Y_3$ ,  $c$  is  $Z_3$ , then  $fn_3 = p_3a + q_3b + t_3c + r_3$ .

Below expression defines the Fuzzy based neural network mathematical operations.

**Layer I.** An adaptive node with node function is included.

$$O_{1,i} = \mu_{A_i}(x), \text{ for } i = 1, 2, \quad (24)$$

$$O_{1,i} = \mu_{B_{i-2}}(y), \text{ for } i = 3, 4, \quad (25)$$

$$O_{1,i} = \mu_{C_{i-4}}(z), \text{ for } i = 5, 6. \quad (26)$$

Here  $\mu_{A_i}(x)$ ,  $\mu_{B_i}(x)$  and  $\mu_{C_i}(x)$  are any acceptable parameterized *MFs*,  $O_{l,i}$  is a membership grade of a fuzzy set  $A = \{A_1, A_2, B_1, B_2 \text{ or } C_1, C_2\}$  which shows the degree to which the supplied input  $x$  ( $y$  or  $z$ ) satisfies the quantifier. This layer is called "Premise



Parameters”. Furthermore, any suitably parameterized  $MF$ , such as a generalised “bell function”, can be used as the membership function for  $A$ :

$$\mu_A(x) = \frac{1}{1 + |(x - c_i)/a_i|^{2b}}, \quad (27)$$

where  $\{a_i, b, c_i\}$  is the set of parameters.

**Layer II.** Each node in this layer is indeed a fixed node, with an outcome equal to the sum of entire incoming signals.

$$O_{2,i} = \mu_{Ai}(x)\mu_{Bi}(y)\mu_{ci}(z), \quad i = 1, 2, 3. \quad (28)$$

Every output node indicates a rule’s “Firing Strength”.

**Layer III.** The function of normalization has a fixed node labelled  $N$ :

$$O_{3,i} = \frac{w_i}{w_1 + w_2 + w_3}, \quad i = 1, 2, 3. \quad (29)$$

The results is generally known as “Normalized Firing Strengths”.

**Layer IV.** Adaptive nodes are included:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + t_i z + r_i). \quad (30)$$

Since each node in this layer is a multiplication of the third layer’s Normalized Firing Strength and the outcome of DNN, it is said to be “Consequent Parameters”.

**Layer V.** The layer contains a signal fixed node denoted by  $S$  with a summing function that calculates the DNN network’s total output as a sum of all incoming signals.

$$\text{Overall output } O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \quad (31)$$

Finally, the enhanced DNN model provides the best results.

### 3. Modified Extreme Gradient Boost (MXGB) Classifier

XG Boost (Extreme Gradient Boosting) is a ML approach formed on Gradient Boosting Decision Tree (GBDT) [16] for classification and regression problems. Number of data is referred to as  $m$ , while the number of features is referred to as  $n$ . The prediction before sigmoid function is represented as  $z_i$ , as well as the probabilistic prediction is  $y_i = \sigma(z_i)$ , in which the sigmoid function is represented by  $\sigma(\cdot)$ . It is crucial to remember that the notations differ and  $y_i$  in their analysis are denoted by  $z$ . The true label is denoted by  $y_i$ , while the parameters for the two loss functions are denoted by  $\alpha$  and  $\gamma$ , respectively. Gradients/hessians expressions are recorded in a merged form that is independent on the value of  $y_i$ , as this simplifies program implementation and facilitates vectorization in other programs. In practice, the additive learning goal is:

$$L^{(t)} = \sum_{i=1}^n l(y_i, z_i^{(t-1)} + f_t(x_i)) + \Omega(f_t). \quad (32)$$

The  $t$ -th iteration of training process is denoted by  $t$ . In the equation, note that the notations were replaced. When the second-order Taylor expansion is applied to equation (32), the following results are obtained:

$$L^{(t)} \approx \sum_{i=1}^n \left[ l(y_i, z_i^{(t-1)} + g_i f_t(x_i)) + \frac{1}{2} h_i (f_t(x_i))^2 \right] + \Omega(f_t) \quad (33)$$

$$\propto \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i (f_t(x_i))^2 \right] + \Omega(f_t). \quad (34)$$

The last line is due to the fact that the term  $l(y_i, z_i^{(t-1)})$  may be omitted from learning goal because it has no effect on model fitting in the  $t$ -th iteration. The activation for both loss functions is sigmoid, as well as the following basic characteristic of sigmoid is applied consistently in derivatives:

$$\frac{\partial y}{\partial z} = \frac{\partial \sigma(z)}{\partial z}, \quad (35)$$

$$\frac{\partial y}{\partial z} = \sigma(z)(1 - \sigma(z)), \quad (36)$$

$$\frac{\partial y}{\partial z} = y(1 - y). \quad (37)$$

- Regularization based Adaptive factor (RA). Calculate the optimum weight  $RAw_j^*$  of the leaf  $j$  for a fixed structure  $q(x)$  by

$$RAw_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad (38)$$

where  $h$  is a factor of approximation.

This implies, obviously, that there are approximately choose points. Every data point is weighted by  $h_i$  that indicates the weight; hence, equation (34) may be rewritten as

$$\sum_{i=1}^n \frac{1}{2} h_i (f_t(x_i) - g_i / (h_i))^2 + \Omega(f_t). \quad (39)$$

Define  $I_j = \{i | q(x_i) = j\}$  as an example set of the leaf  $j$ . Then by extending  $\Omega$  as follows, equation (33) may be rewritten as

$$L^{(t)} = \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \quad (40)$$

$$L^{(t)} = \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2. \quad (41)$$

With labels  $g_i/h_i$  and weights  $h_i$ , this is precisely weighted squared loss. Finding candidate splits that meet the requirements is difficult in large data sets.

## 4. Logistic Regression

LR model is used, that is a popular machine learning algorithm, which is frequently used in real-world applications like data mining [17]. The following is a representation of a logistic regression model:

$$\text{prob}(Y = 1) = \frac{e^z}{1 + e^z}. \quad (42)$$

Here,  $Y$  is a binary dependent variable ( $Y = 1$  if an event happens;  $Y = 0$  else),  $e$  is foundation of natural logarithms, whereas  $Z$  is

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p,$$

where  $\beta_0$  is a constant,  $\beta_j$  are coefficients and  $X_j$  are predictors, for  $p$  predictors ( $j = 1, 2, 3, \dots, p$ ).

## 5. Results and Discussion

Database comprises the motion data of fourteen healthy older people aged 66 to 86 who conducted widely scripted tasks while wearing a battery-less, sternum-level wearable sensor. Owing to using a passive sensor, the data is scarce and noisy. Participants were randomly assigned to one of two clinical rooms (S1 and S2). The room setting S1 (Room1) collects data using four RFID reader antennas (1 on ceiling and 3 on the walls), while the room setting S2 (Room2) collects motion data using three RFID reader antennas (2 on the ceiling and 1 at wall level). The proportion of accurately obtained positive observations to all predicted positive observations is known as precision.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}. \quad (43)$$

The proportion of properly detected positive observations to total number of observations is known as sensitivity or recall.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \quad (44)$$

Weighted average of Precision and Recall is described as the F-measure. As a consequence, false positives and false negatives are taken.

$$\text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (45)$$

The accuracy is computed with regards to positives and negatives as follows:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{False Positive}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}. \quad (46)$$

Table 1 presents the performance comparison results for the proposed and existing methods for the data set I.

Figs. 4, 5 show the performance comparison results for the data set I. This graph clearly identifies the comparison performance among suggested and current techniques. Fig. 4 (a) illustrates the precision comparison results of the proposed ICDFT- EMLM for the health care data. From the results we conclude that the proposed ICDFT- EMLM technique has high precision results. Fig. 4 (b) illustrates recall outcome comparisons of the suggested

ICDFT- EMLM model for the health care data. From the results we conclude that the proposed ICDFT- EMLM technique has high recall results. Fig. 5 (a) shows F-measure comparison outcomes of the proposed ICDFT- EMLM model for the health care data. From the results we conclude that the proposed ICDFT- EMLM technique has high F-measure results. Fig. 5 (b) illustrates the accuracy comparison results of the proposed ICDFT- EMLM model for the health care data. From the results we conclude that the proposed ICDFT- EMLM technique has high accuracy results. Figs. 6,7 illustrate the performance comparison results for both data sets. Based on the results, we note that the suggested method has higher results than current methods for both data sets. Table 2 presents the performance comparison results for both data sets. Based on Table 2, we conclude that the introduced approach has greater performance results for data set I whereas the data set II has less. Fig. 6 (a) shows accuracy and precision comparison outcomes for both data sets. And it is clearly identified that the proposed method has higher results than the exiting methods. Fig. 6 (b) illustrates the comparison results for both data sets. The graph shows the high performance results for the data set I whereas the data set II has less for the given metrics.

**Table 1**

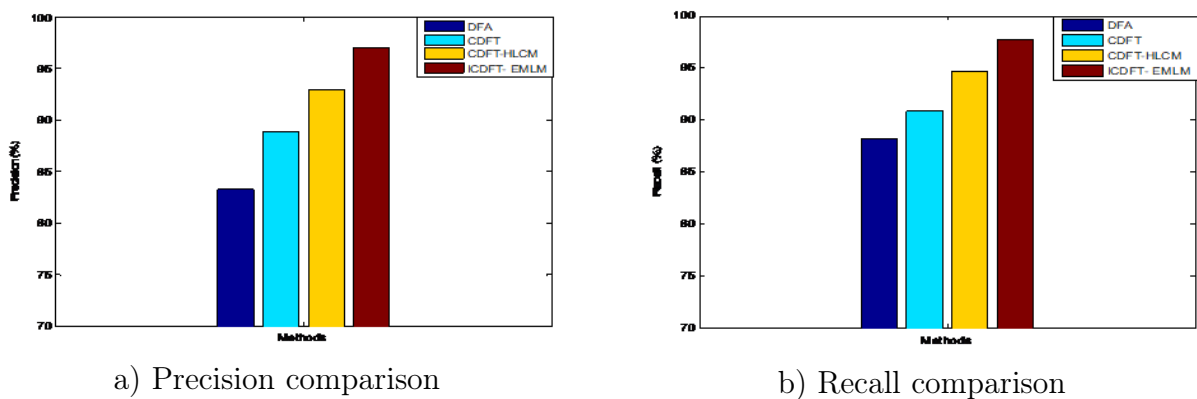
Comparison Results of Performance for Proposed and Existing Approaches

Metrics	DFA	CDFT	CDFT-HLCM	ICDFT-EMLM
Accuracy	86,059	90,950	94,400	97,9003
Precision	83,29	88,92	93,01	96,9710
Recall	88,19	90,86	94,67	97,7730
F-measure	85,67	89,88	93,83	97,3704

**Table 2**

Performance Comparison Results for Proposed and Existing Methods for Both Data sets

Metrics	Dataset – I	Dataset – II
Accuracy	97,9003	95,800
Precision	96,9710	93,420
Recall	97,7730	96,153
F-measure	97,3704	94,766



**Fig. 4.** Results of the proposed ICDFT- EMLM model for the health care data

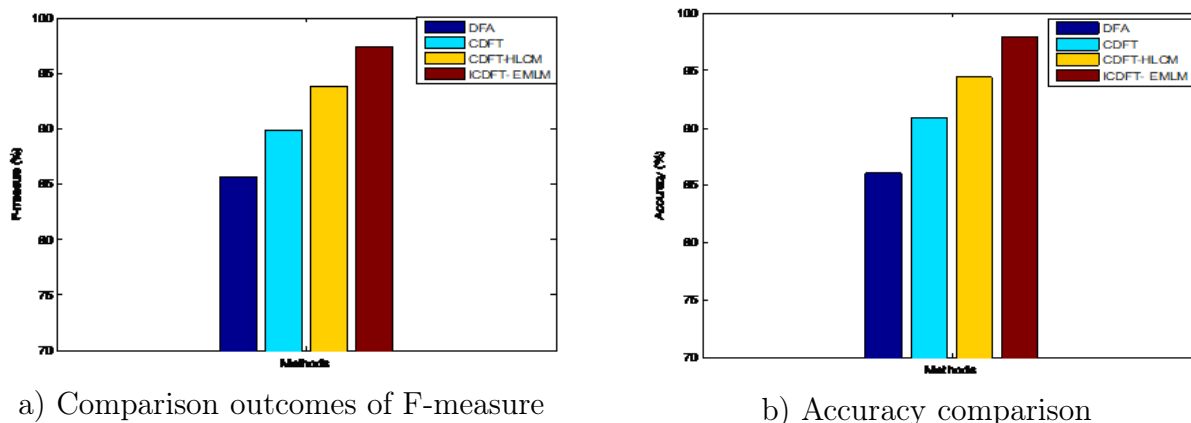


Fig. 5. The proposed ICDFT- EMLM model for the health care data

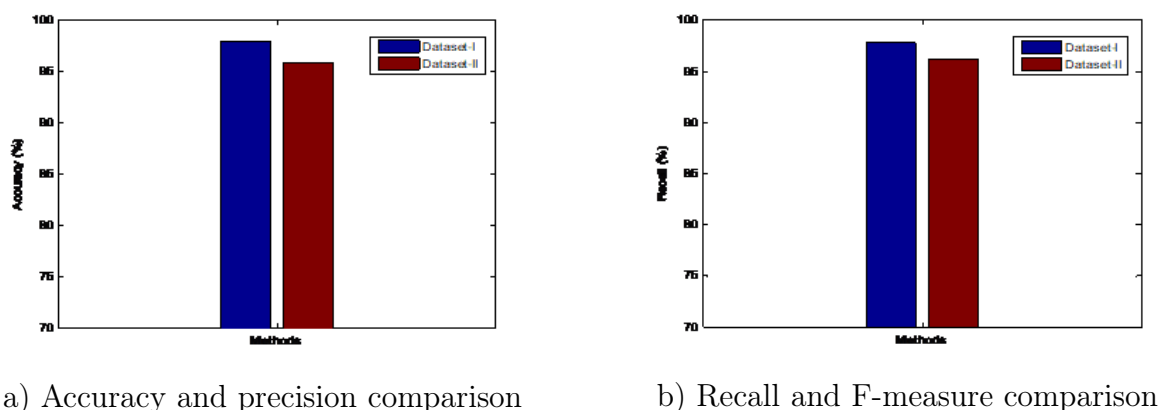


Fig. 6. Comparison results for both data sets

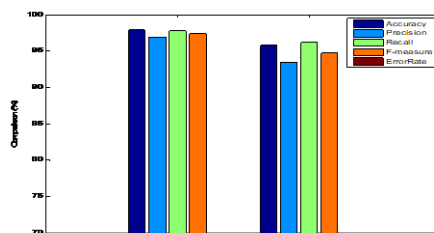


Fig. 7. Performance comparison results for both data sets

## Conclusion

In this work, we introduce a dual filtering method for data pre-processing, which attempts to label the unlabelled attributes in the data that are gathered, so that data fusion can be done accurately. The improved Dynamic Bayesian Network (IDBN) is a good trade-off for tractability becoming a tool for ICDF operations and the inference problem is handled using the Hidden Markov Model (HMM) in DBN model. Thus the proposed HMM method improves the fusion process which increases the prediction performance. After that the Improved Principal Component Analysis (IPCA) is used for feature extraction

as well as dimension reduction. The feature selection process is done by using Enhanced Recursive feature Elimination (ERFE) method for eliminating the irrelevant data in a data set. Finally this data is learnt using the Ensemble based Machine Learning Model (EMLM) for performance checking. Here the Enhanced Neural Network (ENN), Modified Extreme Gradient Boost Classifier (MXGB) and Logistic regression model (LR) are combined to construct a predictive model (ensemble model) for predicting the health care data. Thus the results indicate that the proposed ICDFT-EMLM model improve the prediction performance of the health care data compared to the existing health care applications. Further this work focuses on improving the security level of the health data using cryptography technique as a future work.

## References

1. Hao Jin, Yan Luo, Peilong Li, Jomol Mathew. A Review of Secure and Privacy-Preserving Medical Data Sharing. *IEEE Access*, 2019, vol. 7, pp. 61656–61669. DOI: 10.1109/ACCESS.2019.2916503
2. Perez S., Hernandez-Ramos J.L., Pedone D., Rotondi D., Straniero L. et al. A Digital Envelope Approach Using Attribute-Based Encryption for Secure Data Exchange in IoT Scenarios. *Global Internet of Things Summit*, 2017, pp. 1–6. DOI: 10.1109/GIOTS.2017.8016281
3. Mohanta B.K., Jena D., Sobhanayak S. Multi-Party Computation Review for Secure Data Processing in IoT-Fog Computing Environment. *International Journal of Security and Networks*, 2020, vol. 15, no. 3, pp. 164–174. DOI: 10.1504/IJSN.2020.109697
4. Xueping Liang, Juan Zhao, Sachin Shetty, Jihong Liu, Danyi Li. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications. *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, Montreal, 2017, pp. 1–5. DOI: 10.1109/PIMRC.2017.8292361
5. Theodouli A., Arakliotis S., Moschou K., Votis K., Tzouvaras D. On the Design of a Blockchain-Based System to Facilitate Healthcare Data Sharing. *17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering*, New York, 2018, pp. 1374–1379. DOI: 10.1109/TrustCom/BigDataSE.2018.00190
6. Mikula T., Jacobsen R.H. Identity and Access Management with Blockchain in Electronic Healthcare Records. *21st Euromicro Conference on Digital System Design*, Prague, 2018, pp. 699–706. DOI: 10.1109/DSD.2018.00008
7. Rajesh Kumar, WenYong Wang, Jay Kumar, Ting Yang, Abdullah Khan, Wazir Ali, Ikram Ali. An Integration of Blockchain and AI for Secure Data Sharing and Detection of CT Images for the Hospitals. *Computerized Medical Imaging and Graphics*, 2021, vol. 87, no. 1, article ID: 101812. DOI: 10.1016/j.compmedimag.2020.101812
8. Gilula Z., McCulloch R.E., Rossi P.E. A Direct Approach to Data Fusion. *Journal of Marketing Research*, 2006, vol. 43, no. 1, pp. 73–83. DOI: 10.1509/jmkr.43.1.73
9. Uddin M.Z., Hassan M.M., Alsanad A., Savaglio C. A Body Sensor Data Fusion and Deep Recurrent Neural Network-Based Behavior Recognition Approach for Robust Healthcare. *Information Fusion*, 2020, vol. 55, no. 3, pp. 105–115. DOI: 10.1016/j.inffus.2019.08.004
10. Dautov R., Distefano S., Buyya R. Hierarchical Data Fusion for Smart Healthcare. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 1–23. DOI: 10.1186/s40537-019-0183-6
11. Begum S., Barua S., Ahmed M.U. Physiological Sensor Signals Classification for Healthcare Using Sensor Data Fusion and Case-Based Reasoning. *Sensors*, 2014, vol. 14, no. 7, pp. 11770–11785. DOI: 10.3390/s140711770
12. Muzammal M., Talat R., Sodhro A.H., Pirbhulal S. A Multi-Sensor Data Fusion Enabled Ensemble Approach for Medical Data from Body Sensor Networks. *Information Fusion*, 2020, vol. 53, no. 1, pp. 155–164. DOI: 10.1016/j.inffus.2019.06.021

13. Ando B., Baglio S., Lombardo C.O., Marletta V. A Multisensor Data-Fusion Approach for ADL and Fall Classification. *IEEE Transactions on Instrumentation and Measurement*, 2016, vol. 65, no. 9, pp. 1960–1967. DOI: 10.1109/TIM.2016.2552678
14. Sun-Chong Wang. Artificial Neural Network. *Interdisciplinary Computing in Java Programming*, Springer, Boston, 2003, pp. 81–100. DOI: 10.1007/978-1-4615-0377-4\_5
15. Kondratenko Y.P., Klymenko L.P., Al Zu'bi E.Y.M. *Structural Optimization of Fuzzy Systems' Rules Base and Aggregation Models*, Kybernetes, 2013.
16. Chen T., He T., Benesty M., Khotilovich V., Tang Y., Cho H. Xgboost: Extreme Gradient Boosting. *R Package Version*, 2015, vol. 1, no. 4, pp. 1–4.
17. Chao-Ying Joanne Peng, Kuk Lida Lee, Ingersoll G.M. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 2002, vol. 96, no. 1, pp. 3–14. DOI: 10.1080/00220670209598786

*Received 25 February, 2022*

УДК 519.217

DOI: 10.14529/mmp220308

## ДИНАМИЧЕСКАЯ БАЙЕСОВСКАЯ СЕТЬ И СКРЫТАЯ МАРКОВСКАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ДАННЫХ IOT ДЛЯ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ РАСШИРЕННОГО РЕКУРСИВНОГО ИСКЛЮЧЕНИЯ ПРИЗНАКОВ

*С. Нойягдам<sup>1,2</sup>, С. Баламуралитаран<sup>3</sup>, В. Говиндан<sup>4</sup>*

<sup>1</sup>Иркутский национальный исследовательский технический университет, г. Иркутск, Российская Федерация

<sup>2</sup>Южно-Уральский государственный университет, г. Челябинск, Российская Федерация

<sup>3</sup>Институт высшего образования и исследований Бхарата, г. Ченнаи, Индия

<sup>4</sup>DMI Центральный университет Святого Иоанна Крестителя, г. Мангочи, Малави

В рамках исследовательской работы разработано слияние данных с учетом контекста с моделью машинного обучения на основе ансамбля (CDF-EMLM) для улучшения обработки данных о здоровье. Эта исследовательская работа сосредоточена на разработке улучшенного слияния данных с учетом контекста и алгоритма эффективного выбора признаков для улучшения процесса классификации для прогнозирования данных здравоохранения. Первоначально данные с устройств интернета вещей (IoT) собираются и предварительно обрабатываются, чтобы сделать их понятными для обработки слияния. В этой работе построен метод двойной фильтрации для предварительной обработки данных, который пытается пометить немаркированные атрибуты в собранных данных, чтобы можно было точно выполнить объединение данных. Кроме того, динамическая байесовская сеть (DBN) является хорошим компромиссом для манипулирования и становится инструментом для операций CADF. Здесь проблема вывода решается с использованием скрытой марковской модели (НММ) в модели DBN. После этого анализ основных компонентов (РСА) используется для извлечения признаков, а также для уменьшения размеров. Выбор признаков выполняется с использованием метода расширенного рекурсивного исключения признаков (ERFE) для устранения нерелевантных данных в наборе данных. Наконец, эти данные изучаются с использованием модели машинного обучения на основе ансамбля (EMLM) для проверки производительности слияния данных.

*Ключевые слова:* динамическая байесовская сеть; скрытая марковская модель; IoT данные здравоохранения; машинное обучение; анализ главных компонентов; расширенное рекурсивное устранение признаков.

Самад Нойягдам, PhD, лаборатория промышленной математики, Байкальский институт БРИКС, Иркутский национальный исследовательский технический университет (г. Иркутск, Российская Федерация); кафедра прикладной математики и программирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), snoei@istu.edu.

Сундарапан Баламуралитаран, PhD, кафедра математики, Институт высшего образования и исследований Бхарата (г. Ченнаи, Индия), balamuralitharan.maths@bharathuni.ac.in.

Ведияпан Говиндан, PhD, кафедра математики, DMI Центральный университет Святого Иоанна Крестителя (г. Мангочи, Малави), govindoviya@gmail.com.

*Поступила в редакцию 25 февраля 2022 г.*