Information Fusion 53 (2020) 155–164

Contents lists available at ScienceDirect

Information Fusion

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

Full Length Article

A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks

Muhammad Muzammal [a], Romana Talat [a], Ali Hassan Sodhro [b], Sandeep Pirbhulal [c,*]

[a] Department of Computer Science, Bahria University, Islamabad 44000, Pakistan
[b] IDA-Computer and Information Science Department, Linkoping University, Linkoping SE58183, Sweden
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518118, China

ARTICLE INFO

Keywords:
Multi-sensor data fusion
Body sensor network
Ensemble methods
Disease prediction
Fog computing

ABSTRACT

Wireless Body Sensor Network (BSNs) are wearable sensors with varying sensing, storage, computation, and transmission capabilities. When data is obtained from multiple devices, multi-sensor fusion is desirable to transform potentially erroneous sensor data into high quality fused data. In this work, a data fusion enabled Ensemble approach is proposed to work with medical data obtained from BSNs in a fog computing environment. Daily activity data is obtained from a collection of sensors which is fused together to generate high quality activity data. The fused data is later input to an Ensemble classifier for early heart disease prediction. The ensembles are hosted in a Fog computing environment and the prediction computations are performed in a decentralised manners. The results from the individual nodes in the fog computing environment are then combined to produce a unified output. For the classification purpose, a novel kernel random forest ensemble is used that produces significantly better quality results than random forest. An extensive experimental study supports the applicability of the solution and the obtained results are promising, as we obtain 98% accuracy when the tree depth is equal to 15, number of estimators is 40, and 8 features are considered for the prediction task.

1. Introduction

Recent technological advancements in Wireless Body Sensor Networks (BSNs) [1] and the emergence of data processing techniques such as data fusion [2] have enabled a wide spectrum of possibilities for human-centred applications. The idea of data and information fusion is to predict events from a multitude of sources with higher confidence, which otherwise could not have been detected individually. Thus, observations recorded about the same phenomenon from different sources are combined in a way such that the individual potentially unnoticeable events are fused into meaningful observable events. The fused information is considered to be of more collective value than the individual raw observations and thus, the significance of the readings improves as a result of fusion even if the individual readings have low confidence values.

The design of BSN-based fusion applications is not obvious due to technological and equipment limitations. Typically, such systems work with a continuous stream of data and have to incorporate limited network bandwidth and battery consumption considerations. The sensing hardware is not usually cheap and is constrained by sensor accuracy, limited data storage capabilities, sampling rate, and other concerns. The communication is usually by way of a single-hop star topology network where a set of wearable sensors coordinate with the help of a central device. A variety of BSN frameworks for BSN data processing and coordination have been proposed (see [2] and references therein).

BSNs provide ways to monitor individual activity in a variety of scenarios, for example, activity analysis for detection or prevention of disease [3], rehabilitation after a medical procedure [4], emotion detection during driving [5], integration of BSN data with the environment data [6], integration of BSNs data with social network data [7], and others [8,9]. An illustration of real-time data fusion for wearable body sensor network is shown in Fig. 1. Other notable applications of BSNs and data fusion include vehicle tracking with multimodal data fusion [10], and multi-sensor data fusion enabled smart home [11]. Similarly, uncertainty measures in multi-sensor data [12,13] and stream processing algorithms [14] have also been proposed.

Machine learning techniques such as support vector machine and ensemble methods such as random forest [15] have been employed in a variety of applications domains including e-health for applications such as activity recognition [4], and disease prediction [3]. Recently, it has been shown that the accuracy of better performing machine learning algorithms such as random forests can be improved by the

* Corresponding author.
E-mail addresses: muzammal@bui.edu.pk (M. Muzammal), romanatalat.buic@bahria.edu.pk (R. Talat), ali.hassan.sodhro@liu.se (A.H. Sodhro), sandeep@siat.ac.cn (S. Pirbhulal).

https://doi.org/10.1016/j.inffus.2019.06.021
Received 15 March 2019; Received in revised form 13 May 2019; Accepted 15 June 2019
Available online 19 June 2019

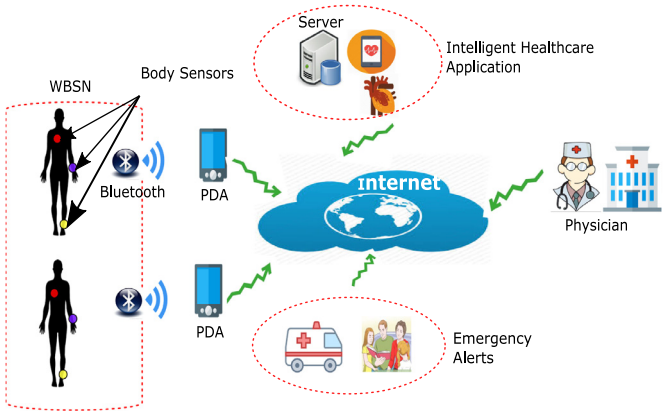

Fig. 1. An illustration of a wireless body sensor network for a smart health-care system.

incorporation of kernel methods. The study [16] suggests that with little effort, random forests could be transformed to kernel random forests which are superior to random forests in overall prediction accuracy. The computing environment for the ensemble is a fog computing environment [17,18] where the individual predictors are hosted on fog nodes and the results are aggregated to obtain the final output [19].

1.1. Highlights

We consider daily activity data which is collected from multiple sensor for an interesting activity recognition scenario to predict the heart disease. The highlights of this work are as follows:

1. The data from multiple sensors is fused to prepare higher quality data which is input to the ensembles for heart disease prediction.
2. The ensembles are placed in a fog computing environment and the results form the individual predictors are combined to produce a unified output.
3. For the prediction task, we consider the novel kernel random forest in a fog computing environment. Kernel random forest are shown to be superior to prevalent ensembles in prediction accuracy.

1.2. Our contribution

Our contributions are as follows:

(1) We propose a multi-sensor data fusion framework for generating unified activity data for the purpose of heart disease prediction.
(2) We propose an ensemble framework for heart disease prediction from the sensor activity data.
(3) We develop a fog-based computing environment, and employ sophisticated Kernel random forest ensembles for heart disease prediction from the sensor activity data.
(4) An extensive empirical study is performed to evaluate the effectiveness of the proposed system. The results show the effectiveness of the fog-based ensemble framework for heart disease prediction from the sensor activity data.

The rest of the paper is organised as follows. Section 2 covers related concepts, and the background information is presented in Section 3. Details about data fusion and ensembles are presented in Section 4. The effectiveness and applicability of the proposed approach is evaluated in Section 5, and Section 6 concludes this work.

2. Related work

Wireless Body Sensor Network (BSN) comprises wearable sensors which communicate with personal devices. The sensors have different sensing and transmission capabilities, storage and computation power [20]. In literature, various aspects of BSNs have been extensively explored. Some of them are multi-sensor fusion and prediction of early health disorders.

2.1. Data and feature level fusion

Based on processing of data, various centralised and distributed data fusion approaches have been proposed in literature. In centralised data fusion, all of the processing is done on a central node, while in distributed environments, all sensors process their own data and transmit it to a central node for data analysis. In BSNs, multiple sensors are placed on human subjects to capture interesting phenomena [21]. Generally, if a system contains multiple sensors which record homogeneous data, then data can be fused directly. On the other hand, if it provides heterogeneous data then some feature and decision level fusion is required [22]. Generally speaking, it is considered that processing unit and storage are reliable; and there is a need for efficient data fusion algorithms which couple data obtained from multiple sensors [23]. Furthermore, various feature-level fusion algorithms have been proposed which extract features from data to create a high dimensional feature vector which is used for classification [24] and recognition [25] purposes.

2.2. Activity recognition

According to world health organisation, approximately 3.2 million people suffer from diseases due to lack of physical activity. The method of assessment of physical activity provides the health awareness and can be used to improve quality of life. A significant effort has been devoted to activity recognition domain. Features are extracted from data obtained from multiple sensors such as magnetometer [26,27], microphone [28], accelerometer [29], and light sensor [30]. Consequently, the optimal subset of feature is selected using methods including correlation-based methods [30], kernel discriminating analysis [26], and relevance heuristics [31]. The selected features are input to a classifier such as Naive Bayes, Neural Network, SVM, or a Decision Tree, to detect the activity.

2.3. Emotion recognition

Many data fusion techniques for emotion recognition have been proposed [32,33]. There are two ways to recognise emotion, (i) through voice or facial [6,34], and (ii) physiological signals [35,36]. Covello et al. [37] proposed a cardiac defence response system which takes heart beat as input to detect fear emotions of human beings using electrocardiogram (ECG) signals with the help of a novel algorithm. The signals in ECG represent heart beat which is used to detect human emotions. Later on, john et al. [38] proposed feature oriented emotion recognition model. This model utilises minimum-Redundancy-maximum relevance selection method in order to improve the kernel based classifier. The main contribution of this model is to combine the statistical feature selection with the classification task. Recently, Hassan et al. [39] introduced human emotion recognition model using deep belief network. The idea is to extract features from data using belief network that can be fused with some statistical features. The extracted features are fed into a Fine Gaussian Support Vector Machine to classify human emotions. Later on, Gravina et al. [40] presented an integrated system which contains body-worn sensors and pressure system to detect in-seat activities. For recognition purpose, frequency and time domain features are extracted which give promising results.

2.4. Fall detection

With the rapid development of Internet-of-things (IoT) technology [41], body sensor networks are capable of monitoring the human vital signs [42,43,44]. Yi et al. [45] presented a system architecture

which integrates and analyses data obtained from multiple body sensors to detect fall detection. With the advancement in bio-medical sensors and network protocol, body sensor network are capable of monitoring human activity, vital signs, and surrounding activities of people. Felisberto et al. [46] presented multi-agent based distributed architecture for monitoring human movement and detecting hazardous activities.

2.5. Heart disease prediction

In recent years, heart failure is one of the major causes of death in the world. With the advancement in machine learning approaches, it is possible to improve the healthcare decision making systems [47,48,49]. Many e-health systems utilise classification techniques to classify the risk levels corresponding to heart disease [50,51]. Kim et al. [51] proposed a fuzzy-based heart disease prediction model which provides content recommendation for heart patients. An electrocardiogram plays a significant role in heart disease detection. For a cardiologist, any disturbance in heart rate or pattern of recorded ECG can be an indication of heart disease. Kiranyaz et al. [52] applied 1-D convolutional neural network (CNN) to classify the ECG for real-time ECG monitoring. Mustaqeem et al. [53] introduced a prediction model for classification of heart disease. In order to select best features, they considered a wrapper algorithm and applied a number of machine learning algorithms.

3. Preliminaries

In this section, we present the fundamental concepts requisite to the methodology presented in the subsequent section. A list of useful notation is in Table 1. We first discuss sensor data fusion.

3.1. Sensor data fusion

Let $\mathbb{A} = \{\alpha_1, \ldots, \alpha_m\}$ be the sensors and $\Gamma = \{\gamma_1, \ldots, \gamma_m\}$ be the corresponding weights such that $\sum \gamma = 1$. Under the assumption that the sensors are working without failure, the output of the system is computed as,

$$\mathbb{O} = \{\alpha_1\gamma_1, \ldots, \alpha_m\gamma_m\}.$$

If a sensor fails during any time in operation, the output of the system is adjusted as,

$$\mathbb{O}' = \{\beta_1\gamma_{n1}, \ldots, \beta_n\gamma_{ni}\}, \tag{1}$$

where γ_i are the working sensors and $\sum \beta = 1$ is maintained. For simplicity, the weights are distributed uniformly for all the sensors, i.e., $1/n$, for the working sensors. When a sensor fails, the data from the sensor is discarded. The data received from other sensors is input to an updated configuration (Eq. 1) for data fusion. For example, if calories consumed from a particular sensor are not available, the output from other sensors is assigned updated weights for the adapted fusion settings.

Table 1
A List of frequently used notation.

Symbol	Meaning
D	Dataset in the form of (X, Y) pairs; X: random variable; Y: response;
E(X)	Entropy of a random varible X
P	Number of trees in the forest
x	Prediction instance
Δ	Randomness factor
Θ	Information gain
ω	Weight factor
τ	Number of x instances for Δ_j
ϕ	Number of data points in forest
σ	Feature set
λ	Normalised information gain

3.2. Computing environment

The computing environment for the proposed framework is at the core of the solution design. As mentioned already, the proposed solution comprises of the following three components:

1. Wearable body sensors and a body sensor network is responsible for collecting the object data at pre-specified intervals. Note that wearable sensors have limited memory, power, storage, and communication capabilities and therefore, can't handle real-time data. Consequently, the data collection is either on the detection of an activity or at pre-specified intervals.
2. Data collected from the sensors is stored on a cloud server which serves as a backup for data storage and for the communications between the data and the fog computing environment.
3. The third component is the fog computing environment which facilitates communication between the wearable sensors and the system to detect activities by way of fog computing services. In essence, the fog processing unit performs online processing. However, for the study, we consider offline data processing due to limitations in obtaining live data from commercial sensors. The idea of the fog computing nodes is to exploit the inherent parallelism in the ensembles by hosting individual ensembles on different fog nodes. Note that a fog computing unit has limited storage and processing power as compared to processing units on cloud. Also, the computing which essentially is performed on the fog node, i.e., algorithms and prediction methods are assumed to be of moderate complexity, computationally. As the fog node serves as an intermediary, data is deleted from the fog node once it has been processed or already sent to the cloud storage.

3.3. Data fusion enabled ensemble

In this section, we present two ensembles that we consider for this study, (i) Random Forest and (ii) Kernel Random Forest. We first discuss Random Forest.

3.3.1. Random forest

Let $X = \{X_1, \ldots, X_n\}$ be the set of random variables, and $Y = \{Y_1, \ldots, Y_n\}$ be the set of responses. A function $f(\mathrm{x}) = \mathbb{E}[\mathrm{Y}|\mathrm{X} = \mathrm{x}]$ predicts the response Y for the random variable X. The records in the dataset $D = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of $[0, 1]^d \times \mathbb{R}$-values are independent pairs of the form (X, Y), where $\mathbb{E}[Y^2] < \infty$. We use infinite random forest to compute $f_{\infty,n} : [0, 1]^d \longrightarrow \mathbb{R}$ of f, for the dataset D. For a collection of P random trees, the predicted value for the k-th tree in the collection at point x is $f_n(\mathrm{x}, \Delta_j)$, where $\Delta_1, \ldots, \Delta_P$ are independent random variables of the dataset D. A unified finite forest is obtained by aggregating the outputs from the individual trees:

$$f_{P,n}(\mathrm{x}, \Delta_1, \ldots, \Delta_P) = \frac{1}{\mathrm{P}} \sum_{j=1}^{P} f_n(\mathrm{x}, \Delta_j). \tag{2}$$

For all $\mathrm{x} \in [0, 1]^d$, let the expectation with respect to Δ on D be $\mathbb{E}_\Delta$, then law of large number suggests that the finite forest case is almost equivalent to the infinite case, i.e.,

$$f_{\infty,n}(\mathrm{x}) = \mathbb{E}_\Delta[f_n(\mathrm{x}, \Delta)].$$

We are interested in the RHS of the above equation, and we use the same in this work.

3.3.2. Kernel random forest

For all $x \in [0, 1]^d$, let $\phi_n(X, \Delta_j)$ contain x, Δ_j be the randomness factor, the following holds for the dataset D,

Algorithm 1 An outline of the Random Forest algorithm.

Require: Dataset D in the form of (X, Y) pairs, number of trees P, $f \in \{1, \ldots, P\}$, $\tau_n \in \{1, \ldots, n\}$, $\mathrm{x} \in [0,1]^P$
Ensure: Prediction of the random forest at x

```
for each j ∈ M do
    Choose τ_n points from D
    For all l ∈ τ_n, set ρ_l = φ, ρ_0 = [0,1]^P        ▷ root partition
    Set η_v = 1, ψ = 0                                  ▷ η: number of vertices; ψ: level
    while η_v < τ_n do
        if ψ ≠ φ then
            ρ ← point x
            if ∑ρ = 1 then
                ρ_ψ ← ρ_ψ ⋃ ρ
            else
                Generate and split the set ρ into ρ_A, ρ_B
                ρ_{ψ+1} ← ρ_{ψ+1} ⋃ ρ_A ⋃ ρ_B            ▷ ρ updated as a result of split into ρ_A and ρ_B
                η_v ← η_v + 1
            end if
        else
            ψ ← ψ + 1
        end if
    end while
    Compute f(x, Δ_j, D) for x                          ▷ local prediction for x
end for
Compute f_{P,n}(x, Δ_1, ..., Δ_P, D)                   ▷ global prediction for x
return
```

Algorithm 2 An outline of the ensemble framework.

Require: Raw Dataset D', feature set σ, k
Ensure: Prediction of class label for record x

```
D ← D'                                  ▷ (Data fusion of raw data as in Section 3)
σ' ← FeatureSelection(D, σ, k)
M ← TrainModel(D, σ')                   ▷ Train model using RF/KeRF
y ← M(x)                                ▷ predict class label for record x
Output class label y for x
return

function FeatureSelection(D, σ, k)
    for each feature S in σ do
        Compute importance score σ_S                    ▷ by Eq. 7
        Compute E and Θ for all σ in D                  ▷ by Eqs. 9 and~10
        Compute normalised gain λ                       ▷ by Eq. 11
        return σ_k ⊆ σ                                  ▷ return top-k features
    end for
end function
```

$$f_{P,n}(\mathrm{x}, \Delta_1, \ldots, \Delta_P) = \frac{1}{P} \sum_{j=1}^{P} \left(\sum_{i=1}^{n} \frac{\mathrm{Y}_i \mathbb{I}_{\mathrm{x}_i \in \phi_n(\mathrm{x}, \Delta_j)}}{\tau_n(\mathrm{x}, \Delta_j)} \right).$$

$$\tau_n(\mathrm{x}, \Delta_j) = \sum_{i=1}^{n} \mathbb{I}_{\mathrm{x}_i \in \phi_n(\mathrm{x}, \Delta_j)},$$

where $\phi_n(\mathrm{x}, \Delta_j)$ is the number of data points. For each observation Y_i, we compute the weights $\omega_{i,j,n}(\mathrm{x})$ by

$$\omega_{i,j,n}(\mathrm{x}) = \frac{\mathbb{I}_{\mathrm{x}_i \in \phi_n(\mathrm{x}, \Delta_j)}}{\tau_n(\mathrm{x}, \Delta_j)}.$$

As the factor $\tau_n(\mathrm{x}, \Delta_j)$ determines the weights, fewer data points in a pre-determined region contribute relatively higher to the weight factor. Therefore, when randomness is at the core of the forest, higher weights to observations in regions with fewer data points may potentially lead to rough set estimates. In particular, empty regions do not contribute to the classifier and may cause prediction error.

The prediction error introduced by the sparse regions and the forest weight problem is addressed by considering Kernel Random Forests (KeRF). For all $\mathrm{x} \in [0,1]^d$,

$$\tilde{f}_{P,n}(\mathrm{x}, \Delta_1, \ldots, \Delta_P) = \frac{\sum_{j=1}^{P} \sum_{i=1}^{n} \mathrm{Y}_i \mathbb{I}_{\mathrm{x}_i \in \phi_n(\mathrm{x}, \Delta_j)}}{\sum_{j=1}^{P} \tau_n(\mathrm{x}, \Delta_j)}. \tag{3}$$

The left hand side of Eq. 3 is equivalent to the mean of the data points in a region that contain x and Y_i. Thus, a region that does not contain x does not participate in prediction and consequently does not contribute to the prediction error.

4. Data fusion enabled fog computing-based ensembles

In this section, we give the framework for data fusion enabled ensembles in a fog-computing environment. We first discuss data collection, followed by data fusion, and the prediction framework. An overview of the proposed framework is given in Fig. 3.

4.1. Data collection

We collect activity data with the help of wearable sensors, as shown in Table 2. The sensors are placed on human body as the proposed framework utilises 5G network to transfer data from sensors to a paired personal devices. Table 2 gives an illustration of the kinds of sensors used in this study and their description. The dataset is created from a local hospital where the subjects are selected with even distribution, i.e., the percentage of positive and negative classes are almost the same. As shown in Table 2, not all the sensors record all the attributes, hence, we take a union of the attributes which are recorded by the individual sensors. There are a total of 11 attributes collected by all the sensors which include number of steps taken in a day, calories burned, cholesterol level, sleeping hours, sleep quality, calorie intake, and others. The prediction task is a binary classification task where the presence of heart disease is predicted in a person based on the recorded activity.

4.2. Data fusion

In this section, we describe the sensor data fusion and behaviour of bio-sensor node as shown in Fig. 2. We deploy multiple bio-sensor nodes on human objects. We assume that each sensor records single physiological sign, referred to as a feature. BSN collects measurements in a periodic manner and the data is later transferred to the coordinator for data fusion. A number of smart devices including medical devices or mobile phones may be employed as coordinators.

4.2.1. Noise removal and segmentation

After raw data is obtained from body sensors, it is pre-processed prior to further computation. As the signals are sensitive to low and high level frequency noise, we consider Finite Impulse Response (FIR) filter for noise removal (Eq. 4).

$$\begin{aligned} f[n] &= c_0 x[n] + c_1 x[n-1] + c_2 x[n-2] + \ldots\ldots + c_N x[n-N] \\ &= \textstyle\sum_{i=0}^{N} c_i x[n-i] \end{aligned} \tag{4}$$

where $x[n]$ is the input signal, $f[n]$ is the output signal, N is the filter, and $c_i, i = 0 \rightarrow N$ represents coefficient of filter. The output signals are used for feature extraction and normalisation of values for assisting in the prediction framework.

Another important step is of data segmentation which could be (i) Overlapping, or (ii) non-overlapping, based on the sliding window. In this study, we consider non-overlapping segmentation for activity recognition. The choice of segmentation is appropriate when the data is retrieved at varying time intervals [54].

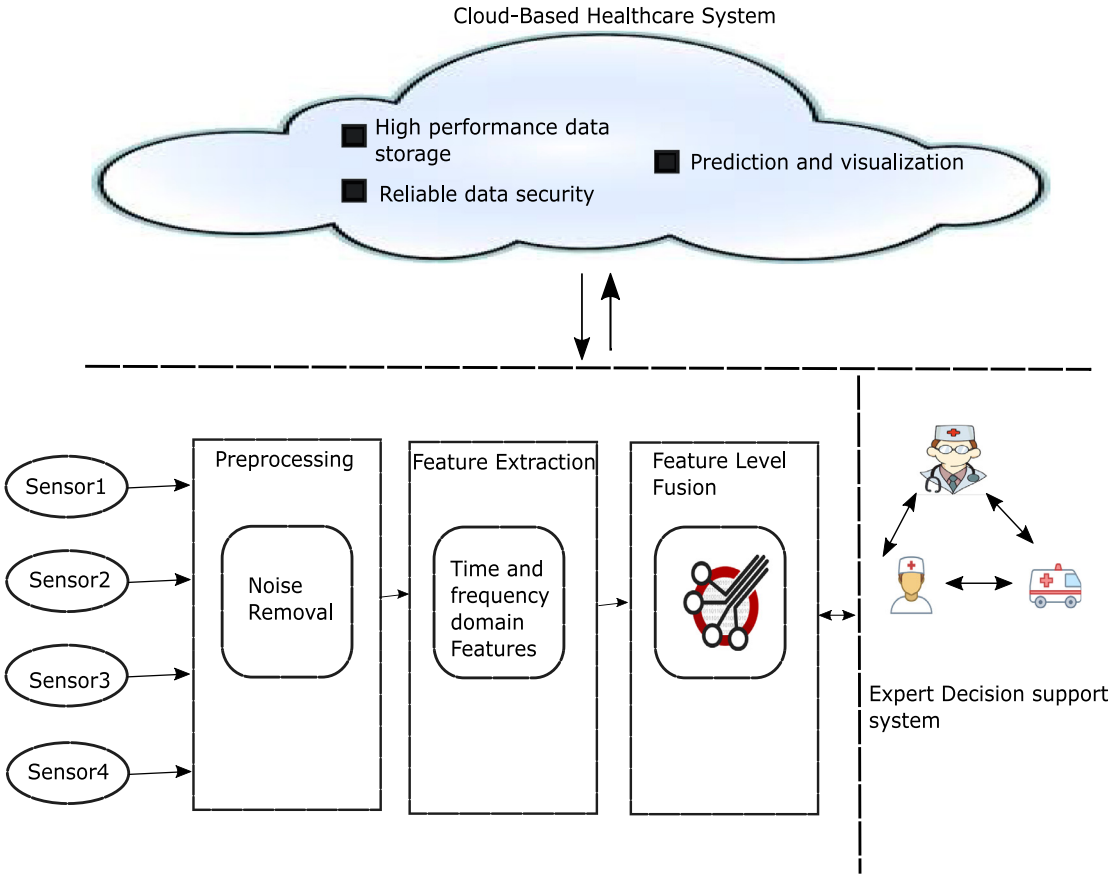

Fig. 2. The proposed real-time data fusion framework in body sensor networks.

Table 2
A list of sensors, features from individual sensors, and data fusion possibilities.

Sensor Type	Steps Count	Distance Covered	Calories Intake	Calories Consumed	Activity Type	Sleep Quality	Sleep Hours	...	Heart Rate
Misfit Shine 2	✓	✓	✓	✓	✓	✓	✓	...	
Whitings Puke	✓	✓			✓		✓	...	✓
Jawbase	✓	✓	✓	✓	✓			...	
Fitbit 3	✓	✓		✓	✓	✓	✓	...	✓

4.2.2. Feature extraction and normalisation

The intuition behind feature extraction is to transform the signal into a feature vector depicting activity details. The features relevant to this study can broadly be categorised into time-domain and frequency domain features.

(1) Time-domain Features: are used to extract statistical measurements that describe the signal characteristics. In this study, we consider time-domain features such as measure of central tendency, and measure of dispersion. Table 3 shows a list of time-domain features.

(2) Frequency-domain Features: are important for analysis of repetitive activities which greatly influence the heart disease prediction. However, raw signals are to be transformed to frequency-domain features using Fast Fourier Transformation. Let a continuous function $f(x)$ having interval $T \in (0, a)$ is defined with period α:

$$f(t) = \sum_{k=\infty}^{\infty} c_k e^{i*2\pi k \frac{T}{\alpha}} \tag{5}$$

The function $f(t)$ is sampled at time t_j, $j = 0 \rightarrow N$ such that:

$$f(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} c_k e^{i*2\pi k*j \frac{1}{N}} \tag{6}$$

Table 3
A list of time-domain features.

Feature	Equation
Mean	$\frac{1}{N}\sum_{i=0}^{N} x_i$
Standard Deviation	$\sqrt{\frac{1}{N}\sum_{i=0}^{N}(x-\mu)^2}$
Maximum	$max(x_i)$
Minimum	$max(x_i)$
Amplitude	$A(x_i) = \max(x) - min(x)$
Zero Crossing	$ZC(x_i) = \sum_{i=0}^{N} \frac{sign(x_i) - sign(x_{i-1})}{2}$ where: $sign(x) = \begin{cases} 1, & \text{if } x_i > 0. \\ 0, & \text{otherwise.} \end{cases}$
Root Mean Square	$\sqrt{\frac{1}{N}\sum_{i=0}^{N}(x_i)^2}$

From the frequency-domain data, features such as energy, entropy, binned distribution, and time between peaks are extracted. The total energy is considered as a global activity and is useful for heart disease prediction. The details about feature computation are as below:

(1) Energy: The sum of the squared FFT magnitude is used to compute the energy feature.

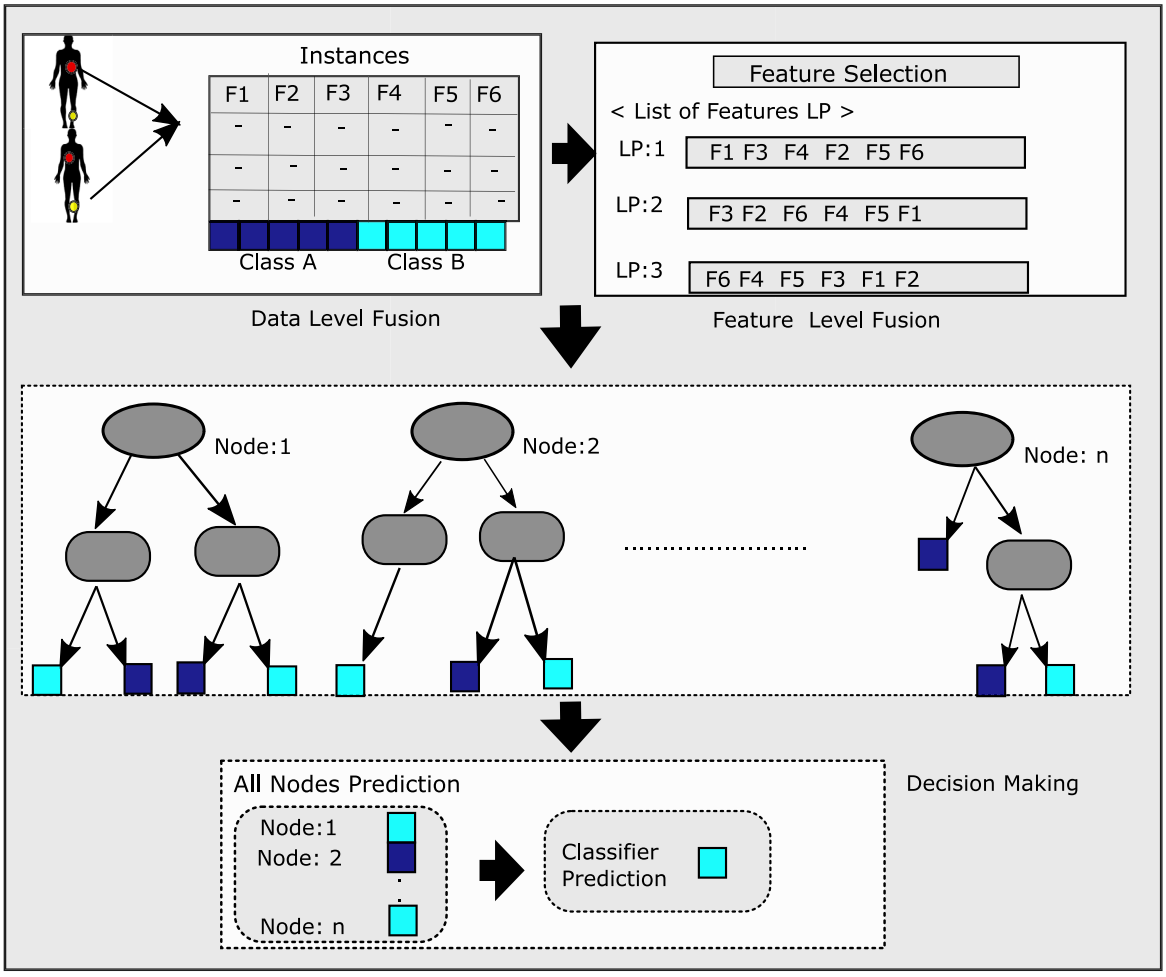

Fig. 3. An illustration of the proposed Fog computing-based ensemble prediction framework.

(2) Entropy: The entropy discriminates among the activities having same values for the energy feature. The entropy is computed as normalised entropy of FFT components.

(3) Binned Distribution: The binned distribution is computed by estimating the histogram of FFT. This is achieved by identifying the range of values and by calculating the fraction of values within a specific range.

4.3. Feature selection

Feature selection is of primary importance in data preprocessing to find optimal subset of features from available feature set. Feature selection helps in reducing the variance of the training model, thus, avoiding the over-fitting problem. Furthermore, computational cost of training model can also be minimised. In this work, we employ Correlation-based Feature Selection (CFS) [55]. The importance of features is predicted using correlation-based importance score.

4.3.1. Correlation-based feature selection

CFS finds the optimal subset of features relevant to the prediction task. The CFS algorithm works on an underlying importance score which represents the usefulness of each feature for predicting the response variable.

$$iScore_S = \frac{mr_{cf}}{\sqrt{m + m(m-1)r_{ff}}} \tag{7}$$

where *iScore* is the importance score for subset S, r_{cf} is response-feature correlation, and r_{ff} represents feature-feature correlation. The importance score is used to pick the most significant features, σ, and disregards irrelevant features as they increase the classification error rate. The CFS discretises the numeric features [56] and computes symmetrical uncertainty to estimate the relationship among the features. Entropy is used to measure impurity in the features. Eqs. 8 and 9 give the entropy and conditional entropy of two discrete random variables X and Y.

$$E(Y) = -\sum_{y \in Y} p(y) \log p(y) \tag{8}$$

$$E(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log p(y|x) \tag{9}$$

Entropy and consequently information gain is used to estimate the usefulness of each feature for the classification task. Information gain is computed as follows:

$$\Theta = E(Y) - E(Y|X) = E(Y) + E(X) - E(X,Y) \tag{10}$$

As information gain is biased towards the features having lower entropy, symmetric uncertainty is used to normalised the values as follows:

$$\lambda = 2.0 * \left[\frac{\Theta}{E(Y) + E(X)} \right] \tag{11}$$

CFS computes the correlation between response variable and features; and based on the correlation values, optimal subset of features is computed using a greedy search algorithm. The optimal feature subset is input to the prediction algorithm for early heart disease prediction.

4.4. Ensemble-based disease prediction

We now present the ensembles for the disease prediction task which we have considered for this study. We first discuss random forest and then present kernel random forest.

Random forest. Random forest is an ensemble of different trees. It can be considered as a multi-way classifier where each tree is grown with some form of randomisation. The idea of random forest is to combine the predicted values obtained from individual trees. This strategy makes the random forest a powerful model for prediction. Suppose that T represent collection of trees, C_1 and C_2 are class label, L and I show the set of leaf node and internal node of tree. Each leaf node of given tree is labelled with posterior probability of a class, while each internal node is responsible to find the best split. The concept of randomisation in random forest is applied at two different stages during learning process. Firstly, samples of training data is selected randomly which ensures that each tree is growing using different subsets. Secondly, at each internal node, features are selected randomly to find the best split, which improves the stability of the random forest. In addition, other parameters such as number of trees, and max depth, also effect the performance of random forests. In this method, each tree is a binary tree and is constructed using a top-down approach. At each internal node, there are two methods of choosing the feature for best split, i.e., (i) random selection, and (ii) greedy selection. As the algorithm chooses the feature that finds best split of training data, the criteria to find the best split is information gain which is computed using Eq. 10. For the prediction task, test data is transferred to each individual tree until a leaf node is reached. Note that we compute average of posterior probabilities to obtain the final prediction value for test data.

Kernel Random Forest. Generally, kernel random forest uses traditional random forest model. The fundamental difference between the random forest and Kernel random forest is the application of randomness in the forest step. In a Kernel random forest, randomness is employed for selecting features at each internal node, training sub-samples are also randomly selected, and the associated kernel functions are consider as best split criteria to separate the space of data that needs to be classified.

5. Experiments and evaluation

We now present the evaluation of the proposed solution. The data preparation and the computing environment has already been elaborated. In this section, we present the results and the analysis.

We now discuss the results and analysis for the ensembles we have considered for this study. We first discuss the feature analysis to show the most significant features which should be considered for the study followed by the experimental analysis. The experimental analysis is focused on the performance of the proposed solution in terms of accuracy, training time analysis, and scalability analysis. We begin with the feature analysis.

5.1. Feature analysis

In order to provide analytical model, we used Correlation-based Feature Selection method to asses the importance of features as shown in Fig. 4. The objective was to determine the predictor variables crucial for prediction of the disease. The attribute analysis shows interesting results, e.g., a negative correlation between heart rate and age, and a positive correlation between exercise and heart disease. The relationship between other attributes can also be seen from the Fig. 4. Note that, not only the important features are to be identified, but the optimal number of features are to be selected to have the best performance in terms of accuracy, training and execution times. The details are presented in the following text.

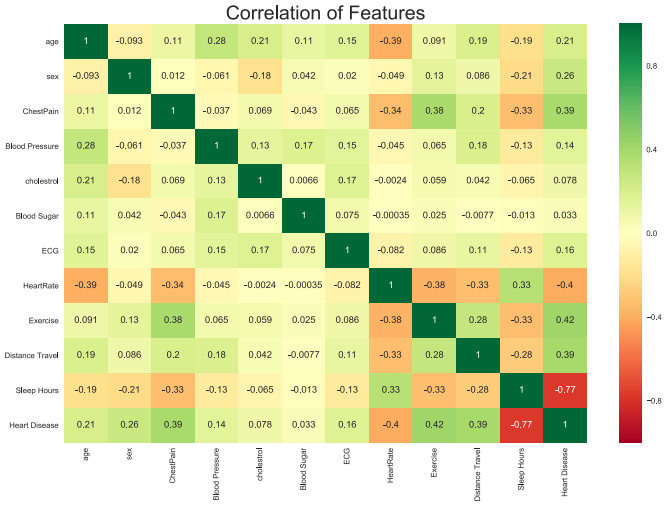

Fig. 4. The relative importance of features based on a correlation matrix.

5.2. Accuracy analysis

In the first set of experiments, we study the accuracy of the two classifiers, i.e., RF and KeRF, for increasing number of estimators with different tree depths. As the data is labelled, the presence or absence of disease was known apriori, for the evaluation purpose. From Fig. 5 and Fig. 6, we see that the accuracy of both the classifiers increases when the number of estimators is increased upto 40 and remains stable af-

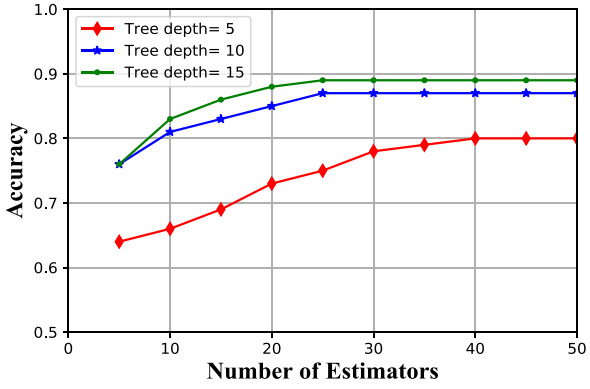

Fig. 5. RF accuracy analysis for increasing number of estimators for increasing tree depth.

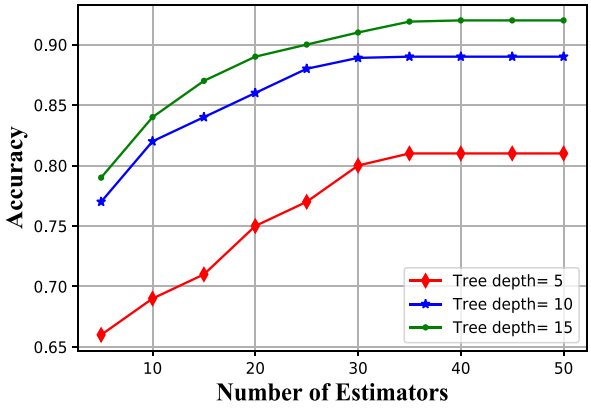

Fig. 6. KeRF accuracy analysis for increasing number of estimators for increasing tree depth.

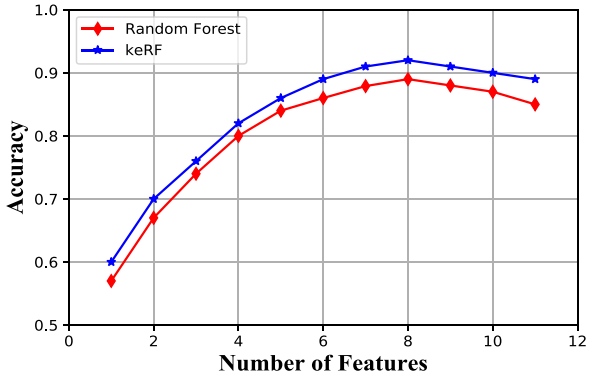

Fig. 7. Accuracy analysis of RF and KeRF for increasing number of features.

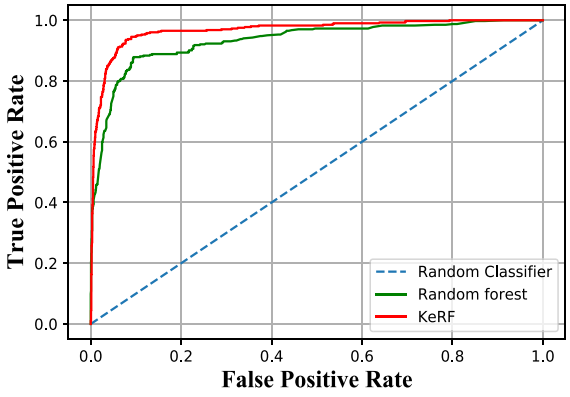

Fig. 8. Accuracy comparison of RF and KeRF using ROC curve.

terwards. For example, in case of random forest (Fig. 5), the accuracy remains stable when the number of estimators is 40. The accuracy is highest when tree depth is 15. It can also be observed that for high tree depth (tree depth = 15), the classifier converges sooner (at 25 estimators) than for lower tree depths (tree depth = 5), at 40 estimators. Similar observations can be noticed about KeRF, Fig. 6. Notice that for KeRF, the highest accuracy achieved is around 98% whereas for RF the maximum accuracy is 90%. Thus, the results support the choice of KeRF as a superior classifier for the considered task. We also study the accuracy of the two classifiers for increasing number of features, Fig. 7, as we set tree depth = 10 and number of estimators = 30. It can be seen that the maximum accuracy is achieved when the number of features is 8, for both the ensembles. Notice that if higher tree depth and number of estimators is utilised, KeRF performs significantly better than the random forest. As the dataset is balanced, we evaluate the accuracy performance of RF and KeRF using an ROC curve. It can be seen in Fig. 8 that the KeRF ensemble is better than the random forest and the results from KeRF show a significantly higher true positive rate as compared with the random forest. In summary, the accuracy analysis of the two classifiers clearly establishes the superiority of the KeRF ensemble for the prediction task.

5.3. Training time analysis

In another set of experiments, we report the training time for increasing number of estimators for different tree depths both for random forest, Fig. 9, and KeRF, Fig. 10. It can be seen from Fig. 9 and Fig. 10 that the training time increases with the increase in the number of estimators and the increase is more visible when the number of estimators is higher than 40. Also, for both the classifiers, for the set of experiments we consider, the training time is highest when the tree depth is maximum, i.e., 15. This is rather expected as the higher tree depth or the increase in

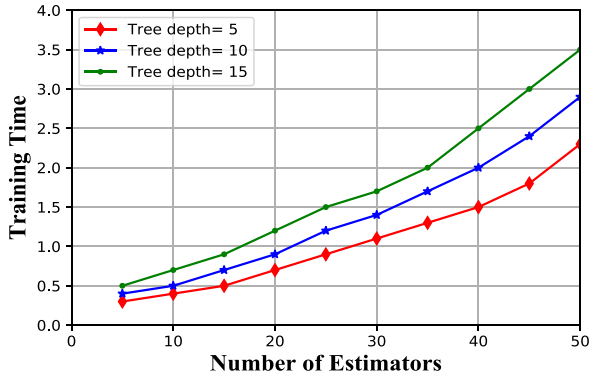

Fig. 9. RF Training time analysis for increasing number of estimators and for increasing tree depth.

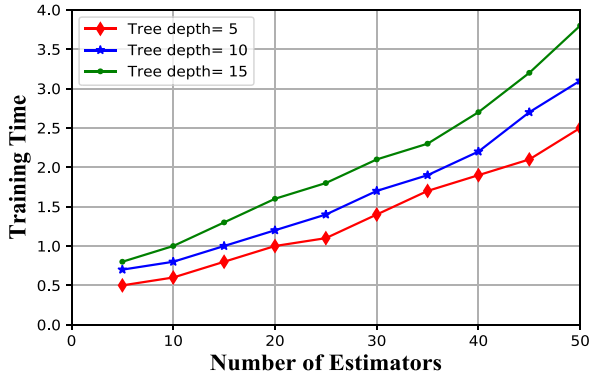

Fig. 10. KeRF Training time analysis for increasing number of estimators and for increasing tree depth.

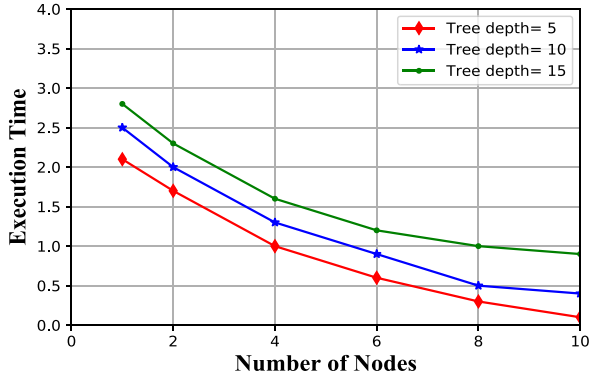

Fig. 11. RF running time analysis for increasing number of nodes for different tree depths.

number of estimators although improves accuracy but comes at a higher computational cost.

5.4. Scalability analysis

As the computing environment is a fog computing environment, we also study the scalability of the system for increasing number of nodes for different tree depths. We set the number of queries to 100K and study the execution time when tree depth is between 5 to 15 and number of computing nodes are increased upto 10. Similar to the training time results presented in the previous text, the scalability analysis shows that for both the ensembles, i.e., Figs. 11 and 12, the execution time decreases for increasing number of nodes and is highest for the maximum tree depth, i.e., 15. The results are consistent with the training time

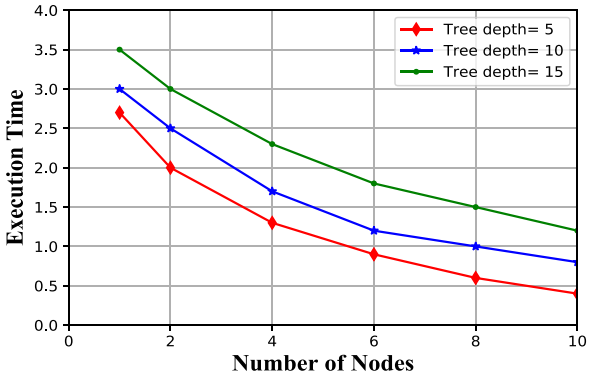

Fig. 12. KeRF running time analysis for increasing number of nodes for different tree depths.

results and it can be seen that the system scales well for increasing tree depth and for higher number of execution nodes. It can also be seen that KeRF is slightly expensive as compared to the random forest approach which is understandable as KeRF is a superior classifier with slightly more computational cost.

In summary, we can conclude that the KeRF is a considerably better solution in terms of accuracy and offers promising results when the number of estimators is set to 40 and tree depth is 15. However, it requires more computational effort for the training and prediction tasks as compared to the random forest due to inherent higher complexity in contrast with the random forest.

6. Conclusion

In this work, a we have proposed a novel data fusion enabled ensemble approach for medical data obtained from BSNs in a fog computing environment. To the best of our knowledge, the use of kernel random forest in a fog computing environment for heart disease prediction is a novel solution. We have considered daily activity data from multiple sensors which is input to an ensemble classifier after performing data fusion. The ensembles are hosted in a fog computing environment and the prediction computations are performed in a decentralised manners. A novel kernel random forest ensemble has been used and high accuracy results are obtained for the heart disease prediction. A number of open directions still remain, for example, the choice of optimal number of sensors to obtain high quality fused data is an interesting future work. A customised fusion mechanism based on the individual sensor capabilities is also an interesting future direction. The consideration of deep neural networks for the heart disease prediction to future improve the prediction accuracy is also an interesting future work.

References

[1] H. Durrant-Whyte, T.C. Henderson, Multisensor data fusion, in: Springer Handbook of Robotics, 2016, pp. 867–896.
[2] G. Fortino, S. Galzarano, R. Gravina, W. Li, A framework for collaborative computing and multi-sensor data fusion in body sensor networks, Inf. Fusion 22 (2015) 50–70.
[3] C.F.C. Junior, A.G. Uría, C. Strumia, M. Koperski, A. König, et al., Online recognition of daily activities by color-depth sensing and knowledge models, Sensors 17 (2017) 1528.
[4] X. Yu, M. Paul, C.H. Antink, et al., Non-contact remote measurement of heart rate variability using near-infrared photoplethysmography imaging, in: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2018, Honolulu, HI, USA, July 18–21, 2018, pp. 846–849.
[5] G. Rebolledo-Mendez, A. Reyes, S. Paszkowicz, M.C. Domingo, L. Skrypchuk, Developing a body sensor network to detect emotions during driving, IEEE Trans. Intell. Transp. Syst. 15 (2014) 1850–1854.
[6] E. Kanjo, E.M.G. Younis, N. Sherkat, Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach, Inf. Fusion 40 (2018) 18–31.
[7] M.C. Domingo, A context-aware service architecture for the integration of body sensor networks and social networks through the IP multimedia subsystem, IEEE Commun. Mag. 49 (2011) 102–108.
[8] G. Aloi, G. Caliciuri, G. Fortino, R. Gravina, P. Pace, W. Russo, C. Savaglio, Enabling iot interoperability through opportunistic smartphone-based mobile gateways, J. Netw. Comput. Appl. 81 (2017) 74–84.
[9] S. Pirbhulal, H. Zhang, M.E. Alahi, H. Ghayvat, S. Mukhopadhyay, Y.T. Zhang, W. Wu, A novel secure iot-based smart home automation system using a wireless sensor network, Sensors 17 (2017) 69.
[10] Y. Zhang, B. Song, X. Du, M. Guizani, Vehicle tracking using surveillance with multimodal data fusion, IEEE Trans. Intell. Transp. Syst. 19 (2018) 2353–2361.
[11] Y. Hsu, P. Chou, H. Chang, et al., Design and implementation of a smart home system using multisensor data fusion technology, Sensors 17 (2017) 1631.
[12] Y. Tang, D. Zhou, S. Xu, Z. He, A weighted belief entropy-based uncertainty measure for multi-sensor data fusion, Sensors 17 (2017) 928.
[13] M. Muzammal, M. Gohar, A.U. Rahman, Q. Qu, A. Ahmad, G. Jeon, Trajectory mining using uncertain sensor data, IEEE Access 6 (2018) 4895–4903.
[14] D.V. Huy, N.D. Viet, Df-swin: Sliding windows for multi-sensor data fusion in wireless sensor networks, in: 9th International Conference on Knowledge and Systems Engineering, KSE 2017, Hue, Vietnam, October 19–21, 2017, pp. 54–59.
[15] N. Yuan, W. Yang, B. Kang, S. Xu, C. Li, Signal fusion-based deep fast random forest method for machine health assessment, J. Manuf. Syst. 48 (2018) 1–8.
[16] E. Scornet, Random forests and kernel methods, IEEE Trans. Information Theory 62 (2016) 1485–1500.
[17] A.H. Sodhro, S. Pirbhulal, A.K. Sangaiah, Convergence of iot and product lifecycle management in medical healthcare, Future Gener. Comp. Syst. 86 (2018) 380–391.
[18] A.H. Sodhro, S. Pirbhulal, G.H. Sodhro, A. Gurtov, M. Muzammal, Z. Luo, A joint transmission power control and duty-cycle approach for smart healthcare system, IEEE Sens. J. (2018).
[19] A.H. Sodhro, S. Pirbhulal, M. Qaraqe, S. Lohano, G.H. Sodhro, N.U.R. Junejo, Z. Luo, Power control algorithms for media transmission in remote healthcare systems, IEEE Access 6 (2018) 42384–42393.
[20] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, R. Jafari, Enabling effective programming and flexible management of efficient body sensor network applications, IEEE Trans. Hum. Mach. Syst. 43 (2013) 115–133.
[21] A.H. Sodhro, Z. Luo, A.K. Sangaiah, S.W. Baik, Mobile edge computing based qos optimization in medical healthcare applications, Int. J. Inf. Manage. (2018).
[22] A.H. Sodhro, A.K. Sangaiah, S. Pirphulal, A. Sekhari, Y. Ouzrout, Green media-aware medical iot system, Multimed. Tools Appl. 78 (2019) 3045–3064.
[23] D. Schuldhaus, H. Leutheuser, B. Eskofier, Towards big data for activity recognition: A novel database fusion strategy, in: 9th International Conference on Body Area Networks, BODYNETS 2014, London, UK, September 29, - October 1, 2014.
[24] C. Chen, R. Jafari, N. Kehtarnavaz, A survey of depth and inertial sensor fusion for human action recognition, Multimedia Tools Appl. 76 (2017) 4405–4425.
[25] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[26] L. Gao, A.K. Bourke, J. Nelson, A system for activity recognition using multi-sensor fusion, in: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2011, Boston, MA, USA, August 30, - Sept. 3, 2011, pp. 7869–7872.
[27] C. Chen, R. Jafari, N. Kehtarnavaz, Improving human action recognition using fusion of depth camera and inertial sensors, IEEE Trans. Hum. Mach. Syst. 45 (2015) 51–61.
[28] A.M. Khan, A. Tufail, A.M. Khattak, T.H. Laine, Activity recognition on smartphones via sensor-fusion and kda-based svms, IJDSN 10 (2014).
[29] N.A. Capela, E.D. Lemaire, N. Baddour, Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients, PLoS ONE 10 (2015) e0124414.
[30] U. Maurer, A. Smailagic, D.P. Siewiorek, M. Deisher, Activity recognition and monitoring using multiple sensors on different body positions, Wearable and Implantable Body Sensor Networks, 2006. BSN2006. International Workshop on, IEEE. p. 4.
[31] S. Liu, R.X. Gao, D. John, J. Staudenmayer, P.S. Freedson, Multisensor data fusion for physical activity assessment, IEEE Trans. Biomed. Engineering 59 (2012) 687–696.
[32] W. Wu, H. Zhang, S. Pirbhulal, S.C. Mukhopadhyay, Y.T. Zhang, Assessment of biofeedback training for emotion management through wearable textile physiological monitoring system, IEEE Sens. J. 15 (2015) 7087–7095.
[33] G. Fortino, S. Galzarano, R. Gravina, W. Li, A framework for collaborative computing and multi-sensor data fusion in body sensor networks, Inf. Fusion 22 (2015) 50–70.
[34] M.M. Hassan, M.S. Huda, J. Yearwood, H.F. Jelinek, A. Almogren, Multistage fusion approaches based on a generative model and multivariate exponentially weighted moving average for diagnosis of cardiovascular autonomic nerve dysfunction, Inf. Fusion 41 (2018) 105–118.
[35] N. Martini, D. Menicucci, L. Sebastiani, R. Bedini, A. Pingitore, N. Vanello, M. Milanesi, L. Landini, A. Gemignani, The dynamics of EEG gamma responses to unpleasant visual stimuli: from local activity to functional connectivity, Neuroimage 60 (2012) 922–932.
[36] R. Gravina, G. Fortino, Automatic methods for the detection of accelerative cardiac defense response, IEEE Trans.Affect. Comput. 7 (2016) 286–298.
[37] R. Covello, G. Fortino, R. Gravina, A. Aguilar, J.G. Breslin, Novel Method and Real-time System for Detecting the Cardiac Defense Response Based on the ECG, in: IEEE International Symposium on Medical Measurements and Applications, MeMeA 2013, Gatineau, QC, Canada, May 4–5, 2013, pp. 53–57. Proceedings.
[38] J. Atkinson, D. Campos, Improving bci-based emotion recognition by combining EEG feature selection and kernel classifiers, Expert Syst. Appl. 47 (2016) 35–41.
[39] M.M. Hassan, M.G.R. Alam, M.Z. Uddin, S. Huda, A. Almogren, G. Fortino, Human emotion recognition using deep belief network architecture, Inf. Fusion 51 (2019) 10–18.
[40] R. Gravina, Q. Li, Emotion-relevant activity recognition based on smart cushion using multi-sensor fusion, Inf. Fusion 48 (2019) 1–10.

[41] G. Fortino, W. Russo, C. Savaglio, W. Shen, M. Zhou, Agent-oriented cooperative smart objects: from IoT system design to implementation, IEEE Trans. Systems Man Cybern. 48 (2018) 1939–1956.
[42] H. Banaee, M.U. Ahmed, A. Loutfi, Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges, Sensors 13 (2013) 17472–17500.
[43] G. Fortino, R. Gravina, W. Russo, C. Savaglio, Modeling and simulating internet-of-things systems: a hybrid agent-oriented approach, Comput. Sci. Eng. 19 (2017) 68–76.
[44] S. Iyengar, F.T. Bonda, R. Gravina, A. Guerrieri, G. Fortino, A. Sangiovanni-Vincentelli, A framework for creating healthcare monitoring applications using wireless body sensor networks, Proceedings of the ICST 3rd international conference on Body area networks, ICST (Institute for Computer Sciences, Social-Informatics and …8.
[45] W. Yi, O. Sarkar, S. Mathavan, J. Saniie, Wearable sensor data fusion for remote health assessment and fall detection, in: IEEE International Conference on Electro/Information Technology, EIT 2014, Milwaukee, WI, USA, June 5–7, 2014, pp. 303–307.
[46] F. Felisberto, R. Laza, F. Fdez-Riverola, A. Pereira, A distributed multiagent system architecture for body area networks applied to healthcare monitoring, Biomed. Res. Int. 2015 (2015).
[47] H.C. Koh, G. Tan, et al., Data mining applications in healthcare, J. Healthcare Inf. Manage. 19 (2011) 65.
[48] I. Kadi, A. Idri, J. Fernandez-Aleman, Knowledge discovery in cardiology: a systematic literature review, Int. J. Med. Inform. 97 (2017) 12–32.
[49] R. Hamza, K. Muhammad, N. Arunkumar, G. Ramírez-González, Hash based encryption for keyframes of diagnostic hysteroscopy, IEEE Access 6 (2018) 60160–60170.
[50] F. Huang, S. Wang, C.C. Chan, Predicting disease by using data mining based on healthcare information system, IEEE, 2012IEEE International Conference on Granular Computing, 191–194.
[51] J.K. Kim, J.S. Lee, D.K. Park, Y.S. Lim, Y.H. Lee, E.Y. Jung, Adaptive mining prediction model for content recommendation to coronary heart disease patients, Cluster Comput. 17 (2014) 881–891.
[52] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ecg classification by 1-d convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (2016) 664–675.
[53] A. Mustaqeem, S.M. Anwar, M. Majid, A.R. Khan, Wrapper method for feature selection to classify cardiac arrhythmia, IEEE, 201739th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3656–3659.
[54] L. Wang, T. Gu, X. Tao, J. Lu, A hierarchical approach to real-time activity recognition in body sensor networks, Pervasive Mob. Comput. 8 (2012) 115–130.
[55] M. Mursalin, Y. Zhang, Y. Chen, N.V. Chawla, Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier, Neurocomputing 241 (2017) 204–214.
[56] U.M. Fayyad, K.B. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, in: Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28, - September 3, 1993, pp. 1022–1029.