Heart Disease Prediction using Fog Computing based Wireless Body Sensor Networks (WSNs)

Dr. Subarna Shakya,

Professor, Department of Electronics and Computer Engineering
Central Campus, Institute of Engineering, Pulchowk
Tribhuvan University, Pulchowk
Lalitpur Nepal - 44600
Email: drss@ioe.edu.np

Dr. P P Joby,

Professor and Head,
Department of Computer Science and Engineering,
St. Joseph's College of Engineering and Technology,
Kerala.
jobypcse@gmail.com

Abstract- Wireless Body Sensor Network (BSNs) are devices that can be worn by human beings. They have sensors with transmission, computation, storage and varying sensing qualities. When there are multiple devices to obtain data from, it is necessary to merge these data to avoid errors from being transmitted, resulting in a high quality fused data. In this proposed work, we have designed a data fusion approach with the help of data obtained from the BSNs, using Fog computing. Everyday activities are gathered in the form of data using an array of sensors which are then merged together to form high quality data. The data so obtained is then given as the input to ensemble classifier to predict heart-related diseases at an early stage. Using a fog computing environment, the data collector is established and the computation process is done with a decentralised system. A final output is produced on combining the result of the nodes using the fog computing database. A novel kernel random data collector is used for classification purpose to result in an improved quality. Experimental analysis indicates an accuracy of 96% where the depth is about 10 with an estimator count of 45 along with 7 features parameters considered.

Keywords: Fog computing, disease prediction, ensemble methods, body sensor network, multi-sensor data fusion

1. Introduction

In the recent years, Wireless Body Sensor Networks (BSNs) have expanded to a wide spectrum resulting in advancement of data processing techniques like data fusion, incorporating it in numerous applications [1]. The basic conception behind fusion is to determine the events based on information gathered from various sources. It is not possible to gather this information from an individual node. Hence, data from various sources are combined together in order to determine the way in which these observations can be made into a clear event [2-3]. When compared with the individual RAW observations, the fused data is said to have higher value. Hence as the fusion process increases with additional individual readings, the significance of the readings will also increase. In general success system which executes using a stream of continuous data will require optimal battery consumption and limited network bandwidth. The hardware that is used for sensing it is generally costly and has a lot of constraints like sampling rate, storage capabilities, sensor accuracy etc [4]. Communication is established using single-hop star topology network where a central unit will coordinate all the BSNs. Over the years a number of BSN architectures have been proposed for data coordination and processing [5].
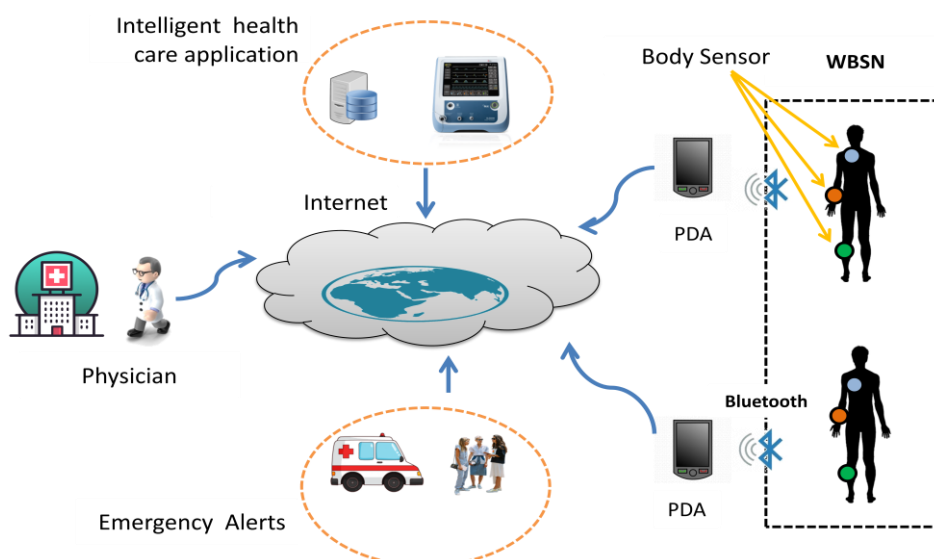

Fig.1. Smart Health Care using Wireless Body Sensors

It is possible to monitor the individual activities by the BSNs in many environments. Some examples are in integration of social network data and BSNs data, integration of environment data with BSN data, emotion detection during driving, rehabilitation after medical procedure, prevention or detection of disease, smart home with multi-sensor data fusion, vehicle tracking with multimodal data fusion etc [6]. Stream processing algorithms and multi-sensor data with uncertainty measures have also been proposed. For applications like disease prediction and activity recognition, a number of machine learning methodologies such as random ensemble algorithms and support vector machines have been used in a many domains, including health. Recent research indicates that it is possible to improve the accuracy of random forest (a machine learning algorithm) with the infusion of kernel methods [7]. In [x] kernel method is fused with random forests to enhance the prediction accuracy of the work. Using a fog environment, the computational operations are carried out such that fog nodes are used to host the individual predictors and the results are collected as the output [8].

The following are the contributions of this proposed work:

1. Heart disease prediction using multi-sensor data fusion framework where the gathered information is unified.
2. Using sensor activity data for the proposed ensemble framework
3. Kernal random forest ensemble in line with fog-based computing environment for predicting the heart disease based on sensor activity.
4. The effectiveness of the proposed work is evaluated using an extensive empirical study.

Experimental analysis indicates that the proposed fog-based [9] ensemble frame will be useful in determining the heart disease in a more effective manner. The paper is organised such that section 2 illustrates a survey of similar methods and concepts, section 3 lists out the proposed ensembles and data fusion. Section 4 studies the outputs obtained and section 5 concludes the work.

SWS

2. Related Work

Wireless body sensor network are made up of sensors that can be fixed on the human to communicate with a device. Each sensor in a particular BSN will have the capacity to transmit the information that it senses. Each sensor is assigned with varying transmission and sensing capabilities, computation power and storage. In this section, a number of previously existing BSNs are surveyed to identify the drawbacks in the already existing methodologies. Using many distributed and centralized data fusion, many approaches have been proposed recently. As per studies conducted by the world health organisation, a total of 4.6 million people are affected by diseases that are caused due to decrease in physical activity [10]. Hence a proper monitoring system is essential to make people aware of their health and to further improve the quality of life. Hence much research is being conducted in the domain of activity recognition. To extract the features a many sensors like light sensor, accelerometer, microphone and magnetometer are used. On the other hand, the right feature subsets are narrowed down using other methodologies like relevance heuristics, kernel discriminating analysis and correlation based methods. SVM, Neural Network, Naïve Bayes and Decision Tree are some of classifiers used to select the right feature. In [11], the authors introduced a novel methodology which uses the heart beats of a human being as the input to determine fear. Here the ECG is required to find the emotions experienced by a person. Similar other methods such as fine Gaussian support vector machine, human emotion recognition model, minimum-redundancy-maximum relevance have been introduced over the years [12-13].

3. Proposed Methodology

3.1. Data Collection

The proposed framework is represented in Fig.2. As the first step, data is collected using wearable sensors fitted on to the human body. The collected data is then transferred to the personal devices using a wireless network. A local hospital was approached for creating dataset where each subject is chosen using even distribution (resulting in equal negative and positive classes). The various attributes of the sensors will not be able to give a complete record of data. Hence a combination of the data collected from the different sensors is used.

SWS

Using the sensors, a total of 11 attributes have been collected such as calorie intake, sleep quality, sleeping hours, cholesterol level, calories burned etc. Here, based on the activity recorded, prediction is made to determine the possibility of heart disease.
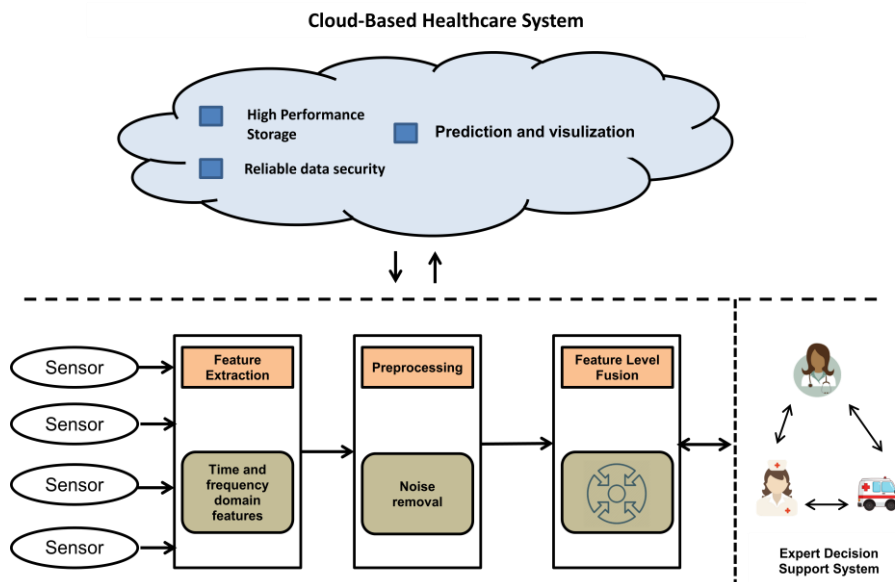

Fig.2. Proposed Framework for BSNs

3.2. Data Fusion

In data fusion session, a number of bio-sensor nodes are used on human objects. Every sensor is capable of saving a feature which is a physiological sign. Each feature is measured in a consistent manner using the body sensors which is later on sent to the main node for fusing with the other data. To coordinate these aspects, smart devices such as mobile phones or medical devices can be used. The gathered data is in the raw state and needs to be processed before being used for further computation. An FIR filter which is used to remove the noise signal as shown in eq (1)

$$f[n] = c_0 x[n] + c_1 x[n-1] + c_2 x[n-2] + \cdots + c_N x[n] \qquad (1)$$

Where N is the filter, the output signal is denoted as f[n] and x[n] is the input signal. Here the coefficient of the filter varies between 0 and N. To normalise the values and to perform

feature extraction, the output signals are used. Similarly, depending on the sliding window data segmentation can be performed as non-overlapping or overlapping, which is a crucial step. In this proposed work, activity recognition is done for non-overlapping segmentation. This process is aptly done when data is obtained at varying time intervals. The proposed framework uses a fog-computing environment where data fusion and data collection is performed.

3.2. Noise Elimination

Using wearable sensors, the data from the human body is gathered using 5G network. This data gathered from the body is then transferred to personal devices. A local hospital was approached to obtain the dataset from and using even distribution, the subjects were chosen. A union of all the attributes obtained from the various sensors are taken and recorded. In the proposed work, we collected 11 attributes using the sensors such as calorie intake, sleep quality, sleeping hours, cholesterol level, calories burned etc. Depending on the recorded activity, it is possible to predict heart diseases using the binary classification task. The human objects are fitted with multiple bio-sensor nodes. To coordinate proper transfer of data, to f use them, smart devices are also used. Once raw data is gathered from the body sensors, they are properly processed using further analysis. The signals are said to be sensitive to high and low noise frequency. Hence an FIR is employed to remove the unnecessary noise from the input signal x[n].

3.2. Feature Extraction and Feature Selection

To convert the obtained signal into a feature vector, feather extraction is a crucial aspect. It can be divided into two categories based on frequency domain and tine-domain features.

- Frequency-domain features: They are used to predict heart diseases based on the statistical measurements of their signal characteristics. Measure of dispersion and central tendency are some of the features considered.
- Time-domain features: Signal characteristics are determined using the extraction of statistical measurements in time domain. Measure of dispersion and measure of

SWS

tendency are some examples of the type of time-domain features considered in this work.

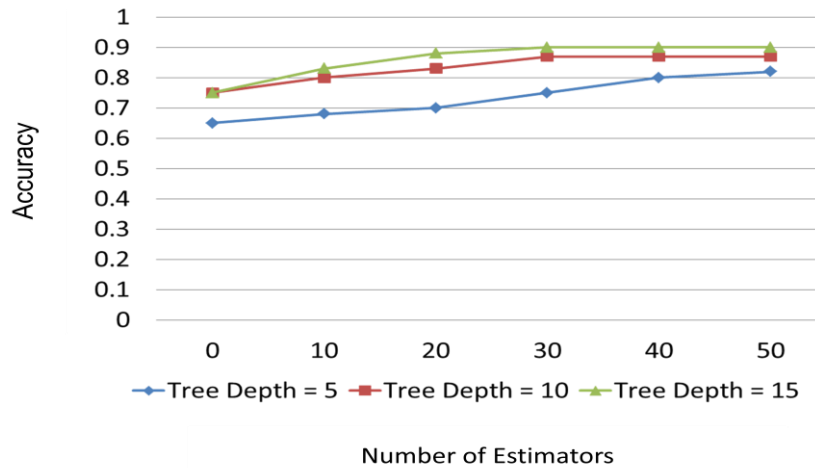## 4. Results and Discussions

Fig.3. Tree depth increase to determine RF Accuracy

The two classifiers RF and KeRF are initially are studied for different depths of tree and the corresponding increase of estimators as seen in Fig.3 and Fig.4 respectively. It is observed that as the depth of the tree increases there is also a subsequent increase in the number of estimators. When the depth of the tree is 15, the accuracy level is found to be very high. Hence the classifiers tend to converge at a quicker rate when the depth is very high.
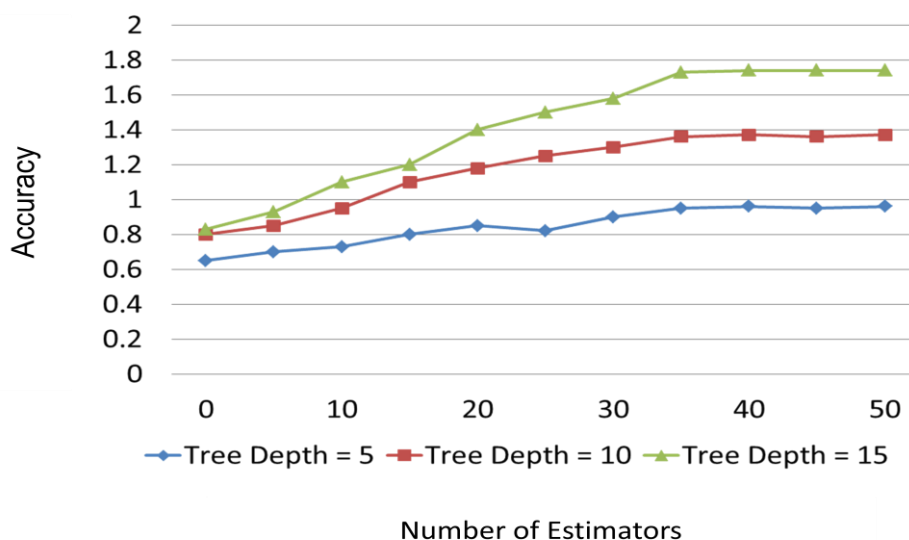
SWS

Fig.4. Tree Depth Increase to determine KeRF accuracy

A comparison of the two methodologies RF and KeRF in Fig.5 indicate that KeRF has better accuracy results when compared with the latter.
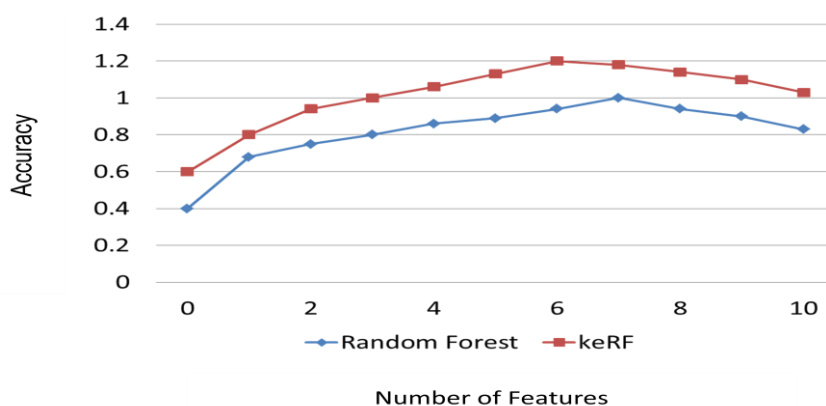

Fig.5. Accuracy Analysis of both methodologies

5. Conclusion

In this proposed work, we have introduced a body sensor network that is used to detect heart diseases using fusion enabled data ensemble technology in a fog computing surrounding.

SWS

Using the different sensors that are used to gather data about the everyday activities, data fusion is performed. A fog computing environment is used to host the data fusion and prediction of heart disease is done in a decentralized manner. To ensure high accuracy, a forest ensemble is used to predict and there are still a number of options for future enhancing the proposed work. They can be used with deep neural network to further enhance the accuracy of prediction. Fusion mechanism that is customised depending on the sensor's capacity is also a possible future venture.

References

[1] Estrin, D., Girod, L., Pottie, G., & Srivastava, M. (2001, May). Instrumenting the world with wireless sensor networks. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)* (Vol. 4, pp. 2033-2036). IEEE.

[2] Slijepcevic, S., & Potkonjak, M. (2001, June). Power efficient organization of wireless sensor networks. In ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No. 01CH37240) (Vol. 2, pp. 472-476). IEEE.

[3] Pathan, A. S. K., Lee, H. W., & Hong, C. S. (2006, February). Security in wireless sensor networks: issues and challenges. In *2006 8th International Conference Advanced Communication Technology* (Vol. 2, pp. 6-pp). IEEE.

[4] Ai, J., & Abouzeid, A. A. (2006). Coverage by directional sensors in randomly deployed wireless sensor networks. Journal of Combinatorial Optimization, 11(1), 21-41.

[5] Abrams, Z., Goel, A., & Plotkin, S. (2004, April). Set k-cover algorithms for energy efficient monitoring in wireless sensor networks. In Proceedings of the 3rd international symposium on Information processing in sensor networks (pp. 424-432).

[6] Wu, F. J., Kao, Y. F., & Tseng, Y. C. (2011). From wireless sensor networks towards cyber physical systems. Pervasive and Mobile computing, 7(4), 397-413.

[7] Zhang, H., & Hou, J. (2004, May). On deriving the upper bound of α-lifetime for large sensor networks. In *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing* (pp. 121-132).

SWS

[8] Cardei, M., MacCallum, D., Cheng, M. X., Min, M., Jia, X., Li, D., & Du, D. Z. (2002). Wireless sensor networks with energy efficient organization. *Journal of Interconnection Networks*, *3*(03n04), 213-229.

[9] Manoharan, S. (2019). Image detection classification and recognition for leak detection in automobiles. *Journal of Innovative Image Processing (JIIP)*, *1*(02), 61-70.

[10] Pandian, A. P. (2019). Artificial intelligence application in smart warehousing environment for automated logistics. *Journal of Artificial Intelligence*, *1*(02), 63-72.

[11] Shakya, S. (2019). Virtual restoration of damaged archeological artifacts obtained from expeditions using 3D visualization. *Journal of Innovative Image Processing (JIIP)*, *1*(02), 102-110.

SWS