Patient monitoring and health management using wearable devices

By
Arbettu Tejas Kamath
16014022041
Mihir Lodaya
16014022049
Ketaki Mahajan
16014022050
Swapnil Ranadive
16014022067

Stress Monitoring Using Machine Learning, IoT and Wearable Sensors

PAPER 1

Abstract

- **Introduction to IoT in Healthcare:** IoT devices, such as wearable sensors and smart medical devices, play a crucial role in continuously monitoring vital signs, medication adherence, and overall patient well-being. By leveraging IoT, healthcare providers can monitor patients remotely, streamline workflows, and deliver more efficient and effective care.

- **Integration of Machine Learning:** Emphasize the integration of machine learning and advanced technology in addressing stress levels, acknowledging its potential impact on healthcare.

- **Introduction of "Stress-Track":** Chronic stress can negatively impact physiological parameters, leading to conditions such as hypertension, cardiovascular disease, and mental health disorders. Introduce the novel machine learning-based system, "Stress-Track," designed to track stress levels through body temperature, sweat, and motion rate.

- **Performance:** Highlight the impressive accuracy rate of 99.5% achieved by the proposed model, showcasing its potential to enhance stress management and healthcare.

Literature Review

- **Essential Characteristics of Ideal Sensors:**
    - Precision, sensitivity, linearity, repetition, reproducibility, drift, calibration, and fast response are crucial.
    - ML techniques create a positive feedback loop for ongoing therapeutic improvements.
- **ML and Signal Processing for Mental Health Monitoring:**
    - ML techniques and signal processing enable continuous mental health monitoring.
    - Various stressors and computing locations (edge, fog, cloud) are considered.
- **Biomarkers and Use Cases for Stress Detection:**
    - Biomarkers for stress detection include ECG, skin conductance, respiration, heart rate variability, and fMRI.
    - Use cases for IoT in healthcare include autonomous insulin infusion, sleep monitoring, and mental health monitoring.

Block Diagram / Flowchart

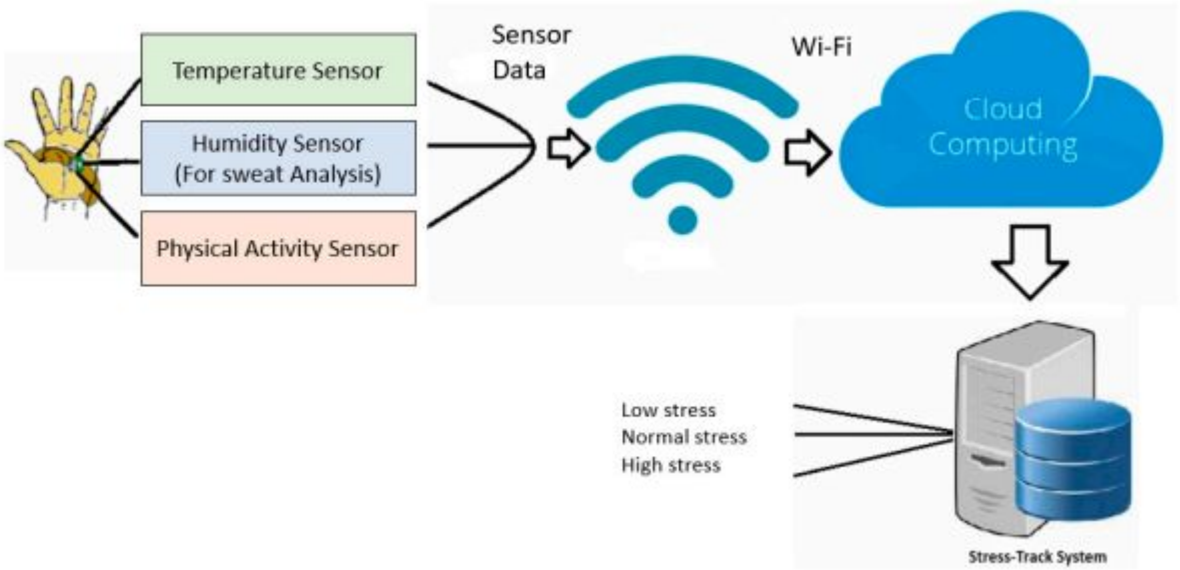


Figure 1. Proposed model.

Methodology

- **Introduction to Ensemble Methods:**
  - Random Forest (RF)
  - Gradient Boosting (GB)
  - Stacked Ensemble Method (SEM)
- **Proposed Methodology:**
  - Overview of the proposed methodology for stress level detection using the Stacked Ensemble Method.
  - Data acquisition and preprocessing steps, including dataset loading, feature-target split, and data encoding.
  - Components of the Stress-Track Sensing Wristband, including body temperature measurement, humidity analysis, and step count analysis.
  - Outline the steps for building the ensemble model, including model initialization, dataset splitting, training individual base models, and training the meta-model.
  - Highlight the evaluation metrics used to assess the model's performance, such as accuracy, confusion matrix, precision, recall, and F1 measure.

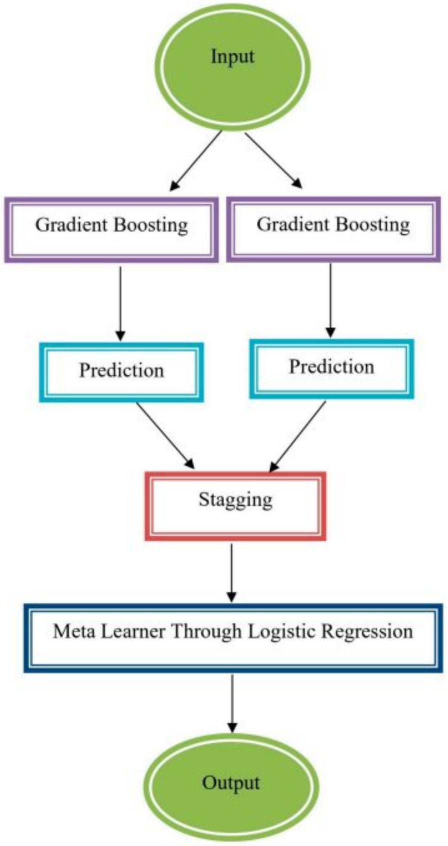


Figure 3. Stacking ensemble method.

Result & Conclusion

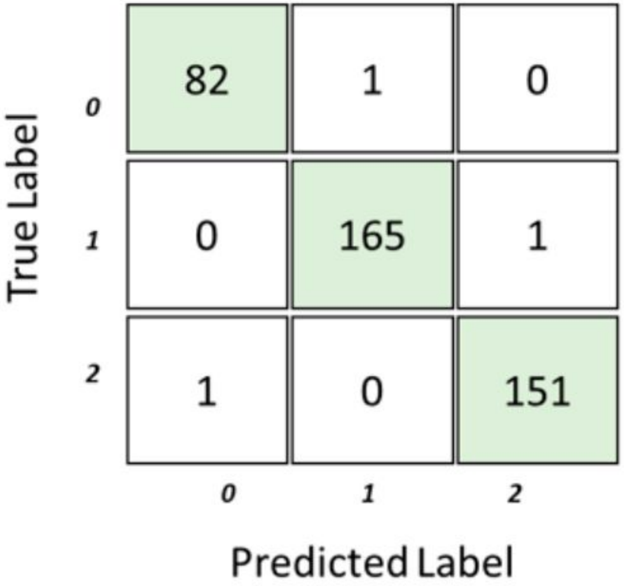


Figure 4. Confusion matrix.

Accuracy	Precision	Recall	F1 Score
99.5	0.99	0.99	0.99

Scope for Future Work

- Challenges and Considerations
- Future Directions
- Closing Statement

References

- **Stress Monitoring Using Machine Learning, IoT and Wearable Sensors** Abdullah A. Al-Atawi 1, Saleh Alyahyan 2, Mohammed Naif Alatawi 3, Tariq Sadad 4, Tareq Manzoor 5,Muhammad Farooq-i-Azam 6 and Zeashan Hameed Khan 7
- https://www.researchgate.net/publication/375058758_Stress_monitoring_using_machine_learning_IoT_and_wearable_sensors/link/6544e767b86a1d521bb42f32/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms

Research Paper 2

Abstract

<u>Keywords</u>: Internet of Things, Diabetic Patient Monitoring, Machine Learning, Data Classification, Healthcare

Continuous monitoring of diabetic patients improve their quality of life. Multiple technologies such as the Internet of Things (IoT), embedded systems, communication technologies, artificial intelligence, and smart devices can reduce the economic costs of the healthcare system. Different communication technologies have made it possible to provide personalized and remote health services.

This article presents an intelligent architecture for monitoring diabetic patients by using machine learning algorithms. The architecture elements included smart devices, sensors, and smartphones to collect measurements from the body. The Intelligent system collected the data received from the patient, and performed data classification using machine learning in order to make a diagnosis.

The proposed prediction system was evaluated by several machine learning algorithms, and the simulation results demonstrated that the sequential minimal optimization (SMO) algorithm gives superior classification accuracy, sensitivity, and precision compared to other algorithms.

Introduction

Healthcare's continuous evolution, driven by IoT technologies, ICTs, sensors, big data, machine learning, and AI, focuses on continuous monitoring of chronic illnesses, including diabetes, which poses significant mortality risks. Patients with chronic diseases require prolonged hospitalization for daily monitoring due to the long-term nature of these diseases.

The rise in diabetic patients has led to increased adoption of monitoring systems aimed at periodic blood glucose checks, enhancing patient, relative, and doctor awareness, and facilitating swift response to abnormalities, thereby improving patient quality of life through reduced hospitalization time with portable monitoring devices.

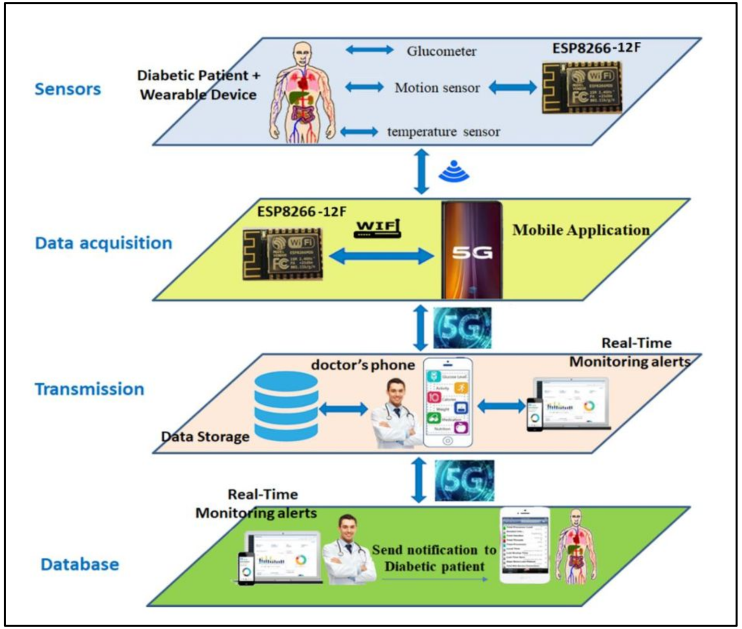
Introduction

The paper presents an architecture for smart continuous monitoring of diabetic patients, integrating a machine learning algorithm for data classification. Portable sensors detect the patient's blood sugar level, temperature, and physical activity, with collected data analyzed for classification and prediction using multiple algorithms. Diabetic patients benefit from future predictions of their blood sugar levels. Tests employ various classification algorithms such as naïve Bayes, random forest (RF), ZeroR, simple logistic, sequential minimal optimization (SMO), and J48 to determine the most effective risk assessment method.

The solution utilized simple, low-power, and cost-effective sensors for glucose monitoring, updating patient data in the cloud daily. Doctors leverage this data to monitor glucose variation and provide necessary medical care. The prediction was reached by using several machine learning algorithms. In order to provide the best accuracy, different classification algorithms were analyzed, tested, and compared using different parameters.

Literature Review

Proposed Architecture



The proposed architecture for monitoring diabetic patients comprises of four main layers:

- Sensors include blood glucose, temperature, and motion sensors connected to an ESP8266 module for wireless data transmission to the patient's smartphone.
- The data acquisition layer involves the patient's smartphone and an application to collect and display sensor data, which is also transmitted via 5G to a base station.
- In the transmission layer, the smartphone sends data to the database for processing and forwards it to the doctor's phone for examination.
- The database layer processes and classifies sensor data using machine learning algorithms to detect abnormal situations, generating notifications for the doctor's review and advice, displayed on the patient's smartphone.

Literature Review

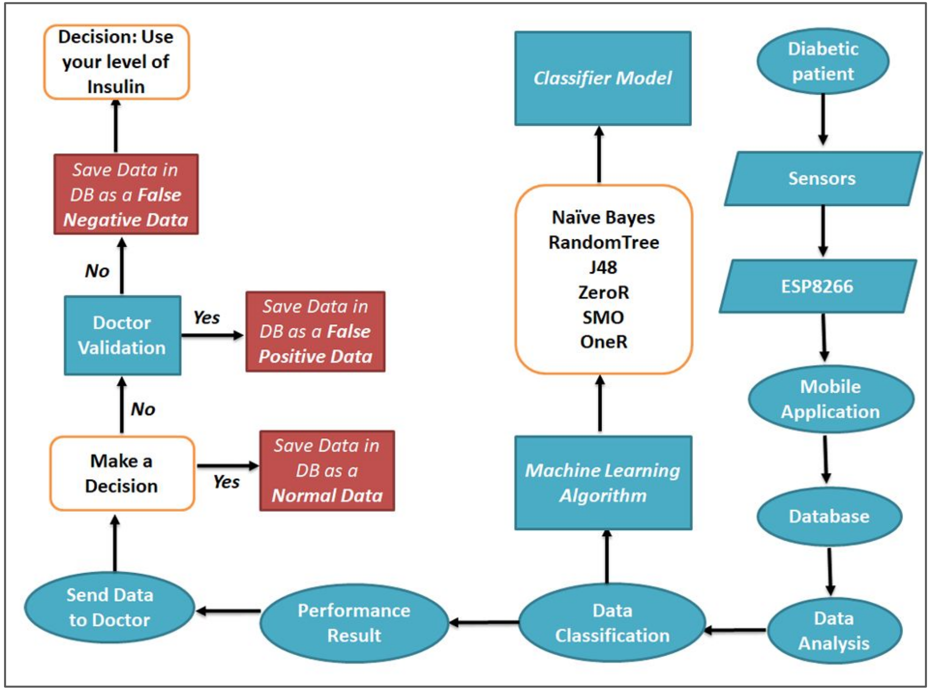
Data Collection

The dataset used consisted of 10,807 data values, containing data on several diabetic patients. We used this dataset to try the different machine learning algorithms to detect and make predictions of diabetes. The dataset includes following attributes: gender, age, day of measure were taken, blood glucose level, insulin used, body temperature and physical activity.

Day	Blood Sugar Level (mg/dL)			Temperature	Number of Steps
	Morning	Afternoon	Evening		
Day1	98	102	111	37	5423
Day2	166	153	124	36	6322
Day3	103	112	114	37	4876
Day4	134	102	98	37	4657
Day5	161	72	88	38	8511
Day6	150	147	123	36	4690
Day7	69	78	82	38	8768
Day8	100	104	111	37	4121
Day9	98	87	86	37	7823
Day10	61	70	77	38	8543

The table presents the dataset being used, which includes Glucose Level, Temperature, Physical Activity etc.

Methodology



Proposed methodology flowchart

Results

Data classification performance is measured by accuracy, sensitivity, specificity, and precision. We define accuracy using the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} (\%)$$

Precision is estimated as the ratio between the value of true positives and the sum of the values of true positives and false positives:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Specificity is defined as the ratio between the value of true negatives and the sum of the total value of true negatives and false positives:

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP})$$

Results

Sensitivity is defined as the ratio between the value of true positives and the sum of the total value of true positives and false negatives:

$$\text{Sensitivity} = TP/(TP + FN)$$

Recall is defined as the ratio between the value of false negatives and the sum of the total value of true positives and false negatives:

$$\text{Recall} = FN/(TP + FN)$$

F-measure is a combination of precision and recall, and it is defined by the following equation:

$$F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

[1] The table presents the accuracy level and the training time of different algorithms

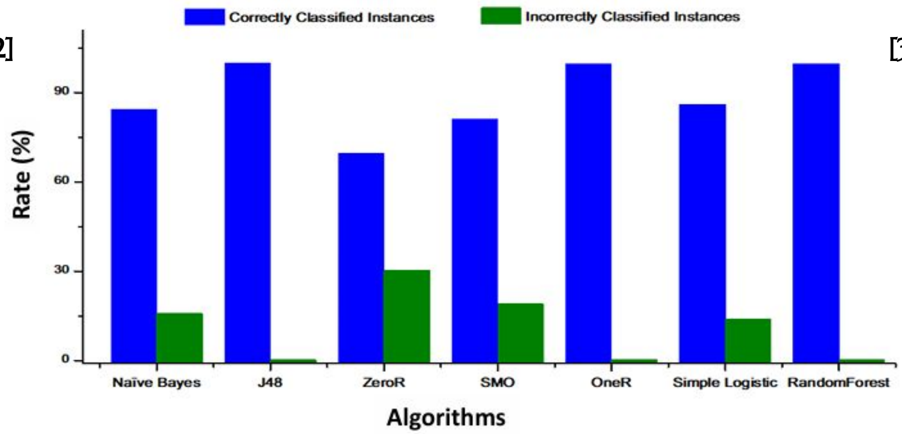
[1]

Algorithms	Correctly Classified Instances	Incorrectly Classified Instances	Training Time (s)
Naïve Bayes	84.1369%	15.8631%	0.03
J48	99.7619%	0.2381%	0.13
SMO	80.9524%	19.0476%	1.02
ZeroR	69.6875%	30.3125%	0.02
OneR	99.5685%	0.4315%	0.01
Simple Logistic	85.9077%	14.0923%	1.56
Random Forest	99.6577%	0.3423%	1.21

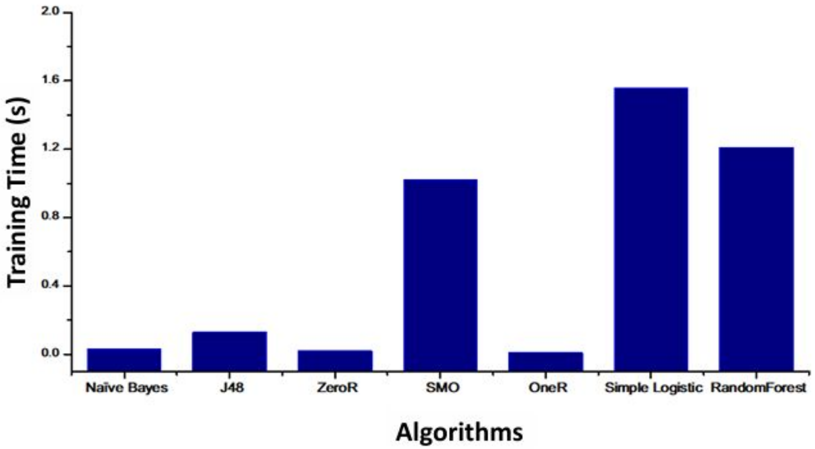
[2] Rate of correctly and incorrectly classified instances of algorithms

[3] Training time results for different algorithms

[2]



[3]



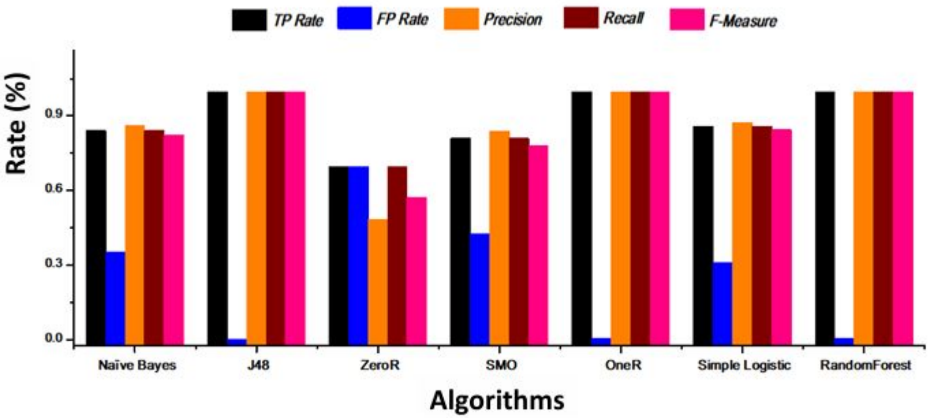
Results

[1]

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure
Naïve Bayes	0.841	0.353	0.862	0.841	0.824
J48	0.998	0.004	0.998	0.998	0.998
SMO	0.810	0.426	0.838	0.810	0.781
ZeroR	0.697	0.69	0.486	0.697	0.572
OneR	0.996	0.008	0.996	0.996	0.996
Simple Logistic	0.859	0.312	0.875	0.859	0.846
Random Forest	0.997	0.006	0.997	0.997	0.997

[1] The table shows the performance of each classifier in terms of precision, recall and F-measure

[2]



[2] Performance results of TP, FP, precision, recall and F-measure of different algorithms

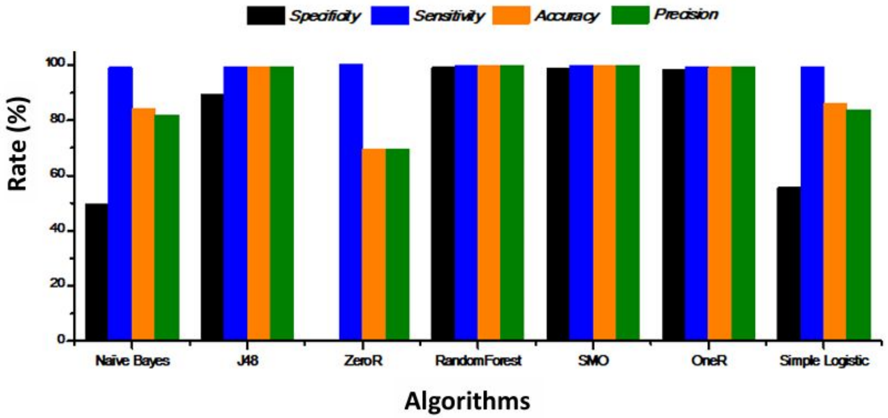
Results

[1]

Algorithms	Specificity	Sensitivity	Accuracy	Precision
Naïve Bayes	49.78%	99.08%	84.14%	81.94%
J48	89.49%	99.47%	99.17%	99.32%
SMO	98.92%	99.85%	99.66%	99.66%
ZeroR	0%	100%	69.69%	69.69%
OneR	98.32%	99.47%	99.11%	99.25%
Simple Logistic	92.32%	99.47%	99.11%	99.25%
Random Forest	55.67%	99.06%	85.91%	83.71%

[1] The table shows the values of specificity, sensitivity, accuracy, and precision for different algorithm

[2]



[2] Performance results of specificity, sensitivity, accuracy, and precision for different algorithms

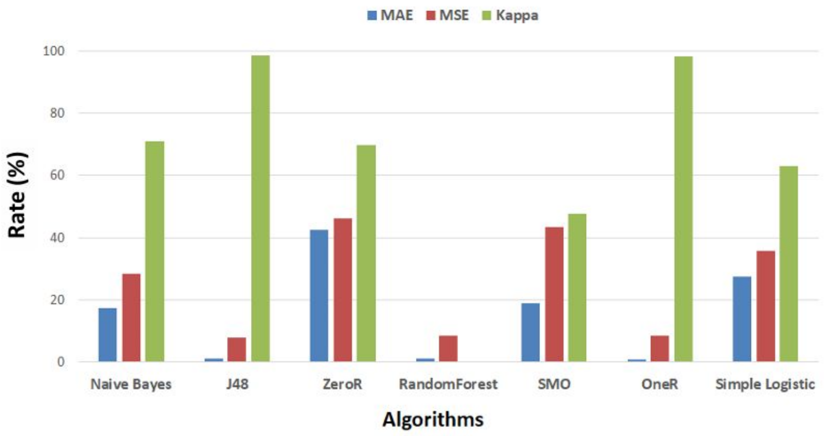
Results

[1]

Algorithms	MAE	MSE	Kappa
Naïve Bayes	17.5%	28.38%	70.89%
J48	1.2%	7.93%	98.53%
SMO	18.75%	43.3%	47.85%
ZeroR	42.65%	46.18%	69.69%
OneR	0.75%	8.63%	98.25%
Simple Logistic	27.49%	35.64%	62.89%
Random Forest	1.08%	8.56%	97.79%

[1] The table shows the values of Mean Absolute Error (MAE), Mean Squared Error (MSE) and Kappa for all algorithms

[2]



[2] Performance results of MAE, MSE and kappa for all algorithms

Results

Precision:	Correctly Classified Instances:	True Positive (TP) Rate:
Precision refers to the accuracy of positive predictions made by a classification model. A higher precision value means fewer false positives, meaning the model is better at correctly identifying true positive instances. In this case, the Support Vector Machine (SMO) algorithm achieved the highest precision of 99.66%. This suggests that SMO has a high accuracy in classifying diabetic patients correctly, minimizing the chances of misclassification.	This represents the percentage of instances correctly classified by each algorithm. Random forest achieved the highest value of correctly classified instances, indicating its effectiveness in accurately predicting the class labels of the data points. With a training time of 1.21 seconds, random forest demonstrates both high accuracy and efficiency in classification tasks.	The TP rate indicates the proportion of true positive instances correctly identified by the model. Both the J48 and random forest algorithms obtained the highest TP rate of 99%, implying that they are highly effective in correctly identifying diabetic patients who are at risk or experiencing health complications.

Conclusion

Predictive analytics in healthcare can help doctors and medical researchers obtain information from medical data and make intelligent and scientific decisions.

For this study, we proposed a monitoring system for diabetic patients using 5G technology and machine learning algorithms. We created an intelligent algorithm based on artificial intelligence on big data capable of analyzing the data of diabetic patients and sending a notification in case of emergency.

For this study, we employed a classification of diabetic patients using the WEKA tool according to six classifiers based on machine learning algorithms, i.e., naïve Bayes, J48, ZeroR, SMO, OneR, random forest, and simple logistic.

These algorithms were compared in terms of precision and accuracy. The proposed system was evaluated by several machine learning algorithms (naïve Bayes, SMO, J48, ZeroR, OneR, simple logistic, and random forest) and the simulation results demonstrated that the SMO algorithm exhibited excellent classification with the highest accuracy of 99.66%, a sensitivity of 99.85%, and a precision of 99.66%.

Scope for Future Work

For future work, we will examine the classification of each patient by adding other health parameters that should be taken into account to better measure the diabetes. Specifically, it could be interesting to add a galvanic skin response (GSR) sensor because when a person is suffering (or minutes before) a problem due to the low or high blood glucose level, usually experiment sweating. So, it could be a good indicator to predict episodes of hyperglycemia and hypoglycemia.

A thermocouple could also be stuck to the skin to better measure the body temperature.

Finally, we will work with other mathematical approaches and use new algorithms to improve the obtained results.

References

https://www.mdpi.com/2227-9032/8/3/348

Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System

Research Paper 3

Abstract

- Human Activity Recognition (HAR) using Wearable Sensors in Healthcare field: This research paper delves into HAR in Healthcare and emphasises its importance for tracking activities of daily life with the elderly population given the most attention.
- Technology and Challenges: Technological aspects of wearable sensors, data preprocessing, and recognition techniques in addition to the difficulties encountered, including data collection, privacy concerns, and the complexity of human behaviour are also emphasised.
- Systematic Review: The study provides classification and comparison of HAR systems by looking at its applications, methodology, datasets, and component parts.

Introduction

- To prepare for the ageing population, communities are concentrating on their healthcare systems.
- For monitoring and emergency response, Human Activity Recognition (HAR) using sensors and computer systems is becoming more popular.
- Because of its advantages in terms of accuracy, privacy, and continuous monitoring, wearable-based HAR is preferred over cameras.
- Wearable sensor drawbacks include privacy, energy consumption, size, look, and personal satisfaction.
- Handcrafted and deep learning techniques for preprocessing, feature extraction, and classification/recognition are some of the solutions.
- K-nearest neighbour, support vector machines, decision trees, and deep learning techniques like CNN, DBN, and RNN are some of the suggested techniques.
- The study offers a cogent architecture for HAR systems and analyses the parts and challenges from both a hardware and software standpoint.

Literature Review

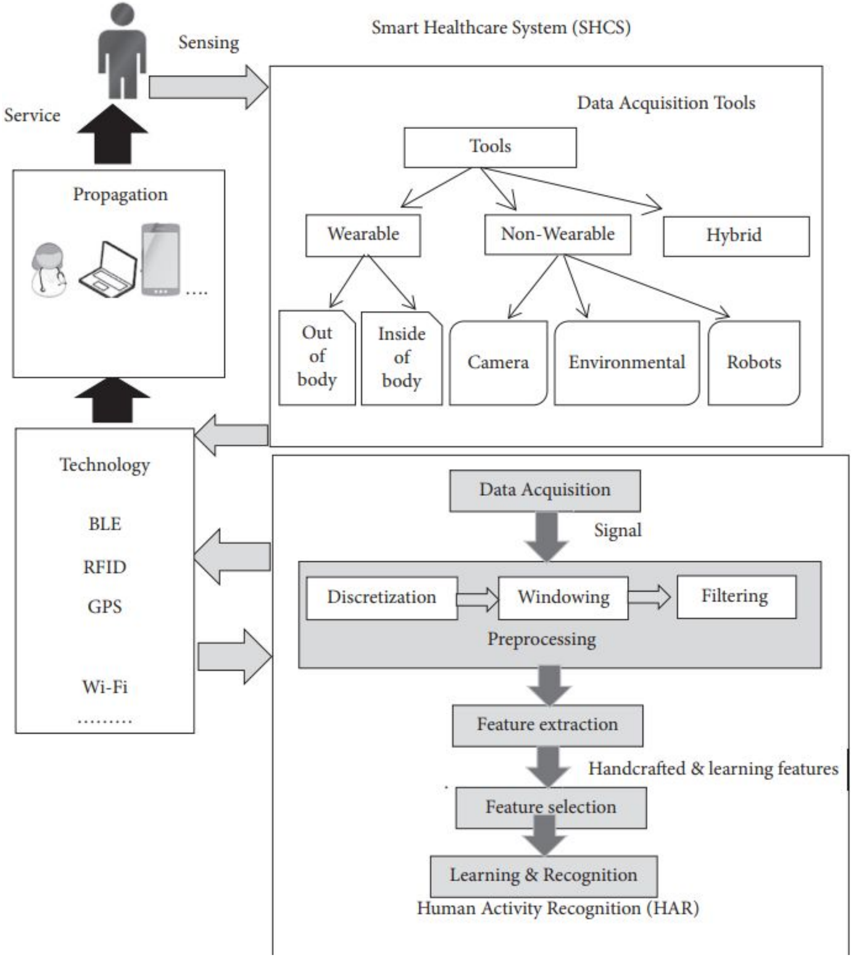
Implementation

- Data Acquisition: Wearable sensors are preferred for their mobility, continuous monitoring, and privacy preservation. Challenges include thermal damage risk, size, water resistance, and power consumption.
- Preprocessing: Involves windowing techniques like time-based, sensor-based, and activity-based to prepare data for feature extraction. The choice of windowing affects the accuracy of activity recognition.
- Feature Extraction: Techniques like PCA, LDA, ICA, and FA are used to reduce dimensionality and extract relevant features from raw data. Deep learning methods like CNNs can also automatically extract features.

Literature Review

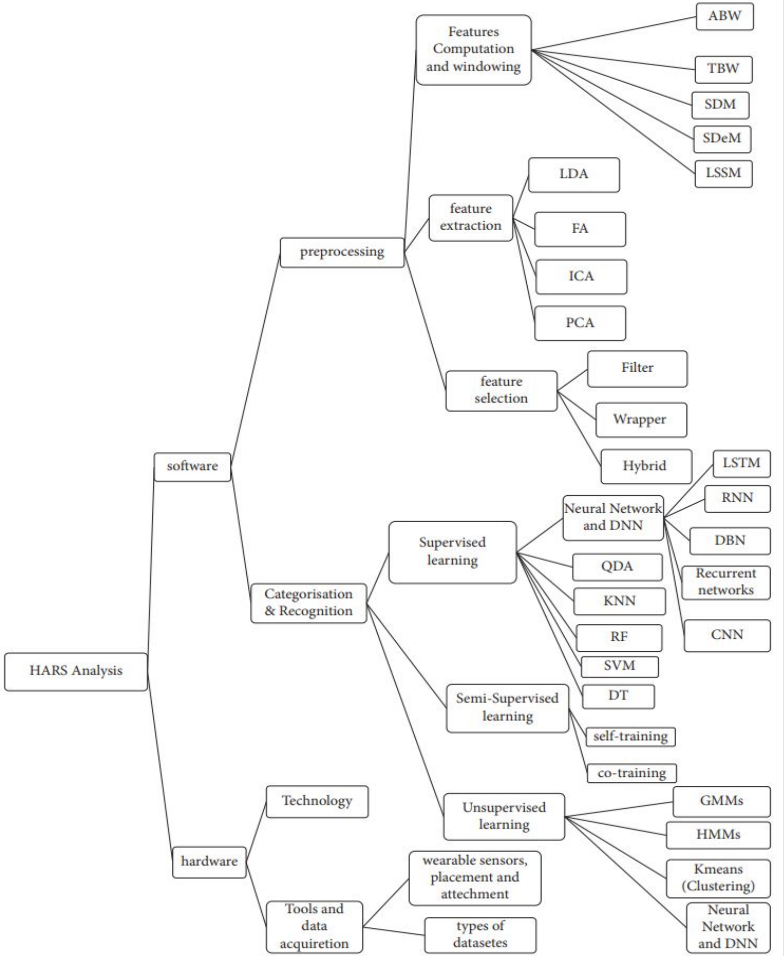
- Feature Selection: Aims to reduce feature space and improve classification performance. Methods are categorized into filter, wrapper, and hybrid approaches, each with its own advantages and computational considerations.
- Learning and Recognition: Involves supervised, semi supervised, or unsupervised learning to classify and recognize activities. The choice of method impacts the system's ability to generalize and adapt to new data.

Diagram of Application of Health Activity Recognition in Healthcare System



Flowchart

A detailed breakdown of HAR and its different components

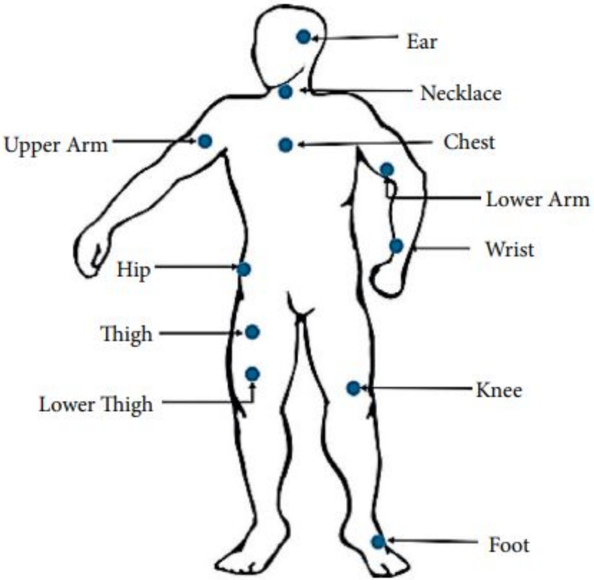


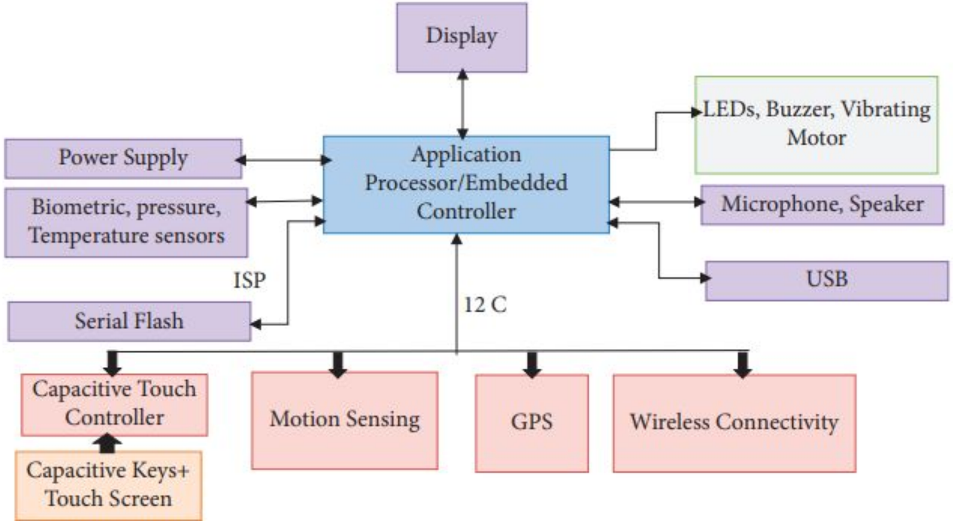
AN overview on Dataset used

Dataset name/ Publicly available	Year	Source	# Classes	# Actors	Sensor location	Single/ multiple person	Size
Skoda/Pub	2008	WS	10 gestures in car maintenance	1 subject	19 sensors on both arms	Single	_
USC-HAD/Pub	2012	IMU with Acc, Gyro, Mag	12 activities	17 (7 M, 7 F)	Front right hip	Single	_
PAMAP/Pub	2012	WS (IMU, 3 Colibri), HR monitor	18 activities	9 subjects	_	Single	3,850,505 (52 attributes)
Opportunity/Pub	2012	WS:(7 IMU, 12 Acc,7 Loc), OS (12), AS (21)	6 runs per subject (5 ADL and 6th for drill)	4 subjects	Upper body, hip and leg	Single	2551 (242 attributes)
UCI-HAR/Pub	2012	SPS (Acc, Gyro)	6 activities	30 (19–48 years)	Samsung galaxy SII mounted on waist	Single	10,299 (561 attributes)
Heterogeinity (HHAR)/Pub	2015	SPS and SW Acc, Gyro	6 activities	9 users	8 SPS & 4 SW	Single	43,930,257 (16 attributes)
MobiAct/Pub	2016	SP (Acc, Gyro)	9 ADL activities and 4 types of falls	57 subjects (42 M, 15F) of (20–57 years)	Samsung Galaxy S3 SP in trousers' pocket	Single	2500
UniMibShar/Pub	2017	SPS	17 activities (9 ADL and 8 fall)	30 of (18-60 years)	_	Single	11,771 samples
WISDM/Pub	2019	SPS and SW's Acc, Gyro	18 activities	51	_	Single	15,630,426

Methodology

Placement of Wearable sensors





Block diagram of wearable sensor

Result

Accuracy is the number of Computational Intelligence and Neuroscience 25 classified activities correctly (diagnosed) to the total number of activities classified:

Recall and precision: total number of correctly identified samples known as recall and accuracy in the HARS are expressed below:

F1 result includes two precision and recall measures that express the system's trust under evaluation to identify the agent's activities. F Measure is used for this criterion by weight injection:

$$\text{accuracy} = \frac{\text{items classified correctly}}{\text{all items classified}}.$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$F - \text{measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}},$$

$$F_1 - \sum_i 2 \times w_i \frac{\text{precision}_i \times \text{recall}_i}{\text{precision}_i + \text{recall}_i},$$

Conclusion & Scope for Future Work

- Activity Recognition: Methods for recognizing human activities, emphasizing the variability in how activities are performed.
- Fuzzy Systems & Networks: Use of fuzzy systems and recurrent networks, such as LSTM, for online Human Activity Recognition Systems (HARS).
- Sensor & Camera Integration: Combining wearable cameras with sensors to enhance accuracy, leveraging deep learning techniques.

In conclusion the potential of integrating advanced computational methods and technologies to create robust and accurate system is the future of HARS.

References

1) https://link.springer.com/article/10.1007/s10462-021-10116-x

2) Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System (hindawi.com)

SOMAIYA VIDYAVIHAR UNIVERSITY

K J Somaiya College of Engineering

Smart wearable model for predicting heart disease using machine learning

Research Paper 4

Abstract

The research aims to develop a wearable biomedical prototype for predicting the risk of heart disease using machine learning algorithms, focusing on the use of wearable technology to address heart disease, a leading cause of death. The prototype uses an electrocardiogram (ECG) sensor to obtain heart rate data, employing the R-to-R method to monitor variations in ECG patterns. The study leverages the Cleveland heart disease dataset, which includes 13 attributes, such as resting ECG results, depression in ST-segment induced by exercise relative to rest, and the slope of peak exercise segment.

The research uses the Random Forest Algorithm to predict the presence of heart disease with an efficiency rate of 88%. For testing purposes, the prototype does not involve live human subjects but uses static real data, sending results to a mobile app for further action. This proof-of-concept prototype has potential applications for elderly people, serving as assistive equipment for monitoring heart health outside hospital environments, which is especially beneficial in regions with low doctor-to-patient ratios.

Introduction

Periodic health exams are vital for early disease detection and prevention, but global healthcare faces challenges, with over 45% of WHO countries having fewer than 1,000 people per doctor. Cardiovascular disease (CVD) is a leading cause of death, with 31% of global deaths, largely from heart attacks. Most CVD deaths occur in low-income countries.Real-time health monitoring can help by allowing continuous patient tracking, especially where healthcare is scarce. A survey suggests many opportunities to detect heart disease early are missed, emphasizing the need for effective monitoring systems. Electrocardiograms (ECGs) are the best method for detecting cardiovascular issues in biomedical prototypes, offering a non-invasive, cost-effective solution.

Need for wearable biomedical devices and its cost

Wearable biomedical devices could improve healthcare access in rural India, where medical infrastructure is limited. However, commercial devices like the Fitbit Surge (Rs. 19,990), Apple Watch 4 (Rs. 52,900), and QardioCore ($220) are often too expensive for many people. The cost is driven by extra features not needed for basic health monitoring.

The product aims to lower costs by focusing only on the essential components for disease prediction, making wearable technology more affordable and accessible.

Introduction

Need for Machine Learning

Machine Learning (ML) is crucial for deriving insights from data, with applications across various fields, including wireless sensor networks, social media analysis, and medical diagnostics. Algorithms can process large datasets to identify patterns and make predictions, useful in healthcare for prognosis and assisting radiologists.

Several machine learning algorithms have been explored for predicting heart disease. For example, Support Vector Machines, XGBoost, decision trees, and k-nearest neighbors were used to detect ischemic heart disease.

The paper presents a wearable biomedical prototype that measures ECG patterns through sensors and calculates heart rate using the R-to-R interval. The data is used to train machine learning algorithms for heart disease prediction, with the Random Forest algorithm outperforming others. This algorithm, consisting of multiple decision trees, was trained on the standard Cleveland dataset and integrated into a mobile app. The app, connected to the wearable prototype, predicts the risk of heart disease, offering a practical tool for health monitoring.

Literature Review

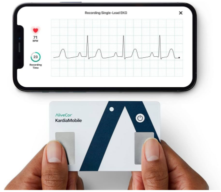
The surge in human population and imbalance in the patient-to-doctor ratio necessitates real-time health monitoring, leading to the emergence of IoT-based wearable sensors for this purpose. Wearable devices like Smart Healthcare Monitoring System (SW-SHMS) and Wearable IoT Cloud-based Health Monitoring System (WISE) offer remote health monitoring through sensors for heart rate, blood glucose, and more. While most research has focused on health monitoring, there is a growing trend toward using wearable sensors combined with machine learning to predict diseases, signifying a shift toward proactive healthcare.

Heart rate calculator

- The Garmin Forerunner 735XT is an exercise tracking device designed for runners, cyclists, and swimmers. It uses an optical sensor to measure blood flow and calculate heart rate. The device is primarily used to monitor various forms of physical activity.
- The FitBit Surge is an improved version of the Garmin Forerunner with enhanced accuracy in heart rate calculation. Its popularity stems from its watch-like design, making it accessible and easy to wear. It has been noted to provide more accurate heart rate measurements during periods of rest compared to other leading wrist-worn activity trackers.

Literature Review

ECG calculators



Kito+ is a credit-card-sized biomedical device used for ECG readings. It connects to an iPhone and displays heart activity through an app. It's recommended for individuals with irregular heart rates, but its limitation is that it's only compatible with iPhones.

Apple Watch has a new upgrade that allows users to take a 30-second ECG reading. This feature is an addition to the older Apple Watches, which could only measure heart rates. However, the Apple Watch is not designed for people with Atrial Fibrillation or other known arrhythmias, and it does not replace medical advice.

QardioCore includes a chest band and a smart app for measuring ECG. While it tracks heart activity, it does not analyze or interpret ECG patterns. It is mainly an instrument for checking ECG but does not provide diagnostic information.

Alivecor Kardia is an ECG monitoring device that reads and interprets ECG signals. It includes filters to smoothen ECG waves, offering more reliable results. However, like Kito+, it's limited to iPhone compatibility, potentially restricting its user base.

Literature Review

Heart disease predicting systems

- **Skin patch**:A skin patch designed to detect the release of negatively charged fatty-acid-binding protein (FABP3), which indicates a potential heart attack. It is typically placed behind the ear or on the wrist to measure the concentration of this protein, thereby estimating the likelihood of a heart attack.

- **CVD risk score calculator**:An AI-driven Cardiovascular Disease (CVD) risk score calculator developed collaboratively by Apollo Hospitals and Microsoft. It uses various lifestyle factors such as diet, smoking habits, physical activity, stress, and hypertension to calculate a more accurate risk score for heart disease, specifically tailored to the Indian population.

- **Heart attack prediction system**:A heart attack prediction system based on machine learning algorithms such as K-Nearest Neighbour (KNN), Decision Trees (DT), Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF). This system is effective for heart attack prediction but can suffer from overfitting and high-dimensional data issues.

- **Human heart disease prediction system using data mining techniques**:An approach for predicting heart disease using various data mining algorithms, including Naive Bayes, KNN, Decision Tree, and Neural Networks. While these algorithms can promise accurate results, they require large datasets to deliver consistent and reliable outcomes.

- **Real-time classification technique for early detection and prevention of myocardial infarction on wearable biomedical device**:A research study proposing a real-time event-driven classification technique using Support Vector Machines (SVM) and statistical outlier detection for early detection and prevention of myocardial infarction. This approach is noted for its energy efficiency and good battery life, with a reported accuracy of 90%, but its performance in real-world settings may still need further validation.

Proposed System

Table 1 Attributes of Cleveland dataset

S.no	Name of the attribute
1.	Age
2.	Sex
3.	Chest pain type
4.	Resting blood pressure
5.	Cholesterol
6.	Fasting blood sugar
7.	Rest ECG
8.	Maximum heart rate achieved
9.	Exercise induced angina
10.	ST depression due to exercise relative to rest
11.	Slope of the peak exercise ST segment
12.	No. of major vessels colored by fluoroscopy
13.	Thalassemia
14.	Diagnosis of heart disease

The proposed system is a wearable heart disease prediction monitor, designed as a belt with probes to measure ECG at key spots. It sends biosignal measurements to a server where a Random Forest machine learning algorithm processes the data to predict heart disease, with results returned to the wearable device. This proof-of-concept model uses static data from a university clinic and hasn't been tested on human subjects.

System overview: The dataset is divided into training and testing samples, with pre-processing using Principal Component Analysis (PCA) before being fed into a Random Forest algorithm for prediction. The system architecture involves an ECG sensor that sends signal data to a Flask server, where the Random Forest algorithm processes it to predict heart disease risk. The prediction results are sent back to a mobile application, notifying the wearer about their risk level.

Dataset: The Cleveland Dataset from the Heart Disease Data Set, containing a subset of 14 attributes, is used for training and testing the learning model.

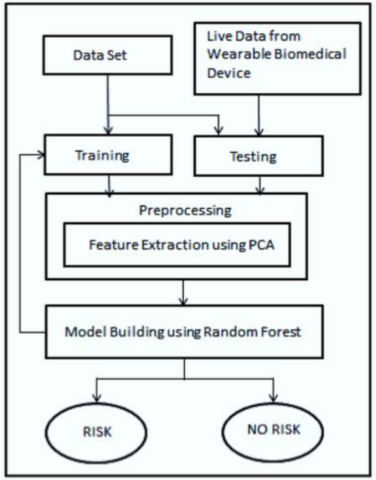


Fig. 1 System architecture

Proposed System

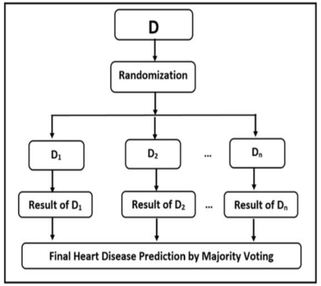


Fig. 4 Random forest

Flask Server

A Flask server running a risk prediction algorithm receives data from a mobile application and creates a JSON array using the numpy package. The data is de-serialized from a byte-stream using the pickle package to load a trained model for prediction. The model predicts using the predict() method, taking the JSON object as input.

Random Forest algorithm

The Random Forest algorithm creates an ensemble of decision trees, each working by splitting branches based on criteria like Gini Impurity or Chi-Square. It builds multiple decision trees by randomly selecting subsets of data, known as Bagging or Bootstrap Aggregating. The final prediction is made by combining the outcomes of all trees and selecting the most popular decision.

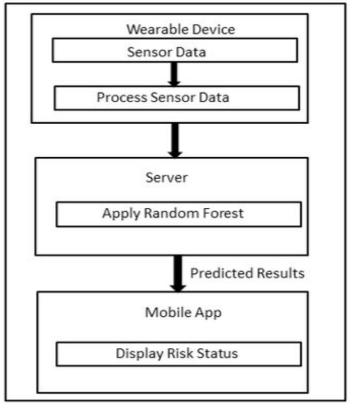


Fig. 2 Process flow diagram

Implementation

The system is a wearable biomedical prototype designed to monitor heart health and predict cardiac risk. It uses an AD8232 ECG sensor to measure heart electrical activity, connected to an Arduino UNO for signal processing. The data is transmitted to a smartphone via a Bluetooth module (HC-O5/ESP8266). The smartphone's mobile app collects user data like age, gender, cholesterol, and sends it to a Flask-based server. This server runs a machine learning algorithm to predict cardiac risk, with a focus on random forest and PCA. Results are sent back to the mobile app for user notification, providing heart health insights.

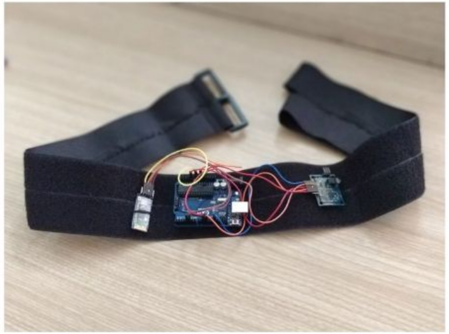


Fig. 8 Wearable biomedical prototype

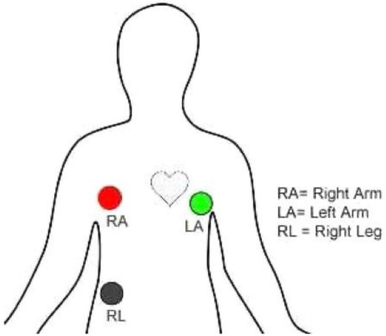


Fig. 5 Sensor pad locations on the chest

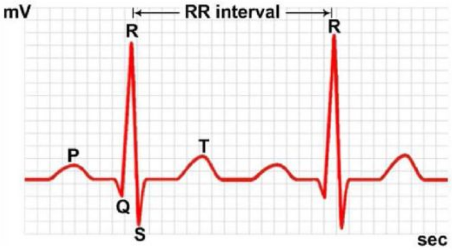


Fig. 7 Interval between two R waves

Result

The model uses Principal Component Analysis (PCA) to reduce data dimensionality for visualization, creating a scatter plot with PCA1 and PCA2 to distinguish between individuals at cardiac risk and those who are safe. Several classifier algorithms were tested, including Linear SVM, Radial Basis Function (RBF) kernel SVM, and Random Forest, with Random Forest proving most accurate.

To optimize the Random Forest classifier, 3 sets of 20-run trials were conducted, resulting in parameters with 10 estimators, "gini" criterion, and random state 0. The data was then split into a train and test set using Stratified K-Fold Cross Validation with 5 folds. The classifier was designed to predict if an individual is at risk or not.

The model's performance was evaluated using a confusion matrix, yielding the accuracy, precision, recall, and F1 scores for the Random Forest classifier. This evaluation confirms that Random Forest outperformed other classifiers in terms of accuracy, making it the chosen algorithm for this model.

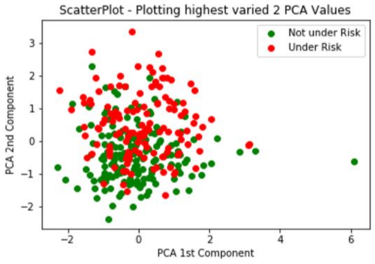


Fig. 9 Visualization of Dataset using PCA

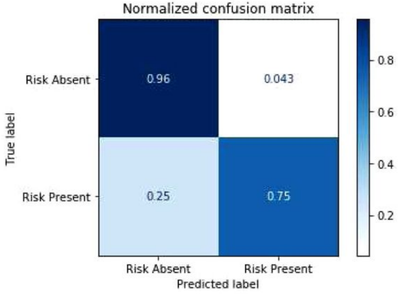


Fig. 10 Normalized confusion matrix

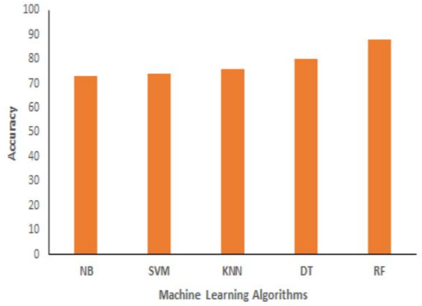


Fig. 11 Comparison of accuracy

Conclusion

The Random Forest Algorithm was found to be the most suitable for predicting heart disease, achieving 88% accuracy. This model was implemented in a wearable biomedical prototype that alerts users of potential heart disease risk via a mobile app, encouraging them to seek medical attention. This innovation addresses the rising fatalities from heart disease and the shortage of physicians in some areas, emphasizing the need for affordable IoT-based heart disease prediction systems.

Future research could focus on enhancing the device's wearability by using body area networks and smaller components like Arduino Nano. Additional sensors such as temperature, galvanic, and motion sensors could broaden the prototype's applicability. Expanding the dataset could further improve the model's performance, leading to more accurate predictions and a wider range of health monitoring applications.

References

- Jansi Rani, S.V., Chandran, K.R.S., Ranganathan, A. *et al.* Smart wearable model for predicting heart disease using machine learning. *J Ambient Intell Human Comput* **13**, 4321–4332 (2022). https://doi.org/10.1007/s12652-022-03823-y

- Adhikari, Das NC, Alka A, Garg R (2017) Hpps: Heart problem prediction system using machine learning. In: Computer science and information technology, SIGI, pp 23–37

- Al-khafajiy M, Baker T, Chalmers C, Asim M, Kolivand H, Fahim M, Waraich A (2019) Remote health monitoring of elderly through wearable sensors. Multimedia Tools Appl 78(17):24681–24706

- Banerjee S, Paul S, Sharma R, Brahma A (2018) Heartbeat monitoring using iot. In: 2018 IEEE 9th annual information technology, electronics and mobile communication conference (IEMCON), IEEE, pp 894–900

Turnitin Plagiarism Check

› Internal Assessment 2 (IA2) ?

Paper Title	Uploaded	Grade	Similarity
ADS Honors IA-II Final.pdf	17 Apr 2024 17:59	--	10%

Thank You!