International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072

Real Time Health Analysis and Prediction of Obesity and vulnerability to Heart Disease

Siddhant Salvi[1]

[1]*Department of Electronics and Telecommunications Engineering, Dwarkadas. J. Sanghvi College of Engineering, Mumbai, India*

Abstract – *The prevalence of junk food is increasing day by day and hence the vulnerability of younger generations to become obese is highly likely. Fostering health is of utmost importance for the society to grow. Many people have been suffering from this chronic disease of obesity. According to the World Health Organization (WHO), there are almost 1 billion people in the world facing the problem of obesity. Obesity can cause severe damages to the person's health and even sometimes lead to death. Early detection of obesity would help the patient fight the disease and completely eradicate the possibility of becoming obese. However, early diagnosis is not as easy as it seems, there are no traits that the person is cognizant about and hence it is highly unlikely that the person would take efforts to diagnose them and also availability of a specialist on 24 hours basis is not possible. The preliminary concept of cloud-based diagnosis to identify potential obesity using machine learning techniques was suggested in this paper. Various machine learning algorithms were used for precision and the one with highest accuracy was implemented in diagnosis of the disease.*

For highly accurate results, the predictive model was trained on the existing datasets of obese and overweight patients. Along with the machine learning, IOT sensor network was used for real time diagnosis of the patient thus achieving early detection and helping in eradicating and lower the problems related to the obesity.

Key Words: **K-NN, Random Forest (RF), SVM, Decision Tree, IoT, Sensor Network, Obesity detection**

1. INTRODUCTION

Major advancements in the field of medical have helped solve the problem of various patients. With the help of cellular technology, patients can communicate with the doctors at distance and the doctor, accordingly, can forge prescriptions to the patients. Readily available 24 hours healthcare facilities have helped in mitigating severe problems of the patients and without wasting much of a time, the patient can be cured. Rapid advances in health systems and low-cost cellular connectivity have significantly helped deal with the issue of fewer hospital facilities over the past decade. The introduction of mobile connectivity with wearable sensors has enabled the transition from clinic-centric to patient-centric healthcare facilities and is referred to in the literature as telemedicine. Telemedicine has allowed healthcare specialists to diagnose and treat the patients using the telecommunication technologies.

On similar basis, the telemedicine can help in the treatment of obesity. According to the WHO, the vulnerability to obesity is burgeoning and is going to flourish in the near future with the prevalence of junk food. With younger children being susceptible to eating junk food, they are highly likely to become overweight or even obese. Early detection of obesity is important, if not detected in the inchoate stage, there is a high possibility of being a diabetic and even cause cholesterol problems, leading to heart attack. The WHO states that, Worldwide obesity has nearly tripled since 1975. In 2016, more than 1.9 billion adults, 18 years and older, were overweight. Of these over 650 million were obese.

39% of adults aged 18 years and over were overweight in 2016, and 13% were obese. Most of the world's population live in countries where overweight and obesity kills more people than underweight. 38 million children under the age of 5 were overweight or obese in 2019. Over 340 million children and adolescents aged 5-19 were overweight or obese in 2016. Doctors have found that, if obesity is treated as soon as possible, there is a high chance of assuaging the possibility of heart disease and even the possibility of cancer which can cause death. As obesity is preventable, and with the advent of technology, it is comparatively easy to detect possibility of obesity in the early stage, it should be treated in the embryonic stage itself.

As the humans are incognizant about the symptoms of obesity, they are highly unlikely to detect the possibility of obesity in the later stage and thereby making them vulnerable to heart disease and even cancer leading to death. Sensor technology along with machine learning techniques can be used for early detection and treatment of obesity. Wearable sensors help in determining the various symptoms that are humans are facing and with the help of machine learning techniques, the vulnerability of the patient to the obesity is determined and accordingly he can be treated to help solve the problem of obesity.

There is plethora of datasets available on obesity. The machine learning model is trained on the available dataset. Various machine learning algorithms are applied for the training of the model and the one with highest accuracy is implemented for better results in prediction. Using the sensor network real time data of the patient can be sent to the cloud and the data sent undergoes the predictive model for determining the possibility of obesity in the patient.

The description of an obese and overweight patient management device using the idea of the Internet of Things (IoT) with numerous physiological signal sensors and Arduino microcontrollers is another contribution to this article. The Internet of Things (IoT) technology is currently being used by sensor networks to capture, process and transfer information from one node to another. After a certain period, the sensors gather data, interpret it and use it to trigger the necessary action, and create an intelligent cloud-based monitoring, planning and decision-making platform. IoT-developed devices, such as embedded systems, allow information to be shared between other nodes or over the Internet, and it has been projected that about 8 to 50 billion devices will be connected by 2020.

2. SYSTEM ARCHITECTURE

It is considered that, IoT in healthcare along with interconnected medical sensors, both wearable and implantable, is capable of providing smart, precise and cost-effective personalized healthcare solution to the patient whenever required at any time of the day. The implementation of sensor networks using IoT and the use of machine learning techniques for the prediction of obesity is shown in the figure 1:

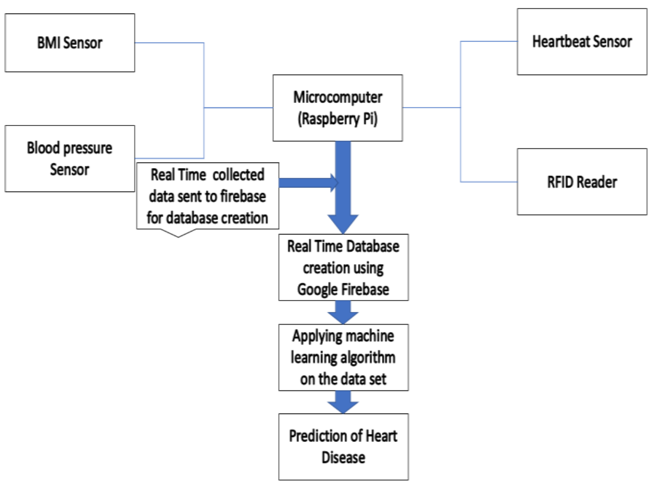

Fig -1: Real Time Obesity and Heart Disease Prediction

In the sensing layer, various sensors are connected to check the health of the patient. Various sensors include blood pressure sensor, sensors for calculating weight, heart beat sensor, ECG. The sensors included are wearable sensors and thus, whenever the patients want to know the status of his health can wear the sensors and check whether he is vulnerable to obesity or even any heart disease which are caused by obesity. These sensors capture the biomedical parameters and send the data to the cloud for real time database creation and further, on board predictive model is made to run on the real time data set by the sensors thereby predicting the vulnerability to obesity and heart diseases. The following sensors were used:

- The **heartbeat sensor** is a quick way to research the role of the heart, which tracks blood ow through the area of contact. The amount of blood in the region of touch varies over time, as the heart pushes blood into the blood vessels through the skin. The sensor shines through the skin patch with a light lobe (small incandescent lamp) and tests the light that is transmitted. The clip may be used on the thumb and index finger on a fingertip or on the network of tissue.
- The **blood pressure sensor** is a non-invasive device developed for human blood pressure monitoring purposes. Using the oscillometric form, it tests systolic, diastolic, and mean arterial pressure. All the time, blood pressure does not remain the same. It moves to satisfy the needs of the body. Various variables, including body location, breathing or mental condition, exercise and sleep, influence it.
- A **load cell sensor** is used for determining the weight of the person which can be fed into the database to check for vulnerability to obesity. It is a transducer which converts the force into measurable electrical output. The load cell converts mechanical force into measurable digital values and integrating it with the ultrasonic sensor can help us determine the height of the person further calculating the BMI of the patient to determine whether the patient falls in the category of obese or not.
- An **RFID reader** is connected to the processor board, which facilitates the identification of different users, thus a RFID tag should be amounted to each individual patient for optimum data logging and easier user-interfacing.

Secondly, the network layer allows the transfer of data to the corresponding data process units to be effective and secure. A variety of short-range communication protocols, such as ZigBee, have been widely implemented. The third layer is the data processing module, which is responsible for collecting useful information from the sensor data collected from the first layer. The model implemented uses a Raspberry Pi as the processing unit for sensor data integration and cloud capabilities. The system takes the patient's health information from sensing layer to Google Firebase for database creation. The most promising approach for data mining has been learning-based methods. The initial dataset is improved with patient data integration to increase machine learning model accuracy for determining the vulnerability of the individual to obesity and also checking for possible heart disease that could cause obesity and viceversa. Finally, intelligent resources and software will be provided, such as disease detection based on the actions of the top three levels. With improved sensor feedback and real-time monitoring and cloud access capabilities, the system can be very useful across remote locations with heart disease critical patient history vulnerable to obesity.

3. DATASET

The training of the model is done on the basis of the available datasets pertaining to the heart disease records of the patients and also the datasets of the patients containing the weight, height, BMI of the patient. Based on the parameters in the dataset such as the cholesterol levels of the patients, maximum heart rates, blood pressure rates, vulnerability to diabetes, weight, height, BMI, blood sugar, the model is trained to determine the vulnerability to obesity and heart disease because of obesity. The following histogram graphs shows the various parameters in the dataset that helped in development of the model.

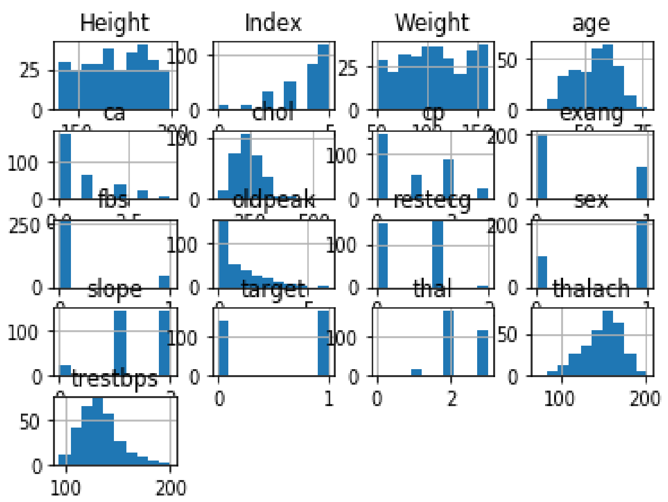

Fig -2: Histogram of the dataset

Table -1: Dataset features for the model

DATASET DESCRIPTION	
Parameters	Description
age	Patient age in years
Cp	Chest Pain Type (typical angina = 1, atypical angina = 2, non-anginal pain = 3, asymptomatic = 4)
restecg	Resting electrocardiographic results (normal = 0, ST-T wave abnormality = 1, left ventricular hypertrophy = 2)
thal	Normal = 0, fixed defect = 2, reversible defect = 3
Thalach	Maximum heart rate
oldpeak	Exercise based ST depression
Index	1= underweight, 2= normal weight, 3= over weight, 4= Obese

4. MACHINE LEARNING ALGORITHMS

The performance of four algorithms is considered for the diagnosis of the patient to detect the probability of the occurring of cancer disease to the patient. The algorithm with the highest accuracy was implemented for the detection of the cancer. The four algorithms implemented on the dataset were K-NN algorithm, Support Vector Machine (SVM) Algorithm, Random Forest (RF) algorithm and the Decision Tree algorithm.

4.1. K-NEAREST NEIGHBOUR

The k-nearest neighbour algorithm (k-NN) is a method for the classification of an object among its k-nearest neighbours based on the majority class. k-NN is a type of lazy learning where the function is only locally approximated and all calculations are postponed until classification. The Euclidean distance is commonly used by the k-NN algorithm. However, it is also possible to use some other point, such as the Chebyshev norm or the Mahalanob distance. Euclidean distance is being used in this experiment. Suppose the question example of co-ordinates (a, b) and the training sample coordinate is (c, d), then the Euclidean distance square is:

$$x^2 = (c - a)^2 + (d - b)^2 \qquad (1)$$

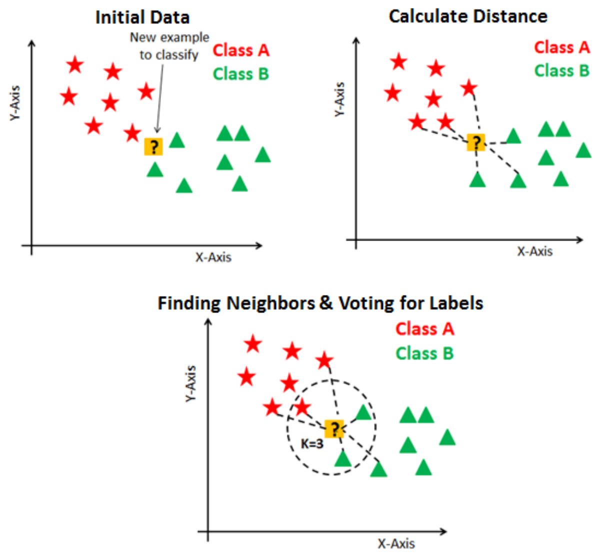

Fig -3: K- nearest neighbor algorithm

4.2. Support Vector Machine (SVM)

By constructing an n-dimensional, SVM performs classification and maximizes the margin to achieve the best result in classification. SVMs are based on the concept of hyper-plane or linear separability classifiers. Supposing we have n training data points (x1, y1); (x2, y2); ……... (xn, yn)

where $x_i \in R^m$ and $y_i \in \{-1, +1\}$. Considering a hyper plane given by (w, b), where w is weight and b is bias.

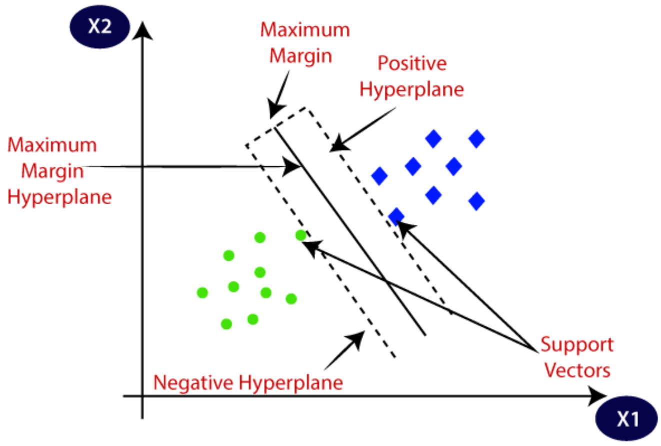

Fig -4: Decision Boundary with hyperplane

Classification can be given by:

$$D(x) = \text{sign}\ (w.x + b) = \text{sign}\ ({}_i\sum^n a_i y_i\ (x_i.x) + b)$$

where w represents the hyper-plane, and weight vector direction gives us the class expected. The data points that are similar to the hyper plane, which are called the support vectors have a minimum distance to the decision boundary as shown in Fig. 2.

SVM has limitations that it experiences lots of computational expenses and produces inconsistent results when data set is characterized by wide space of features and train data set is small. This can be solved by adding a kernel function in the feature space, instead of the inner product of two transformed data vectors. A function of the kernel is set to that corresponds to a dot product of two characteristic vectors in some extended space. In such processes there are some widely used kernel functions which have been implemented for comparison of accuracy:

- Linear Kernel function:
$$K(x_i, x_j) = x_i . x_j \qquad (2)$$
- Sigmoid Kernel function:
$$K(x_i, x_j) = \tanh(a x_i . x_j + b) \qquad (3)$$
- Polynomial Kernel function:
$$K(x_i, x_j) = (x_i . x_j + 1)^p \qquad (4)$$
- RBF Kernel function:
$$K(x_i, x_j) = \exp[-\Upsilon|x_i - x_j|^2] \qquad (5)$$

4.3. Random Forest

Random Forest is an ensemble of learning algorithms based on methods. RF consists of a series of classifiers for tree. Every tree is composed of nodes and edges. The received group classifies new data points through a majority within each classification model's predictions, as shown in Fig. 4.

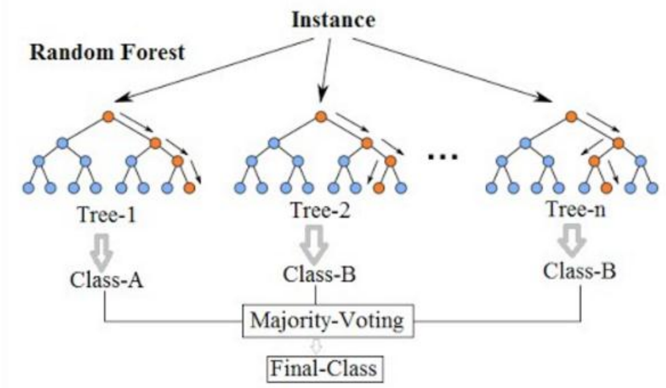

Fig -5: Random Forest Classifier

This approach incorporates a bagging cycle (bootstrap aggregation) and a set of random splits. Each tree is extracted from the data set from a separate bootstrap sample, and each tree categorizes the data. The final outcome is a majority vote between the trees. The random forest algorithm is defined by the following steps:

- Construct samples of the data from k trees bootstrap.
- For each of the bootstrap samples grow an unpruned tree.
- Randomly sample n-try of the predictors at each node, and pick the best split among those factors.
- Predict new data through a combination of the k tree predictions.

4.4. Decision Tree

The tree is a system of data that consists of nodes and edges. Sub-trees on a main tree path can be split into three, namely the root node, branch / internal nodes and leaf nodes. For a finite number of levels, the decision tree is a basic representation of a classification strategy. The internal node and the root node are marked with an attribute name; the edges are marked with potential attribute values; and the leaf node is marked with various classes. The decision tree is one of the most common classification models because the results obtained are easy to explain and easier to understand.

In order to represent the decision to make the decision linked to the decision tree, trees are used.

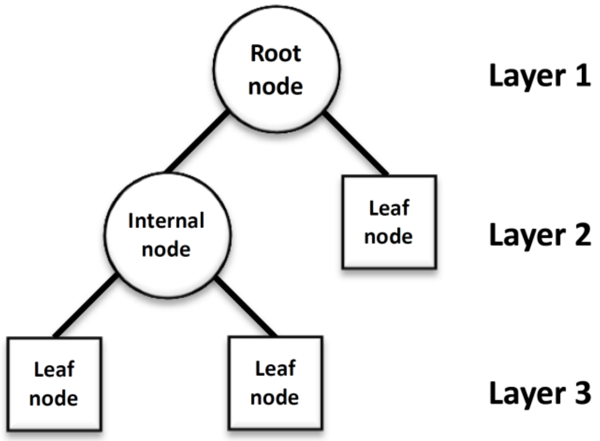

Fig -6: Random Forest Classifier

The decision tree method is to adjust the form of the data (label) into the model of the tree, transforming the model tree into a law. The biggest advantage of using the decision tree is that it becomes possible to break down complicated decision-making mechanisms so that decision-makers perceive the problem-solving solution. The decision tree is also useful for exploring the details, discovering hidden relationships with the input variables of the target variable with certain candidates.

5. RESULTS

The results of the implementation of the different machine learning algorithms on the dataset is discussed below. For providing the highest precision, the algorithm with the highest accuracy for the determination of obesity and its implications causing heart diseases is selected.

5.1. Parameter optimization of machine learning models

Various machine learning algorithms have been implemented to get the best performance as shown in above section. To further analyse and optimize the performance before conducting a comparative analysis, the algorithm parameters are iterated across to find the best match. As shown in Fig. 7: a), four kernel functions are used and trained with the dataset for the heart disease monitoring system. The performance of each kernel function i.e. polynomial (0.56), linear (0.84), radial basis (0.87) and sigmoid (0.82) is presented from which we go forward with sigmoid kernel function for support vector machine algorithm.

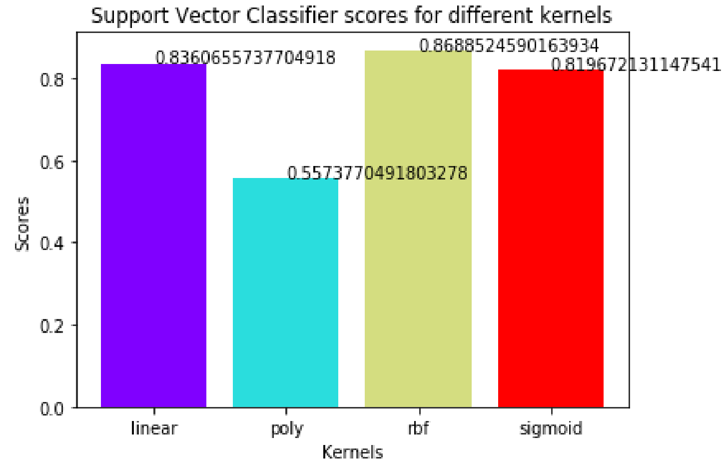

Fig -7(a): Support Vector Classifier for different kernels

Next, we consider the k-Nearest Neighbour algorithm across various values of k i.e. number of neighbours to analyse the variation in performance and choose the optimum value for comparative analysis across algorithms for dataset. As shown in Fig. 7: b), the accuracy is highest on k = 8, 12, 14 which will be implemented.

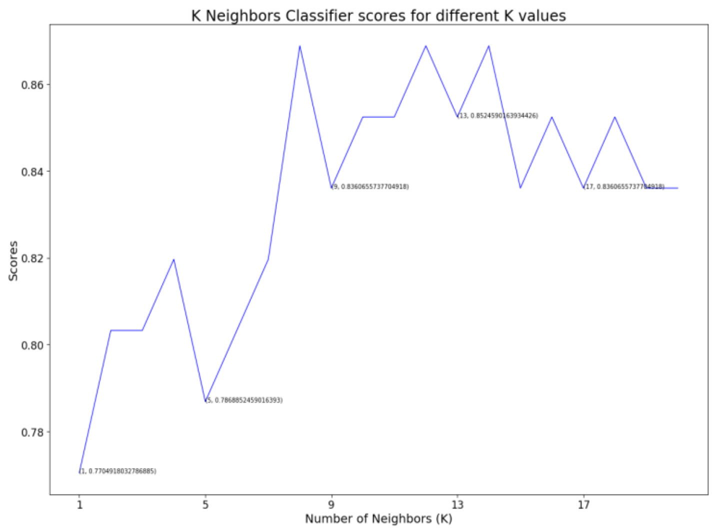

Fig -7(b): K-nearest neighbor scores for different k values

Similarly, the optimum number of features have been computed for decision tree as shown in Fig. 7: c), with 10 features.

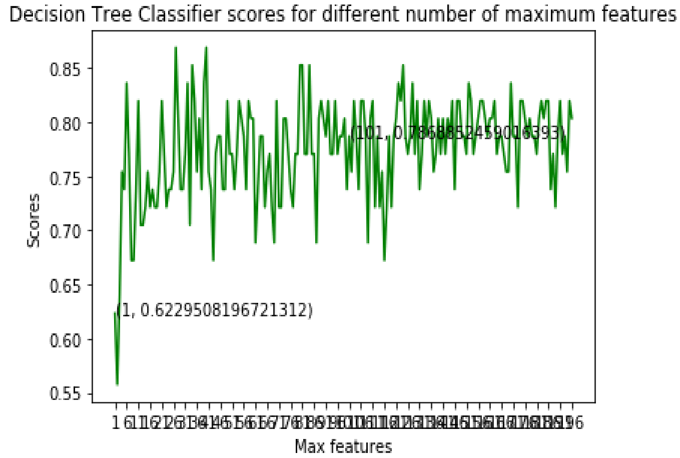

Fig -7(c): Decision tree classifier scores for maximum features

The number of estimators show variation in the performance of random forest algorithm although minimally. The highest scores are noticed on 200 estimators with 100 estimators being a close second as shown in Fig. 7: d). With 10 estimators, the accuracy was 0.85, for 100 it was 0.8, for 200 it was the highest 0.885.

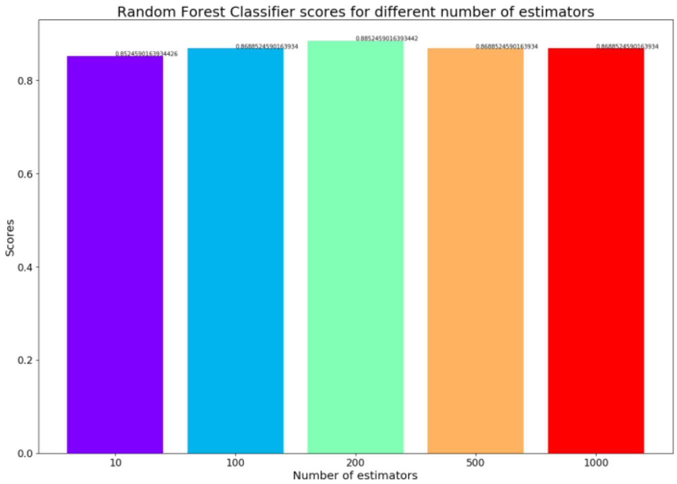

Fig -7(d): Random Forest algorithm for different estimators

5.2. Implementation

The optimum parameters are chosen for each algorithm and the dataset is training using these specifications. The training and testing data is segregated by a k-cross fold validation of 80 (train):20 (test) on the combined dataset. Out of the four algorithms implemented, Random Forest (0.885) gives the best performance in terms of accuracy followed by SVM (0.819), k-NN (0.836) and decision tree (0.803) as shown in Fig. 8.

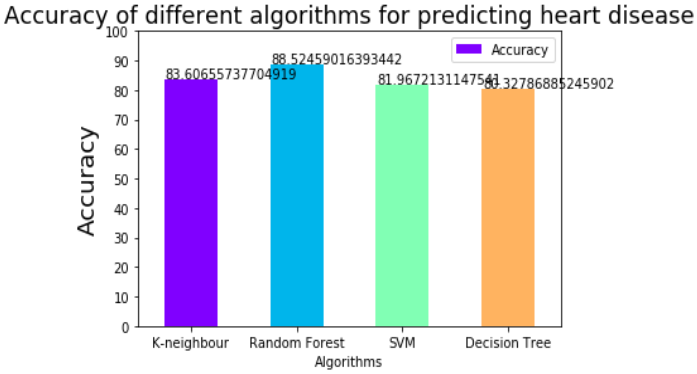

Fig -8: Accuracy of different machine learning algorithms

Of the four machine learning algorithms used, the random forest algorithm had the highest accuracy and hence was implemented on the model for the determination of obesity and possible vulnerability to heart diseases.

Further, the dataset is further built upon with the help of real-time sensor monitoring which is updated into the original dataset on which machine learning models are trained to further optimize the overall accuracy as well as give health and cardiac activity support to the user. The output on heartbeat sensor, blood pressure sensor, the values of weight, height and BMI are extracted and pre-processed to satisfy the original dataset features and add to the generalisation of system. The RFID tag is additionally implemented to keep a log of the patient's health with a unique identification number which keeps the system user-friendly.

6. CONCLUSION

With the dissemination of machine learning, artificial intelligence and IoT technologies, it has become possible for the health industry to treat patients remotely and accurately. This paper provides a succinct summary of the same technologies that could help solve one of the issues related to the health industry that is the prediction of obesity and the implications of obesity on heart and thereby predicting the vulnerability of heart disease in the near future. Furthermore, it demonstrates a modern technology architecture for health tracking, which requires real-time surveillance of patients or elderly users and requires access to data from the cloud along with the creation of a network to further enhance outcomes.

7. ACKNOWLEDGEMENT

I would like to acknowledge Prof. Tushar Sawant who has helped me during various aspects of the research, teaching techniques related to machine learning and IoT and providing support as when required.

8. REFERENCES

[1] W. Ping, W. Jin-Gang, S. Xiao-Bo, and H. Wei, \The research of telemedicine system based on embedded computer," in Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS'05), pp. 114-117, IEEE, Shanghai, China, January 2006.

[2] A. Gani, A. V. Gribok, Y. Lu, W. K. Ward, R. A. Vigersky, and J. Reifman, Universal glucose models for predicting subcutaneous glucose concentration inhumans," IEEE Transactions on Information Technology in Biomedicine, vol. 14,no. 1, pp. 157-165, 2010.

[3] World Health Organization, Health Statistics and Health Information Systems, World Health Organization, 2012.

[4] Vázquez, M.; Jimenez, E.; Nieto, J.I.; Sánchez, J.D.D.; Garcia, A.; Torres, J.P. Development of a Mobile Health Architecture to Prevent Childhood Obesity. IEEE Lat. Am. Trans. 2015, 13, 1520–1527. [CrossRef]

[5] Alloghani, M.; Hussain, A.; AI-Jumeily, D.; Fergus, P.; Abuelma'atti, O.; Hamden, H. A Mobile Health Monitoring Application for Obesity Management and Control Using the Internet-of-Things. In Proceedings of the 2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC), Beirut, Lebanon, 21–23 April 2016; pp. 19–24. [CrossRef]

[6] Wibisono, G.; Astawa, I.G.B. Designing Machine-to-Machine (M2M) Prototype System for Weight Loss Program for Obesity and Overweight Patients. In Proceedings of the 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, Thailand, 25–27 January 2016; pp. 138–143.

[7] Shin, S.-A.; Lee, N.-Y.; Park, J.-H. Empirical study of the IoT-learning for obese patients that require personal training. In Advances in Computer Science and Ubiquitous Computing; Park, J.J., Pan, Y., Yi, G., Loia, V., Eds.; Springer: Singapore, 2017; Volume 421, pp. 1005–1012. [CrossRef]

[8] Ahmed, M.U.; Bjorkman, M.; Causevic, A.; Fotouhi, H.; Lindén, M. An Overview on the Internet of Things for Health Monitoring Systems. Internet Things IoT Infrastruct. 2016, 429–436. [CrossRef]

[9] Hussain, A.;Wenbi, R.; Da Silva, A.L.; Nadher, M.; Mudhish, M. Health and emergency-care platform for the elderly and disabled people in the Smart City. J. Syst. Softw. 2015, 110, 253–263. [CrossRef]

[10] Maharaj, B.T.; Gupta, P.K.; Malekian, R. A novel and secure IoT based cloud centric architecture to perform predictive analysis of users activities in sustainable health centres. Multimed. Tools Appl. 2016, 76, 18489–18512.

[11] Pattanapairoj, S.; Silsirivanit, A.; Muisuk, K.; Seubwai, W.; Cha'On, U.; Vaeteewoottacharn, K.; Sawanyawisuth, K.; Chetchotsak, D.; Wongkham, S. Improve discrimination power of serum markers for diagnosis of cholangiocarcinoma using data mining-based approach. Clin. Biochem. 2015, 48, 668–673. [CrossRef] [PubMed]