A Cloud IoT Based Framework for Diabetes Prediction

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

Submitted By
Neha Sharma
(Roll No: 801631015)

Under the supervision of
Dr. Ashima Singh
(Assistant Professor)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA, PUNJAB, INDIA
June, 2018

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "A *Cloud IoT based Framework for Diabetes Prediction*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Ashima Singh* and refers other researchers' work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Neha Sharma

Signature:
(Neha Sharma)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Ashima Singh)
Assistant Professor
CSED

Acknowledgement

I would like to thank God for blessing me with all the strength and resources required to complete this task. I would like to express my deepest gratitude to **Dr. Ashima Singh** for guiding me through the whole process and providing me with your knowledge and experience.

I am also heartily thankful to **Dr. Maninder Singh,** Professor and Head, Computer Science & Engineering Department for motivation and providing uncanny guidance and support throughout the preparation of the thesis report.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection.

I would like to thank my friend and Phd scholar **Ms. Parampreet Kaur**, for being there for me always. Last but not the least, I would like to thank my family for their wonderful support and encouragement without which none of this would have been possible.

Neha Sharma
(801631015)

Abstract

A healthcare system using modern computing techniques is the highest explored area in healthcare research. Researchers in the field of computing and healthcare are persistently working together to make such systems more technology ready. Recent studies by World Health Organization have shown an increment in the number of diabetic patients and their deaths. Diabetes is one of the basic sicknesses which have long-haul complexities related to it. A high volume of medical information is produced. It is important to gather, store, learn and predict the health of such patients using continuous monitoring and technological innovations. An alarming increase in the number of diabetic patients in India has become an important area of concern. With the assistance of innovation, it is important to construct a framework that store and examine the diabetic information and further see conceivable dangers. Its early detection and analysis remain a challenge among researchers.

In this research a prosperous/advanced and skillful method is presented in this research including IoT to gain a better result from the diabetic dataset. The proposed system is evaluated using “Pima Indians Diabetes” data set

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

The lives of human beings are altered by the evolution of Information and Communications Technologies (ICT). Economy, policy and other areas of the society are greatly affected. Now, healthcare is tremendously associated with technology. This connection has been turned out to be more grounded in the course of the most recent two decades. Devices of all types can be simply established in most health centres that are the main reason for this growth. Moreover, telemedicine, which was first specified quite a few years prior, is currently a fact and has been highly expanded and development like this has reached out to other healthcare sectors.

1.1 IoT

Internet of Things (IoT) is contemplated like biological group affiliated corporal items which are obtainable on the internet. The 'thing' in IoT can be a man with heart screen or vehicle including worked in-sensors. It is capable of assembling and sharing data on the network with no physical aid. The implanted technology in the devices causes to associate with the inner situation or with outside situation, which then affects the conclusion. There are an expanding awareness and commitment of individuals with respect to their health. Medical examination depletes a huge part of hospital bills. Technology can change by sending medical checks to the patient's residence from a hospital (home-driven) so, the healthcare business is between the fastest to take on the Internet of Things. Figure1 exemplifies how this dramatic change in the practice of medicine will examine in IoT hospital. A patient suffering from diabetes will have an ID card that when examined will be linked to secure cloud that holds their electronic health documents and medical treatment chronicles. Doctors and attendants will easily utilize the record on tablet or computer.

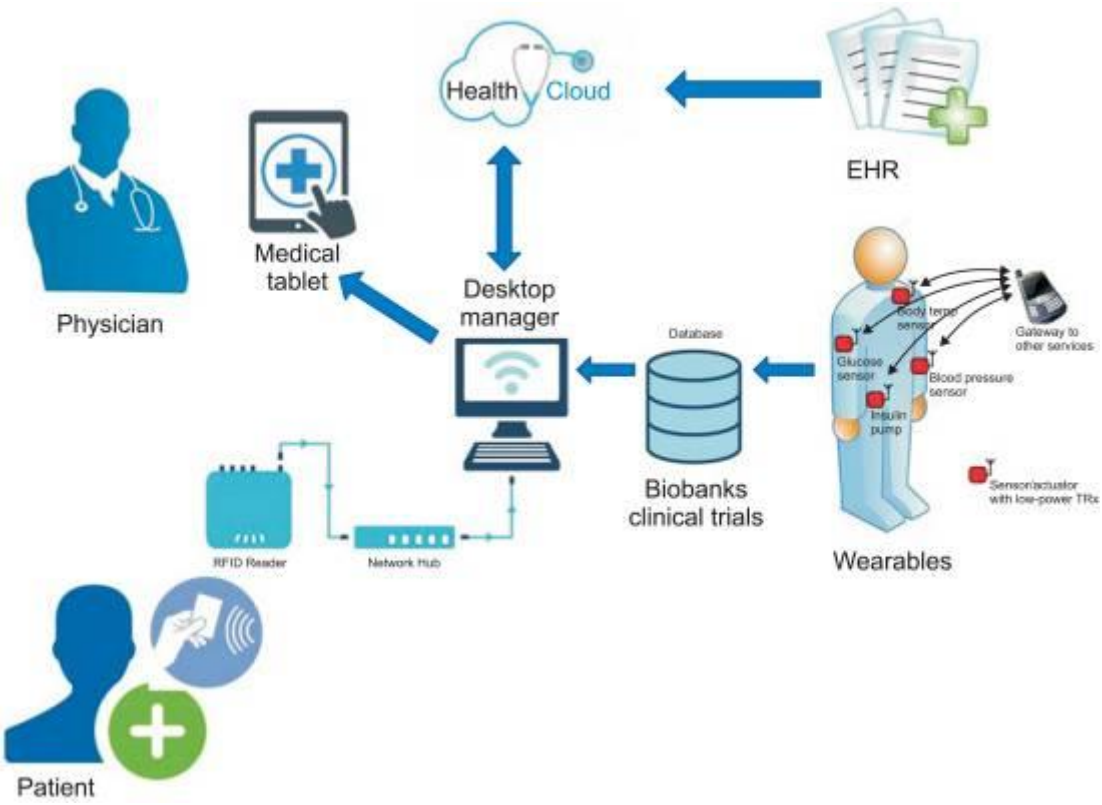

Figure1.1 Exemplifies the dramatic change in the practice of medicine will be examined in IoT hospital [1].

1.1.1 IoT infrastructure

Every "thing" beginning along tiny sensor and finishing along robotic manufacturing arm, it has the following attributes demonstrating its fit into IoT infrastructure:

- Sensors– For measuring activity in reality sensors are there in devices and systems.
- Connection – IoT system is first connected via gadget directly or indirectly via a hub, smart phone or base station and then it is further connected to the internet in any case. The utilization of different wireless technologies helps to maintain connections between gadgets.
- Processing unit – As any calculating unit, the gadget in the IoT background has some handling power for inspecting information and forwarding it. There is some exclusion and various remotely attainable appliances or sensors have no calculating potentiality but they still have argumentation for managing, receiving or forwarding the information.

- Energy efficiency – Big electricity consumption must not be required by all IoT devices to perform the required task.
- Economic efficiency – Sensors and small actuators must be inexpensive in producing and assisting because big solutions like smart phones appliances are high priced.
- Quality and reliability – User devices must be powerful and well assemble and IoT devices should be able to work in any environment.

1.1.2 Different concepts related with IoT

A schematic diagram of all the different concepts is provided in Figure 1.2
In IoT framework cloud computing and edge computing technologies are utilized.

Smart objects are utilized by an IoT framework. Smart objects are just gadgets which are attached to the internet and with the furnished interface for few predefined capacities for communicating, they are not same as smart gadgets. The structures are overlapping in the overview because a smart object can provide the identical capacities as a smart device. When the utilized objects are sensors attached to the internet it is called as wireless sensor network (WSN). Every smart object is not WSN so these structures in addition to overlap.

IoT and Cyber physical systems (CPS) are two unlike abstraction but there is an overlap. IoT technology is utilized by CPS application in most cases, for the interaction of smart objects and the accumulation of information.

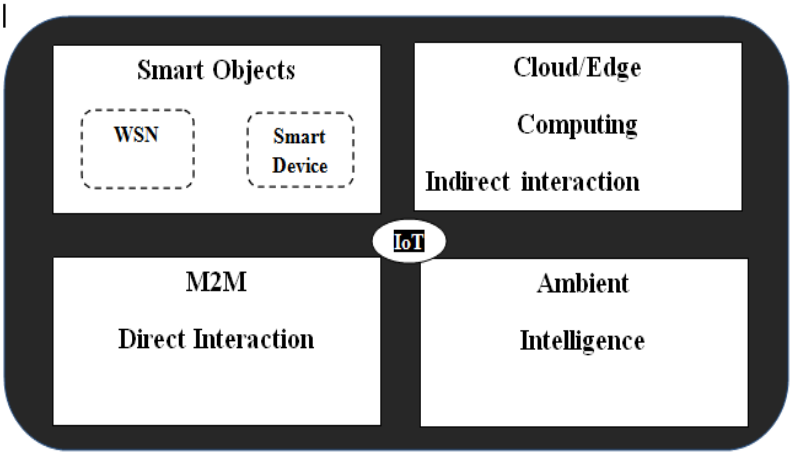

Figure 1.2 Different concepts related to IoT

To analyze and recognize medical conditions from IoT system patient data may be used and in this manner, we can mitigate frequent hospitalizations, and envision a beginning of attacks in chronic patients. Moreover, IoT frameworks may be used to give health administrations to retirement homes or homes for eldercare, geographically remote or difficult to reach regions.

1.1.3 Architecture of IoT

Network layer, Perception layer, and application layer are the layers in which internet of things can be detached. The perception layer comprises of Global position systems (GPS), two-dimension code follow alongside code reader, sensor network, radio frequency identification devices tag (RFID), sensor gateway, camera and so on. The primary reason for this layer is examination and finding of things, gathering and getting of data [2].The network layer is composed of different types of communication networks and converged networks. It has been generally acknowledged that this is growing element of network architecture. Moreover, the loT administration center and data center can be the elements for the network layer. The network layer that is mostly liable for changing the data work comprises of wireless and communication channels, wired networks, intelligent processing and network interface. The application layer is composed of the Internet of Things innovation joined with industry capacity to attain an expansive arrangement of smart application outputs. Throughout the application layer, IoT can achieve the task of reconciliation of information innovation with the business which can be utilized for a long period of time. It will greatly affect financial and social improvement. The elementary matter of the application layer is allocation and safety of information. Figure 1.3 clarifies the layered architecture of IoT and work related to each layer. Internet of Things (IoT) includes several tiny gadgets related together to outline a shared processing framework. IoT powers specific objectives to the extent connectivity, computational power and vitality spending plan, which make it exceptional in connection to those considered by the standard convention of security in the distributed system.

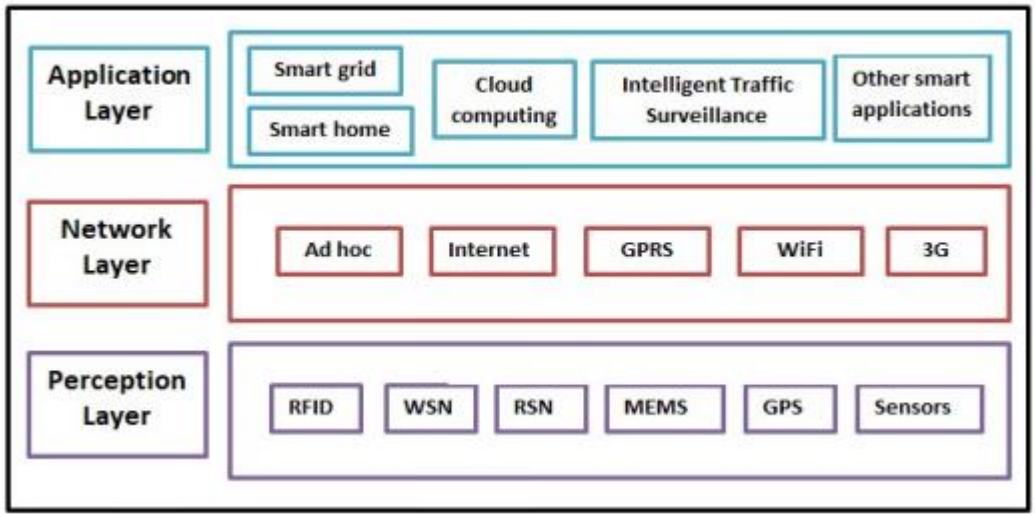

Figure 1.3 Architecture of IoT

1.2 Machine Learning

Electronic health records are becoming progressively usual in ambulatory care practices and in patient care. The transformation to digital records also describes an important transformation in how patient data are in order and made approachable. Early detection and screening play the main role ineffectual prevention of diabetes. The procedure of learning starts with surveyor data, such as examples, exact occurrences, or commands, in order to gaze for patterns in data and make satisfactory decisions in the future based on the examples that we provide as it shows in Figure 1.4. Machine learning is an informal investigation procedure that instructs PCs to do what falls into place without any issues for people. The purpose of machine learning for the majority is to understand the form of data and place that data into models which can be grasped and utilized through persons.

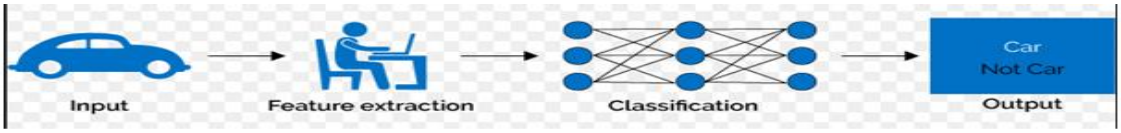

Figure 1.4 Machine Learning [5]

1.2.1 Machine learning methods

a) Supervised Learning: In supervised learning, the PC is given an illustration input which is decided with their coveted yields. The reason for this technique is for the

calculation to have the capability to "discover" by distinction between its genuine yield and the "instructed" yields to find out blunders, and change the model as needed. Supervised learning consequently used examples to anticipate mark esteem on extra unlabeled data.

b) Unsupervised Learning: In unsupervised learning, data is unlabeled, so the learning calculation is left to find out mutual features. The purpose of unsupervised learning might be as straight as discovering concealed examples inside a dataset. However, it might likewise have an aim of highlight learning, which enables the computational machine to accordingly discover the portrayals that are accepted to group crude data.

c)Semi-supervised Learning: In the above two specified types, either for all the observation in the dataset there is no labels or just labels are available for total the observing. In many empirical conditions, the amount of label is completely high, since it needs an experienced human specialist to do that. In this way, in the lack of description in the larger number of the inspections but available in hardly, semi-supervised algorithms are the most appropriate applicant for the model building. These techniques utilize the fact that even though the group memberships of the unlabeled information are not known, this data carries significant information about the group parameters.

d) Reinforcement learning: Reinforcement learning, with regards to computerized reasoning, is a kind of powerful programming that trains calculations utilizing an arrangement of reward and discipline. A reinforcement learning calculation, or specialist, learns by associating with its condition. The operator gets remunerates by performing effectively and punishments for performing mistakenly. The operator takes in without mediation from a human by augmenting its reward and limiting its punishment.

In Figure 1.5 different machine learning applications are showcased depending upon the type of learning.

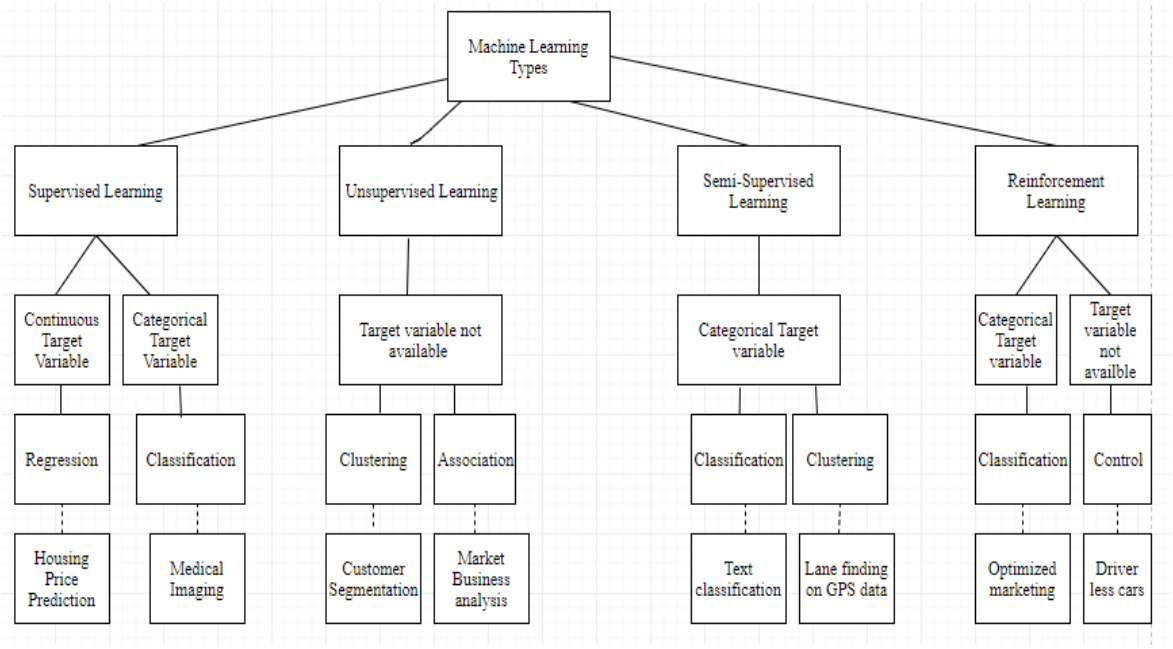

Figure 1.5 Applications of different types of Machine Learning

1.3 Ensemble Methods

Ensemble method is the blend of different classification algorithms to improve prediction accuracy and the ability of generalization. Most prominent ensemble-based strategies are packing and boosting. In boosting, order algorithm uses past one and besides focuses on its mistakes, while packing trains every characterization algorithm by a subset of the preparation set. The different ensemble methods are discussed in Table 1.1.

Table 1.1 Ensemble techniques

S.No	Techniques	Architecture	Trainable	Adaptive	Information level
1.	Voting	Parallel	No	No	Abstract
2.	Bagging	Parallel	Yes	No	Confidence
3.	Boosting	Parallel hierarchical	Yes	No	Abstract
4.	Random subspace	Parallel	Yes	No	Confidence
5.	Borda Count	Parallel	Yes	No	Rank
6.	Stacking	Gated Parallel	Yes	Yes	Confidence

The blend of Boosting, Bagging, and stacking is called “Meta-Algorithm approach” is utilized to diminish the variance or enhance the predictive force of the ensembles. Some of the above techniques are explained below.

1.3.1 Bagging

Bagging (Bootstrap aggregation) [6] is a voting system where construct models are discovered in light of different adjustments of learning informational records that are made by bootstrapping (bootstrap sampling) [7]. Utilizing an unstable learning algorithm is utilized to use bagging (e. g., neural networks or decision trees), result into great extent extraordinary classifiers when little changes in the learning set [8]. There is a proposal to recognize a novel intrusion in view of on the wearable method of Machine Learning. As a base 5 class, the bagging method of the dress with REPTree is used to implement intrusion detection system [9].

1.3.2 Boosting

Boosting [10] has an entire group of equal family members, for example winning, using voting in for coalitions to join the forecasts of a base model learned by means of a solitary learning algorithm. The complexity in among two methodologies with the goal of the built-in base models are dropped on the occasion of completing, while we attempt to model the supplementary model by learning further models, keeping in mind the mistakes of the previous model. With the same learning examples, learning the respectable starting point display on the whole showing set begins the procedure. For the following base model, we have to get an exact gauge of the illustrations which have not been appropriately anticipated by past base models. Accordingly, we increment the heaviness of these cases (or shed pounds of exact prescient illustrations) and take in another base model. At the point when some stop criteria are fulfilled, we quit adapting new base models.

1.3.3 Stacking

Stacking [11] or stacked speculation is the system for consolidating odd base machine models, i.e., models were discovered by different learning strategies, for example, the nearest neighbour strategy, choice tree, unconstrained bayas and so forth. Base models are not combined with an unequivocal arrangement like voting, rather an extra model known as meta or stage model is discovered and used for the blend of the base or stage-0 models. There are two stages at the same time. First of all, we create meta-learning informational collections using forecasts of the base model. Second, by using a meta-learning set, and learns meta-models that can incorporate desires of the first model to the last conjecture.

1.4 Performance Measurement Parameters

Sensitivity, accuracy and specificity are the usual measures of performance metrics in medical recognition. Accuracy discovers capacity of the classifier to compose correct disease recognition. Sensitivity calculates the capacity of the model to recognize the existence of target class correctly. Specificity process the capacity of the model to distinguish the target class.

The Accuracy, Sensitivity, and Specificity are measured as follows.

1.4.1 Accuracy

Accuracy in classification problems is the number of correct predictions prepared by the model**.**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.1)$$

1.4.2 Sensitivity

Sensitivity is a measure which discloses to us what extent of patients that really had diabetes was analyzed by the algorithm. The actual positives (People having diabetes are TP and FN) and the people analyzed by the model having diabetes are TP.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1.2)$$

1.4.3 Specificity

Specificity is a measure which reveals to us what extent of patients that did not have diabetes, was anticipated by the model as non- diabetic. The actual negatives (People actually not having diabetes are FP and TN) and the people analyzed by us not having diabetes are TN.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (1.3)$$

Here, the True Positives (TP) and True Negatives (TN) are correct classifying. When the output is inaccurately predicted as yes (or positive) at that time it is really no (negative) then occurring is of False Positive. When the output is incorrectly predicted as no when it is really yes then occurring is of False Negative (FN).

1.5 Classification Algorithms

1.5.1 Decision tree

A decision tree helps for classifying categorical data based on their attributes and it is also supported tool that uses a tree-like graph or model of decisions. It is also efficient for processing a large amount of data.

1.5.2. Support Vector Machine

Support vector machine is binary classification algorithm. It is base on the abstraction of decision planes which defines decision boundaries. A set of objects having dissimilar class memberships is split with the help of division plane.

1.5.3 Naive Bayes Classification

Naive Bayes is based on Bayes' theorem and it is suited at that time when input dimensionality is more than average. Sophisticated classification methods are outperformed by it.

1.5.4 Random Forest

Random forest is an adaptable, simple to utilize machine learning algorithm which produce even without hyper-parameter tuning. It is additionally one of the mainly utilized algorithms as of its simplicity and gives estimates of important variables in the classification.

1.5.5 Neural Network

The way of a biological nervous system inspired the neural network. The main element is the novel structure of data processing framework. A group of highly interconnected processing elements worked together to solve particular problems.

1.6 Applications of machine learning in IoT

- Understanding Medical Data: With the healthcare-focused Internet of Things (IoT) gadgets such as the Fitbit on the rise, the number of ways in which to gather vast amounts of medical information from mysterious sources is expanding. Machine learning is helping to make sense of all that data. For example, there are apps and services accessible that assistance to assemble

information in order to aid research into certain conditions such as Parkinson's disease or Asperger's syndrome by gathering information from users over time using machine learning for facial recognition.

- Outbreak Prediction: The ability of machine learning and artificial intelligence to translate information and utilize it to predict future outcomes is also being used in other fields of healthcare that could demonstrate to profit all of humanity. In one example, neural networks were utilized alongside support vector machines to help predict outbreaks of malaria in the Indian state of Maharashtra. The project took into account rainfall, temperature, reported cases and other details in order to generate its predictions.

- Treatment Personalization: Machine learning technologies can be utilized to translate the high volumes of patient information gathered by IoT and healthcare devices and after that utilizing these interpretations to predict conditions or suggest treatments.

- Cost Savings in Industrial Applications: Predictive capabilities are significantly useful in mechanical surroundings. By drawing data from a variety of sensors in or on machines, machine learning calculations can "realize" what's commonplace for the machine and afterward identify when somewhat unusual start to occur.

1.7 Chapterization

The organization of rest of the thesis is as given below:

Chapter 2 describes the literature review of the related work done on IoT framework and machine learning algorithms.

Chapter 3 describes the research gaps encountered from the literature review and problem statement of this work. Also, the methodology to be followed in this research is discussed in details.

Chapter 4 discussed the proposed IoT and machine learning framework for diabetes. Also, the algorithms of machine learning are explained.

Chapter 5 describes implementation and results achieved by applying machine learning to diabetes dataset.

Chapter 6 presents the conclusion along with the unique contribution and limitation. In the end, the future scope of this research is described.

1.8 Summary

This chapter provides detailed information about IoT, IoT infrastructure, IoT architecture, machine learning, machine learning types, ensemble method and different machine learning algorithms.

Chapter 2

Literature Survey

In the following chapter, the analysis is perform on Diabetes, IoT, Machine Learning, Ensembling, and various diversity ensemble methods and techniques. Below are abstracts of those analyses performed by various research fellows.

2.1 Diabetes

Diabetes is a disease which is detected on a blood test when the blood sugar is higher than normal value i.e. between (72 to 99 mg/dL) when fasting and up to (140 mg/dL) 2 hours subsequent to eating. Naturally, the pancreas emancipates insulin to assist the body to stock and use the sugar fat from the food eaten. Periodically body doesn't make sufficient insulin or doesn't take insulin well. Glucose then remains in the blood and doesn't stretch out at cells. There are 3 categories of diabetes Type1, Type2 and Gestational diabetes. About 10% of all diabetes cases are type1 the body doesn't generate insulin in this compose and about 90% of every of cases of diabetes global are of Type 2 the body doesn't provide sufficient measure of insulin for the absolute purpose. Diabetes effect females during pregnancy are known as Gestational. Diabetes will be the seventh driving reason for death in 2030 as predicted by WHO.

2.1.1 Causes of diabetes

Though, all the factors causing diabetes have not been discovered yet. The discovered factors have been listed below:

i) Heredity

It is seen that above 46 percent of aggregate diabetes show a genetic problem of diabetes. A few analysts trust that diabetes creates not on account of defective chromosome of parents which has been inherited by the individual but the chromosome which imparts resistance to this disease has not been received that from his/her parents. One might say that despite the fact that genetic factors do assume a part in the evolution of diabetes, to what extent and how these characteristics demonstration is as yet a secret. One might say that genetic factors can end up viable just functional at that time when definite other exciting stimulating aspects like obesity, defective dietary manners, and deficient bodily exertion is at work.

ii) Obesity

Overweight individuals turn out to be simple casualties to diabetes. Research has demonstrated that overweight individuals about 60 to 85 percent are diabetic. The more noteworthy is the death rate because of the problem of diabetes including the factor obesity. Bodyweight which is 30% underneath the perfect is a relatively certain assurance opposed to diabetes.

iii) Incorrect dietary habits

Life can keep up, spare or devastate because of food. If the Proper food is intake it will serve the need of medicine while ill-advised food fills in a toxic substance and can be a source of disease. For the source of diabetes, unnecessary food is as great to be faulted as inappropriate (i.e. refined and process) food. The body needs to deliver additional digestive juices and insulin for absorbing intemperate food. Beneath the weight of such intemperate workload, the pancreas organ debilitates and eventually breaks down, prompting to diabetes. A philosopher has correctly supposed that 'not especially many individuals pass away because of starvation; the rest pass away because of overeating'. It would not be an embellishment to express that we burrow our grave with our teeth.

iv) Inadequate physical work

In light of the industrialization, men have floated away from physical hard work. Amid physical work, muscles go through a considerable measure of glucose display in the blood. Therefore, the effort stack on the pancreas is diminished. Additionally, bodily work likewise anticipates or decreases obesity, which is personally associated with diabetes.

v) Viral infection

A conceivable part of some viral infection as an aetiological factor for diabetes is likewise being considered by numerous a scientist. A few kids in the wake of affliction from mumps have been believed to contract diseases which are a viral infection. Beta cells of the pancreas are delivered by the viruses by annihilating the insulin. Additionally, to battle the virus the antibodies created by the body which also hit the beta cells and aggravate the disease.

vi) Effects of certain hormones

A few hormones delivered in the body have an activity inverse so as to of insulin, i.e., they increment the measure of glucose in the blood. These hormones incorporate glucagon, growth hormone, adrenaline, and thyroxin. On the off chance that the emission of these hormones is unreasonable, the viability of proficiency of insulin declines and blood glucose level ascents.

vii) Side effects of certain drugs

Long haul utilization of specific drugs like cortisone (used for asthma, respiratory diseases, and arthritis), contraceptive pills and thyroid grouping of drugs is able to likewise deliver diabetes by means of harming the pancreas.

2.1.2 Symptoms of diabetes

Diabetes influences different organs or systems of the body to offer ascent to these symptoms as would now and again delude even a doctor. Maturity-onset diabetes creeps into the body so quietly that the victim usually remains unknowing and symptomless.

The accompanying symptoms point towards a probability of diabetes:

i) Polyurea (intemperate and frequent urination)

The sugar getting away in the urine drags alongside itself, an expansive amount of water. A lot of urine is often passed by the diabetic.

ii) Polydipsia (dryness of mouth and immoderate thirst)

This side effect is the after effect of efforts by the body to adjust for the fluids wasted by immoderate urine.

iii) Polyphagia (intemperate hunger)

In diabetes, glucose can't enter the different body cells. Therefore the cells keep in demonstrate hatred for from being showered by the glucose-rich serum. They experience the ill effects of 'neediness amidst bounty'. To defeat this cellular starvation, the body offers to ascend to strange and over the top hunger.

iv) Loss of weight

At the point when the cells can't use glucose, the body breaks down to put away fats to give the cells the vital nourishment. Therefore, the individual gets thinner.

v) Weakness, fatigue and body ache

The body for sustaining the starving cells it crumbles put away the muscle protein. For undue weakness and fatigue, it is the reason.

vi)Mental fatigue and lack of concentration

The brain cells for their nourishment need to dependent predominantly on glucose. Anyway, they can't use the accessible glucose, because of which the individual encounters undue mental exhaustion, can't focus and winds up abstracted.

vii) Wound infection and delayed healing

For the formation of pus forming microorganism's glucose rich blood is a decent reproducing medium. In addition, diabetes additionally influences the small blood vessels and nerve prompting a reduction in the blood supply of the skin and confusion of skin sensations. This is the motivation behind why even a small twisted on a diabetic individual's body effectively gets contaminated and neglects to recuperate in time.

2.2 Machine Learning

Machine Learning is a characteristic outcome of the crossing point of computer science and statistics [4]. Machine Learning is a branch of science that courses of action with programming the systems to such an extent that they normally learn and upgrade with understanding. Here, learning infers perceiving and understanding the data and settling on insightful choices in light of the provided information [12, 13].

It is exceptionally hard to cover every one of the choices in view of every single conceivable information. To deal with this issue, algorithms are made. These algorithms fabricate data from specific data and past inclusion with the measures of likelihood hypothesis, rationale, measurements, fortification learning, look, combinatorial optimization, and control hypothesis [12, 13].

There are a few approaches to execute Machine Learning procedures, be that as it may, the most usually used ones are regulated and unsupervised learning.

2.2.1 Supervised Learning

Supervised learning manages to learn a capacity from accessible preparing data. A supervised learning algorithm separates the preparation information and makes construed work, which can be used for mapping new illustrations [12, 13]. Supervised learning as relapse (for ceaseless yields) and classification (for distinct yields) is an essential element of insights and Machine Learning, also for examination of informational indexes or as a sub goal of a more difficult issue [14].

Here some examples are shown below

i) Classifying e-mails as spam

ii) Labelling web pages based on their content

iii) Voice recognition.

2.2.2 Unsupervised Learning

Unsupervised learning comprehends unlabeled information without having any predefined dataset for its planning. Unsupervised learning is a to a great degree effective instrument for separating open data and search for examples and patterns. It is most usually utilized for bunching comparative contribution to consistent gatherings. Basic ways to deal with unsupervised learning incorporate [12].

i) k-means

ii) Self-organizing maps

iii) Hierarchical clustering

2.2.3 Semi-supervised learning

To defeat the drawback of supervised learning algorithms with the intention, they can't make utilization of unlabeled information, semi-supervised learning (SSL) has been anticipated to use equally marked and unlabeled information [15]. Common approaches to semi-supervised learning include-

i) Generative Models

ii) S3VMs

iii) Graph-Based Algorithms

iv)Multi-view Algorithms

v)Self Training

2.2.4 Diabetes prediction using machine learning

Diabetes prediction is done with the existing models by various research fellows as discussed below.

W. Xu et al. [16] initiated random forest prediction algorithm which focuses on inspecting some accessible measures like waist, hip, age and weight that impacts on diabetes. Random forest utilizes decision trees to prepare the examples and combine significance of every tree to acquire last outcomes.

K.K. Gandhi et al.[17] applied information digging innovation for anticipating diabetes and performed a pre-handling venture to manage dataset like feature determination strategy, standardization and evaluated the machine-learning method for example SVM.

M. Panwar et al. [18] proposed the methodology, in view of novel pre-processing systems and K-closest neighbor classifier.

K.Sowjanya et al. [19] utilize the portable/android application for responses to overcome the absence of care about disease diabetes. Four machine learning algorithms such as J48, naïve Bayes, support vector machine, and multilayer perceptron were used to arrange the assembled information.

Artificial Neural Networks (ANN) is used in diabetes prediction by the various research fellows as discussed.

M. Heydari et al.[20] for type 2 diabetes they discussed different data mining methods and calculations which were utilized and connected to an arrangement of displaying information. The execution of machine learning methods has been contracted to get the best result like artificial neural network (ANN), support vector machine (SVM), decision tree, nearest neighbor and Bayesian network. They achieved a precision rate of 97.44 % using Artificial Neural System.

M. Komi et al. [21] explored the early prediction of diabetes. The experiment result proves that ANN (Artificial Neural Network) provides the highest accuracy than other techniques.

A. Swain et al. [22] utilized hybrid adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) they investigate the forecast and characterization of Diabetes mellitus. In terms of accuracy this explorer uncovers that after comparing ANFIS approach with ANN, ANFIS is more satisfactory.

Perceptron algorithm is used in diabetes prediction. M. Jahangir et al. [23] displayed a utilization of Automatic Multilayer Perceptron which is joined with an outlier detection strategy enhanced class outlier detection by separation base calculation to make a prediction framework named as enhanced class outlier with automatic multilayer perceptron (ECO-AMLP). A progression of investigations is performed on openly accessible pima india diabetes dataset to contrast ECO-AMLP and other individual classifiers and also ensemble-based methods.

Fuzzy cognitive maps (FCM) have been employed to model knowledge-based systems and is been used by various research fellows as discussed.
N. Doualia et al. [24] depicted procedure on gestational diabetes forecast and utilizing case-based fuzzy cognitive maps decision support system.

N. Bhatia et al. [25] discovered the nearness or nonappearance of diabetes mellitus using specific fuzzy cognitive maps (FCM).The product instrument was tried on 50 cases, indicating comes about with an exactness of 96%.

Owing to the use of detection through iris image is a remarkable progress in medical imaging technology in the recent decade: P. Samant et al. [26] the capability of the determination utility of the iridology alongside machine learning modes. The proposed model was connected to a systemic disease with ocular effects. 200 subject information of 100 every diabetic and non-diabetic was assessed, 89.66% accuracy was achieved with using random forest classifier.

2.3 IoT

The appearing paradigm of the Internet of Things (IoT) is uniquely centred on gathering and preparing information all over the place and time. IoT is utilizing all kinds of objects for a wide range of applications exceeding their basic purpose. Due to advancement in the area of healthcare, now the patient can monitor chronic conditions devices are being in designing phase. Related to this idea, detection of diabetes without pricking of the blood sample is a new concept to create effective healthcare assistance and resolution for a diabetic patient.

The challenge that peoples are presenting is handling a huge amount of information produced by each entity on the planet. The massive amount of information is produced over the internet as a billion of peoples are associated with it. Appropriate use of information is necessary for extracting some knowledge. Healthcare applications can be benefited by using machine learning algorithms. The part of doctors in disease detection can be replaced if the execution of the classification algorithm results well in terms of testing time and overall accuracy. Other than that classification algorithm will be utilized for low-cost and large-scale screening.

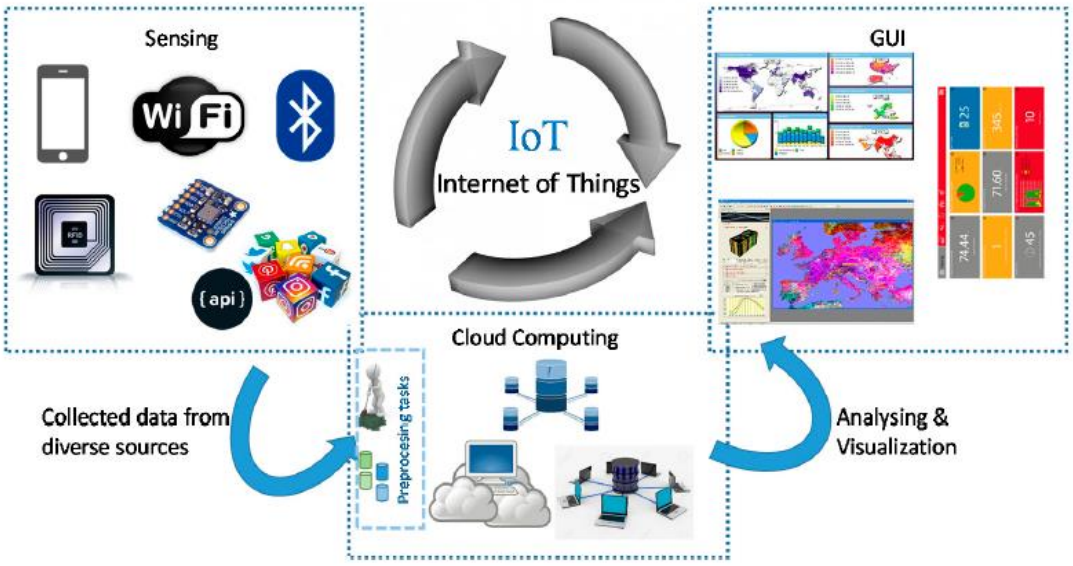

Figure 2.1 General schema of IoT [27]

2.3.1 Different Techniques applied to predict diabetes

IoT starts to become the main method of information procurement and transmission and would turn an important technology across several kinds of sensors to gather, investigate and deal with the whole data and it has been used by various research fellows as discussed.

S.H. Chang et al. [28] the intuitive to m-well being system (ImHS) gives continuous, two-route communication between diabetes patients and guardians by using Internet of Things innovation.

J. Gomez et al. [29] the system created quiet monitoring in illumination of Internet of things which can be utilize to assist persons having constant infections.

Smart devices associated with a system or different gadgets and they have the capability to interact. Over a recent couple of years, there has been gained remarkable advancement in the innovation of sensors. They are more inexpensive, simple to install and affordable and it has been used by various research fellows as discussed.

P. K. Gupta et al. [30] the engineering proposed depends on the installed sensors of the hardware as opposed to utilizing wearable sensors and smart phone sensors to utilize the estimation for the essential well being interconnected variables. The design presented here was assessed for appropriation, forecast investigation of physical activities, proficiency, and safety.

D.A. Antonovici et al. [31] developed an application on the android platform which trains for recording data which estimates the diastolic blood pressure (DBP), systolic blood pressure (SBP) and heart Rate by electronic sphygmomanometer which conveys utilizing Bluetooth TM procedure. The application provides the probability of sending medical information by means of mobile internet or remote.

M.A. Al-Taee et al.[32] The mobile well being (mHealth) approach takes into consideration multiple care dimensions of diabetes by means of remote gathering and monitoring of patient information and arrangement of information on a smart phone

platform. Such help to self-management of diabetes empowers ongoing clinical communication and input customized to the individual needs of the patient.

Nataraja et al. [33] the normal outcome is raspberry Pi gathers and stores the medical information through the sensors joined. The gathered information is exchanged to the client through applications.

M. Rathore et al. [34] proposed a real-time response system of the medical emergency which includes light of the human body by IoT-construct medical sensors sent in it. Additionally, the system proposed comprises of the information investigation building called BIntelligent building which is depict by the layered proposed engineering.

A. J. Jara et al. [35] proposed arrangement depends on the internet of things all together to on the one hand bolster a patient's report organization engineering in view of individual Radio-frequency identification (RFID) cards.

Different frameworks aim to improve care for people with diabetes as discussed below by various research fellows
R. Ab. et al. [36] this examination introduces a technique for checking ketone level by utilizing breath estimation. The method comprises to encourage the procedure of patient's diagnosis by advancement of equipment association with the IoT framework. Here, to sense the breath Arduino board was utilized.

G.M Bhat et al. [37] the embraced methodology uses IR spectroscopy to choose the blood glucose level in a person. A section based quickly adaptable care framework with a Wi-Fi module, gives prepared administration in unending thought condition. The areas in their framework consolidate patients, specialists, relatives and medicinal administrations providers. The framework was created for the home use of patients that ought to be ceaselessly or intermittently watched.

A. Winterlich et al. [38] demonstrated how the proving ground will give the framework to mobile well-being self management devices including quiet instruction, wearable sensors and supporting applications and administrations.

D.S Rajput et al. [39] proposed an appropriate self- management model using IoT framework for perpetual sickness, for example, hypertension, diabetes. 5 segments are included in the proposed stage. The principal component is a gadget which quantifies and sends the therapeutic information known as medical sensor and the second component is a software sensor having a wise conclusion calculation and it combines data from various physical medical sensors and server known as virtual medical sensor. They also created a portable application which is about searching medical information about patient or customer from medical IoT device and patient can use it for self-administration.

M. Bhatia et al. [40] proposed model made up of four distinct layers that is data abstraction layer, data accumulating layer, predictive layer and data categorization layer. The target of giving powerful prescient healthcare services was done by each layer which has assigned some predefined tasks. Initially, a mixture of IoT gadgets such as actuators, bio-sensors and smart sensors were utilized to obtain real-time health-oriented and non-health oriented qualities. The information was then transmitted to cloud storage for classification and feature extraction.

A robot is intelligent in the sense that it has embedded observing, detecting capabilities and at the same time can get sensor information from other sources which are fused for the 'acting' reason of the device. S. Mall et al. [41] introduced the eHealth mind stage by utilizing Robots which was associated through IoT for giving customized various care approaches especially to a diabetic patient. The robot contains sensors for medical and dietary observing of the diabetic person which subsequently gives them complete multidimensional-care.

Table2.1.Significant studies in the prediction of diabetes using Machine Learning / IoT

Author	**Year**	**Technique**	**Accuracy**	**Datasets**
K.K. Gandhi et al.[17]	2014	Support vector Machine	98%	PIMA India
M. Panwar et al. [18]	2016	K-nearest neighbor		
K .Sowjanya et al. [19]	2015	Mobile application and J48, naive bayes, Support	J48 algorithm	

		vector machine and Multilayer perceptron	gives preferred outcomes over different algorithms	
M. Heydari et al. [20]	2013	Artificial neural network	97.44%	Tabriz, Iran
A. Swain et al. [22]	2016	Hybrid adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN)	Best accuracy with Adaptive neuro-fuzzy inference	
N. Doualia et al. [24]	2015	Fuzzy cognitive maps decision support system	90.2%	
N. Bhatia et al. [25]	2015	Fuzzy cognitive maps	96%	
P. Samant et al. [26]	2017	Random forest	89.66%	Iris image
S.H Chang et al. [28]	2016	IoT by using GPRS BGM and a smartphone		
J. Gomez et al. [29]	2016	IoT		
P. K. Gupta et al. [30]	2016	Internet-of-Things on Cloud-centric architecture		
D.A. Antonovici et al. [31]	2014	Electronic sphygmomanometer, Bluetooth TM technique		
M.A. Al-Taee et al. [32]	2015	Mobile health (mHealth) approach		
K. Natarajan et al. [33]	2016	IoT		
M.M. Rathore et al. [34]	2016	IoT medical sensors		
A.J. Jara et al. [35]	2011	RFID cards		
G.M. Bhat et al. [36]	2017	Spectrophotometry		
A.Winterlich et al. [37]	2016	IoT		
S.Mall et al. [41]		By Internet of things assisting Robot		
R.S.H. Istepanian et al. [42]	2011	Non-Invasive glucose level sensing by m-IoT		
Ephizbah et al. [43]	2011	Fuzzy logic	78%	UCI

2.4 Summary

The part of literature review has clearly brought about the fundamental overview of the study on clinical decision support system, IoT, missing value in datasets and also the security of information and other authentications. Likewise, study about many kinds of algorithms from machine learning (ML). There are numerous difficulties related to the current trend of development procedure and can discover a new procedure for further research. By analysis the existing techniques and research, our research aims to propose an intelligent system for improving the performance of clinical decision support system using IoT by removing missing and abnormal values from datasets.

Chapter 3

Research Gaps and Problem Statement

3.1 Research gap

This section elaborates the gaps encountered during the research by reviewing the already existing literature in the area of Diabetes by using IoT and machine learning.

i) The varieties of ensembles are a pivotal issue within machine learning, which have proved to be an unfolding to research space in past few years. The variety and variability is a challenge in classification algorithms when particular dataset is established on regression [8].

ii) Security and privacy concerns being the major problems in huge amounts of data. Data is examined first and then excavate for particular order or arrangement called pattern [45, 46].

iii) Timeliness is an additional concern for large datasets in machine learning. The size of dataset increases with the increase in time taken during analysis [41, 46].

iv) Currently, in the technological development area such as IoT security is one of the most restrictive to deal with [47]. When the data started to be shared or uploaded on cloud [48] or mobile cloud computing (MCC) [49], it acquires more significance because of processing load factor.

v) The IoT application requires excessive infrastructure to maintain and organize the interconnected devices orderly [50].

vi) A difficult situation is created every day for the IoT practitioners because of fragmentation of standards including new ones [51].

vii) Most of the services are related to mobile users' mobility is examined as the main problem for the IoT adoption [51].

3.2 Problem Statement

There are millions of people over the world suffering from diabetes disease due to unavailability of early detection and continuous monitoring system. Finger pricking is also painful to test blood sugar. One needs to put a drop of blood by pricking fingertip and calculate the glucose level by device. This create mental trauma between most of

the diabetes patient particularly children. To overcome this problem, a non-invasive way of measurement is necessary. A cloud computing part is also required for automated monitoring of blood glucose level and storage of data. Diabetes diagnosis is based on various epidemiological and genetic factors, it is necessary to collect all the factors to get accurate prediction.

3.3 Objectives

i. To study existing cloud & IoT technologies implemented in health care for diabetes.

ii. To compare and analyze machine learning algorithms to be used in diabetes prediction.

iii. To propose A Cloud IoT based Framework for Diabetes Prediction.

3.4 Methodology

Block diagram for A Cloud IoT based Framework for Diabetes Prediction is given below in Figure 3.1 and working approach is further explained.

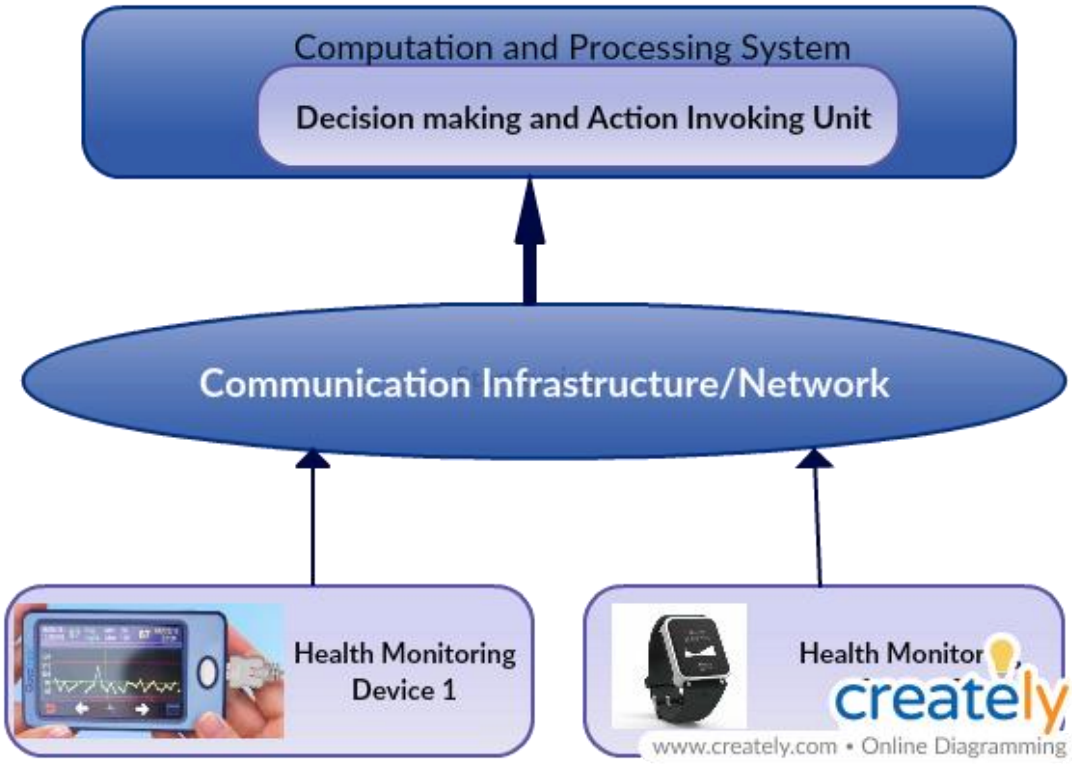

Figure 3.1 Block Diagram

The working approach to be followed is given below:

Step 1: Through the use of IoT device clients can send attributes value of diabetes disease, into cloud infrastructure layer with the help of communication protocols.

Step 2: As Cloud computing is necessary for analyzing all the (big) information and to aggregate it. On-demand network access to computing services is permitted. This will work at the backend and is utilized to analyze and interpret the information from the smart things and for the user, it also provides the visualizations.

Step 3: For prediction of data stored in cloud, machine learning algorithms are utilized. Whenever users will send information, it can be unstructured information. Then it can change those data sets into "structured data". In that way, the framework replaces the abnormal and missing values and stores them. At that point when there is no abnormal or missing value in the system, it will start analysis all data according to a decision tree, random forest, support vector machine, naive bayes or any other algorithm and predict the result for patients. The predicted result will be saved.

Step 4: Users and doctors can see the predicted result.

3.4 Summary

In this chapter, the research gaps in the study are discussed first. The problem statement and objectives of the work are defined. The chapter explains the methodology which is to be followed to perform the research work.

Chapter 4

Proposed Methodology

This chapter discusses proposed framework for A Cloud IoT based Framework for Diabetes Prediction. The algorithmic approach adopted for prediction is explained in detail. In this research, we have used "Pima Indians Diabetes Data Set" [52] as the data set.

Different layers are used to made up this architecture for easy detection of glucose levels , then storage of data and usage of machine learning for prediction of data.

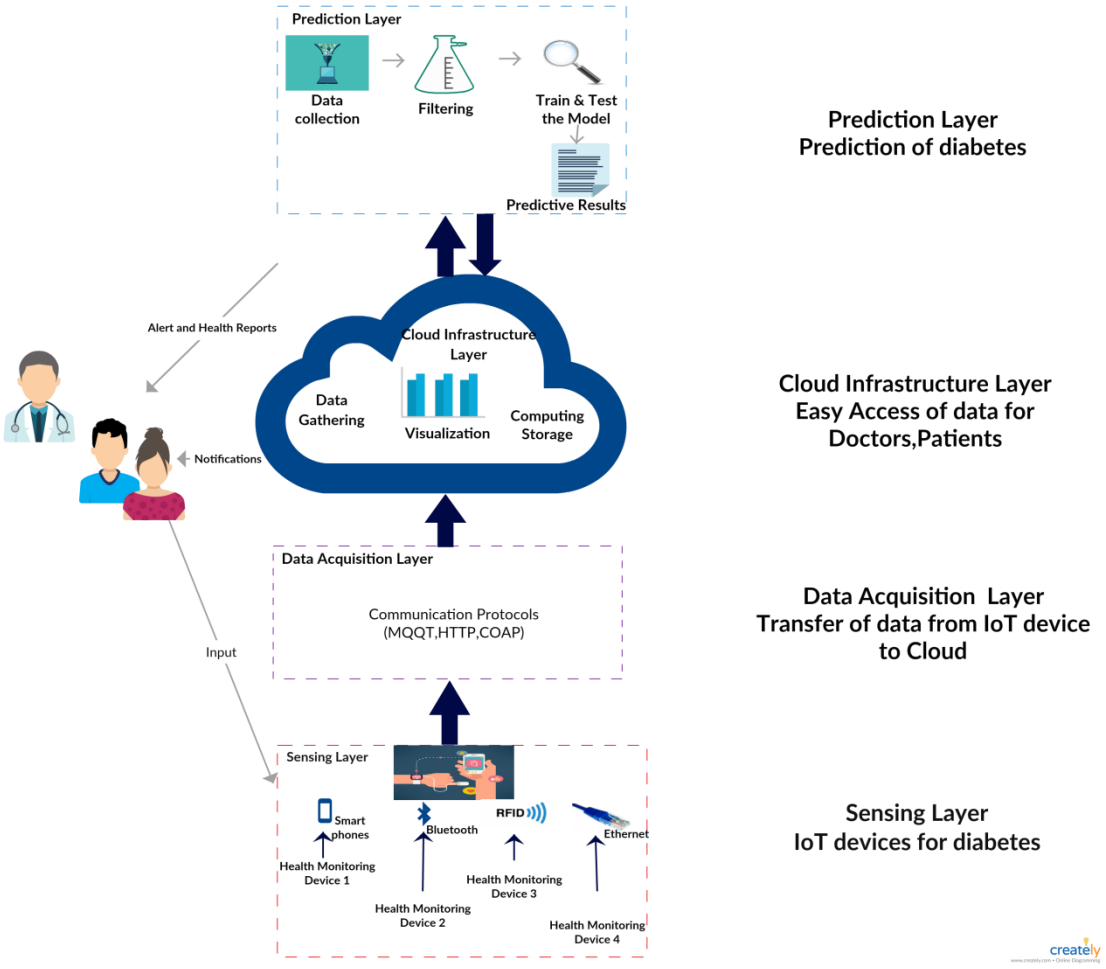

Figure 4.1 A Cloud IoT based Framework for Diabetes Prediction

4.1 Sensing Layer

This layer describes about Non-invasive devices. Non-invasive devices are those in which break in the skin is not involved. There is no contact with the mucous membrane or internal body cavity other than through a natural or artificial body orifice.

4.1.1 IoT Devices

Smart healthcare technology is one of the emerging technologies in the current healthcare system. By making use of smart wearable devices, the patient-generated data can be sent to electronic devices or any health records so that the doctors/caregivers can directly monitor the patient activity in real-time. Every time diabetic patient has to pierce the finger with the sharp needle so we considered upon devices which will not hurt the patient i.e. without piercing or injecting something into the body they can detect their glucose level. These devices can reduce the healthcare costs. There are several connected smart devices that have been developed in the healthcare industry.

i. Gluco track [53]: Gluco track clips to your ear cartilage keeping in mind the end goal to test your glucose level. The gadget contains 2 sections: Main Unit (MU) and ear cut. The ear cut doesn't hurt. Simply cut it on and tap! Your glucose level shows up on the MU. Gluco track utilizes three free advancements, all the while: ultrasonic, electromagnetic and warm. All estimations are joined by a one of a kind exclusive calculation, which computes the weighted normal and returns the client's glucose level.

ii. Dia-vit [54]: Dia-vit is a non-obtrusive glucose self-observing gadget. It gauges the glucose level in your blood. Their cell phone application keeps a diary of your day by day information, so you can track designs in your variance glucose level. By observing your condition, you are then ready to be more mindful of your condition.

iii. Sugar beat[55]: Sugar beat is a non-obtrusive fix. It contains an electronic sensor that distinguishes ongoing estimations. The fix is expendable and is around 1mm

thick. Your glucose level is estimated through the skin like clockwork. Sugar Beat is associated with an application where additionally every one of the readings is sent.

iv.Siren Care's [56]: Diabetic foot ulcers are preventable with Siren Care's savvy stun. As temperature rises cause skin aggravation, Siren savvy stun can screen foot temperature to get wounds right on time before they transform into an ulcer. Also, there is no compelling reason to charge the socks. Each sock is worked with a 6-months life expectancy and is additionally launderable.

v. K'Track [57]: K'Track just requires a straightforward press signal on the watch to shows the glucose level. K'Track urges individuals to wear everything day long to screen blood glucose amid game and exercise when glucose levels are inclined to spike. In addition, K'Track Glucose can likewise track steps taken, separated voyaged and calories consumed.

vi. Johnson and Johnson [56]: Johnson and Johnson likewise offer us an insulin fix named One Touch Via. At times individuals with diabetes can feel humiliated when they have to infuse insulin at mealtimes. As an infusion free insulin fix, clients simply need to squeeze two catches on the gadget to convey the insulin bolus. This activity can be possible even through apparel, along these lines; patients feel urged to measurement all the more regularly which prompts better treatment.

Figure 4.2 Different devices for diabetes detection

4.1.2 Sensors

Now with the help of sensors, we can transmit information. A sensor is an electronic hardware to identify and react to a number of resources as of the neighbourhood. The particular information could be light, heat, motion, moisture, pressure, or some other environmental phenomena. The product is normally a digital or analog signal to change over to comprehend capacity of human through different electronic displays at the sensor position or transferred electronically over a network for studying or additional preparation. Information managing applications can be considered as high-level sensors. They indeed merge different sensors measures in order to extract information.

4.1.3 Protocols

The protocols used are smart phone and ethernet for connectivity with Internet with the help of that we can interface the devices wherever it is found.

4.2 Data Acquisition Layer

This layer will revolve just about the Internet and utilizations of Internet services as the principal center to transfer information. When the glucose level is predicted it needs to be stored on the cloud for further analysis and study. So, we need some way of communication between IoT device and cloud with help of communication protocol we can establish that.

4.2.1 Communication Protocols

The data will be stored on the device and will be shared on the internet. There are different protocols like MQTT, HTTP and CoAP for sending information from the device to server.

a) MQTT [58]: MQTT (Message queue Telemetry Transport) is a machine-to-machine (M2M)/"Internet of Things" connection protocol. The general architecture of MQTT is shown in Figure 4.3. It is a lightweight messaging protocol which gives resource-constrained network clients with an easy process to issue telemetry information.

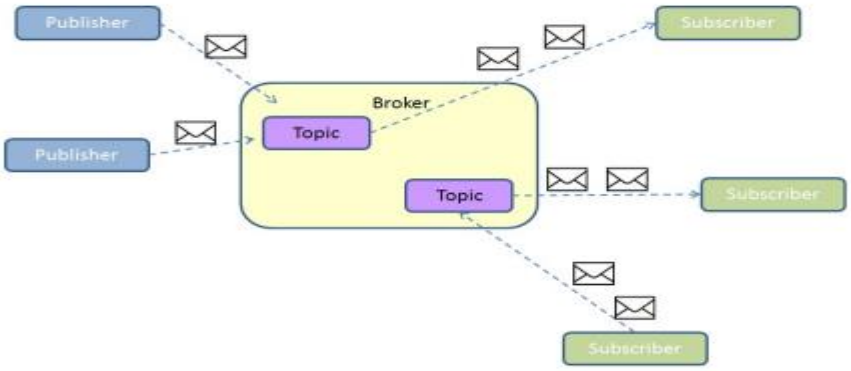

Figure 4.3 General Architecture of MQTT [58]

b) HTTP [59]: HTTP (Hypertext Transfer Protocol) for moving the files on the internet it provides the rules. The framework of Internet Protocol Suite, it is an application layer protocol. Its definition expects a fundamental and solid transport layer protocol accordingly Transmission Control Protocol (TCP) is frequently use with HTTP.

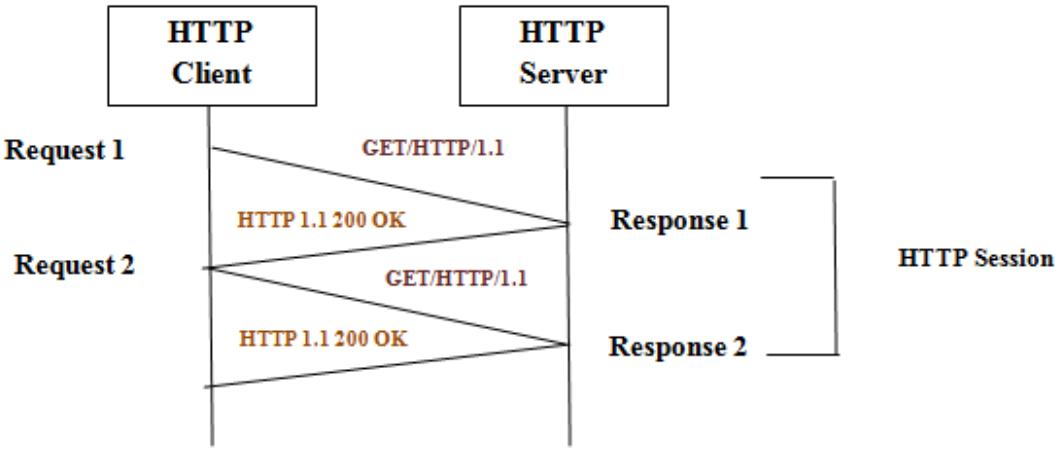

Figure 4.4 HTTP Session [59]

The HTTP client begins a request by setting up a TCP connection with a particular port on the server as shown in Figure 4.4. A HTTP server tuning on that port wait for the client's request message. Once getting the request, the server responds by sending back a status line for example "HTTP/1.1 200 OK" and a message body. In the present version, HTTP/1.1 reuses the TCP connection a couple of times to send and get different HTTP request/response instead of to create a new TCP connection for every single request/response pair.

c) CoAP [60]: The Constrained Application Protocol (CoAP) is a particular web exchange protocol for utilizing with constrained nodes and networks in the IoT. A client sends a CoAP request to request an action on a resource on a CoAP server. After accepting a request, the server sends back a response with a response code; this response may comprise of a resource representation. There are four sorts of CoAP messages: Confirmable (CON), Non-confirmable (NON), Acknowledgement (ACK) and Reset (RST). CoAP supports multicast IP destination addresses. CoAP offer two messaging models for different QoS requirements: Reliable message transmission and unreliable message transmission, as described in the following Figure 4.5.

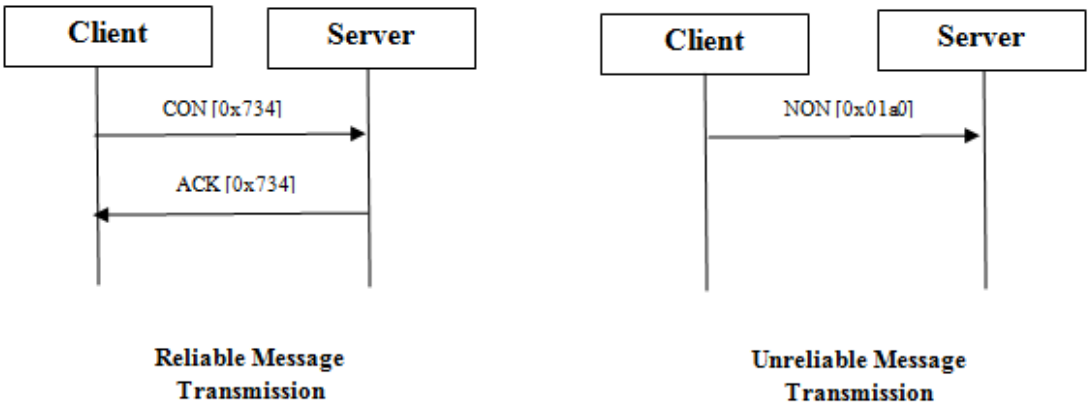

Figure 4.5 CoAP Messaging Models [60]

Each of these protocols helps taking information or updates from the individual device and to send it over to a central location.

4.3 Cloud Infrastructure Layer

In the framework, sensors connected to the network output information that is keep in cloud storage and then machine learning algorithms are performed to change raw information into comprehension and understanding.

A platform is something that can run applications and store information. In an organization's data centres such as, it may have computers running Windows Server and other software that give a platform to in-house applications. A cloud platform is a similar object: It's an establishment for running applications and storing information. The biggest dissimilarity is that it runs in data centres own by an external service

provider, such as Microsoft, Amazon, Google, IBM etc. and it is access via the Internet. For example, Microsoft's cloud platforms run Windows Azure relatively than Windows Server. Storage on cloud is most appropriate for such kinds of system.

4.4 Prediction Layer

This layer is used for prediction of diabetes which will be helpful for doctors as well as patients.

4.4.1 Workflow of machine learning algorithms

To make experiments more interesting the first part is to select new data sets that have different attributes, different instances, and different parameters. The second part has pre-processed data which includes data format adaptation and data sampling. In data format adaptation first formats of datasets must be converted to into.CSV files which are required by R interface. Data sampling depends on two parameters - percentage and bias. Figure 4.6 shows the workflow of the algorithm.

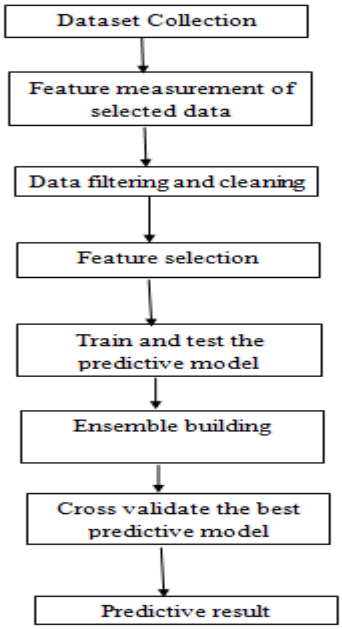

Figure 4.6 Workflow of machine learning algorithms

We performed ensembling on various Machine Learning methods for example Random Forest, Neural Network, Decision tree, Support Vector Machine, Naive Bayes with Stepwise Selection, on each dataset under following measures: Accuracy, Specificity, Sensitivity and we compared obtained results to determine best Machine

Learning algorithm for each dataset. The research starts by selection of the variation in datasets. Classes are not represented uniformly in an imbalanced dataset.

4.4.2 Data Selection

Data selection is considered as the way toward deciding the appropriate information sort, source and in addition reasonable instruments to collect information. Data selection goes before the real routine with regards to data gathering. The essential goal of data selection is the assurance of fitting data sort, source, and instrument(s) that enable specialists to sufficiently respond into inquiries. This assurance is regularly discipline-specific and is primarily determined by the means of the examination, existing writing, and accessibility to necessary data sources. For this construction we have taken the dataset as input from the UCI Machine learning.

4.4.3 Data cleaning and filtering

A data set is a gathering of information that portrays attribute values (variables) of various real-world objects (units). A data set is an accumulation of data that portrays characteristic esteems (factors) of different genuine items (units). To begin with it can be specifically supposed as having a place with a particular variable, and second is put away in a data class that speak to the esteem space of this present reality variable. As such, for every unit, a content variable ought to be put away as content, a numeric variable as a number and this in a configuration that is predictable over the data set. Data filtering is done in three steps Screening, Diagnosis and Editing as shown in Figure 4.7. Respectability issues can emerge when the choices to choose "fitting" data to gather are constructing principally in light of cost and comfort contemplations as opposed to the capacity of data to enough answer examine questions.

4.4.4 Feature Selection

Feature selection is essential measure of separating insignificant and pointless characteristics for the model construction [61]. Feature selection incorporates and rejects attribute there in the data with no updates [46, 49]. Feature selection act as a filter, as it quiets out those characteristics that aren’t helpful in accumulation to the current features [48]. Feature selection enhances performance of classification algorithms. Feature selection methods are Filter method, Wrapper method, and embedded method [41, 47]. These methods choose those features which are

superlative for execution of model. Filter method measure quality of important selected features, from machine learning algorithms, while wrapper methods needs application of classification algorithm to measure quality of selected features. For learning of optimal parameters, embedded methods perform feature selection.

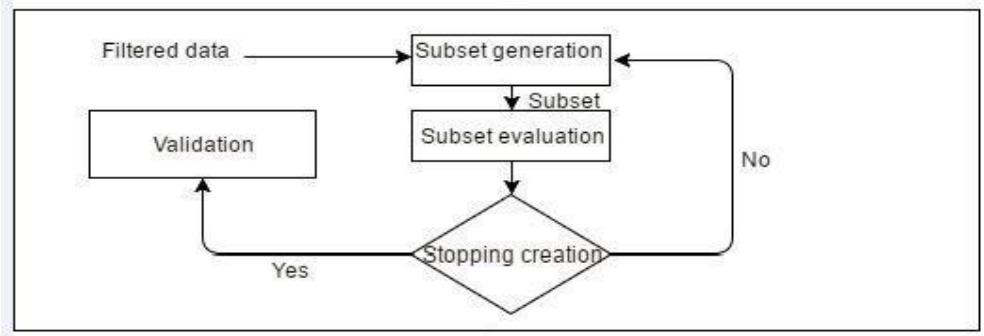

Figure 4.7 Feature Selection

4.4.5 Training and testing of the dataset

A preparation set is a preparation of data utilized to find possibly predictive connections. A test set is a preparation of data utilized to survey the quality and value of a prescient relationship. Testing and preparing sets are used as components of frameworks, machine learning, genetic programming and statistics. During the implementation in first step we have to do preparation of the data then explore the data to use in predictive models for final evaluation, and deliver the data for ensemble building. The following four steps are shown in Figure 4.8.

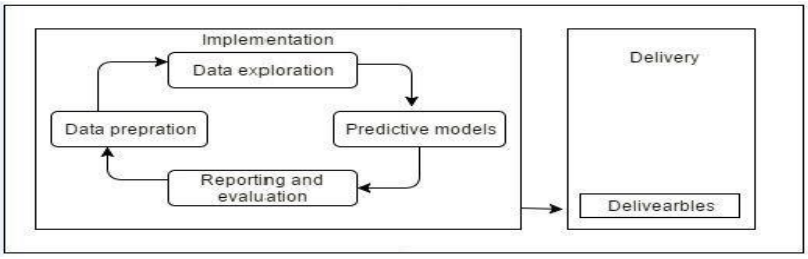

Figure 4.8 Training and Testing

4.4.6 Ensemble building

Ensemble strategies are systems that make different models and then join them to deliver enhanced outcomes. An ensemble strategy, as a rule, delivers more precise arrangements than a solitary model would. This has been the situation in various machine learning rivalries, where the triumphant arrangements utilized ensemble

techniques. Set of a predictive model is created with the help ensemble method, which is a method of learning and combines their output in the same forecast. The objective of combining multiple models together, to reach a better future performance, and in many cases, it is shown that ensembles can be more accurate than a model [9]. Ensembling is defined with the help of different machine learning algorithm as shown in Figure 4.9.

4.4.7 Cross-validation

In the simple words, testing sets are built with taking a few unique dataset into more than one section. Yet, the assessments acquired for this situation have a tendency to mirror the specific method the information is isolated up. The arrangement is to utilize factual examining to acquire more precise estimations. This is known as cross-validation. The purpose of cross-validation is to certify as to every example from the original dataset has the equal probability of appear in the training and testing set. The final output from the cross-validation is used to check robustness.

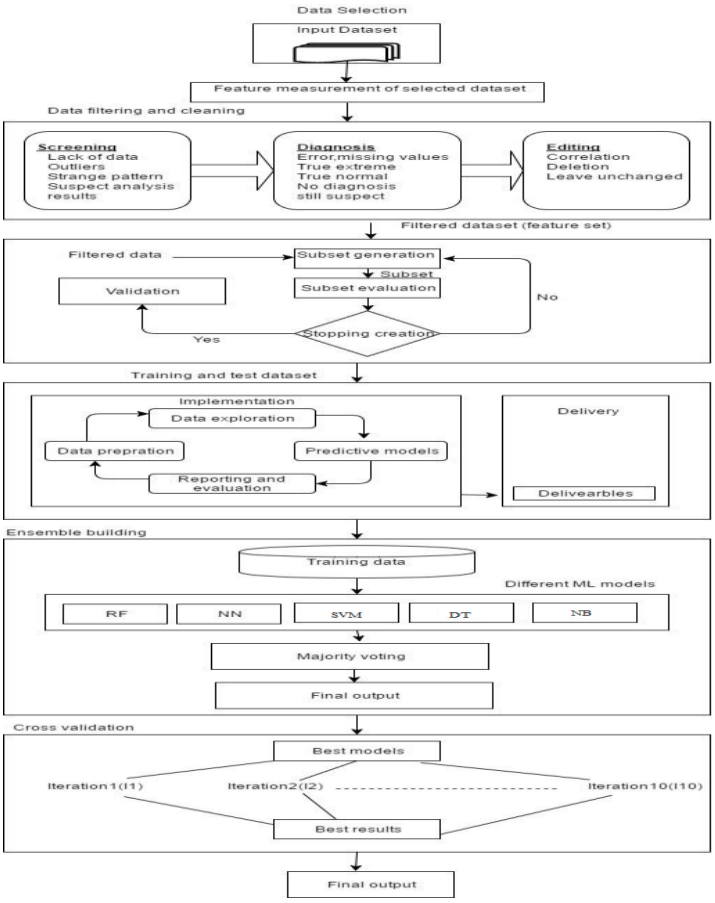

Figure 4.9 Learning approach for A Cloud IoT based Framework for Diabetes Prediction

The concept of feature selection is the concept of selecting a subset of suitable features like predictors and variables. K-cross validation is a process to calculate the accuracy of a system. For example, take the dataset, D, which is randomly divided into K equally exclusive subsets called folds of the same size (D1, D2, …, Dk) and K classifiers are built. The ith classifier is skilled on the addition of all value of j on D and checked on Di. The accuracy of the calculation is the on the whole number of the correct classification, which is separated by the number of events occurring in the dataset. We have applied the five Machine Learning methods for predicting, testing and training, namely Decision Tree, Support Vector Machine, Random Forest, Neural Network and Naive bayes.

i) Decision tree: A decision tree helps for classifying categorical data based on their attributes and it is also supported tool that utilize a tree-like graph or model of decisions. It is also efficient for processing a large amount of data. It is utilized as a process for classification and prediction with an illustration utilizing nodes and internodes. The root and internal nodes are the test cases that are utilized to divide the instances with distinct characteristics. Internal nodes themselves are the outcome of attribute test cases. Leaf nodes indicate the class variable. Figure 4.10 show a sample decision tree structure.

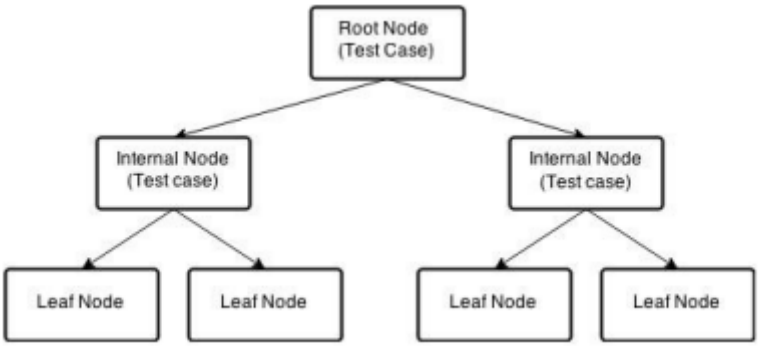

Figure 4.10 Decision Tree Structure

ii) Support Vector Machines: Support vector machine is an arrangement of interrelated to supervised learning method utilized in medical diagnosis for classification and regression [62, 63]. A set of objects having different class memberships is unconnected with the help of decision plane. It is base on the idea of finding a hyperplane that best divide a dataset into two classes, as shown in the

Figure 4.11. SVM concurrently reduce the empirical classification error and maximize the geometric margin. Therefore SVM is called Maximum Margin Classifiers.

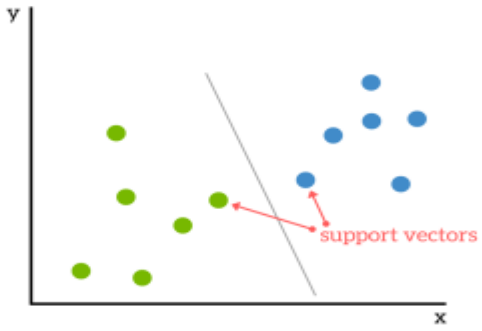

Figure 4.11 Support Vector Machine

iii) Random Forest

Random Forest is one of the largely popular machine learning algorithms which was created by Breiman [47]. Random Forest is nothing but a group of many simple decision trees and all these trees are able to predict the outcome of any input. These trees are able to predict in which class a particular input belongs to if our problem is of classification and if the problem is of regression these trees are able to predict a continuous number. In case of classification, each tree in random forest votes for a particular class and the class which have most votes is given as output for that particular input. On the other hand in regression output of every tree is averaged to obtain the output for that particular input. In below Figure 4.12 how a random forest would look like with two trees is displayed. Random Forest can be seen as an ensemble of many simple decision trees. Ensembling of many decision trees in the random forest has shown dramatic improvement in performance of the model. Random Forest is also able to overcome the issue of over fitting which is one of the biggest problems in the single decision tree.

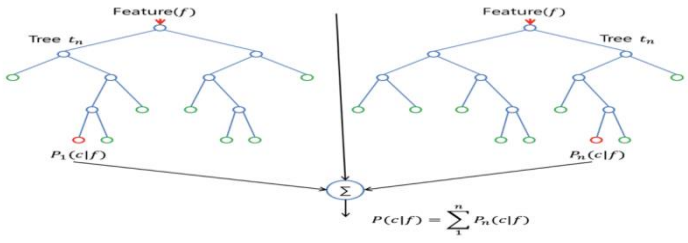

Figure 4.12 Random forest

iv) Neural Network: The process of a biological nervous system inspired the neural network. The main element is the novel structure of data processing framework. A group of highly interconnected processing elements worked together to solve particular problems.

v) Naive Bayes Classification: Naive Bayes is base on Bayes' theorem and it is suited at that time when input dimensionality is more than average. sophisticated classification methods are outperformed by it. The algorithm works on the simple Naïve Bayes formula.

$$\text{Posterior Probability} = \text{Likelihood} \times \text{Class Prior Probability}$$

$$P(c|x) \qquad \frac{P(x|c) \qquad P(c)}{\text{Predictor Prior Probability}} \qquad (4.1)$$

$$P(x)$$

These methods propose enhanced execution outcomes for the measurement parameters for example Accuracy are calculated as follows:
In medical diagnosis accuracy, sensitivity and specificity are the basic measures of performance metrics. Accuracy determines the capacity of the classifier to generate precise disease diagnosis. Sensitivity measures the capacity of the model to distinguish the occurrence of target class correctly. Specificity measures the capacity of the model to divide the target class.

The Accuracy, Sensitivity, and Specificity are measured as follows [64].

- TN is the number of correct predictions that an instance is negative.
- FP is the number of incorrect predictions that an instance is positive.
- FN is the number of incorrect of predictions that an instance negative.
- TP is the number of correct predictions that an instance is positive.

$$\text{i) Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4.2)$$

ii) Sensitivity = $\frac{TP}{TP+ FN}$ (4.3)

iii) Specificity = $\frac{TN}{TN + FP}$ (4.4)

Confusion matrixes mainly describe the performance of a classifier model and it contains data about actual and predicted classifications completed by a classification system [55]. For evaluating the performance of such systems we have to use the information in the matrix. The following table 4.1 show the confusion matrix for a two-class classifier.

Table 4.1 Class classifier

		Predicted	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	TP

The fundamental technique is to train dissimilar classifiers like Random Forest, Decision tree on different subsets of the features and then ensemble all the model by the combination. Hereafter, the output is used for training and testing by dissimilar classification algorithms and results are shown in tables and graphs. The yield of all the five algorithms is compare and analyze.

4.5 Summary

In this chapter proposed architecture A Cloud IoT based Framework for Diabetes Prediction is explain in detail with the help of algorithms. By using IoT devices and diabetes prediction method a new model is proposed.

Chapter-5

Implementation and Experiment Result

This chapter discusses the experimental test bed for the proposed framework. Implementation results are also presented and discussed.

5.1 Experimental setup

5.1.1 Software and Hardware requirements (minimum)

Table 5.1 H/W and S/W Requirement (minimum)

1.	Processor	32 bit
2.	RAM	2 GB
3.	Hard Disk	80 GB
4.	Operating System	Windows 7
5.	Programming Language	R (Rattle)
6.	Platform	R Studio

5.1.2 Random Forest Implementation

We implemented machine learning algorithm with the help of R language. For Random Forest implementation we install random forest package and libraries like hmeasure. After that, we performed read and write operation on the datasets and determines the performance of the random forest in terms of accuracy. The random forest can be implemented using the following function which includes formula, trainDataset, and method in R.

Code snippet in R for Random Forest

```
install.packages("randomforest")
library(randomForest)
library(hmeasure)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula
model   <- randomForest(formula, trainDataset, ntree=500,mtry=2)
model
```

5.1.3 Neural network Implementation

For implementing neural network for diabetes prediction we added R packages and libraries like nnet and hmeasure. After that, we performed read and write operation on the datasets and determine the performance of the neural network in terms of accuracy.

Code snippet in R for Neural Network

```
library(nnet)
library(hmeasure)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula

model    <- nnet(formula, trainDataset, size=10, linout=TRUE, skip=TRUE,
                 MaxNWts=10000, trace=FALSE, maxit=100)
model
```

5.1.4 Support vector machine Implementation

For implementing support vector machine for diabetes prediction we add various R packages and libraries like kernlab and hmeasure. After that, we performed read and write operation on the datasets and determine the performance of the Support vector machine in terms of accuracy.

Code snippet in R for Support Vector Machine

```
library(kernlab)
library(hmeasure)

formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula

model  <- ksvm(formula, trainDataset, kernel="rbfdot", prob.model=TRUE)
model
```

5.1.5 Decision tree Implementation

For implementing decision tree for diabetes prediction we add various R packages and libraries like rpart. After that, we performed read and write operation on the datasets and determine the performance of the decision tree in terms of accuracy.

Code snippet in R for Decision Tree

```
library(rpart)

formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula
model    <- rpart(formula, trainDataset, method="class", parms=list
(split="information"), control=rpart.control(usesurrogate=0, maxsurrogate=0))
model
```

5.1.6 Naive bayes Implementation

For implementing naive bayes for diabetes prediction we add various R packages and libraries like e1071, rminer and hmeasure. After that, we performed read and write operation on the datasets and determine the performance of the naive bayes in terms of accuracy.

Code snippet in R for Naive bayes

```
library(e1071)
library(rminer)
library(hmeasure)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula

model<-naiveBayes(Outcome~.,data=trainDataset)
model
```

5.2 Results

To predict a person is diabetic or not and also to predict the diversity of the different machine learning models with ensembling by using accuracy parameter. We choose a data set and applied 5 different Machine Learning model, the dataset belongs to classification data so we applied some Machine Learning model based on classification. The summary of dataset is shown below in Table 5.2.

Table 5.2 Summary of Dataset

Attribute	
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test

Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2)
Diabetes Pedigree Function	Diabetes pedigree function
Age	Age (years)
Outcome	Class variable (0 or 1)

Table 5.3 Dataset

No. Of Samples	No. Of features
768	9

5.2.1 Methodology

Step 1: PIMA India Diabetes dataset is taken.

Step 2: The elimination of duplicates and absent value entries (data cleansing and filtering) from the dataset is conceded out.

Step 3: Decision on different parameters for diabetes is calculated in this phase.

Step 4: The Feature selection is done which is going to make the forecast of model efficient, fast and truthful.

Step 5: Five Machine Learning algorithms chosen: Random Forest, Neural Network, Naive bayes, Decision tree, Support Vector machine and ensemble of the algorithms are performed.

Step 6: Selected algorithms are evaluated for parameters such as accuracy, sensitivity, specificity is used to evaluate the efficiency of the method.

Step 7: K-fold cross-validation is used to compute the sturdiness of the best predictive techniques.

Step 8: Finally, we get the results in terms of accuracy.

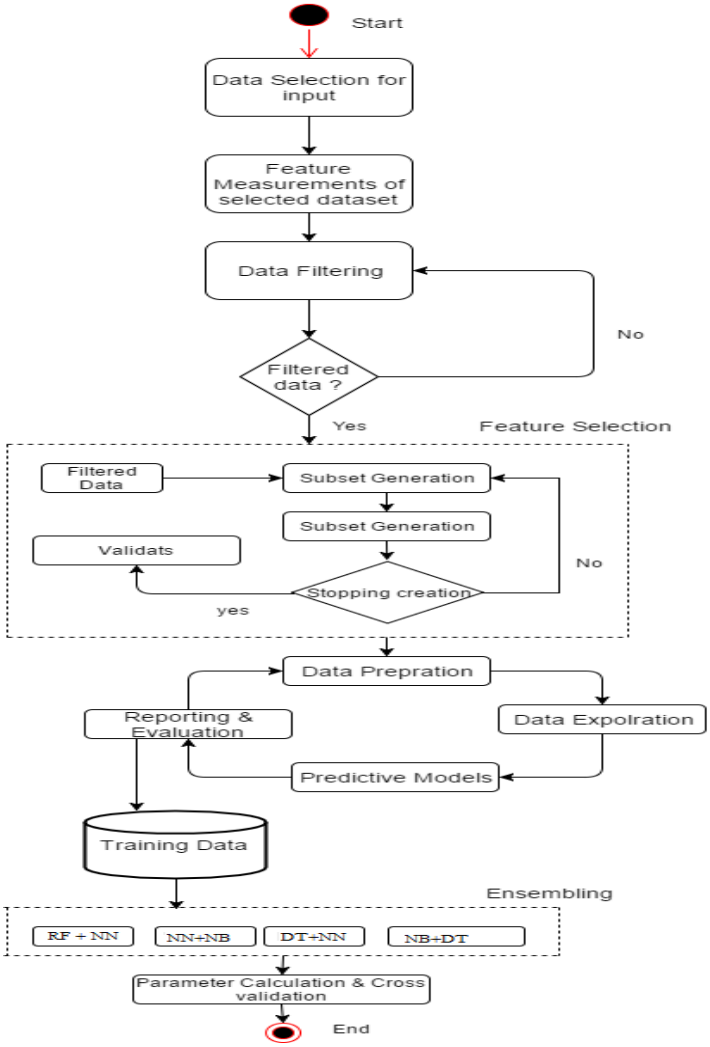

Figure 5.1 Activity Diagram for machine learning algorithms

The table contain outcome of 5 different Machine Learning models all along with the value of accuracy, sensitivity, specificity. The suggested framework is applied to PIMA India diabetes dataset. As already explained we have applied the classification algorithms on the dataset. These algorithms are as follow Random Forest, Neural Network, Decision tree, Naive Bayes. We have compared the all these algorithms on the basis of their accuracy. We have applied 10-fold cross-validation in order to check the results of every classifier for unknown instances. The experimental results are shown below in the Table 5.4.

Table 5.4 Single model results

Model No.	Model Name	Accuracy	Sensitivity	Specificity
1	Random Forest	82.25	71.4	88.4
2	Neural Network	80.95	66.7	87
3	Decision Tree	78.79	57.8	90.5
4	Naive Bayes	56.00	38.8	63.4
5	Support Vector Machine	80.52	52.8	92.1

The results show that Random forest gave the best result for a dataset with the highest accuracy and Naive Bayes gave the worst results, so we performed ensemble to get better performance.

Evaluation of different machine learning algorithms on parameters like specificity, sensitivity and accuracy.

The comparative performance of different models in terms of specificity is shown in Figure 5.2.

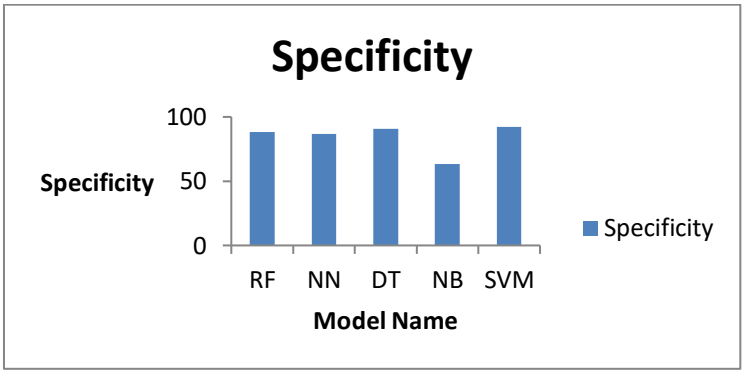

Figure 5.2 Specificity

The comparative performance of different models in terms of sensitivity is shown in Figure 5.3.

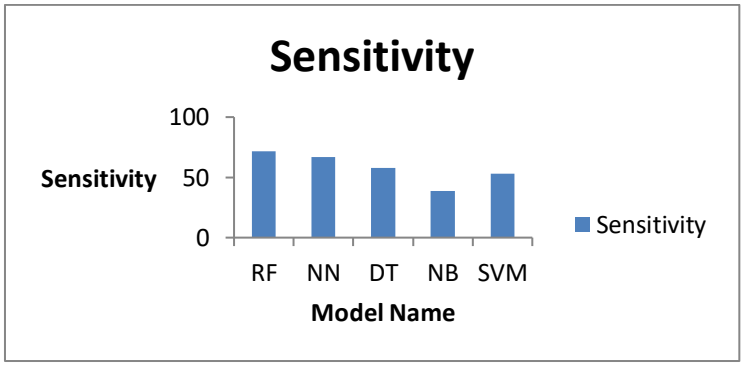

Figure 5.3 Sensitivity

The comparative performance of different models in terms of accuracy is shown in Figure 5.4.

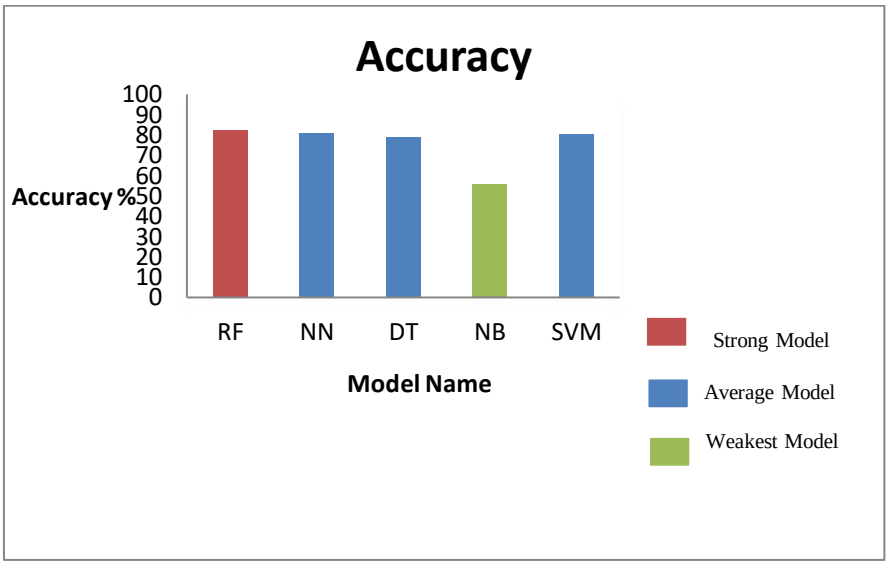

Figure 5.4 Accuracy

Now to improve the accuracy we have applied the ensembling algorithm, in which we combine two models and run for both the dataset. The improved experimental results are shown in Table 5.5

Ensembling Results

Table 5.5 Ensemble Results

Model No.	Model Name	Specificity	Sensitivity	Accuracy
1	Naive Bayes +Decision tree	90.2	93.1	90
2	Decision tree+ Neural network	94.5	79.5	83.12
3	Decision tree+ Random forest	76.6	71.4	83.7
4	Decision tree+ Support vector machine	78.5	94.4	82.25
5	Support vector machine+ Random Forest	35.8	80	84.1
6	Random Forest+ Neural Network	76.1	76.4	83.5
7	Random Forest+ Naive Bayes	76.5	80.8	77.49
8	Neural network+ Naive Bayes	87.5	94.5	89.18
9	Support vector machine+ Neural network	87.9	96.5	90.04

10	Naive bayes+ Support Vector machine	84	93.5	86.58

After performing ensembling on the above models we get the highest accuracy of the ensemble model no. 1(Naive Bayes+ Decision tree) and model no.2 (Decision Tree+ Neural Network).They individually gives the worst performance when we use separately but after ensembling combination of two worst models, accuracy gets increased as displayed in Table 5.5.

The comparative performance of different models after ensembling in terms of accuracy is shown in Figure 5.5.

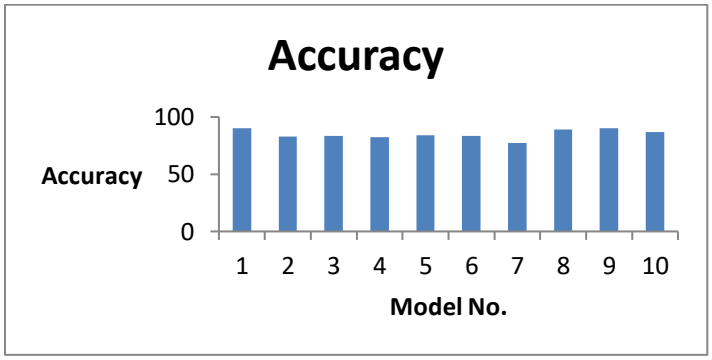

Figure 5.5 Ensembling Accuracy

The comparative performance of different models after ensembling in terms of specificity is shown in Figure 5.6.

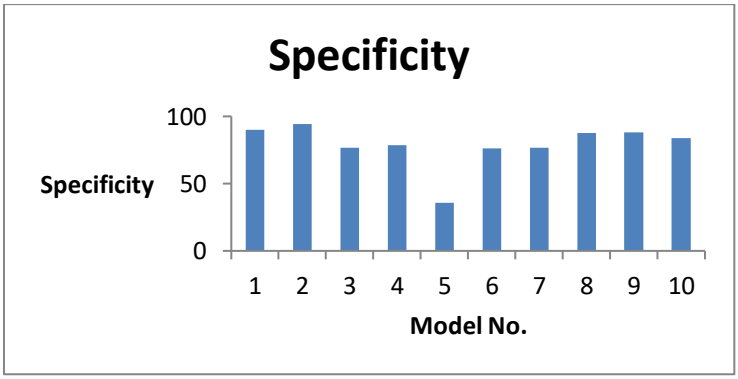

Figure 5.6 Ensemble Specificity

The comparative performance of different models after ensembling in terms of accuracy is shown in Figure 5.7

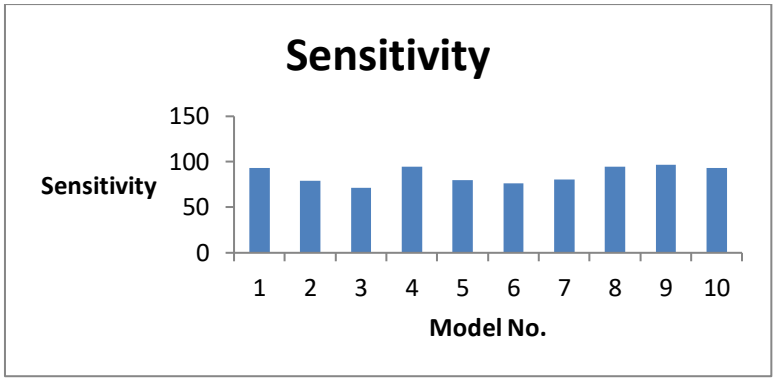

Figure 5.7 Ensemble Sensitivity

5.3 Summary

This chapter provides implementation of machine learning algorithms and also discussed about various parameter like accuracy, sensitivity, specificity. When we run the different machine learning algorithms we obtain the less accuracy of the model Decision tree and Naive bayes, then we ensemble both the models and get higher accuracy.

Chapter-6

Conclusion and Future Scope

6.1 Conclusion

In this thesis, we proposed A Cloud IoT based Framework for Diabetes Prediction to support the blood glucose management. The important aspect of this solution is that most of the measurements and interactions with the patient are done at home, which enhances the self-monitoring blood glucose solutions, allowing the interaction of the doctors with the proposed platform, which is connected to the different non-invasive IoT devices through communication protocols and cloud to keep it updated. In particular, we utilized five machine learning algorithms naive bayes, decision tree, random forest, support vector machine and neural network. We found that random forest proven to give best results and naive bayes proven to give worst results as compared to other algorithms. For improving the accuracy of naive bayes we perform ensembling on the models and got accuracy 90.2%.

6.2 Future Scope

This work can be extended on gestational diabetes because gestational diabetes mellitus (GDM) is a severe and neglected threat to maternal and child health. Numerous women with GDM experience pregnancy-related complications including high blood pressure, large birth weight babies and obstructed labour.

The coming age of eHealth is intrinsically linked to the successful deployment of a safe and privacy-preserving M2M/IoT infrastructure. We emphasize that security and privacy for eHealth in the emerging IoT landscape offers serious challenges as well as exciting opportunities.

References

[1] Dimitrov, D.V," Medical internet of things and big data in healthcare," *Healthcare informatics research*, pp.156-163.

[2] Yun, Miao and Bu Yuxin, "Research on the architecture and key technology of Internet of Things (IoT) applied on smart grid," *In Advances in Energy Engineering (ICAEE), 2010 International Conference on*, pp. 69-72. IEEE, 2010.

[3] Jing, Qi, Athanasios V. Vasilakos, Jiafu Wan, Jingwei Lu, and Dechao Qiu, "Security of the Internet of Things: perspectives and challenges,"*Wireless Networks 20*, no. 8 (2014) pp. 2481-2501.

[4] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, 2015.

[5] Aihow. Com, "Difference between machine learning and deep learning" [Online] Available:https://artificialintelligencehow.com/2017/10/18/difference-machine-learning-deep-learning/ [Accessed: 5 January, 2018]

[6] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[7] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[8] Dietterich, Thomas G, "Ensemble methods in Machine Learning," *In International workshop on multiple classifier systems, Springer Berlin Heidelberg*, pp. 1-15, 2000.

[9] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. Kegelmeyer, "Ensemble diversity measures and their application to thinning," *Information Fusion*, vol. 6, no. 1, pp. 49–62, 2005.

[10] Freund, Yoav, and Robert E. Schapire, "Experiments with a new boosting algorithm," *In icml*, vol. 96, pp. 148-156. 1996.

[11] Wolpert DH, "Stacked generalization," *Neural Network* ,vol. 5, no. 2, pp. 241–259,1992.

[12] Huang, Gao, Shiji Song, Jatinder ND Gupta, and Cheng Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Transactions on Cybernetics* ,vol. 44, no. 12, pp. 2405-2417, 2014.

[13]"IOSPress,"IOSPress. [Online] Available: http://www.iospress.nl/catalogue/Book series/frontiers in artificial intelligence and applications [Accessed: 10-January-2018].

[14] C. E. Rasmussen, "Gaussian Processes in Machine Learning," *Advanced Lectures on Machine Learning Lecture Notes in Computer Science*, pp. 63–71, 2004.

[15] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien, "Semi-supervised Learning," IEEE Transactions on Neural Networks vol. 20, no. 3, pp. 542-542, 2009.

[16] Xu, W., Zhang, J., Zhang, Q., & Wei, X, "Risk prediction of type II diabetes based on random forest model," *Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* ,pp. 382-386,IEEE, 2017.

[17] Gandhi, K. K., & Prajapati, N. B, "Diabetes prediction using feature selection and classification," *International journal of advance Engineering and Research Development*, 2014 May.

[18] Panwar, M., Acharyya, A., Shafik, R.A. and Biswas, D, "K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus". In Embedded Computing and System Design (ISED), pp. 132-136. IEEE,2016.

[19] Sowjanya, K., Ayush Singhal, and Chaitali Choudhary, "MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices," *In Advance Computing Conference (IACC)*, pp. 397-402. IEEE, 2015.

[20]Heydari, M., Teimouri, M., Heshmati, Z. and Alavinia, S.M., "Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran," *International Journal of Diabetes in Developing Countries*, pp.167-173.

[21] Komi, M., Li, J., Zhai, Y., & Zhang, X, "Application of data mining methods in diabetes prediction," *Image, Vision and Computing (ICIVC)*, pp. 1006-1010. IEEE, 2017.

[22] Swain, Aparimita, Sachi Nandan Mohanty, and Ananta Chandra Das, " Comparative risk analysis on prediction of Diabetes Mellitus using machine learning approach," *In Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 3312-3317.

[23] Jahangir, Maham, Hammad Afzal, Mehreen Ahmed, Khawar Khurshid, and Raheel Nawaz, "ECO-AMLP: A Decision Support System using an Enhanced Class Outlier with Automatic Multilayer Perceptron for Diabetes Prediction"

[24] Douali, Nassim, Julien Dollon, and Marie-Christine Jaulent, "Personalized prediction of gestational Diabetes using a clinical decision support system," *In Fuzzy Systems (FUZZ-IEEE)*, pp. 1-5.

[25] Bhatia, Nitin, and Sangeet Kumar, "Prediction of Severity of Diabetes Mellitus using Fuzzy Cognitive Maps," *Advances in Life Science and Technology*, pp.71-78, 2015.

[26] Samant, Piyush, and Ravinder Agarwal, "Diagnosis of Diabetes Using Computer Methods: Soft Computing Methods for Diabetes Detection Using Iris," *Power*, 651, pp.63915.

[27] Mora, H., Gil, D., Terol, R. M., Azorín, J., & Szymanski, J, "An IoT-Based Computational Framework for Healthcare Monitoring in Mobile Environments," *Sensors*, 17(10), pp.2302,2017.

[28] Chang, S.H., Chiang, R.D., Wu, S.J. and Chang, W.T, "A context-aware, interactive M-health system for diabetics," IT Professional, 18(3), pp.14-22,2016.

[29] Gomez, Jorge, Byron Oviedo, and Emilio Zhuma, "Patient monitoring system based on internet of things", *Procedia Computer Science*, pp: 90-97.

[30] Gupta, Praveen Kumar, B. T. Maharaj, and Reza Malekian, "A novel and secure IoT based cloud centric architecture to perform predictive analysis of users activities in sustainable health centres," *Multimedia Tools and Applications*, pp. 18: 18489-18512.

[31] Antonovici, Dorin-Andrei, "Acquisition and management of biomedical data using Internet of Things concepts," *Fundamentals of Electrical Engineering (ISFEE), International Symposium on*. IEEE, 2014.

[32] Al-Taee, M. A., Al-Nuaimy, W., Al-Ataby, A., Muhsin, Z. J., & Abood, S. N, "Mobile health platform for diabetes management based on the Internet-of-Things," *Applied Electrical Engineering and Computing Technologies (AEECT*) pp. 1-5. IEEE, 2015.

[33] Natarajan, K., B. Prasad, and P. Kokila, "Smart health care system using internet of things," *Journal of Network Communications and Emerging Technologies (JNCET) 6*, pp. 3.

[34] Rathore, M.M., Ahmad, A., Paul, A., Wan, J. and Zhang, D, "Real-time medical emergency response system: exploiting IoT and big data for public health," *Journal of medical systems*, p.283.

[35] Jara AJ, Zamora MA, Skarmeta AF, “An internet of things–based personal device for diabetes therapy management in ambient assisted living (AAL),”*Personal and Ubiquitous Computing, pp.431-440*.

[36] Rahman, Ruhani Ab, “IoT-based personal health care monitoring device for diabetic patients,” *Computer Applications & Industrial Electronics (ISCAIE),* pp. 168-173. IEEE, 2017.

[37] Bhat, Guruprasad M., and Nayana G. Bhat, “A novel IoT based framework for blood glucose examination,” *In Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT),* pp. 205-207.

[38] Winterlich, Anthony Iain Stevenson, Andrew Waldren, and Tom Dawson, “Diabetes Digital Coach: Developing an infrastructure for e-health self-management tools,” *Developments in systems Engineering (DeSE), 9th International Conference*, pp. 68-73. IEEE, 2016.

[39] Rajput, Dharmendra Singh, and Rakesh Gour, “An IoT framework for healthcare monitoring systems,” *International Journal of Computer Science and Information Security 14.5* ,pp. 451.

[40] Bhatia, M., & Sood, S. K, “A comprehensive health assessment framework to facilitate IoT-assisted smart workouts: A predictive healthcare perspective,” *Computers in Industry*, pp. 50-66.

[41] Mall, Shreejay, Mansi Gupta, and Rahul Chauhan,” Diet monitoring and management of diabetic patient using robot assistant based on Internet of Things,” *Emerging Trends in Computing and Communication Technologies (ICETCCT)*, pp. 1-8, IEEE, 2017.

[42] Istepanian, R. S., Hu, S., Philip, N. Y., & Sungoor,“ The potential of Internet of m-health Things “m-IoT” for non-invasive glucose level sensing,”. *In Engineering in Medicine and Biology Society, EMBC,* pp. 5264-5266.

[43]Ephzibah, E.P,“ Cost Effective Approach on Feature Selection using Genetic Algorithms and Fuzzy Logic for Diabetes Diagnosis,” *International Journal on Soft Computing (IJSC), 2,* pp.1-10.

[44] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien, “Semi-supervised learning” *IEEE Transactions on Neural Networks* ,vol. 20, no. 3, pp. 542-542, 2009.

[45] Ron Kohavi and David H Wolpert,“Bias plus variance decomposition for zero-one loss functions,” *In Machine Learning: Proceedings of the Thirteenth International*, pp. 275–283, 1996.

[46] Krogh, Anders, and Jesper Vedelsby, “Neural network ensembles, cross validation, and active learning,” *Advances in neural information processing systems* vol. 7, pp. 231-238, 1995.

[47] Guo J., Chen, I.-R., Tsai Jeffrey, J.P ,“A survey of trust computation models for service management in internet of things systems,” *Comput. Commun*. 2017, pp.1–14, 97.

[48] Stergiou, C.,Psannis, K.E.,Kim, B.-G.,Gupta, B, “Secure integration of IoT and Cloud Computing,” *Future Gener. Comput.* Syst. 2018, 78, pp. 964–975.

[49] Mollah, M.B., Azad, M.A.K.,Vasilakos, A, “Security and privacy challenges in mobile cloud computing: Survey and way ahead,” *J. Netw. Comput. Appl*. 2017, 84, pp. 38–54.

[50]Botta, Alessio, Walter De Donato, Valerio Persico, and Antonio Pescapé, “Integration of cloud computing and internet of things: a survey,” *Future Generation Computer Systems,* pp.56: 684–700,2016.

[51]Al-Fuqaha, Ala, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and MoussaAyyash, “Internet of things: A survey on enabling technologies, protocols, and applications.” *IEEE Communications Surveys & Tutorials*, pp.17(4): 2347–2376,2015.

[52]UCI "PIMA India Diabetes dataset" [Online] Available: https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes [Accessed: 10 December,2017]

[53] Glucotrack. Com, "Gluco Track device". [Online] Available: http://www.glucotrack.com/about-glucotrack/ [Accessed: 10 January, 2018]

[54] "Dia-Vit", [Online] Available: https://dia-vit.com/ [Accessed: 10 January, 2018]

[55] "Sugar beat",[Online] Available: http://sugarbeat.com/ [Accessed: 10 January, 2018]

[56] "Wearable technologies",[Online] Available: https://www.wearable-technologies.com/2017/01/wearables-for-diabetic/ [Accessed: 10 January, 2018]

[57] "K watch", [Online] Available: https://www.pkvitality.com/ktrack-glucose/ [Accessed: 10 January, 2018]

[58] Paolo Patierno, "Iot protocols landscape" Available: http://www. slideshare. net/paolopat/io-t-protocols-landscape,2014.

[59] Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee, "Hypertext transfer protocol–http/1.1", 1999.

[60] Zach Shelby, Klaus Hartke, and Carsten Bormann, "The constrained application protocol (coap)," 2014.

[61] U. DBD, "Confusion Matrix", Www2.cs.uregina.ca, 2017. [Online] Available:http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html. [Accessed: 23- Dec- 2017].

[62] Cortes, C., Vapnik, "Support-vector networks", Machine Learning, 20(2),pp. 273-297, 1995.

[63] V. Vapnik, "The Nature of Statistical Learning Theory", NY: Springer- Verlag. 1995.

[64] K. Raja, G. Tholkappia Arasu, Chitra. S. Nair, "Imputation Framework for Missing Values" ,International Journal of Computer Trends and Technology, vol.3, no.2, pp.215-219, 2012.

List of Publications

[1] Neha Sharma, Dr. Ashima Singh.”Diabetes detection and prediction using machine learning and IoT: A survey” *International Conference on Advanced informatics for computing research ICAICR 2018* [Accepted]

Nehasharma_plag

ORIGINALITY REPORT

13%	8%	8%	5%
SIMILARIT Y INDEX	INT ERNET SOURCES	PUBLICAT IONS	ST UDENT PAPERS

PRIMARY SOURCES

1	doria.fi Internet Source	1%
2	Pawan Lachheta, Seema Bawa. "Combining Synthetic Minority Oversampling Technique and Subset Feature Selection Technique For Class Imbalance Problem", Proceedings of the International Conference on Advances in Information Communication Technology & Computing - AICTC '16, 2016 Publicat ion	1%
3	Weifeng Xu, Jianxin Zhang, Qiang Zhang, Xiaopeng Wei. "Risk prediction of type II diabetes based on random forest model", 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Publicat ion	1%
4	Submitted to National Institute of Technology, Patna St udent Paper	1%