Global Journals $\LaTeX$ JournalKaleidoscope™

Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

A Proposed Method to Identify the Occurrence of Diabetes in Human Body using Machine Learning Technique

Tanvir Rahman

Received: 7 September 2021 Accepted: 1 October 2021 Published: 15 October 2021

Abstract

Advanced machine-learning techniques are often used for reasoning-based diagnosis and advanced prediction system within the healthcare industry. The methods and algorithms are based on the historical clinical data and factbased Medicare evaluation. Diabetes is a global problem. Each year people are developing diabetes and due to diabetes, a lot of people are going for organ amputation. According to the World Health Organization (WHO), there is a sharp rise in number of people developing diabetes. In 1980, it was estimated that 180 million people with diabetes worldwide. This number has risen from 108 million to 422 million in 2014. WHO also reported that 1.6 million deaths in 2016 due to diabetes. Diabetes occurs due to insufficient production of insulin from pancreas. Several research show that unhealthy diet, smoking, less exercise, Body Mass Index (BMI) are the primary cause of diabetes. This paper shows the use of machine learning that can identify a patient of being diabetic or non-diabetic based on previous clinical data. In this article, a method is shown to analyze and compare the relationship between different clinical parameters such as age, BMI, Diet-chart, systolic Blood Pressure etc. After evaluating all the factors this research work successfully combined all the related factors in a single mathematical equation which is very effective to analyze the risk percentage and risk evaluation based on given input parameters by the participants or users.

Index terms—

1 Introduction a) Background

enerally, Diabetes Mellitus (DM) develops in the body silently when there are higher or uncontrolled blood glucose level exists in the blood-plasma cell for a long time. Food is the prime source of calorie and food is prime the energy generator of the body. Generally, the foods are taken in regular basic contain a lot of glucose or glucose substance. Glucose is the primary and basic unit of energy circulation and energy regulation. Glucose is divided into several substances and then the small cell units are oxidized with the sufficient amount of oxygen. Then, the small subsequent oxygen particles are transmitted through blood circulation and produce sufficient amount of energy and nutrition for all the organs of the body. Insulin is a pancreas produced Hormone which is the key component to synthesis Glucose and divide Glucose into millions of active particles. For a healthy and active person sufficient amount of Insulin is produced and emitted from Pancreas. That is why for a general Non-diabetic patient , Insulin production rate is equal to the glucose Intake of the body .So, For a Non diabetic-patient the all the amount glucose casted from daily food intake ,is sufficiently divided into molecules and produces energy and rest of the unused energy is stored as Fat in the body. In the common scenario, no extra glucose particles are available in the blood plasma. According to several health study, a person is considered to be a Type-1 diabetic patient when his/her pancreas fails to generate sufficient amount of insulin to react with glucose.

The normal range of glucose level is reference value is 3.9 to 5.4 mmol/l (70 to 99 mg/dl) [1] for normal patients at fasting phase and according to American Diabetes Association the reference value at fasting time in the period of Diagnosis of diabetes is considered as 7.0 mmol/l (126 mg/dl) or above [1] and the reference value for non-diabetic patients is under 7.8 mmol/L and preferred value for Type-2 diabetes is under 8.5 mmol/L at Random diabetes testing phase (1.5 hour after food) [1][2] [5].

Scientists have found a significant link between This high blood sugar and several other diseases like catastrophic damage of several nerves, kidney and Renal failure, heart and vein damages, eye-sight. An uncontrolled diabetes level for a long time period can also lead a patient to dead. The growth rate of diabetes patients is enormous around the globe. From a report published in 2013, the International Diabetes Federation (IDF) predicted the probable diabetes patients around the world. It claimed that estimated about 382 million or more people worldwide are carrying excessive amount of glucose in blood and probably had been suffered from diabetes, and the report also predicted that within the year of 2035, it enormous number of diabetes patients can even exceed to 592 million. From the report of various heath surveys it is estimated about the consequent percentage of total 8.5% of the population of South-east Asian region have diabetes where about half of the population of victims even do not aware of that they are carrying diabetes silently in the body .The growth rate of diabetes patients is alarming in several middle income and emerging countries and Asian countries are the major contributor for devasting growth rate of diabetes [2].

From the analysis of several health studies, it is known that the advance and predefined adequate proper knowledge about the consequences of diabetes and better and more compact prediction solution may be very effective to fight against diabetes in more convenient way and help to raise awareness among people.

World Health Organization (WHO) published a static based analysis and research-based report to focus and intensify the real diabetes scenario of Bangladesh. From the recent meta-analysis conducted by WHO reviewers showed that the recent threating pervasiveness of diabetes among Bangladeshi civilian had increased in an alarming rate, the report focused on dramatic increment sequence of growth rate from in 1995 to 2010 [3]. In 1995 the rate was only 4% which increased 9% in 2006 and until the year of 2010.

According to the analysis and prediction of the International Diabetes Federation based on the analysis of several case-studies , the organization predict that the devastating rate will grow further and will be increased about 13% by the year of 2030 [3].

By reviewing the previous documentations and reviewing several journals, it was confirmed that there is a serious limitation in the field of diabetes research and predictive system because there are no available suitable documentations or studies dedicated for this specific region. The prime drawback of Previous studies were that the previous models were not designed properly and combination of attributes were not properly designed. Again, the previous studies were bounded to specific region or focused on specific sex or gender.

The prime emphasis of the study is to obtain a full set of co-relating factors which are the prime responsible attributes for diabetes and to establish a predictive model to predict and identify diabetes at an early age. This model is best specially optimized for the south Asian counties like Bangladesh because to conduct this study a lot of matters and factors regarding for the specific region were taken on consideration based on need and expert opinion.

Therefore, there was a serious need for a specific model. According to expert opinion and WHO's Report guideline, the primary goal was to identify each individual, household and related fixed or specified community factors associated with the conditions. WHO found and expressed a significant connectivity or relationship with diabetes and age. In most common term, older /middle aged people have the more chance to get attached by diabetes because age is one of the prime differentiators for diabetes.

2 Fig. 1.1: Diabetes ratio around the world

From the result of previous study, it was confirmed that most of the diabetic affected population likely (40%) are enough educated, have sufficient knowledge about diabetes and they belong to middle class income level, where almost 13% participants were from lower income family. From the previous study, it was known that about 40% of total diabetes population were receiving regular medical check-up and proper healthcare system.

Diabetes is known as a silent carrier and it is a carrier of several deadly diseases which can cause long term health hazards. To maintain a good health score, it is needed to identify diabetes at primary stage and maintain a proper diet and exercise chart. Again, the devastating growth rate of diabetes in this region, is very harmful for the human resource management for the country as diabetes patients become unable to overtake heavy and handy task [7] [9].

So, the method of prediction of diabetes at an early stage is very important and beneficial for the community. There is no available preventive methods of totally cure the diabetes and root out the disease from the body, but there is a well-defined solution to control the glycemic index and sugar of the body.

Again, By using several data-mining techniques, it is possible to predict the disease far early and assist doctors and healthcare providers to reach in a better diseases management procedure. Analysists and researchers Patient will also get food and exercise recommendation through this system.

3 b) Problem Definition

Diabetes is a wide spread disease and it has some common symptoms and attributes. Family history, Age, Sex, BMI, blood pressure etc were taken into consideration to make a proper evaluation of model. The normal measurement level of diabetes is fall between to 6.0 mmol/L during the time of fasting and it will cross the level of 7.8 mmol/L after 2 hours meal.

Diabetes has 2 different types which include type 1 diabetes and type 2 diabetes. Diabetes has some specific symptoms. These symptoms appear to the people, especially those with patients with type 2 diabetes, some of these symptoms may appear lately. For the type 1 diabetes patients the symptoms may appear quickly and more severe [4].

Some common symbol of type 1 and type 2 diabetes are [4]:

? Increased amount of extreme thirst ? Increased amount of hunger [7] [12].

? Sudden weight loss ? Frequently a chemical substance named ketones is found in the urine ? More Frequent and unexplained urination [5] ? Decrement of vision gradually ? Get attacked by more and Frequent common infections, like skin infections and infection in several sensitive organs [8].

From the result of various surveys and analysis, it was confirmed that Type 1 diabetes can develop in the body at any indigenous period or age, though it often found in childhood stage. where Type 2 diabetes is more widely spread to the middle-aged person and common in people older than 40 [4], though type-2 diabetes can also appear at an early age. Diabetic diseases is classified into four category. Therefore, patient can have these type of diabetes. These are given below:

4 Type 1 diabetes

The prime cause of type 1 diabetes is still unknown today. As per scientific documentations, combination of genetic susceptibility and environmental factors are considered as the primary reason of Type-1 diabetes .In this type, the immune system of the body surprisingly misunderstands and destroys the insulinproducing cells in the pancreas. This action is hazardous for metabolic system because in this case there is only a little or no insulin found in the body which is insignificant for metabolism. As a result, metabolism and energy transmission to the body cell is hampered and extensive level of sugar found in bloodstream [4].

5 Type 2 diabetes

As per scientific documentations, in type 2 diabetes, body cells become resistant to the action of insulin. one the other hand, the organ named pancreas cannot produce sufficient insulin for the body [4]. For a type-2 Diabetes, adequate exercise and proper diet plan is needed to manage the proper blood sugar level. Again, for some specific case, doctors recommend insulin to some specific patients. Doctors and health scientists around the world indicated that genetic factors, nature and environment plays an important role for creating diabetes. Overweight and diabetes Type-2 has also strong relation.

6 Gestational diabetes

During the period of pregnancy, the placenta produces some dedicated hormones. These specified hormones act against insulin [4]. Generally, in all the cases and types of diabetes, patients pancreas releases extra-more insulin to control and manage the diabetes. But in some cases pancreas cannot emit extra insulin which causes gestational diabetes [4].

7 c) Overview of the thesis

The responsible reacting factors were marked and identified based on several important factors and co-related relationship. The findings and the results of Several statics and previous studies were extensively used to make the proper evolution of the model. Then important information and findings from the results of previous successful case studies were identified and sorted for future use. Attributes were selected with extensive care and based on their contributions for developing diabetes in body. The expert opinions and doctors advice enlisted in several health journals were taken into consideration to select the proper attributes. Again, participants were classified and divided into several groups and different categories based on research demand. Several statistical evolution and informative data were the prime source of data and patients real time data depending on the Complete list of foods and different meal plan [0]. Based on the collected samples and evaluations of different attributes of each and individual patients, the desired calorie need was taken into consideration based on patients need and health need. If the diabetes patients follow the recommended diet chart and adequate exercise then it will be beneficial to control diabetes.

8 d) Scope of the thesis

The thesis was done based on the evolution of several related attributes and their contribution towards the thesis. The findings and recommendations of thesis will pave the way to find out a more prominent and trustworthy solution for the diabetes patients to digenesis the disease far earlier than it appear and it will recommend a optimum lifestyle needed for the diabetes patients. The lifestyle, food habit, exercise time and taken insulin can be stored in a database for further analysis. Again, based on the comprehensive analysis and extensive data analysis, prediction of risk factor of diabetes can be calculated by using our developed model. Risk factors and the probability of patient's get attacked by diabetes can be predicted accurately to fight against diabetes in more convenient way.

9 e) Objective of the thesis

The primary goal of thesis is to identify the risk of diabetes at an early age to create awareness among the future diabetes patients and to manage the diabetes in much more pre-planned and organized way to fight against diabetes. As diabetes is a permanent diseases and there is no available solution to up-root diabetes from the body. The proper and well defined safety regulations can be ensured by regular assessments. Diabetic patient's recommended and optimum lifestyle was also suggested in this evaluation. It will also pave the way to reach in a compact solution by predicting diabetes earlier and upcoming diabetes victims will become more conscious about their habit and lifestyle to minimize the risk and to manage a better health index.

10 f) Organization of the thesis

Despite of the advancement of medical science, there is no permanent cure Diabetes and it is growing at an alarming rate which is cautious for the economical development of the country. Most of the times, the patient's health condition becomes worse of Diabetic Patients because they are often ignorant of risk factor and do not maintain a proper diet chart. So, the advance prediction of risk factor of diabetes can be life saving for the ignorant patients.

In the previous study, Many scientists used various kinds of machine learning techniques. Researchers around the world experimented and used several different types of classification algorithms. Several statistical techniques and mechanisms were used to predict diabetes in advance. Again, Doctors and Health experts also analyzed the performance of different algorithms and cross-validated the model. The previous approaches paved the way to reach in a compact solution.

In this paper, a custom designed well defined model along with a refined formula was proposed to identify diabetes at an early age and manage the proper health index in much more convenient way.

A well structured and efficient dataset was collected from the combination of various medium like Internet open data sources, survey results and questionnaires and all the data was stored in a integrated database to use it for further validation and future development of the model.

Extensive Synthetic and analysis of collected dataset was done based on the combination of various attribute. In this research, it was the primary goal to find out a proper relationship status and proportional or inverse-proportional relationship between the each and every reacting attributes. Then, the scientific evaluation and proper cross-validation process were done to recheck highest accuracy of the model.

11 II.

12 Literature Review a) Correlative Factors on Type 2 Diabetes Prevention Efforts of the Senior High School Students in Makassar

This study has found out the most common and successfully indicated related probable factors responsible for diabetes. The primary goal of the paper was to analyze associated factors and clauses related to DM specially for the level of teenage student in the city of Masakkar, Indonesia. In this study, the primary dataset was collected from high school students and age between 11-16 year students were highlighted and focused to determine the diabetes Meletus's impacts and related reasons to analyze the risks and the threads of diabetes for the teenager and to prevent it at an early stage. Data was collected in various methods like questionnaire, survey etc. The study is based on Indonesia, where DM prevalence is estimated increase from 8.4 million to 21.3 million between 2000 and 2030. The devastating growth rate diabetes affected patients is hazardous and it is indicating a highly health disaster. By analyzing the datasheet thoroughly, it was found that only 6 respondents (24.0%) had parent with lower or less standard education level where majority of respondents 189 in numbers had parent with high education level, consciousness in prevention efforts on Diabetes. In this paper Researchers noticed a common fact that among the participants of the study, those parent have higher or sufficient level of education, their children are more aware about the risk factors and they are adopting better prevention efforts to protect and fight against Diabetes. This study suggest that those people with low education level had "1.27 times" at risk of suffering DM than people with high education level From this survey, it was established that high incidence number of DM type 2 because of low of parental education level on prevention of DM type 2 incidence. Again, By analyzing the datasets, case studies and reports of several health sheets, it was seen that parental support plays vital role to maintain the goodheath index (standard) and it is a key promotor for creating prevention effort against Type-2 Diabetes among the teenager group. The result of the several data analysis and data sorting techniques showed a significant relationship between parental support with prevention efforts of DM type 2 among senior high school students. The study strongly found a link of The peer support and DM type 2 diabetes. The prime clause of the peer support was referred as providing relevant information, preventive measures on DM type 2.

In this paper, Authors observed and noted a significant relationship between hereditary parental educational awareness and health consciousness level, the strong relation of benefits perceived, barriers perceived, knowledge, peer support and social awareness and proper informational advantage can create a huge improvements and significant progress of prevention of DM at an early age

13 b) Designing Technological Interventions for Patients with Discordant Chronic Comorbidities and Type-2 Diabetes

In this present decade it is often found that Patients with Discordant-Chronic Comorbidities (DCCs) are likely to attacked by multiple complex DCCs with a set of fully contradictory medicinal requirement, prescription and guidance. This problem is disastrous because a medical professional should minimize the medication as per priority of diseases .So, there was a great demand for a help assistant system which can prioritize based on patient's health index and suggest a optimum solution for patients simultaneously. As the part of the model, authors focused on developing and publish a mobile application to evaluate the risk assessments scores on demand. The prime purpose of the application is to suggest and provide proper medication guidance based on the need and physical condition of every individual patients. The suggested application gathers a ton of useful health-information, heath report, health index etc data from every individual user, then analyses data and enable patients to assist their conditions and treatments. It's often found in the medical data analysis that the Chronic conditions and apparent symptoms last for five or more months such as common diseases like Diabetes, Arthritis, or Depression, are becoming increasingly common in patients. Due to habitat and integrated nature of diabetes and it's typical conditions, patients are asked to play an active role in their treatments schedule and planning. From the health summary and results of several surveys, it is easily understood that the patients who do not follow or maintain the standard life guidance recommended for diabetes patients are at a greater risk .It's often found that the specific fact that, The development of Discordant Chronic Comorbidities with multiple chronic conditions, has become most common and often can be seen with highest rate co-connectivity which creates difficulties for healthcare assistants and desired patients when it comes to term of managing and controlling the impact of the managing conditions. To control and manage the state of diabetes, some plethora of available tools, apps, sensing devices and various sensor tools only support the care and proper management of diabetes diseases. In previous study, it is known the proper diet sheet and diet management is the key factor to manage the proper status of diabetes. From the factbook, it is known that The prime challenge in studying patients with comorbidities arises from their compounding health factors and health assessments issues, which states often leads the affected patients lead to more sicker and more spending time in enrollment of hospital admission. This is the primary barrier of understanding the proper guidance of selfmanagement assessments of their diseases and it's associated risk factors. Based on interviews conducted with patients with Type 2 Diabetes and other Discordant Chronic Comorbidities, researchers designed a mobile application based on the barriers patients faced in successfully managing their treatment as well as some of the solutions they used or wished to use. The overall goal of this mobile application is to encourage patients to inbound in the application assessment exercise to improve their long-term health and quality of life.

By Approaching forward on this certain topics, Researchers emphasized and tried to develop this application and participate in testing with users with the ultimate hope of releasing this application to the general public. In addition, Researchers are extensively looking to find out the optimum ways to manage diabetes in a more convenient way with the supervision of computer intelligence.

14 c) Recurrent Neural Networks with Non-Sequential Data to Predict Hospital Readmission of Diabetic Patients

It's recognized that Hospital readmissions and vulnerable health index rates are the indicators of poor quality of Medicare, such as inadequate discharge planning and care coordination. It's often consider that the frequent readmission and lower health index can be avoided by certain methods and propositions. In this paper, a Recurrent Neural Network model is carefully designed to predict whether a patient would be readmitted in the hospital or his/her health index parameters will be reevaluate to gain the highest productivity with several machine learning algorithm .IN THIS Study, it is found that RNN showed highest prediction precision to target high risk patients and prevent recursive admissions. Hospital readmission and degradation of Health index what will happen when a patient within a specified time interval or timeframe, who had been released from a hospital with vital increment of health condition is admitted again. Again, a lot of research studies and publications proved that healthcare centers can be engage in several activities like clarifying patient discharge instructions, coordinating with patient's health conditional index, handling with post-acute care providers, vibrant cleaning mechanism to reduce the rate of readmissions of patients. In is paper, therefore raises a big question that which patient groups or which type of patients must be targeted to effectively reuse and redesign available resources for preventing readmission and to use the classified information for special case study. Many predictive and specially designed Models that can predict accurately these are of a great help for hospitals all over the world as they can put extra efforts on high risk patients and can decrease their readmission rates.

In this experimental procedure of Research topic, the prime motto was to redesign, analysis and construct a powerful model to predict exact numbers of diagnosis's measurements and different type of machine learning Approaches and models were used to predict with highest accuracy. In this case, especially Recurrent oriented Neural Network outperformed the rest of the machine learning models in the prediction quality in the scale of productivity and accuracy. The knowledge, experimental results and outputs gained from the journal can effectively improve the traditional health system to target high risks patients, reduce rate of readmission and deliver better health care.

15 d) Development of Indian Weighted Diabetic Risk Score (IWDRS) using Machine Learning Techniques for Type-2 Diabetes

Medical experts and scientists have expressed their opinion that detection of diabetes at an early phase can be a lifesaving effort. Advance Diabetes relating factors and different screening tools such as Diabetes Risk Score (DRS) can effectively assist diagenesis and detecting diabetes accurately and help to prevent the diabetes among pre-diabetes phase at an early time before diabetes occurs. In current evaluations and assessments, Researchers have observed certain related issues in the available data and advocate the need to address the same. In this paper it's established a novel South-Asian regional Weighted Diabetic Risk assessments and co-relating factors. Different Machine Learning algorithms such as distance based clustering with Euclidean distance, k-means etc techniques were used by the researchers as a part of establishing a profound diabetes risk assessment tools to analyze the contribution of associated factors like blood pressure, age, stress and life quality BMI, diet, physical activity to boost up high plasma glucose level. In this paper ,the strategy to establish a strong and co-relating relationship between several differentiating factors , establishing a formula and then test and validates the formula with several test datasets to ensure the maximum accuracy. On an research World health organization referred that South-Asian countries citizen's are affecting on diabetes on this last two decades encounters at an devastating rate due to several depending factors. In this paper, the researcher collected datasets from various data sources, conducted surveys and used previously available data and information's to represent informational support. Data is collected form the south-Asian populations mostly from Bangladesh under the supervision of medical professionals. Several collected and trustworthy datasets were also used to strengthen the decision. Different types of Machine learning algorithms and advance data sorting principles are used for determining threshold values for various parameters when it was needed. A proper diabetes evaluation system or function is calculated for each factor like BMI, age, phenotypes, personal medical history, family history, diet, physical activity, stress and life quality. The genetic property, phenotype, lifestyle, working habit and some others factors are seriously related to diabetes. Different individual research, case studies and scientific studies have been proposed earlier by scientists to reduce the risk of diabetes to reduce the risks of diabetes, it's needed to differentiate the relationship between diabetes and different co-relating factors to fight against the risks of diabetes at an very early stage.

In this study, several reacting mechanisms, techniques and elements were successfully sorted which is very important to bring a new dimension in healthcare imagining prediction system. Different type of surveys, questionnaire, data synthesis techniques and computational intelligence were successfully used to identify and analyze the risk factors and their scores.

16 e) Study of Type 2 Diabetes Risk Factors Using Neural

Network For Thai People and Tuning Neural Network Parameters Advanced datamining techniques and analyzing tools are very Efficient to detect and predict diseases and their relating risk factors at an early age. In this paper, Researchers are trying to find out the relating factors which are mainly responsible for Type-2 Diabetes and proposed a relating solution to identify diabetes. In this paper a complete set of related factors which includes blood pressure, weight, body mass index, family history are considered as a primary factors. Again, smoking and alcohol consumption were considered as a strong co-relational factor based on their linked found in several researchers. To analyze and synthesis data BNN algorithm was as used. To collect datasets and sample information for training set about two-thousand samples of various health attributes were managed from BMC Hospital, Thailand. Based on previous learnings and previous research suggestions ,this paper found a strong relationship status and divided the risk level in there consequent stages i) low risk denoted as -1 point, ii) Medium denoted as between the range of -1 to 1 point and iii)High Risk was marked as the cautious level and contributed a single (1) scoring point for each risk based on different scale of measurement (unit) depending on the weighted contribution of linked factors like Family history, Age, Sex, BMI, blood pressure to make a proper evaluation of model. It was also added 1 point to the risk score for smokers and consumer of alcohol with timeframe of 4 weeks or more to summarize higher risk capability. By analyzing the documentations of the paper it established a U-shaped relationship with the consumption alcoholic drinks and smoking habit. The major findings, research analysis and conclusions was divided in two different portions. In this study, authors Initially identified the major related and responsible factors and made a complete a list of the proper corelating factors based on the evolution of collected datasets and previous records. Then , the all concerned factors and related terms were intelligently sorted and divided into three sophisticated categories based on their of different level of contribution to diabetes. Atlast, the study was concentrated on acquiring the learning rate with the tuning of BNN parameters.

This study concentrated in some vital factors and redefined the traditional reasoning methodologies which provides a better performance markup, higher accuracy level and better predictability compared to existing solutions and predictive analysis. From the result analysis of the paper, it was summarized that The prediction accuracy of the proposed strategy was not as good as expected, but in this paper, authors focused on the best optimum strategy to find out a better solution in future to predict diseases in much smarter way.

17 f) Data-Based Identification of Prediction Models for Glucose

From Result of various Health surveys and analysis, it is known that Diabetes mellitus is one of the wide spread diseases in all over the world .There are many co-effective factors which mostly responsible for the appearance of Diabetes, but there is a general or common reason between every single diabetes patients is that they might have deficiency in insulin production or insulin is not functioning well to improve the digestive system . It's advised to all the DM patients to track the regular status of blood glucose to maintain a proper control of the glucose count in the blood to become healthy and active. In this paper, it was observed that common barrier to control the diabetes or glucose level by a semiautomatic model is to monitor the mechanism of glucose levels in blood interact with insulin, diet intake or other factors interact with each other .In this paper, a set of traditional and classical identification techniques such as Holt's smoothing, classical simple smoothing model was compare to genetic programming models and techniques to evaluate the working efficiency of the model. Again, to maintain a proper and balanced autonomous glycemic control, a glucose control and blood sugar level monitoring principles and algorithms is extensively needed to outperform all existing solutions. In this paper, Authors put main emphasis to develop a forecasting or predicting model to the evaluate the level ricks DM based on trustable parameters like the real-time measurement of blood glucose. The Researcher also tried to predict the realtime basics blood glucose monitoring system and this algorithm would successfully measure the blood glucose level on the real time, it will analyze all the details and classified data and refer an insulin inhibitor system to supply the necessary amount of insulin particles based on the patient's need and health condition on the real time. The researchers have collected tons of data and heavily analyzed the data in terms of the space direction and the power spec-trum. for the 10 in-silico patients.

In this study, the researchers have reached in a conclusion that the combine package of both the previous Grammatical evaluation model and genetic programming is the best suitable techniques to predict, identify and manage the issue. This proposed approach will bring a new revolution and new strategy to adopt with next generation diagenesis and prediction modules to predict and fight against diseases at an early stage.

18 g) Improve Computer-Aided Diagnosis with Machine

Learning Techniques Using Undiagnosed Samples Now-a-days, different types of computer aided diagnostic tools, various predication and machine learning algorithms are used to identify the root causes and responsible factors. Again, to predict the risk of several fatal diseases in far advance and several computer aided tools and gadgets are extensive used today to assist human to prevent diseases more effectively or to maintain a good health score. To analyze and to understand thoroughly about a certain disease usually a huge number of diagnostic samples, opinions, surveys etc are needed to be collected, examined and analyzed to sort out the effective responsible factors and it's impossible for expert to analyze, simplify, synthesis this vast amount of information. That's why authors of the paper put emphasis to develop a new technique to analyze data faster. In this study Researchers proposed a effective semi-supervised machine learning algorithms named Co-Forest. Researched marked the new algorithm as an extended and extensively modified version of existing machine learning algorithm named "Random Forest". This algorithm is better for providing the analysis result and giving final hypothesis assessments compactly. [0].

Semi-supervised learning combines the both labeled and unlabeled data to extensively synthesis and extract the required information to establish a reliable and trustworthy hypothesis. The study suggests that, To plan or design a conventional methodology from scratch, the desired "co-training" data should be described by two sufficient and redundant attribute subsets.[0]In this methodology, each of the section of classification-division must be independent or act like as independent attribute and will capable of providing sufficient scopes unique learning capability. In this paper, author denoted L as a tag of labeled set and U denote unlabeled set.

In this co-training mechanism, 2 different sets of classifiers are trained from Labeled data, after that circumstance, each of sets should selects the most confident contents in Unleveled data to label from its point of View[0].

This study extensively focused on the usability of unlabeled data to boost the extensive learning capability, train from the unlabeled data and to save a lot of time in the field of health science. This approach is revolutionary and it will bring more pace in sample data management process, comparing and analyzing a ton of information in a short range of time frame.

19 h) Diabetes Prediction Using Ensemble Perceptron Algorithm

Today's people food habit is largely dependent on ready-made, high sugar and high calorie enriched foods. Medical experts and health scientist's advice the every suspected or affected diabetic affected person to diagnosis the level of glucose in blood in a routine cycle, which is costly and time consuming. The extensive use of data mining and machine learning algorithm with the assistance of computer aided system can effective be used to predict, identify and maintain diabetes in a controlled manner way. In this paper, authors proposed a whole new machine learning methodology and mechanism which will effectively predict the risk of diabetes for the unidentified patient and the working procedure of the new algorithm was tested on 3 different datasets to ensure the effectiveness. Several A broad range of machine learning algorithms, data mining tools and specially designed computer guided equipment are now effectively used to analyze medical data and to reach in a medical solution for any specific diseases. In this paper, researches pave the new effective way to successfully diagnosis of disease in a most convenient, compact and more rapid way. Several and different type of customized Machine learning algorithm is now vastly used to analyze medical data and to reach in a medical solution for any specific diseases, In this paper, authors suggested a new type called "Ensemble Perceptron Algorithm (EPA)" is proposed. This profound attention marked on the algorithm because this methodology is used to utilize the classified method of Perceptron Algorithm method of unseen data by a new proposed method with the help of Boosting algorithm.[0].

In this paper, Authors divided the working principle of the proposed method into 2 consequent phases. At the session of training stage, a broad range of collected samples recognized as the training set are analyzed by the perception algorithm in the cycle of arbitrary iterations and the help of packet algorithm and the cycle of the iteration will come to and an end after identifying the best weight vector. At last, the discovered weighted vector is kept in an array for further use. Then, by analyzing the weight vector, the profound analysis, score and remarks of training sets then data and scores will be reevaluated and extensively calculated by using a described function which was discovered in the paper [0]. Based on the extensive findings on several different domains, the prime factors responsible for DM were placed according to the descending order for further use .In this paper, authors considered "positive" for those resulted values which are greater than zero and rest of the values are referred as negative. After all, the analyzed sample elements are need to be properly labeled and separated by desired divisions as per the analysis of results achieved from the tests [0]. The prime approach of The Machine learning algorithms is that it stores informational attributes of several participants for medical survey and then analyze the data heavily to prepare to construct a model. In this study, the researchers identified the key factors based on the proof of certain medical evidences and then suggested a profound relationship with diabetes and it's associated risk factors. It is expected that, The learning and relational data gathered from the proposed model can effectively be used in near future with certain modifications for medical prediction of undiagnostic patients to accurately identification of the risk of the disease.

20 i) Prediction of diabetes based on personal lifestyle indications

Diabetes Mellitus develops in the body when there are higher or uncontrolled blood glucose level exists in the blood-plasma cell for a long time. Recently, Researchers noticed that an uncontrolled level diabetes for a long period of time can cause serious health hazards including blindness, kidney and renal failure to the affected patients who do not maintain a standard pro-diabetic lifestyle. In this study, it was marked that diabetes has a keen relationship with a person's attitude, lifestyle form factors. That's why the authors of the paper greatly devoted to establish a profound and strong relationship status between diabetes and it's associated risk factors like (age, Blood pressure, sex, Body mass index, waist circumference etc) and put their emphasis to develop a model. In this study, various algorithms like a Chi-Squared Test of Independence and another data analyzing technique named the "CART" (Classification and Regression Trees) were applied to test and analyze data. To integrate this proposed model with computer based Data clustering system, the proper cross validation steps of the process needed to be performed to ensure quality. From the analysis of previous study and research work, it was identified that the people in the age margin of 45 years or above, having high blood pressure, BMI range beyond the 25 and having a common genetic history of diabetes are the most vulnerable participants to be considered and if the participants do not follow the proper diet chart or do not take proper physical exercise (minimum 40 minutes /day) having these described attributes, these group of people have the highest probability to fall in diabetes in the near future. To conduct the research work and to build a relationship model, Authors of the paper collected the primary data about various relationship parameters like (age, BMI, BP, sex, sleep time, Exercise time) from various sources by surveys, questionnaires and categorized and leveled the data in several bounds based on the research requirement. In this study, it is found that for the categorical dataset, an algorithm name "CART "prediction model performed the accuracy level of 75%.

Again, In this paper Researchers have investigated the collected datasets and found that High blood pressure and unbalanced diet habit and consumption of junk food have a deep relationship with diabetes and this assumption and profound relationship will bring a new era in healthcare diagenesis.

21 j) Diabetes prediction using Medical Data

Dr. D. Asir Antony Gnana Sin [1] in their research they presented a diabetes prediction system based on some existing algorithms like Naive Bayes (NB), function-based multilayer perceptron (MLP), decision tree-based random forests (RF). Some specified and custom techniques as well as some well specified algorithms were used to find out a brand new and effective concept of new machine learning techniques and learning to bring out a whole new process of diagnosis of diabetes in advance. Then this model was tested with different testing methods such as 10-fold cross validation (FCV) and furthermore use percentage split with 66% (PS), and use training dataset (UTD) to check the accuracy of the system. Some effective concepts-processing techniques were used by the authors to increase the overall prediction precision level of the proposed model. They concluded that the pre-processing technique produces better average accuracy for NB compared to other machine learning algorithm. They gave the diabetes datasets into the machine algorithm (NB, RF, MLP) and noted the accuracy with different test methods (FCV, UTD, PS). Then for removing the irrelevant feature through the preprocess the dataset is given into the correlation-based feature selection. This is a looping process. They used WEKA software and collected datasets from University of California, Irvine (UCI) machine learning repository.

This proposed approach and the learnings from the study will definitely bring a new revolution and brand new effective strategy to adopt with next generation diagenesis and prediction modules to predict and fight against diseases at an early stage.

22 k) Prediction on Diabetes Using Data mining Approach

Pardha Repalli et al. [2] in their research they predict how likely the different group of aged people are being affected with diabetes based on their life style and for finding out factors responsible for the individual to be diabetic. In this paper, authors considered some statistical datasets and information. Based on the learnings from the datasets, some specialized data sorting techniques were used based on demand in order to understand which group of aged people are being affected by this disease.

To establish a structure of the model and to find the co-relative factors , two algorithmic techniques were used to predict accurately .They are i) binary target variable decision trees and ii) regression models . The best model is selected by running multiple models such as step wise regression , forward regression , back ward regression, decision tree with entropy. They have used the dataset of 50784 records with 37 variables.

Variable selection method was used by the Researchers to identify the target (input) variables for the study. High Blood Pressure, Cholesterol Last check, Heart disease, Los all teeth, Years Education etc are important input variables to predict the binary target variable. In this paper, Researcher used the parameter: age both as nominal and quotative variable. By considering various different attributes like young age, middle age and old age, authors divided and placed them in 3 separate categories. People with age above 45 years mostly affected by diabetes, they concluded. Moreover they are suggested to visit for regular checkup, dental checkup and cholesterol checkup frequently in order to control the diabetes. They also suggested young and middle age people for visiting clinic in order to check whether they have diabetes or not. Age, High blood pressure, last cholesterol check, adult BMI, Last flu shot and heart attack are the factors that also responsible for the individual to be diabetic.

23 l) Predictive Analysis of Diabetic Patient Data using Machine Learning and Hadoop

Diabetes Mellitus generally referred to as Diabetes is one of the form of Non Communicable Diseases. Diabetes is so critical that it forms a long time complication situation associated with other types of diseases. For this purpose a wise and definite way have to be found to reduce the overall impact related to diabetes by doing early prediction of Diabetes patients history that can be datasets related to diabetes patients.

This paper proposed a systematic way that consists of machine learning and datasets analysis procedure includes Hadoop and map reduce approach. This methods are used to analyze the huge amount of datasets and find a pattern matching for it and also implements the missing data during analysis of data and this procedure is followed for predictive analysis. For machine learning purpose supervised machine learning approach is followed-Supervised machine learning is an approach where the overall input types and what sorts of output can be generated or what sort of output can be produced in any of the cases is previously known. For this approach it uses its previous datasets or past experiences to trained up itself and provides an expected result.

Hadoop or Apache Hadoop is one of the open source framework which forms a computer cluster in a distributed way and it is massively used for analyzing massive amount of data in a very easy and less amount of time. For analysis and processing of further data map reducing technique is followed ,it is a way of processing data in a more reliable manner i.e this framework has the capability of processing data in a parallel and distributed way. And it is done in two phases-Firstly it will take input of data (map phase) and will convert it into intermediate data in the form of key value pairs and the next phase is the reduce phase where, by integrating and analysis of all the key values from map phase it is converted to final output.

One of the vital and major factor that is used data analysis is all the attributes that are present in datasets and used for analyzing and results obtained is used for predicting the future risk. During the dataset analysis one of the major factor that causing problem is the values that are missing of any one of the attribute i.e null values that can cause serious affects on results. So to overcome this situation classification clustering is used and by using this technique missing values are replaced with their attribute mean. For this Missing Value Imputation (MVI) algorithm is used by them. This algorithm firstly identify missing values from all attributes and then for each attribute It calculates the attribute mean. Afterwards it impute missing values in dataset with attribute mean and finally it combines missing values and datasets to produce the final result.

24 m) Application of Data Mining Methods in Diabetes Prediction

Medical field refers and deals with accuracy. Without accuracy in this field it can cause serious negative effects on patient.

This paper refers that early diabetes prediction can be done through the use of 5 types of Data mining techniques-GMM, SVM, Logistic regression, Elm and ANN. Among the mentioned techniques ANN (Artificial neural Networks) gives the highest accuracy rate and that result is much more closer to the actual result.ANN is a method where it's consist of multiple layers or a cubical design, here the single path traverses its way from front to back and this helps in resetting weights on the frontal neural units. ANN includes Layers and network functions. The ANN consist of or configured of three layers namely-input, hidden and output. Firstly the input layer or neuron defines all the inputs that will be given and this inputs are non other than all the attributes of the datasets. According to the paper they have used 7 attributes so their neurons is also 7. Hidden layers receives inputs from input layer and provides output to output layer. The most important work of hidden neuron is, it assigns a weight for the input neurons and this assigned weights shows the relevance and importance of particular and specific input to hidden neurons. Mathematically it can be defined as a neurons network function f(X) is a combination and composition of other function gi(x) and this can be again defined as composition of some other function. The most widely composition is the non linear weighted sum where f(x)=k(?i wi gi(x)) where K is the activation function i.e it's a predefined function .The activation function provides a small out change when a small change is made in the input. In this paper they have used ANN to predict the diabetes and the result is 0.89 which is closer to actual result and this result is obtained when the hidden layer number is 2 and hidden neuron is 5. That is it is found that by using ANN method it gives highest accuracy rate of 89%.

25 n) A Clinical Perspective

Diabetes is one of the common type of diseases where the blood sugar level in body become immensely high it generally of two types namely type 1 and other one is type 2.

Type 1: Type 1 is a kind of diabetes in where it is a discontinuation or disorder of glucose regulation and it is characterized by autoimmune destruction of the pancreatic beta cells that produces insulin and it leads to hyperglycemia and it have higher tendency to ketoacidosis. It is more general and seen in among children but in many case it may appear at any age. Genetic marker and the presence of antibodies can assist to identify diabetes. Antibody markers of autoimmunity that is against beta cell includes autoantibodies islet-cell and autoantibodies against insulin, decarboxylase, glutamic acid or tyrosine phosphates IA-2 and IA-2?, and ZnT8.3.Containing at least one or more than one of this are present during fasting hyperglycemia it was initially detected in persons where 85% to 90% of people can eventually contain or may develop type 1 diabetes. It is found that some patients and mostly children and adolescents contains ketoacidosis as the first symptom of this disease. In less common cases and typically in older patients, it can present with the mild fasting hyperglycemia or diminished glucose level tolerance. T1 diabetes is not a linear progression disease but it progress at a variable pace in different patients. Symptoms and sign including higher level insulin deficiency and hyperglycemia include polydipsia, fatigue, weight loss, polyphagia and polyuria. This are causing defective transport of glucose from the blood vessel/stream into body tissues and it results in increased glucose levels in the blood and moreover it elevates glucose in the urine and concomitant calorie and fluid losses with the urine. For this when insulin level falls down to such a low level lipolysis cannot be able to suppressed and products containing fat metabolism naming ketone bodies is accumulated in the blood and due to hyperventilation it leads to metabolic acidosis and compensatory respiratory alkalosis.

26 o) Application of Data Mining Methods in Diabetes Prediction

In any sort of medical field the most important factor is all about accuracy. Without accuracy in this field it can cause serious negative effects on patient. So accuracy is the most important factor.

According to this paper early prediction of diabetes is made through the use of 5 types of Data mining techniques-GMM, SVM, Logistic regression, Elm and ANN. Among all the five techniques ANN (Artificial neural Networks) provides the highest rate of accuracy ANN is a method where it's consisted of multiple layers or a cubical design, here the single path traverses its way from front to back and this helps in resetting weights on the frontal neural units. ANN includes Layers and network functions. The layers are-Input layer, hidden layer, output layer. The input layer or neuron defines all the inputs that will be given and this inputs are non-other than all the attributes of the datasets. According to the paper they have used 7 attributes so their neurons is also 7 Hidden layers receives inputs from input layer and provides output to output layer. The most important work of hidden neuron is, it assigns a weight for the input neurons and this assigned weight shows the relevance and importance of particular and specific input to hidden neurons. Mathematically it can be defined as a neurons network function f(X) is a combination and composition of other function (x) and this can be again defined as composition of some other function. The most widely composition is the non linear weighted sum where f(x)=k(?i wi gi(x)) where K is the activation function i.e it's a predefined function. The most help and useful characteristic of this activation function is that it provides a small out change when a small change is made in the input. In this paper they have used ANN to predict the diabetes and the result that was assuming to be the best is 0.89 and it is obtained when the hidden layer number is 2 and hidden neuron is 5.That is it is found that by using ANN method it gives highest accuracy rate of 89%.

27 p) Blood pressure and ageing

Increase in blood pressure with the increasing of age can of many varied factors and it is also depended on many cases like lifestyle and living environment of different person. BP seems to be rise or fall with age. It is of two types systolic and diastolic blood pressure in short SBP and DBP. With the increase of age the blood pressure is associated mostly with the changes relating with arteries, large artery stiffens and also with increase of risk related to cardiovascular the blood pressure also rises. In case of aged person with the effect of increase of systolic and decrease of diastolic pressure related to blood there causes a risk of increasing pulse pressure that consequences in blood pressure. SBP dramatically and continuously starts to increase between the age of 30>above and in case of DBP it does not show a continuous pattern but it varies with age until fifth decade it starts to rise but suddenly starts falling at the age of 60-84. According to this paper a definite level of age is chosen for identifying the BP, in case if it is classified within different range of ages it would be much more easier to identify the provable causes of increasing or decreasing of BP.

28 III. Diabetic Patient Data Management and Support System a) Introduction

The primary process of the research was to determine the principle co-relating factors and their contribution and impact toward the diabetes. To conduct the research, previous learning and knowledge base of previous health reports were considered to reach in a decision. Health information and datasets are collected from various different sources like direct questionnaires, results of conducted online surveys, previously available datasets and available heath samples of diabetes patients on various health portals and recognized health journals. Samples and essential information or heath data based on several attributes were collected from different sources from the available information of more than 450 participants of various health surveys and questionnaires. Then all the necessary information and parameters were carefully sorted and selected. After extensive sorting and filtering incomplete, less trustworthy and irrelevant information were discarded. After all, relevant information of 300 participants collected from various sources from the time period of (2011-2019) years were placed and stored in a dataset for this research purpose. This dataset was the primary information source of this research. Some principal attributes were taken into consideration. The prime attributes are age, gender, Blood pressure, height, weight, BMI, sleeping time and exercise time of each and individual patient.

The output of the research work is to build a sustainable model which is essential to predict diabetes with highest precision and detect the chance of getting diabetes in near future. This system will also suggest the optimum lifestyle and exercise suggestion to the participants

29 b) Diabetes patient data analysis model

The performance evaluation of a health model broadly dependent on four variables. They are Participant's real time health information, Participant's food habit, Participant's exercise sheet, Participant's medical feedback. A proper health supervision for a diabetic patient is provided by this model as this model is capable of predicting the risk of diabetes in advance and it will help the upcoming diabetes victim by providing advance alert to them. In this paper, the attributes like BMI, height, weight, sleeping time and working hours or weekly bases exercise time, blood pressure were identified as the prime reacting factors. These attributes are the dominant factor.

Calorie intake and exercise time are also important factor for the diabetes.

Proper management of diet system and medication can treat and manage diabetes in proper way.

The patient containing extensive blood sugar may be suggested to take insulin. This approach will help to manage the proper health status and control the weight, blood sugar and calorie consumption for specific patients.

30 c) System Architecture

Modeling of system is consisting of designing the system, processing the system architecture and integrate the proper modules and interfaces based on system requirements.

The approach and process is divided in several consequent steps. First of all, a diabetes dataset is carefully prepared and then proceed the dataset as input to the specified system analyze the data with exact precision. Then, this system is designed to perform in ready state to analyze based on input data. A well specified model and properly guided mathematical equation with proper optimization of the backend calculative format is placed in the backend of the system to analyze data. Certain terms and conditions are also set in the system to work efficiently. After taking input data from the participants, then the system measures the input data based on the developed mathematical equation. At-last, the system provides a prediction with a precise risk estimate in percentage for each individual patient. Then, this system will provide the optimum exercise goal and lifestyle for each and every individual patient. The medical experts and data scientists can use this prediction for further improved diagnosis process. However, the system is quite accurate to analyze data and to predict data for each and every individual patients. Moreover, the effectiveness of the proposed system can further be improved by.

31 Fig. 3.1: System Architecture d) Integrated Database Design

To conduct this research and to prepare the dataset, some conditions were taken into consideration. These conditions and research terms were carefully selected based on the experimental approaches and previous leanings of related research works.

In this Integrated datasheet completely emphasis on various health parameters of the patients. Again, this dataset provides a minimalistic idea about the lifestyle of the participants based on the analysis of various different parameters. By this approach, it is possible to identify the probability of diabetes at an early age.

To prepare the model realistic data was set and higher and lower bound values were carefully selected based on realistic data set, web source and medical fact data sheet. The values are carefully analyzed and not a single input in this range is out of the topic, it was confirmed that the participants having the fasting blood sugar range below the 3 mmol/L have the lowest.

? Possibility to fall in diabetes in near future. The Fasting blood sugar range from 3 mmol/L to 10mmol/L was taken into consideration for this system.

? Participants Blood Sugar (Random): Participant's blood sugar after 2 hours phase is likely an important indicator .The Random blood sugar range from 5 mmol/L to 30mmol/L was taken into consideration for this system. The random sugar should be noted with highest professionalism because any malfunctioned result or data input will change the whole result of prediction probability .The random blood sugar range above 10mmol/L is a serious indicator of getting diabetes. ? Participants Blood Pressure: From the analysis of several medical studies, scientists have found a significant connection of Participant's blood pressure with chance of patient's getting diabetes. The normal range of Systolic blood pressure is less than 120 mmHg and Diastolic blood pressure is less than 80 mmHg. ? Participants Sleeping-time: In recent studies, health scientists have found specific link to sleeping hour with the probability of getting the chance of diabetes. From the analytical reasoning of the dataset, it was found that balanced sleeping time has a inverse-proportional relationship with diabetes. The participants sleeping time were counted in hours on weekly basics. ? Participants Exercise-time: Exercise is the key factor to control the glucose level of the blood. Optimum exercise plan can significantly lower the blood glucose level and chances to get attacked by diabetes in near future. So, Exercise time has a inverse-proportional relationship with the blood glucose level. The participants working or exercise time were counted in hours on weekly basis.

In our database design we have generated following attributes. The generation process has been discussed below: ? Participant's Age: Patient's age is one of prime factor for this study. From the analysis of the dataset and the previous learning suggested that age has a very close relationship with diabetes. From the analysis it was observed that the people of age range belongs to 40 years to 60 years [5] have the highest risk to be get attacked by diabetes and the people of age below 40 years and above 65 years have comparatively the lower risk percentage. ? Participant's BMI: Body Mass Index(BMI) is an important indicator of Health index. To calculate BMI it is needed to collect Height and weight of individual patients.

To calculate BMI it was needed to record height and weight of each and every individual participants. In this system, it was considered the existing "Guinness world records" fact book to find out the tallest and smallest heighted people's height to set the lower and upper bound of the height for the model. Though most of the participants belonged to the height range of 5 feet 3 inch to 5 feet 11inch range.

The generalized formula to calculate BMI: BMI=Weight / (Height) 2

Where Weight is calculated in Kilo-gram (Kg) and Height is calculated in Meter(m). That's why, taller patients with moderate weight have likely to face less risk of getting diabetes than the shorter participants with moderate weight. To calculate BMI, it was needed to collect weight and height of the participants. In this study, it was considered the BMI range from 10 kg/m2 to 50 kg/m2 .Where the participants having the BMI range of 18.5 kg/m2 to 25 kg/m2 are considered to be healthy and participants having BMI above 30 kg/m2 are at a risk of getting diabetes in near future.

? Participant's Gender: Participants gender is a related factor to estimate the risk for individual patients. Patients gender is a discriminating factor for the analysis of dataset. Female patients have different type of diabetes characteristics and many women fall in temporary diabetes which is called gastrointestinal diabetes. So, Data was collected from both Male and Female participants.

Participants Blood Sugar (Fasting): Participant's blood sugar at fasting phase is likely an important indicator .It is one of the prime concern for the analyzing diabetes because patients having higher blood sugar in fasting phase likely to fall in diabetes in most of the times. By analyzing the dataset and previous study In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Age .In this dataset what was used a primary data source for the research , the Age range was between the range of 1year to 80 years of a different groups of male and female. Blue color plotted line is representing the Age (attribute). As per the information of dataset, Age started from the numerical value of 10 years old and finished at the ending point of 80 years. In this graph, Age was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 13.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that age has a proportional relationship with test_result(negative/positive).Diabetes risk occurrence is heavily linked with the age. The persons /participants under the age of 26 years have the lowest risk probability and the age range between 26-40 years have the lower possibility. The age group of above 45 years old people have the highest risk of diabetes occurrences.

From the analysis of the graph and previous studies , it is confirmed that older people have the higher risk of diabetes occurrences. Diabetes Risk Occurrence ? participant's Age. Diabetes Risk Occurrence = K 1 *participant's Age (4.1) Where k 1 is a constant.

32 Body Mass Index (BMI):

33 Fig. 3.3: Body Mass Index (BMI) vs Surveillance participants diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Body Bass Index(BMI) .In this dataset what was used a primary data source for the research , the Age range was between the range of 16.5 kg/m 2 to 42 16.5 kg/m 2 of a different groups of male and female. Blue color plotted line is representing the BMI (attribute). As per the information of dataset, Age started from the numerical value of kg/m 2 and finished at the ending point of 42 kg/m 2 .In this graph, BMI was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "13.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value "7.0" (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that BMI has a proportional relationship with the surveillance participants diabetes status test_result (negative/ positive). Diabetes risk occurrence is heavily linked with with BMI. The persons /participants under the BMI of 21 kg/m 2 have the lowest risk probability and the BMI range between 22 kg/m 2 -28 years kg/m 2 have the moderate risks. Having the BMI above 31 kg/m 2 , such group people have the highest probability of risks of diabetes occurrences. From the analysis of the graph and previous studies, it is confirmed that people having the BMI >32 kg/m 2 people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ? participant's BMI (body mass index).Diabetes Risk Occurrence = K 2 * participant's BMI (4.2)

Where k 2 is a constant Blood Sugar (Fasting): In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Blood Sugar (Fasting) .In this dataset what was used a primary data source for the research , the Blood Sugar (Fasting) range was between the range of 3.9 mmol/L to 9.1 mmol/L of a different groups of male and female. Blue color plotted line is representing the Blood Sugar Fasting (attribute). As per the information of dataset, Blood Sugar Fasting started from the numerical value of 3.9 mmol/L and finished at the ending point of 9.1 mmol/L.In this graph, Blood Sugar Fasting was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "13.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Blood Sugar Fasting has a proportional relationship with test_result (negative/positive). Diabetes risk occurrence is seriously linked with with Blood Sugar Fasting. The persons /participants under the Blood Sugar (Fasting) range of 4.2mmol/L have the lowest risk probability and the Blood Sugar Fasting range between 4.2mmol/L to 5.1mmol/L have the moderate risks.

From the analysis of the graph and previous studies, it is confirmed that people having the Blood Sugar Fasting >6.0 mmol/L people have the higher risk of diabetes occurrences. Diabetes Risk Occurrence ? participant's Blood Sugar (Fasting).

Diabetes Risk Occurrence = K 3 * participant's Blood Sugar (Fasting) **??**4.3) Where k 3 is a constant Blood Sugar (Random): In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Blood Sugar (Random). In this dataset what was used a primary data source for the research, the Blood Sugar (Random) range was between the range of 4.1 mmol/L to 18.5mmol/L of a different groups of male and female. Blue color plotted line is representing the Blood Sugar Random (attribute). As per the information of dataset, Blood Sugar Random started from the numerical value of 4.1 mmol/L and finished at the ending point of 18.5 mmol/L.In this graph, Blood Sugar Random was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 13.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Blood Sugar (Random) has a proportional relationship with the surveillance participants diabetes test_result(negative/positive).Diabetes risk occurrence is seriously linked with with Blood Sugar (Random). The persons/participants under the Blood Sugar (Random) range of 4.4mmol/L have the lowest risk probability and the Blood Sugar Fasting range between 5.4 mmol/L to 6.1mmol/L have the moderate risks. From the analysis of the graph and previous studies, it is confirmed that people having the Blood Sugar (Random) > 7.0 mmol/L people have the higher risk of diabetes occurrences. In this graph ,it represents the risk of diabetes occurrence was compared with the relational attribute Sleeping time. In this dataset what was used a primary data source for the research, the Sleeping time range was between the range of 4 hours/day to 14hours/day of a different groups of male and female. Blue color plotted line is representing the Sleeping time/day (attribute). As per the information of dataset, Sleeping time started from the numerical value of 4 hours/day and finished at the ending point of 14 hours/day .In this graph, Sleeping time was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 5.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value "3.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph , it is clear that Sleeping time has a inversly proportional relationship with the surveillance participants diabetes status test_result (negative/positive). Diabetes risk occurrence is seriously linked with Sleeping time. The persons /participants having sleeping time range between 6 hours/day to 8 hours/day have the lowest risk probability.

34 Diabetes

From the analysis of the graph and previous studies, it is confirmed that people having the Sleeping time > 8 hours/day for a week & Sleeping time < 5.5 hours/day for a week people have the higher risk of diabetes occurrences. Step 4: passing the converted values into desired variables;

Step 5: starting of calculation by using the passed variables values into the derived equation

Step 6: Analysis Report or Result is received.

Step 7: Comparing the predicted calculation with predefined sets of terms and conditions Original result is equal to multiplication of (Constant value , Age, BMI, fasting sugar level, random sugar level, systolic and diastolic bp) which is divided by multiplication of sleep and exercise time. Step 8: Displaying the predicted result and risk evaluation to the user

Step 9: Ready for further analysis of different inputs IV.

35 Result and Analysis

For checking the Diabetes occurrence percentage rate we have used a computer programmed system which is developed according to our mathematical model. All the required attributes that we are using are taken in consideration for giving input into the system and from that we get our diabetes occurrence percentage rate. For the overall procedure 28 sets of data are given input into the system starting from age 11-77 yrs. Afterwards by using the acquired occurrence percentage rate for every individual sets of data, graphs are prepared. The graphs show the comparative analysis of diabetes occurrence rate with individual attributes. Here for every individual graph Blood sugar level(random time) and Blood sugar level(fasting time) are taken in consideration because this two attributes contributes the most crucial part for occurrence rate change because with a small change in these attributes overall occurrence rate changes at a higher or lower rate. From the above Exercise time (min) VS Occurrence percentage rate graph it is found that with higher exercise time and blood sugar level (random/fasting) less than 7 have a less chance of diabetes occurrence and from the graph it is also found when exercise time is 40 min and blood sugar level above 7 have a occurrence rate of 55 % and again with a decrease of exercise time to 20 min with a similar blood sugar level the chance of occurrence increases by 5%. So with less exercise time and having higher blood sugar level increases the chance of diabetes occurrence. For general cases sleeping time is inversely proportional to occurrence rate that is with the increase of sleeping time diabetes occurrence chance will be decreased.

From the above graph it is seen that for sleeping time of 11hrs the diabetes occurrence rate is 4.45% but that occurs if the blood sugar level is less that if below 7. But if the blood sugar level is high sleeping time have a very little effect on the overall occurrence rate as it is found from the graph when sleeping time is 10hrs and blood sugar level(random) is 17 and blood sugar level(fasting) period is 9 in that case it has the highest chance of diabetes occurrence. In this section an overview of our developed system is shown. There are eight input fields for the user. All the inputs will be in standard unit of these parameters. Users will input their data in the text fields in proper formatting. After clicking the check result button the front end will collect the form data and will place them in some variables and it will pass all variables to the back-end system to analyze data and generating a result. In the back-end the equation equates values and calculate the occurrence percentages of diabetes based on the mathematical model. Figure **??**1 shows the User Interface developed in this research to identify the occurrence of Diabetes in human body.

V.

36 Conclusion a) Summary of the thesis

The primary goal of the developed model is to identify the occurrence rate of diabetes at an early stage with highest precision. Therefore, to identify the crucial factors for the thesis work a largesets of attributes were taken into consideration and after extensive analysis and scientific evaluations between the attributes, some attributes were finally selected to establish a scientific based mathematical equation which is combining all the terms, co-relations and all factors in a single mathematical equation for better and fast predictability. Using machine learning and data analysis techniques, it was established that the prediction score from the developed model matches closely with previous results. The model will provide valuable result and it will be helpful to identify diabetes occurrence rate with a less amount of diagnosis time and lowest cost consumption. Though the system has some error tolerance issue but after successful experimental and testing phase, the quality of data analyzing model and software system got better and became more reliable for accurate prediction.

37 b) Findings of the thesis

To conduct the research work , a huge number of case studies were analyzed and 50 more related journals, health science articles, analytics, survey reports were thoroughly studied to find out the actual reason of diabetes occurrences in human body and in this paper 8 co-relational actors were indicated and their co-relationship , bindings and contribution towards the diabetes was identified and marked .Then a complete mathematical term is established based on the previous knowledge, analytical attributes synthesis and based on mathematical terminology. Established mathematical equation and concepts were combined in a single equation with a universal constant formatting All the mathematical terms were reverified in several techniques like plotting different attributes in graph to identify the correct relationship. A dataset of 250 participants of different age, groups and communities were selected for the case study and testing of the developed system. From this study it was confirmed that age is the most dominant factor and then random blood sugar level is a clear indicator of the diabetes status or diabetes level .All the attributes studied in this research like age, BMI, blood sugar, blood pressure, working and sleeping time have some contributions on diabetes risk score. A person can easily minimize the risk score by adjusting his/her life status, daily habits, food-calories intake and scaling an ideal exercise or sleeping time. Though diabetes is not preventable but the blood sugar level of any patient can easily be maintained in ideal level by inducting an ideal food, diet chart, balanced sleep and working hours and a good quality of life. The risk of diabetes will be optimized by an ideal lifestyle recommended by health nutrition experts and medical professionals.

38 ( ) G

Year 2021 c) Future Scope of the thesis In this developed model, an estimated compulsion proportion between all the attributes were selected and all the attributes consist of same weighted values. Some attributes like Age and working time are primary deal breaking factors but in this work , genetics property of diabetes was not considered due to lack of proper evidence ,lack of previous studies .In future work, genetic inheritance factor will be considered for further detailed analysis .In this study, a software system is developed with manual input checking and it shows the output of risk percentage .In future work, a complete data book for every patient will be added. Interface of the computer system will be further modified. Social media's add-ons can also be added so that the system can easily fetch user data from social account for further analysis with less user input, which will become more user friendly .Our system can be also integrated with other health monitoring devices like smart watches like Apple Watch 3 or others which will be very effective to sync user data in real time basics and to store a portfolio for the patients .This system will be ready to sync data from other input sources, health devices and generate results based on the users input .Then the results will also be sent to added IoT gadgets for better health management. In the next edition our software will predict with more precision and accuracy with the extended use of IOT connected devices which will help patients to maintain an optimal lifestyle and balanced diet. In future edition, our developed software and ecosystem will also provide a better health analytic and better health management system.

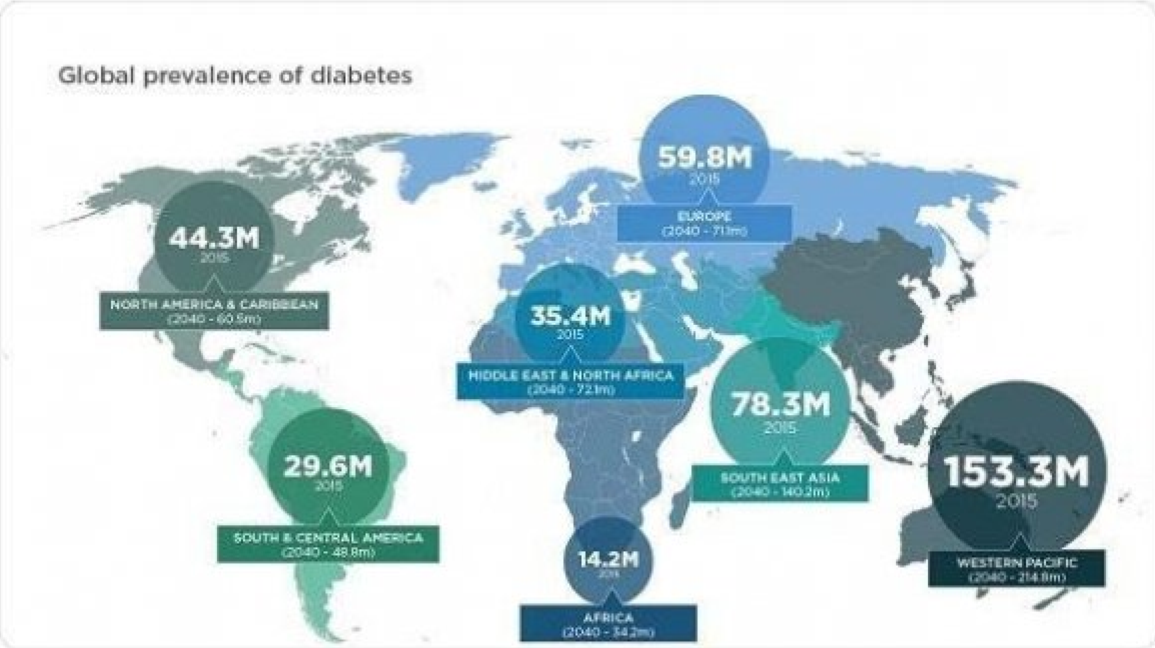
32 

Figure 1: Fig. 3 . 2 :

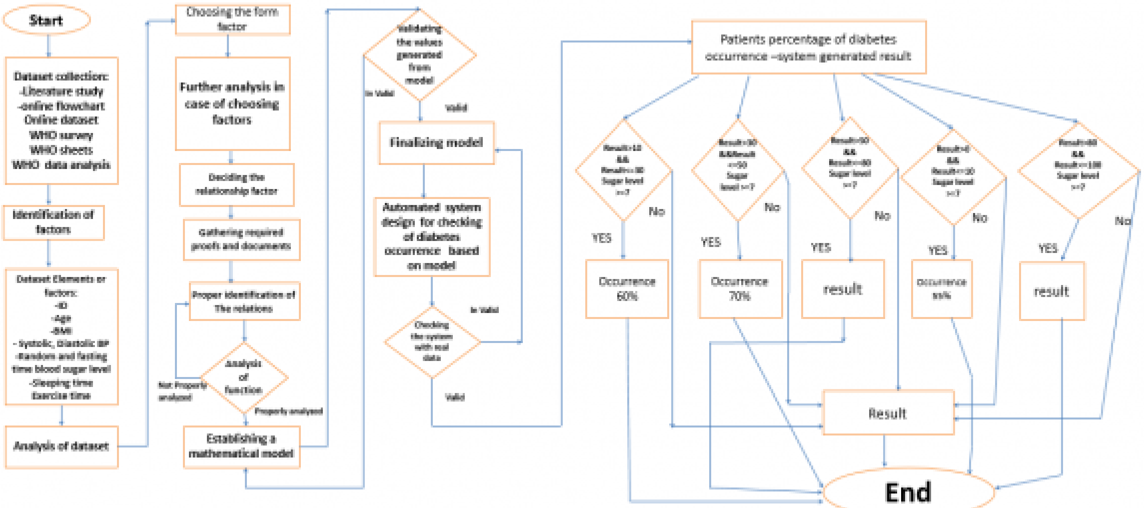
34 

Figure 2: Fig. 3 . 4 :

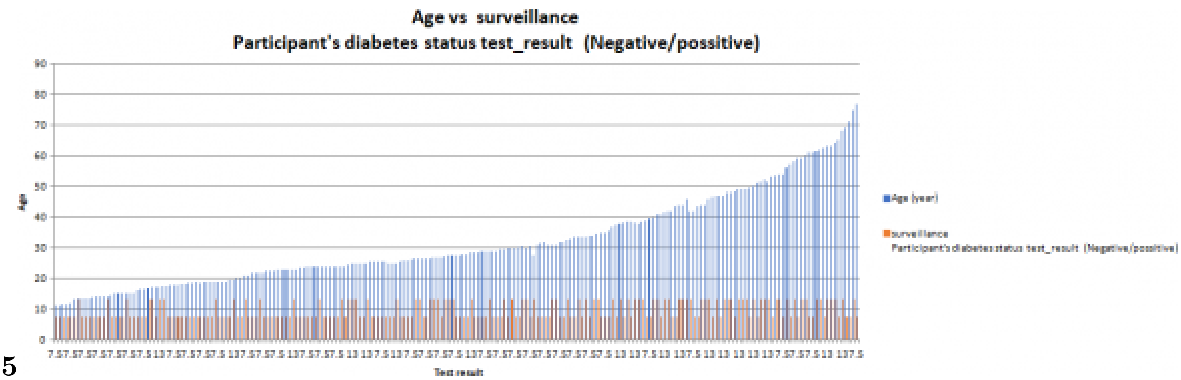
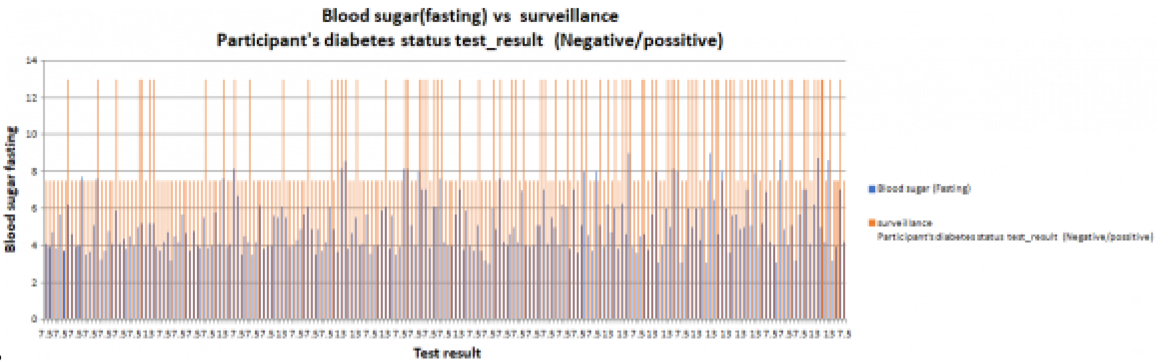
35 

Figure 3: Fig. 3 . 5 :



3638

Figure 4: AFig. 3 . 6 :Fig. 3 . 8 :

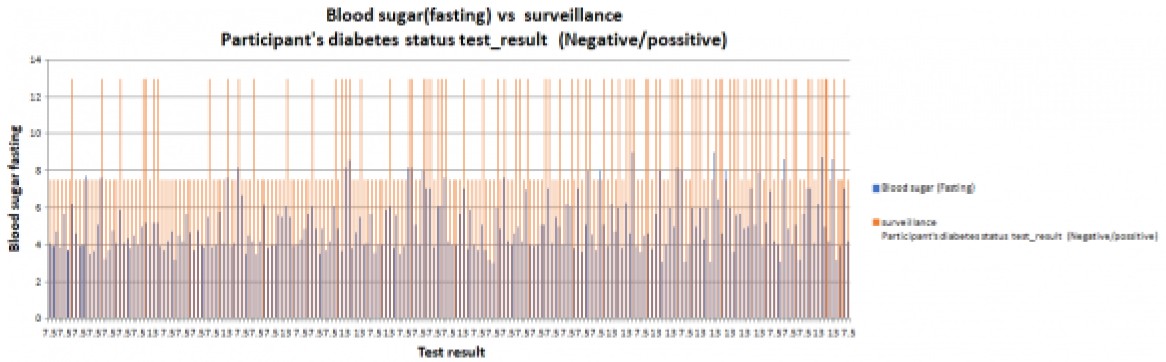
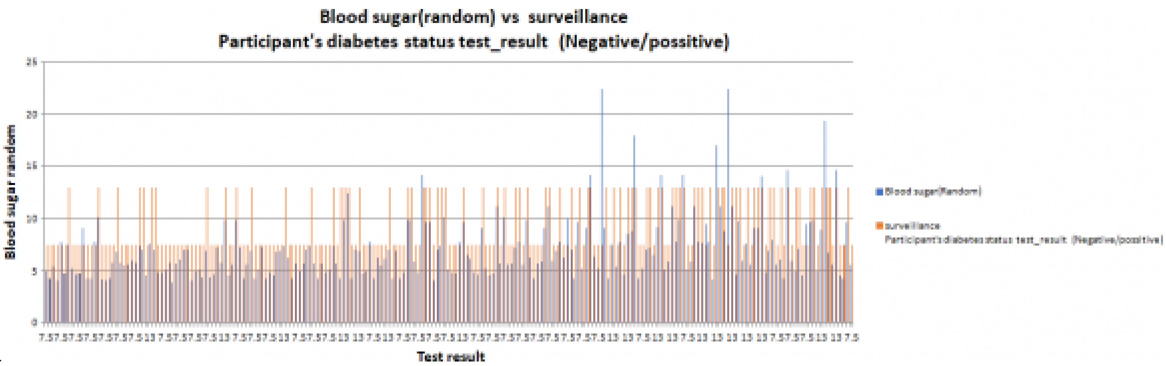
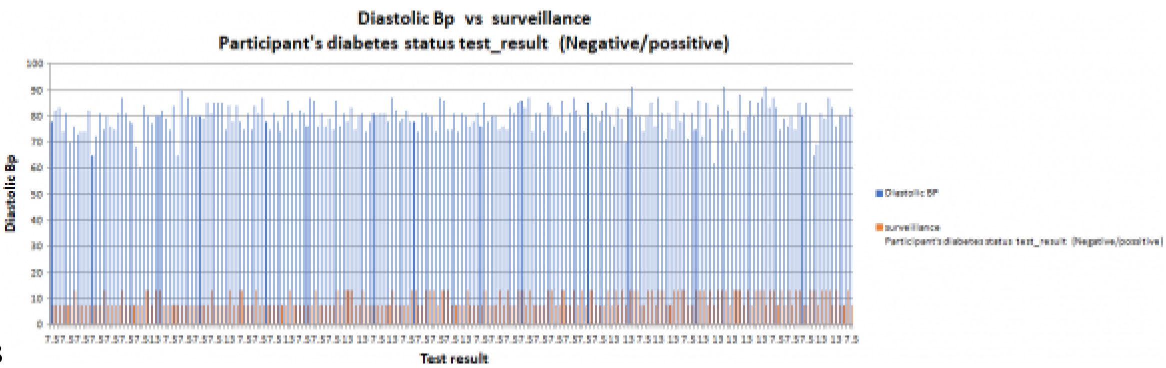


Figure 5:



41

Figure 6: Fig. 4 . 1 :



4243

Figure 7: Fig. 4 . 2 :Fig. 4 . 3 :

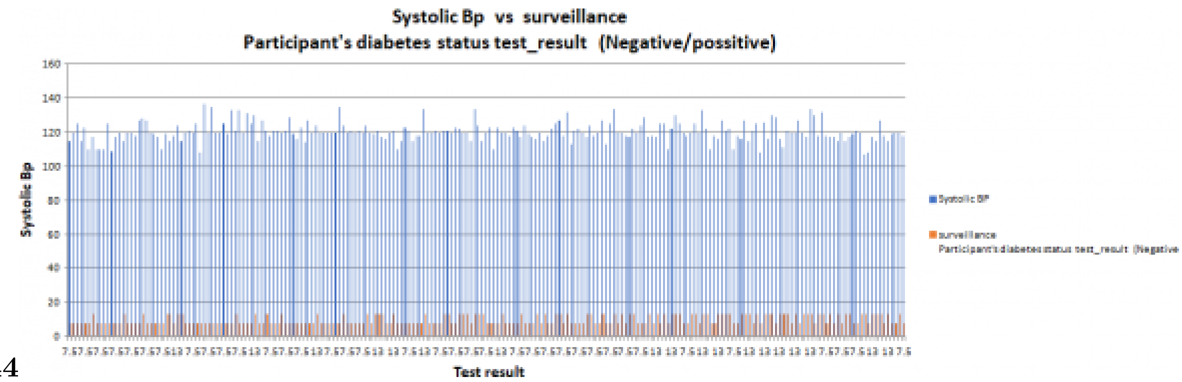
44 

Figure 8: Fig. 4 . 4 :

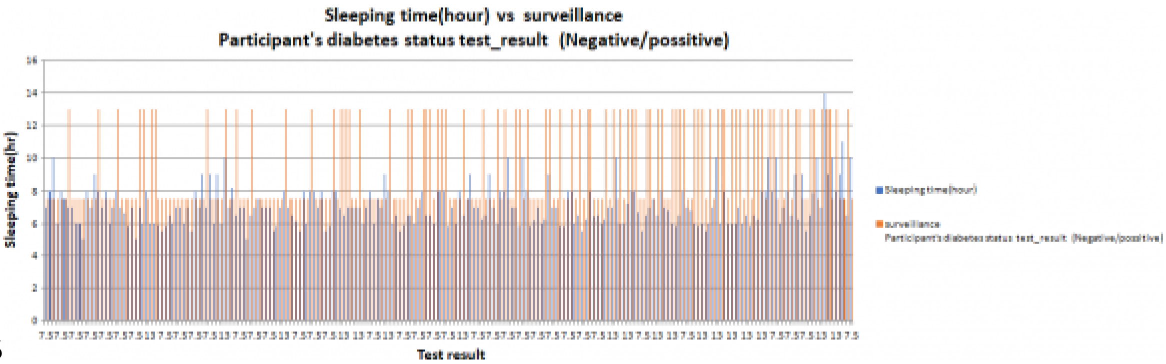
45 

Figure 9: Fig. 4 . 5 :

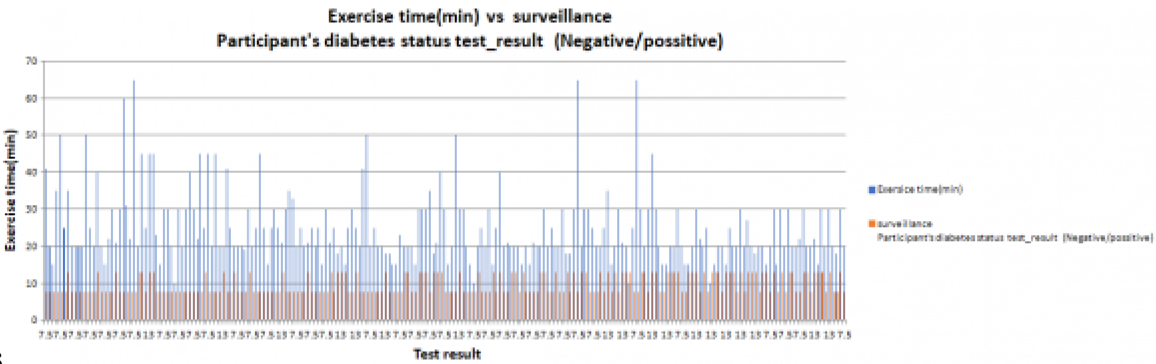
46 

Figure 10: Fig. 4 . 6 :

47 $$\frac{K*a*b*c*d*e*f}{g*h}$$

Figure 11: Fig. 4 . 7 :

48 $$x = \frac{a*b*c*d*e*f}{2a}$$

Figure 12: Fig. 4 . 8 :

$$\frac{Age*BMI*Systolic\ Bp*Diastolic\ Bp*Blood\ sugar\ level(fasting\ period)*Blood\ sugarlevel(random\ period)}{Exercise\ Time*sleeping\ Time}$$

Figure 13:

$$\frac{\text{Occurrence}*\text{Exercise time}*\text{sleeping time}}{\text{Age}*\text{BMI}*\text{Systolic Bp}*\text{Diastolic Bp}*\text{Blood Sugar level(Fasting period)}*\text{Blood Sugar level(random period)}}$$

Figure 14:

$$\frac{100*16200*900}{1.41912*109*37*17331.6*11998.98*(6*10^{-3})*(8*10^{-3})}$$

Figure 15:

$$\frac{K*a*b*c*d*e*f}{g*h}$$

Figure 16:

1

Attribute Name	Lower bound	Upper bound
Age	1 year	123 years
BMI	10 kg/m2	50 kg/m2
Blood Sugar (Fasting)	3 mmol/L	10 mmol/L
Random Blood Sugar	5 mmol/L	30 mmol/L
Systolic Blood pressure	70 mmHg	190 mmHg

Figure 17: Table 1 A

If Random sugar level greater 0.007
Then Output 60 percent
Else output original result
2. Else if original result greater 30 and original result less than or equal to 50
If Random sugar level greater 0.007
Then Output 70 percent
Else output original result
3. Else if original result greater 50 and original result less than or equal to 80
then, Output original result
4. Else if original result greater 0 and original result less than or equal to 10 If Random sugar level greater 0.007 Then, Output 55 percent
Else output original result
5. Else if original result less than 0
If Random sugar level greater 0.007
then Output 51 percent
Else output 0.0001 percent
6. Else if original result greater 80 and original result less than or equal to 100
then, output original result
Else if original result greater 100
output 100 percent
else
Output "invalid input";

© 2021 Global Journals

Figure 18:

Exercise Time: In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Exercise time .In this dataset what was used a primary data source for the research, the Exercise time range was between the range of 10 min/day to 100 min /day of a different groups of male and female. Blue color plotted line is representing the Exercise time/day (attribute). As per the information of dataset, Exercise time started from the numerical value of 10 min/day and finished at the ending point of 100 min/day .In this graph, Exercise time was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 8.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value "3.50 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Exercise time has a inversly proportional relationship with the surveillance participants diabetes status test_result (negative/positive). Diabetes risk occurrence is seriously /day to 60 min/day have the lowest risk probability .

From the analysis of the graph and previous studies, it is confirmed that people having the Exercise time < 20 min/day for a week have the higher risk of diabetes occurrences. Where k is a constant. By considering all the medical datasheets and references it was confirmed that diabetes risk for certain factor has a certain upper bound and risk doesn't increase exceeding that bound.

.1 Considering The worst case scenario

[Priyadarshini et al. ()] 'A Novel approach to Predict Diabetes Mellitus using Modified Extreme Learning Machine, 1 edn'. Jrojalina Priyadarshini , Nilamadhab Dash , Rachita Mishra . *IEEE Xplorer: IEEE Xplorer*, 2014.

[Katzmarzyk et al. (2007)] 'Adiposity, physical fitness and incident diabetes: the physical activity longitudinal study'. P T Katzmarzyk , C L Craig , L Gauvin . *Diabetologia* Mar. 2007. 50 (3) p. .

[Berber et al. (2001)] 'Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population'. A Berber , R Gómez-Santos , G Fanghänel , L Sánchez-Reyes . *Int. J. Obes. Relat. Metab. Disord* Dec. 2001. 25 (12) p. .

[Rajesh and Sangeetha ()] 'Application of data mining methods and techniques for diabetes diagnosis'. K Rajesh , V Sangeetha . *International Journal of Engineering and Innovative Technology (IJEIT)* 2012. 2 (3) p. .

[Adidela et al. ()] 'Application of fuzzy ID3 to predict diabetes'. D R Adidela , D G Lavanya , S G Jaya , A R Allam . *Intr Jrnl Adv Comp Math Sci* 2012. 3 (4) p. .

[Feng et al. (2012)] 'BMI is strongly associated with hypertension, and waist circumference is strongly associated with type 2 diabetes and dyslipidemia, in northern Chinese adults'. R N Feng , C Zhao , C Wang , Y C Niu , K Li , F C Guo , S T Li , C H Sun , Y Li . *J. Epidemiol* May 2012. 22 (4) p. .

[Ara et al. ()] 'Case Study :Integrating IoT, Streaming Analytics and Machine Learning to improve Intelligent Diabetes Management System, 1 edn'. Affreen Ara , Aftab Dr , Ara . *International Conference on Energy, Communication, Data Analytics and Soft Computing*, 2017. IEEE. (ICECDS-2017)

[Cash ()] Jill Cash . *Family Practice Guidelines*, 2014. Springer. 9780826168757 p. 396. (3rd ed.)

[T ()] 'Control and complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin independent diabetes mellitus'. T . *N Engl J Med* 1993.

[Afrandp et al. ()] 'Design and implementation of an expert clinical system for diabetes diagnosis'. Afrandp , N M Yazdani , H Moetamedzadeh , Naderif , M S Panahi . *Global Jrnl of Sci, Engg* 2322- 2441. 2012. p. . (Tech)

[Diabetes Fact sheet N°312 (2013)] *Diabetes Fact sheet N°312*, October 2013. 25 March 2014.

[Bum Ju Lee and Jong Yeol Kim (ed.) ()] *Identification of Type 2 Diabetes RisK Factors using Pheno Type Consisting of Anthropometry and Triglycerides based on Machine Learning, 1 edn*, IEEE xplorer: IEEE xplorer. Bum Ju Lee and Jong Yeol Kim (ed.) 2014.

[Snijder et al. (2004)] 'Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study'. M B Snijder , P Z Zimmet , M Visser , J M Dekker , J C Seidell , J E Shaw . *Int. J. Obes. Relat. Metab. Disord* Mar. 2004. 28 (3) p. .

[Patel ()] 'Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes'. A Patel . *The New England journal of medicine* 2008. 358 p. .

[Suzuki et al. ()] 'Mixture of expert 3D massivetraining ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography'. K Suzuki , H Yoshida , J Nappi , S G Armato , A H Dachman . *Med. Phys* 2008. 35 (2) p. .

[Webb and Zheng (2004)] 'Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques'. G I Webb , Z Zheng . *IEEE Transactions on Knowledge and Data Engineering* Aug. 2004. 16 (8) p. .

[Webb and Zheng (2004)] ‘Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques’. G I Webb , Z Zheng . *IEEE Transactions on Knowledge and Data Engineering* Aug. 2004. 16 (8) p. .

[Kumar and Pranavi ()] *Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics*, P Kumar , S Pranavi . 2017. (IEEE xplorer)

[Varshney ()] *Pervasive Healthcare Computing: EMR/EHR, Wireless and Health Monitoring*, U Varshney . 2009.

[Anand and Shakti (2015)] ‘PREDICTION OF DIABETES BASED ON PERSONAL LIFESTYLE INDICATORS, 1 edn’. Ayush Anand , Divya Shakti . *1st International Conference on Next Generation Computing Technologies (NGCT-2015)*, (Dehradun, India) September 2015. 2015.

[Saravanakumar and Sampath ()] ‘Predictive Methodology for Diabetic Data Analysis in Big Data’. Eswari T Saravanakumar , P Sampath . *Science direct* 2015. 50 p. .

[Kumar and Srivastava (2014)] ‘Some observations on the performance of segmentation algorithms for microscopic biopsy images’. R Kumar , R Srivastava . *Proceedings of the International Conference on modeling and Simulation of Diffusive Processes and Applications (ICMSDPA ’14)*, (the International Conference on modeling and Simulation of Diffusive Processes and Applications (ICMSDPA ’14)Varanasi, India) October 2014. p. . Department of Mathematics, Banaras Hindu University

[Lear et al. (2002)] ‘The relationship between waist circumference and metabolic risk factors: Cohorts of European and Chinese descent’. S A Lear , M M Chen , J J Frohlich , C L Birmingham . *Metabolism* Nov. 2002. 51 (11) p. .

[Misra et al. (2006)] ‘Waist circumference cutoff points and action levels for Asian Indians for identification of abdominal obesity’. N K Misra , R Vikram , R M Gupta , J S Pandey , V P Wasir , Gupta . *Int. J. Obes. (Lond)* Jan. 2006. 30 (1) p. .

[Xu et al. (2013)] ‘Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes’. Z Xu , X Qi , A K Dahl , W Xu . *Diabet. Med* Jun. 2013. 30 (6) p. .