

An Approach for Developing Diabetes Prediction and Recommendation System

Saima Sultana, Abdullah Al Momen, Mohoshi Haque, Mahmudul Hasan Khandaker,
Nazmus Sakib

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology, Dhaka-1208, Bangladesh

ABSTRACT

Diabetes is a severe, enduring disorder with a huge impact on the existence and health of individuals and the people around them. It happens due to insufficient production of insulin in human body. After investigating the dangers of diabetes, it can be said that diagnosing diabetes with basic medical information at early stage of diabetes can help patients to control it and also predicting the probability of having diabetes can really decrease the number of diabetic patients in future. So, for the prediction of this disease, Multiple Linear Regression (MLR) has been used. To implement this model, some basic medical information of a person have been used as parameters. 83% accuracy has been achieved using this model. A set of suggestions using the Reinforcement Learning also have been generated to help the diabetic patients to control this disease.

Keywords

Diabetes Mellitus, Prediction System, Recommendation System, Multiple Linear Regression (MLR), Simple Linear Regression, Age, Body Mass Index (BMI), Blood Pressure, Blood Sugar, Exercise Time and Sleeping Time.

1. INTRODUCTION

Nowadays healthcare monitoring data is growing immensely and it has different properties such as structured, semi-structured and unstructured. In recent days, data-science and information technology play a vital role in providing services in the area of healthcare [1].

Diabetes Mellitus known as diabetes is an organ disorder, occurs when the pancreas does not produce adequate insulin hormone to digest carbohydrate food or the body cells does not respond to insulin anymore. Foods that contain an extensive amount of glucose are oxidized for the production of energy in human body cells. By using this type of energies, the human body operates its physical activities. Insulin controls blood sugar in the human body [2]. It is produced in a natural way in a healthy human body. But in the case of a diabetic patient, the production of this insulin hormone decreases and therefore, blood sugar level increases in comparison with a sound body. This high blood sugar often causes nerve damage, kidney damage, eye damage, heart disease, skin hearing impairment, etc. Loss of limb, blindness is also some of the effects [3].

Diabetes is one of the topmost 10 reasons of death in grownups, and was predicted to have affected four million deaths worldwide in 2017. Severe and deadly medical conditions can easily be developed by diabetic patients which may eventually result in unexpected death. So, diabetes is not curable, but it can be controlled. And to control it should be detected as early as possible. Therefore, a method to detect the disease at an early stage would be really beneficial [4].

In this paper, an approach is suggested to predict the probability of developing diabetes with the help of basic medical information which are directly and highly responsible for the development of diabetes. And MLR has been used to complete this approach. After that, using Q-learning of Reinforcement Learning has been used to provide recommendations to the patients so that, they can control this disease.

The rest of the paper is structured as follows. In section 2 and 3 describe literature review and proposed solution respectively. Section 4 contains detailed impression of the approach. Section 5 covers the result and comparison analysis. Lastly, section 6 and 7 draw the conclusion of this paper with some comments and ideas to improve this approach as future work.

2. LITERATURE REVIEW

2.1 Types of Diabetes

Though diabetes has a number of different types, it is mainly divided into three categories: type 1, type 2, and gestational diabetes.

2.1.1 Type 1 Diabetes (T1D)

Type 1 diabetes (T1D) named as insulin-dependent diabetes, is a severe situation in which the pancreas produces little or no insulin. The actual reason of T1D is unidentified [4]. Generally, the human immune system which normally fights against harmful bacteria or viruses kills the insulin-producing (islet, or islets of Langerhans) cells in the pancreas by mistake. This type of diabetes cannot be cured efficiently with oral medications alone and the patients need to take insulin therapy [5].

2.1.2 Type 2 Diabetes (T2D)

Type 2 diabetes (T2D) is the most widespread type of diabetes among all. Because, 90% of the diabetes are T2D. It is generally described with insulin resistance, where the body is unable to react with insulin. As, insulin is incapable to perform properly, blood sugar levels keep escalating and releasing even more insulin. For some people with type 2 diabetes, this can sooner or later exhaust the pancreas, causing in the body producing less and less insulin, resulting in even higher blood sugar levels (hyperglycemia) [7]. Older adults are very likely to have T2D, but is gradually happening to people of all ages. A lot of people with T2D have shown signs of prediabetes (impaired fasting glucose and/or impaired glucose tolerance) before meeting the criteria for type 2 diabetes. Lifestyle factors and genetics are main reasons of developing T2D. Some lifestyle factors are responsible for T2D, including physical inactivity, obesity, and lack of sleep, stress and urbanization [6].

2.1.3 Gestational Diabetes Mellitus (GDM)

Gestational diabetes mellitus (GDM) is the most severe and neglected threat to the mother and the unborn child. It occurs when a hormone produced by the placenta forces the body from not using insulin efficiently. It is fully curable, but requires careful medical observation throughout the pregnancy [7].

After a thorough research, it can be said that among these three types of diabetes, T2D is the most common as well as severe one and requires more attention than other two types.

2.2 Diabetes Risk Factors

As it is intended to predict the risks of T2D, so, only the risk factors of T2D has been discussed. Several factors combinedly affect the probability of developing T2D such as genes and lifestyle. Although it's impossible to change the risk factors such as age, ethnicity or family history but lifestyle risk factors like eating, physical activity, sleep time and weight can be changed. These lifestyle changes can help to control the chances of developing T2D.

- **Age:** The risk of T2D increases with age, especially after the age 45. That's probably because people incline to exercise less, lose muscle mass and gain weight as they are aging. But type 2 diabetes is also increasing intensely among children, teenagers and younger adults too [8].
- **BMI:** BMI is a numerical value that can be obtained by dividing a person's weight in kilograms (kg) by his /her height in meters squared. Overweight is a BMI of 27.3 or more for women and 27.8 or more for men. Obesity is a BMI of 30 or more for either gender [9]. Obesity has been found responsible for almost 55% of cases of T2D [10].
- **Blood Pressure (BP):** Having high blood pressure or hypertension appears to accelerate the risks of T2D. Hypertension has two types of stages and they are:
 - **Stage 1:** Systolic 130–139 mm Hg and diastolic 80–89 mm Hg
 - **Stage 2:** Systolic 140+ mm Hg and diastolic 90+ mm Hg [11]

Whereas, 120/80 mm Hg is the normal blood pressure which should be maintained to avoid diabetes. A blood pressure (BP) of 130/80 mm Hg or less is the recommended therapeutic BP target for T2D patients. Because, diabetic patients can also develop hypertension and both of them can increase cardiovascular risk [12].

- **Blood Sugar:** High blood sugar (hyperglycemia) is another risk factor which contributes to the possibility of having diabetes. Blood sugar level below 200 mg/dl (randomly taken) and below 100 mg/dl (after fasting for 8 hours) is normal. But, when a person is diagnosed with blood sugar level equal to 200 mg/dl or more (randomly taken) and equal to 126 mg/dl or more (after fasting for 8 hours) have diabetes [13].
- **Sleep:** Sleep deprivation is a habitually ignored but it is a major risk factor for T2D. The link may seem unimaginable. Specifically, with constant sleep loss, fewer insulin is released in the body after eating. Meanwhile, human body secretes more stress hormones which makes it impossible to sleep and makes it tougher for insulin to do its job effectively. To sum up, a lot of glucose remains in the blood, which can raise the risk of developing T2D [14]. It is essential to get 7 to 9 hours of

uninterrupted sleep regularly so that the body can work properly and lessen the risk of developing T2D and other health problems.

- **Physical Activity:** Less physical activity is responsible for 7% of the burden of T2D in the European Region. T2D was until recently seen as a disease of middle-aged and elderly people, but it is now increasingly seen in youngsters and children. Exercise improves blood sugar control in T2D, reduces cardiac risks and helps in to weight loss process. Regular exercise may be able to prevent or slow down the T2D development. Everyday exercise, or at least 5 times a week for 20-30 minutes, is recommended to reduce the insulin resistance [15].

2.3 Regression Analysis

Regression analysis is a statistical method to model the correlation between a dependent (target) and one or more independent (predictor) variables. Mainly, it helps to recognize how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are considered fixed. It predicts continuous values such as age, salary, temperature, price and so on [16].

Linear regression is a statistical regression method which is used for predictive analysis. It is used for resolving the regression problem in machine learning. It is known as linear regression because shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis). If there is only one input variable, then it is called simple linear regression. And if there is more than one input variable, then it is called multiple linear regression.

2.3.1 Multiple Linear Regression

Multiple linear regression tries to model the correlation between two or more independent variables and a dependent variable by fitting a linear equation to experimental data. The implementation of multiple linear regression is quite similar to that of simple linear Regression. The alteration can be seen in the evaluation. It can be used to find which independent variable has the highest impact on the predicted output and how different variables relate to each other. For MLR, the dependent variable have to be the continuous/real, but the independent one can be continuous/categorical. Each independent variable must show the linear relationship with the dependent variable. MLR attempts to fit a regression line using a multidimensional space of data-points. It predicts little or no multicollinearity (correlation between the independent variable) in data. The regression residuals must be normally scattered [17].

2.4 Reinforcement Learning

Reinforcement Learning (RL) is a feedback-based Machine learning technique where an agent learns to act in an environment by executing the actions and perceiving the outcomes of actions. For every good actions, the agent receives positive feedback or reward, and for every bad actions, the agent receives negative feedback or penalty. The agents learns with the process of trial and error using feedback from its own actions and experiences.

There are different types, approaches and representation of reinforcement learning. As, Q-learning is one of the most famous algorithms and has been used to develop the recommendation system so, only this algorithm is discussed here.

2.4.1 Q-learning

Q-learning is a model-free RL algorithm, which is considered for the temporal difference Learning. The temporal difference learning methods are the technique of comparing temporally successive predictions. This algorithm attempts to discover the best action to take at a current state. The goal of the agent in this algorithm is to maximize the value of Q (Q-value specifies which action is more lubricative than others) [18].

It's also considered as an off-policy algorithm because the Q-function learns from actions that are outside the current policy, like taking random actions, and therefore a policy is not needed. More specifically, it tries to learn a policy that maximizes the total reward. The 'Q' in Q-learning means Quality. Quality signifies how valuable a given action is in achieving some future reward [19].

2.5 Related Works

There is a lot of research on diabetes happening all around the world. Diabetes prediction is really tough but now-a-days with the help of healthcare monitoring data and machine learning algorithm prediction is less tough than before.

This paper was concentrated upon the predictive analysis of diabetic cure using a data mining technique to discover patterns that identified the best mode of treatment for diabetes across dissimilar age clusters in Saudi Arabia. Oracle Data Miner (ODM), support vector machine algorithm was used. Dataset was analyzed into five age groups. Six types of treatment of diabetes were described. The pattern of the researchers of the work indicated that drug treatment was operative for both groups of patients but more operative for patients in the old age group [20].

Another work was done to find the possibilities of occurrence of diabetes using multi-linear regression algorithm based on R language. The data was collected from the Kaggle website. The dataset had six attributes and they were Glucose, Blood Pressure, Insulin, BMI, Age, and Outcome. Here, regression algorithm was presented with powerful diagnostic features for the occurrence of diabetics. The performance of regression algorithm was also analyzed using only selected attributes among total number of records from the input dataset [17].

The emerging optimization algorithms and machine learning algorithms were summarized in this paper. Three types of optimization algorithms and four types of machine learning algorithms were discussed. The four applications in the field of diseases diagnosis were considered and the important challenges in the deployment of disease were discussed [21].

Study on various prediction techniques and tools for machine learning in practice were discussed in this paper. How to get pattern from a large dataset and specific steps on how to apply machine learning to data were evaluated. Various algorithms, some tools and libraries were talked over. Some diseases were also considered and how to process its data while doing diagnosis and predictions using machine learning were showed [22].

In another paper, a general framework was presented for recommendation system based on reinforcement learning. This system could learn directly from the customer's behavior and gave customer more chance to explore new pages/products which are not popular which may match with their interests. The learning process was SARSA method and ϵ -greedy policy. The system composed of two models. First, a

global model for all customers that could discover behavior of the system. Second, a local model was used to keep records of user browsing history and make offer from each customers. The objective was to find relationship between ϵ and user click rate. Lastly, it was acknowledged that, the ϵ -greedy policy could give a chance to explore new produces, but the same time it was exploited to the trend of system [23].

The aim for this paper was to introduce, review, and summarize several works and research papers on Reinforcement Learning. The main components of reinforcement learning system: policy, reward signal, value function and model were discussed. Markov Decision Process (MDP) was explained in detail. Different methods of RL algorithms such as: Tabular, Approximate Solutions, Monte Carlo and Temporal Difference, Policy based RL, Deep Q-Learning were summarized. After that, data from a wide collection of works in various reinforcement learning applications were reviewed. Finally, the prospects and the challenges of reinforcement learning were discussed [24].

3. PROPOSED METHODOLOGY

Diabetes is a common but a chronic disease with serious consequences such as long-term damage and disabling body parts. As, it disturbs many parts of the body so, if left untreated, can create severe health issues for example kidney failure, heart disease, blindness, stroke, and lower-limb amputation and so on [4]. It currently affects 200 million people and is the 5th cause of death throughout the world [25]. 1 in 2 (232 million) people with diabetes were untreated. It affected 4.2 million deaths till now. Diabetes caused minimum 760 billion dollars USD as health expenses in 2019 which now makes it an expensive disease too [7].

All this facts and severe consequences of diabetes motivated us to develop a system which can help people to predict the possibilities of having diabetes with the basic medical information and also provide appropriate recommendations to control the disease. Because, if a person knows that he has a chance of being a diabetic patient at an early stage, he may be able to control the severity of the disease by following the suggestions. Also, he will be warned in advance and let himself diagnosed thoroughly. The proposed methods for both systems have been discussed in the following sections.

3.1 Proposed Method for the Diabetes Prediction System

For the first part of this paper, a method has been proposed to develop a diabetes prediction system using regression analysis. Eight risk factors associated with diabetes have been used as independent variables to find the dependent variable predicted risk. Multiple Linear Regression has been used for regression analysis. Steps taken for the proposed method have been discussed below.

- At first, the dataset was finalized which has been used for the method.
- Then, Multiple Linear Regression model has been applied to the selected dataset to predict the risk of diabetes.
- Lastly, the co-relation between the risk of diabetes and the risk factors has been analyzed using Simple Linear Regression to identify the most significant factors which affects the risk of diabetes.

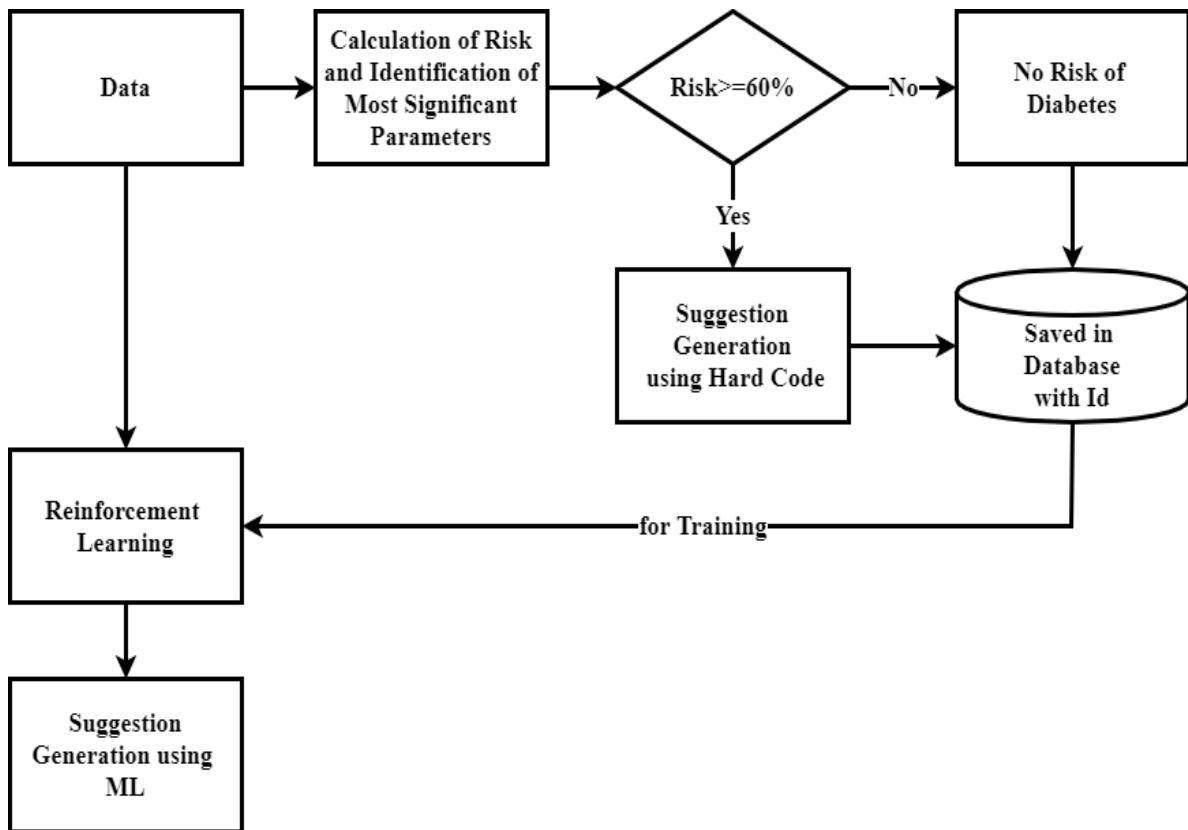


Figure 1: Flowchart for the Proposed Method of Recommendation System

3.2 Proposed Method for the Recommendation System

For the second part of this paper, a recommendation system has been developed that can generate suggestions for the people considered as diabetic patients (people with risk percentage of having diabetes greater than or equal to 60%). To develop this system, Reinforcement Learning has been used. Therefore, a method has been proposed that can generate suggestions. Steps taken for the proposed method are provided below (figure 1).

- At first, using the previous system the risk of diabetes has been calculated and also the three most significant parameters has been identified.
- Then, if the risk percentage is equal to or greater than 60% then some suitable suggestions are generated by considering the most affecting parameters.
- But, if the risk percentage is less than 60%, then the output is “No Risk of Diabetes”. For this case, it is then checked if the parameter **Fasting Blood Sugar** is one of the three most affecting parameters or not. If it is one of them, then it is considered as a case of prediabetes. And suggestions has been generated for this case too.
- For both the cases, the result is stored in the database with ID number.
- After following the given suggestions for a certain period, the risk of diabetes has been calculated and the suggestions has been generated again. Then, the system compares this new suggestions with the previously stored one and the comparison result is saved in the database.
- Then, the stored dataset has been fed to train the

reinforcement learning algorithms used in the recommendation system. Here, instead of passing the saved data one by one, it has been passed as a dataset (a chunk of data) for updating the RL system effectively.

4. IMPLEMENTATION

At first, the proposed diabetes prediction system has been implemented. And, then predicted risk from the prediction system has been used to implement the recommendation system. The implementation of both proposed methods have been described in the following sections.

4.1 The Deployment of the Multiple Linear Regression

To predict the risk of diabetes, Multiple Linear Regression has been applied on the dataset. Before applying the model on a dataset, the dataset has to go through some series of steps. These steps are known as pre-processing steps.

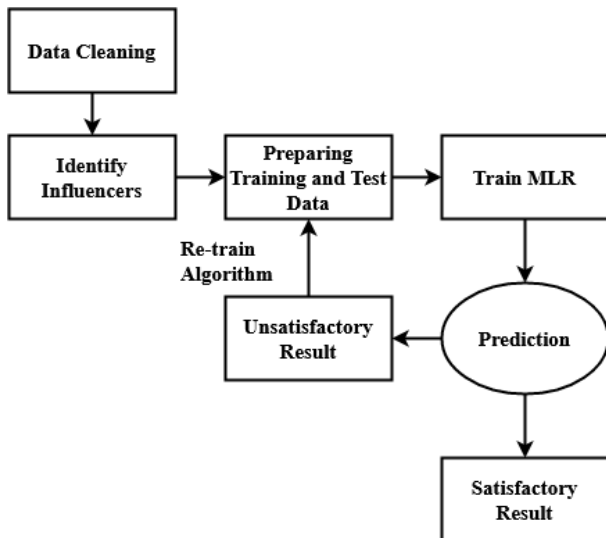


Figure 2: The Workflow of Multiple Linear Regression

At first, the dataset has been cleaned by removing the null and infinite values. From the cleaned dataset, the highly weighted influencers (attributes) has been identified. In the next step, the dataset has been prepared for training and testing process. Then, the dataset has been trained by the MLR model for prediction. If the unsatisfactory results are generated from the prediction then the algorithm needs to be retrained by preparing the dataset for training and testing process. And if satisfactory results are generated then the desired outcome will be achieved (figure 2).

4.2 The Development of the Recommendation System

A standard reinforcement learning (RL) system has been considered where an agent interacts with the environment in discrete time. Here, the clinical parameters have been passed as input. The input parameters list has been considered as an agent. And in the environment, every state is considered as a set of suggestions (based on a specific parameter). Initially, at every timestep, the agent accepts a state from the environment and selects an action based on its policy. The environment returns a reward and the agent moves to the next state.

In the environment, a reward table has been used from which the agent can learn. It chooses actions by examining the rewards to move to the next state. After that, it stores the Q-value in the table. These values are mapped to a (state, action) combination. A Q-value for a specific state action arrangement represents the quality of an action taken from that state. Better Q-values indicate better possibilities of receiving bigger rewards. Q-values has been set to an arbitrary value, and as the agent enters the environment and receives different rewards by performing different actions, the Q-values are updated using the equation:

$$Q(\text{state}, \text{action}) \leftarrow (1 - \alpha) Q(\text{state}, \text{action}) + \alpha (\text{reward} + \gamma \max_a Q(\text{next state}, \text{all actions})) \quad (1)$$

In this equation,

- α (Alpha) is the learning rate $0 < \alpha \leq 1$. It is the extent to which the Q-values are being updated in every iteration.
- γ (Gamma) is the discount factor $0 \leq \gamma \leq 1$. It

determines how much importance should be given to the future rewards. A higher value for the discount factor (close to 1) can catch the long-term effective award. On the other hand, a discount factor of 0 allows the agent to consider only immediate reward. In the implementation of the proposed method, $\gamma = 0.8$ has been used.

The Q-value of the agent's current state and action has being updated by first taking a weight $(1 - \alpha)$ of the previous Q-value, then by adding the learned value. The learned value is a combination of the reward for taking the current action in the current state, and the discounted maximum reward from the next state are to be in after taking the current action. Mainly, the agent is learning the appropriate action to take in the current state by observing the reward for the current state-action arrangement, and the maximum rewards for the next state. Ultimately, the system will consider the path attached with the best rewards.

The whole Q-Learning process can be summarized as follows.

- Initialize the Q-table by all zeros
- Start exploring actions: for every state, select any of possible actions for the current state (S)
- Move to the next state (S') as an outcome of that action (a)
- For all possible actions from the state (S'), select the one with the maximum Q-value
- Update Q-table values using the equation
- Fix the next state as the current state
- If goal state reached, then terminate and repeat the process

After sufficient random exploration of actions, the Q-values has converged helping the agent as an action-value function which it can be exploited to select the most ideal action from a given state. There's a tradeoff between exploration (choosing a random action) and exploitation (choosing actions based on already learned Q-values). To prevent the action from selecting the same path all the time and then result in overfitting, a new parameter ϵ (epsilon) is used during the training. Instead of just choosing the best learned Q-value action, it is essential to explore the action space more at times. Lower epsilon value results in episodes with more penalties which is natural as the action space is explored and random decisions are made.

Through this learning process, the system will gradually achieve the goal. It will generate the best possible sets of suggestions for the patients based on their clinical parameters and keep learning through their feedback results.

5. RESULT ANALYSIS

In this chapter, the experiment's result has been analyzed. For the prediction, a mathematical equation has been formed. To get the predictive risk percentage by solve the equation, a dataset which was suitable for the research was selected.

5.1 Dataset

All the information was collected from different web sources where more than 1000 people submitted their all health indexes in real-time (current status of Diabetes). Only a few participant's data has been considered in this paper.

5.2 Prediction System

5.2.1 Diabetes Prediction using MLR

By implementing Multiple Linear Regression on the dataset,

Table 1: Diabetes Prediction Dataset with Predicted Risk using MLR

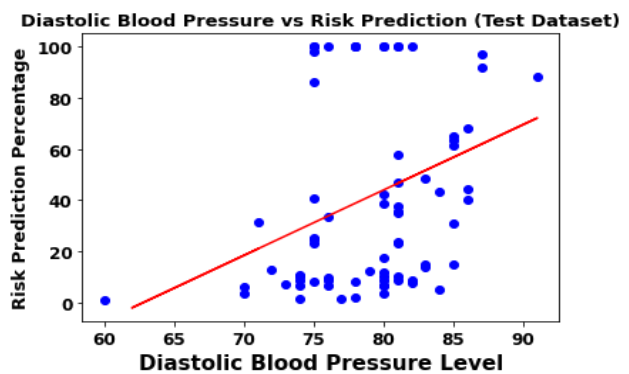
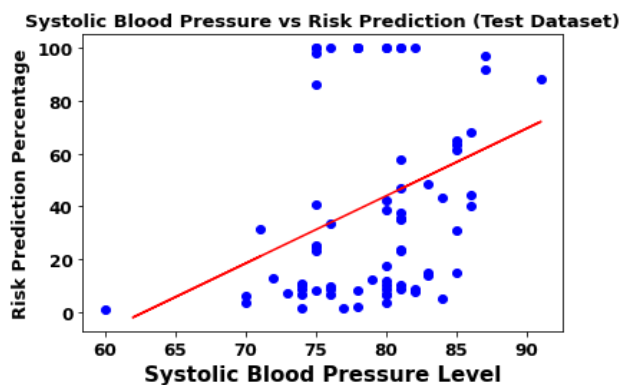
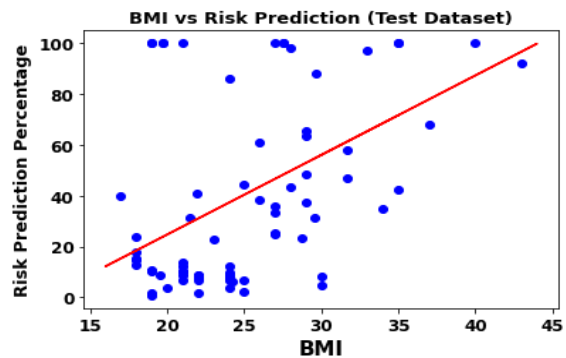
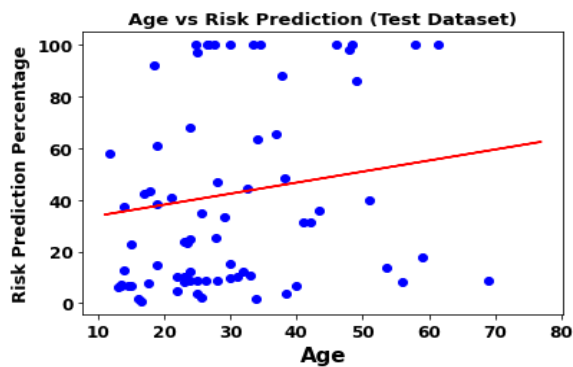
Age	BMI	Systolic BP	Diastolic BP	Blood Sugar (Fasting)	Blood Sugar (Random)	Sleep Time (Hours)	Exercise Time (Minutes)	Test Result	Risk Prediction using MLR
59	18	120	80	4.9	6.0	6.5	20	17.7%	40%
18.5	43	135	87	5.7	7.05	6	20	92.06%	83%
40	21	120	80	4.5	7.02	6.5	30	6.54%	32%
46	19	121	78	8.14	9.89	6.5	20	100%	90%
26.7	35	122	81	8.05	14.1	8	30	100%	100%
24	27	120	75	4.2	5.1	5.8	20	24.59%	26%
26.4	19	121	78	8.14	9.89	8	25	100%	85%
30	24	120	76	4.89	5.65	7.5	25	100%	32%
34	29	117	85	5.08	9.12	6.5	30	6.125%	47%

5.2.2 Co-relation between Diabetes Risk Factors and Risk Percentage

It is important to relate the parameters associated with a disease with the risk percentage of the disease. It can help to identify the most affecting parameters in general. So, to recognize these most affecting parameters by co-relating the

risk factors and the risk percentage, Simple Linear Regression has been used.

The relations between the risk factors and risk percentage are shown in the figure 3. The attribute's columns and the test result column of the dataset has been used to generate these figures.



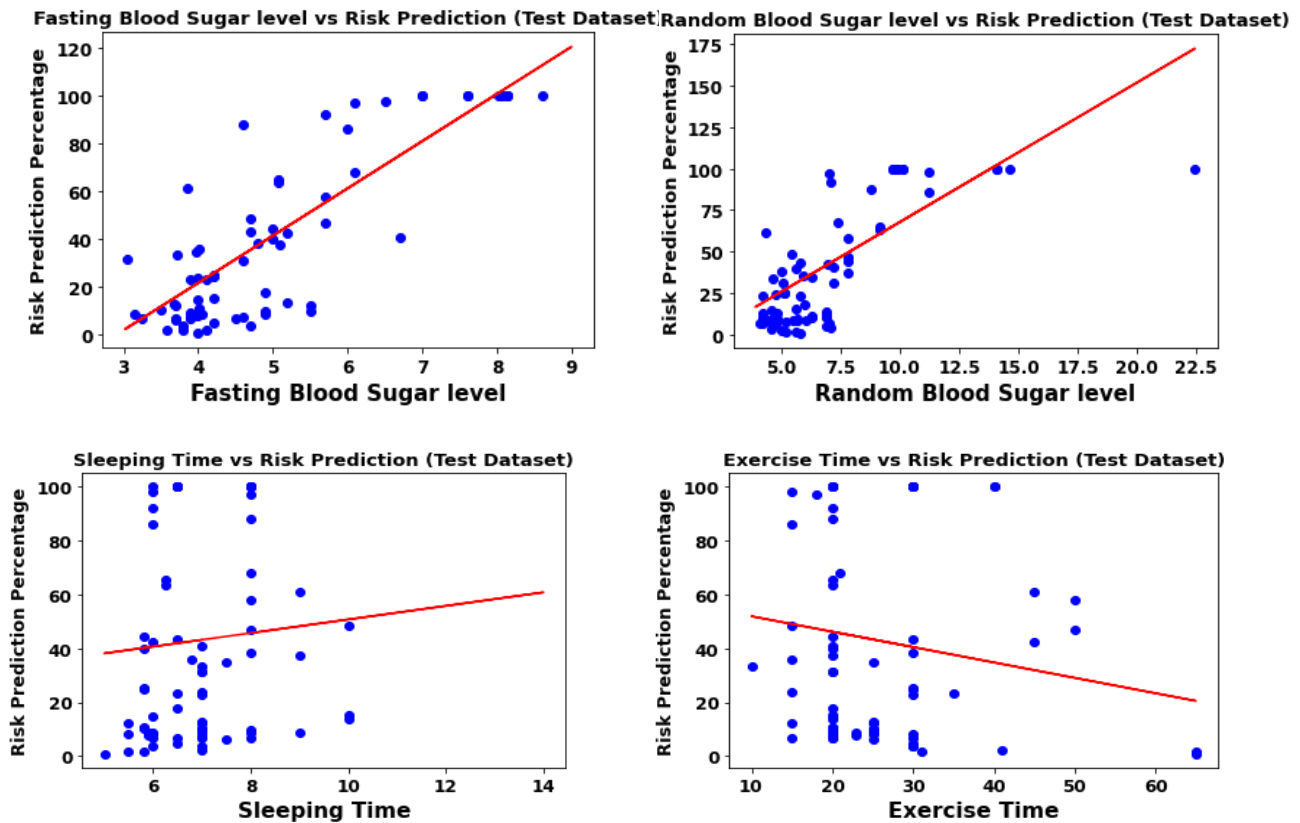


Figure 3: Diabetes Risk Factors vs Diabetes Risk Prediction

5.3 Recommendation System

After obtaining the risk prediction from the MLR model, the risk for every individual parameters have been analyzed and the three most affected parameters have been identified. Based on these parameters, the suggestions have been generated. Then, the IDs of these parameters with the corresponding risk have been stored in the database. Basically, it has been tried to focus on every clinical parameters and generate suggestions based on the most affected parameters so that the patient can follow the suggestions and reduce the risk percentage.

Table 2: Paths for the Suggestions Generated using RL

Output Path	Path (0) Suggestion	Path (1) Suggestion	Path (2) Suggestion
Path [3, 2, 4]	Diastolic BP	Systolic BP	Blood Sugar
Path [5, 1, 6]	Sleeping Time	BMI	Exercise Time
Path [2, 5, 7]	Systolic BP	Sleeping Time	Consultation
Path [5, 3, 4]	Sleeping Time	Diastolic BP	Blood Sugar
Path [4, 3, 5]	Blood Sugar	Diastolic BP	Sleeping Time
Path [2, 7, 1]	Systolic BP	Consultation	BMI
Path [5, 4, 7]	Sleeping Time	Blood Sugar	Consultation
Path [7, 6, 3]	Consultation	Exercise Time	Diastolic BP
Path [5, 2, 3]	Sleeping Time	Systolic BP	Diastolic BP
Path [3, 2, 4]	Diastolic BP	Systolic BP	Blood Sugar

On the other hand, for the recommendation system which has been implemented using RL, the suggestion has been generated based on the Q-value. The RL system automatically generates the best possible path that can generate suggestions

as good as possible. Based on the path the best possible suggestions has generated.

In the table 2, the **Path** represents the most affecting parameters and based on these parameters, the suggestions have been generated. These seven (1-7) paths refers to all seven columns of table 1.

6. CONCLUSION

Being a long-lasting disease, diabetes can make its patients suffer for the rest of his life. After developing diabetes once, it cannot be cured. But, it can be controlled. The earlier it is diagnosed, the easier it will be to control this disease. And, this system can play an important role here. This system will predict the probability of a person's having diabetes and if the probability is near 60%, the person will know that and he can change the lifestyle habits which affects the possibility and he will probably be able to avoid diabetes. But, if unfortunately, the probability is over 60%, the person can go a doctor and be able to control it before it gets too late. It will be a great help to the people who may not have the financial ability to take a diabetes diagnosis whenever he/she wants. Lastly, the hope is to reduce the number of patients who are suffering from diabetes and also the increasing death rate solely because of diabetes.

7. FUTURE SCOPE

Based on a patient's percentage of risk of having diabetes, the system can be developed so that it can provide a set of suggestions using reinforcement learning algorithm for the diabetic patients to control it. Then, it will further generate suggestions based on the patient's negative or positive outcome by following the system's previous generated suggestions. The patient's clinical histories will be stored time to time so that the system can generate appropriate suggestions to gain the best possible outcome.

8. REFERENCES

- [1] N. Jothi, W. Husain, et al., "Data mining in healthcare—a review," *Procedia Computer Science*, vol. 72, pp. 306–313, 2015.
- [2] Rosenstock, J., Park, G., Zimmerman, J., & Glargine, U. I. (2000). Basal insulin glargine (HOE 901) versus NPH insulin in patients with type 1 diabetes on multiple daily insulin regimens. US Insulin Glargine (HOE 901) Type 1 Diabetes Investigator Group. *Diabetes care*, 23(8), 1137-1142.
- [3] Lal, B. S. (2016). Diabetes: Causes, Symptoms And Treatments. book: *Public Health Environment and Social Issues in India, Edition, 1*, 55-67.
- [4] Mellitus, D. (2005). Diagnosis and classification of diabetes mellitus. *Diabetes care*, 28(S37), S5-S10.
- [5] Ndisang, J. F., Vannacci, A., & Rastogi, S. (2017). Insulin resistance, type 1 and type 2 diabetes, and related complications 2017.
- [6] Dariush Mozaffarian, Aruna Kamineni, Mercedes Carnethon, Luc Djoussé, Kenneth J. Mukamal, David Siscovick." Lifestyle Risk Factors and New-Onset Diabetes Mellitus in Older Adults: The Cardiovascular Health Study." *Archives of Internal Medicine*, vol.169, issue.8, Pages.798.
- [7] Atlas, D. (2015). International diabetes federation. IDF Diabetes Atlas, 7th edn. *Brussels, Belgium: International Diabetes Federation*.
- [8] Chatterjee, S., Khunti, K., & Davies, M. J. (2017). Type 2 diabetes. *The Lancet*, 389(10085), 2239-2251
- [9] Ray, D. E., Matchett, S. C., Baker, K., Wasser, T., & Young, M. J. (2005). The effect of body mass index on patient outcomes in a medical ICU. *Chest*, 127(6), 2125-2131.
- [10] Centers for Disease Control and Prevention (CDC). (2004). Prevalence of overweight and obesity among adults with diagnosed diabetes--United States, 1988-1994 and 1999-2002. *MMWR. Morbidity and mortality weekly report*, 53(45), 1066.
- [11] Cheung, B. M., & Li, C. (2012). Diabetes and hypertension: is there a common metabolic pathway?. *Current atherosclerosis reports*, 14(2), 160-166.
- [12] Lipman, M. L., & Schiffrin, E. L. (2012). What is the ideal blood pressure goal for patients with diabetes mellitus and nephropathy?. *Current cardiology reports*, 14(6), 651-659.
- [13] O'Sullivan, J. B., & Mahan, C. M. (1965). Blood Sugar Levels, Glycosuria, and Body Weight Related to Development of Diabetes Mellitus: The Oxford Epidemiologic Study 17 Years Later. *JAMA*, 194(6), 587-592.
- [14] Touma, C., & Pannain, S. (2011). Does lack of sleep cause diabetes. *Cleve Clin J Med*, 78(8), 549-58.
- [15] S. R. Colberg, R. J. Sigal, J. E. Yardley, M. C. Riddell, D. W. Dunstan, P. C. Dempsey, E. S. Horton, K. Castorino, and D. F. Tate, "Physical activity/exercise and diabetes: a position statement of the american diabetes association," *Diabetes care*, vol. 39, no. 11, pp. 2065–2079, 2016.
- [16] Freund, R. J., Wilson, W. J., & Sa, P. (2006). Regression analysis. *Elsevier*.
- [17] V. Madhubala, P. Porkodi, R. Selvapriya and P.Tamilzhchelvi. Diabetics Prediction Based on Multi-Linear Regression Using R Language. *Asian Journal of Computer Science and Technology*, Vol.8 No.S2, 2019, pp. 17-19.
- [18] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.
- [19] Glorennec, P. Y. (2000, September). Reinforcement learning: An overview. In *Proceedings European Symposium on Intelligent Techniques (ESIT-00)*, Aachen, Germany (pp. 14-15).
- [20] A. A. Aljumah, M. G. Ahamad, and M. K. Siddiqui, "Application of data mining: Diabetes health care in young and old patients," *Journal of King Saud University-Computer and Information Sciences*, vol. 25, no. 2, pp. 127–136, 2013.
- [21] K. Chui, W. Alhalabi, S. Pang, P. Pablos, R. Liu, and M. Zhao, "Disease diagnosis in smart healthcare: Innovation, technologies and applications," *Sustainability*, vol. 9, no. 12, p. 2309, 2017.
- [22] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 492–499, IEEE, 2017.
- [23] P. Rojanavas, P. Srinil, and O. Pinngern, "New recommendation system using reinforcement learning," *Special Issue of the Intl. J. Computer, the Internet and Management*, vol. 13, no. SP 3, 2005.
- [24] A. Hammoudeh, "A concise introduction to reinforcement learning."
- [25] Wang, L., Kong, L., Wu, F., Bai, Y., & Burton, R. (2005). Preventing chronic diseases in China. *The lancet*, 366(9499), 1821-1824.
- [26] Alehegn, M., Joshi, R. R., & Mulay, P. Diabetes Analysis And Prediction Using Random Forest, KNN, Naïve Bayes, And J48: An Ensemble Approach.
- [27] Duke, D. L., Thorpe, C., Mahmoud, M., & Zirie, M. (2008, March). Intelligent Diabetes Assistant: Using machine learning to help manage diabetes. In *2008 IEEE/ACS International Conference on Computer Systems and Applications* (pp. 913-914). IEEE.
- [28] Kumar, P. S., & Pranavi, S. (2017, December). Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)* (pp. 508-513). IEEE.
- [29] Mirshahvalad, R., & Zanjani, N. A. (2017, September). Diabetes prediction using ensemble perceptron algorithm. In *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 190-194). IEEE.
- [30] Priyadarshini, R., Dash, N., & Mishra, R. (2014, February). "A Novel approach to predict diabetes mellitus using modified Extreme learning machine." In *2014 International Conference on Electronics and Communication Systems (ICECS)* (pp. 1-5). IEEE.
- [31] M. Adam, E. Y. Ng, S. L. Oh, M. L. Heng, Y. Hagiwara, J. H. Tan, J. W. Tong, and U. R.Acharya, "Automated characterization of diabetic foot using nonlinear features

extracted from thermograms,” Infrared Physics & Technology, vol. 89, pp. 325–337, 2018.

- [32] M. R. Devi and J. M. Shyla, “Analysis of various data mining techniques to predict diabetes mellitus,” *International Journal of Applied Engineering Research*, vol. 11, no. 1, pp. 727–730, 2016.
- [33] S. Bashir, U. Qamar, and F. H. Khan, “Intellihealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework,” *Journal of biomedical informatics*, vol. 59, pp. 185–200, 2016.
- [34] S. R. Colberg, R. J. Sigal, J. E. Yardley, M. C. Riddell, D. W. Dunstan, P. C. Dempsey, E. S. Horton, K. Castorino, and D. F. Tate, “Physical activity/exercise and diabetes: a position statement of the american diabetes association,” *Diabetes care*, vol. 39, no. 11, pp. 2065–2079, 2016.
- [35] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, “Type 2 diabetes mellitus prediction model based on data mining,” *Informatics in Medicine Unlocked*, vol. 10, pp. 100–107, 2018.