Forecasting the Risk of Type II Diabetes using Reinforcement Learning

Most. Fatematuz Zohora*, Marzia Hoque Tania‡, M Shamim Kaiser*¶, and Mufti Mahmud**
*Institute of Information Technology, Jahangirnagar University, Savar, 1340 - Dhaka, Bangladesh.
‡ Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK.
**Dept. of Computing & Technology, Nottingham Trent University, Clifton Campus, NG11 8NS – Nottingham, UK.
‡ORCID:0000-0002-4496-1896
¶ORCID: 0000-0002-4604-5461
**ORCID: 0000-0002-2037-8348

***Abstract*—Type II Diabetes (T2D) is one of the most common lifestyle diseases which is characterized by insulin resistance. Lack of insulin's proper working causes uncontrollable blood glucose rise in the body which leads to life taking situations. Therefore, early detection of T2D is imperative to save many lives. Towards this goal, this work presents a machine learning-based prediction model to detect T2D. The Q-learning algorithm belonging to the Reinforcement Learning (RL) paradigm has been applied to the PIMA Indian Women diabetes dataset in developing the detection model. The model identifies patients with T2D using three factors (such as Body Mass Index, glucose level and age of subject) by generating an off-policy based RL and making the learning agent to find an optimal policy for the factors. The information of a subject can be in any of 330 possible states. The proposed RL model's accuracy, Precision, Recall, F-measure and AUC values have been compared with the state-of-the-art techniques such as K Nearest Neighbors and Decision Tree. The performance of the proposed RL-based T2D prediction outperforms the K Nearest Neighbors and Decision Tree.**

Keywords: Type II Diabetes, Machine Learning, Reinforcement Learning, Q-learning, Healthcare.

I. Introduction

Diabetes is a life-changing medical condition that affects millions of people every year. As stated by the 2017 national diabetic statistics report, Type II Diabetes (T2D) accounts for 90 % to 95% of all diabetes cases [1]. T2D is a serious disease with a high occurrence around the world, and a trend that is still on the increase. T2D is a principal cause of sickness and death and contributes to augmented risks of cardiovascular risk diseases by 2 to 4 times [2]. So, early detection of T2D is very necessary to progress the quality of life of patients and improvement of their life anticipation.

Machine Learning (ML) algorithms are utilized to understand the patterns from the data. In these days, researchers are employing ML in a variety of tasks including biological data mining [3], image analysis [4], financial forecasting [5], anomaly detection [6], disease detection [7], [8], natural language processing [9], assay detection [10] and strategic game playing [11]. Patterns obtained from ML can be utilized to increase our understanding (e.g., identifying risk factors for diabetes) or predict 0the future outcome (e.g., forecasting who will grow diabetes). ML investigates many areas ranging from statistics, optimization to computer science. Regarding data set nearly any ML problems reported and assessed as an optimization problem.

In recent years, Electronic Health Records (EHRs) has become more available for increasing interests to discover personalized healthcare suggestions to optimize clinical decision making and patient management [12], [13]. Thus the study on treatment recommendation moves from knowledge-driven into data-driven. In such data availability, the target is to find a model that best interprets the data. Among numerous different types of ML, most applications fall into 1 of 3 categories: Supervised, Unsupervised, or Reinforcement Learning (RL) [14].

Here, we worked on RL, specially Q-learning, an algorithm in which an agent learns in a dynamic environment by trial and error basis utilizing feedback from its own actions and experiences.

Through experiencing an enough number of episodes the agent learns a mapping technique from a set of data to the final outcome. This learning process is accomplished on the training dataset. The obtained mapping model can be practised to new test data for both pattern discovery and prognosis purposes when the agent is trained. The introduced RL framework can automatically and adaptively predict T2D using EHRs of real-life data like Body Mass Index (BMI), glucose and age.

In this paper, the performance of proposed Q-learning model is compared with simple but powerful supervised ML algorithms like k-nearest neighbors (KNN) and Decision Tree (DT) in predicting either a person has developed T2D or not. The data is chosen from PIMA Indian Women Diabetes database.

II. Literature Review

Several studies had investigated diabetes data and constructed models to predict diabetes. Vijayan and Anjali [15] discussed several classifiers such as Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) as a base classifiers for AdaBoost calculation to order to calculate the accuracy. The accuracy received for AdaBoost calculation with base classifier has found to be 80.72%.

Comparison and prediction of diabetes on Pima Indian data set are performed using different classification techniques like NB, J48 DT, Random Forest (RF), Logistic Regression (LR), Multilayer Perception by Saman *et al.* [16]. Multilayer Perception worked best in terms of accuracy and performance.

Khalil *et al.* [17] found that the SVM classifier can be utilized for predicting T2D patients with the accuracy higher compared to K-Mean, FEMean, Probabilistic Neural Network.

Fox and Wiens [18] explored the efficacy of RL techniques for automated blood glucose control. They compared the performance of various deep RL methods to non-RL methods over a sequence of experiments. The primary result suggested that RL could be useful for improving blood glucose control algorithms based on their result that is RL methods are competitive with the baselines and are better able to handle hidden behavioral patterns [19].

Javad *et al.* introduced a RL based technique to handle the blood glucose level of a patient having Type I diabetic (T1D) with a insulin pump where the RL agent learns from the exploration/experiences and policy to pick the optimal actions [20]. Authors have considered physical activity, Body weight, glycated hemoglobin (HbA1C) level, as the state spaces of a diabetic patient in this method.

Myhre *et al.* [21] proposed the fitted Q-iteration framework to control the glucose level in the blood. The RL agent employed optimal insulin policy for T1D patients. Using a non-parametric regression method with functional features, the person-specific customization has been done using manipulating information confined in the shape of the glucose curve.

Yasini *et al.* [22] discovered agent-based off-line simulation using a model-free Q-learning technique in limiting the blood glucose level of a patient having T1D. The prime attributes in the work was accuracy, unresponsive to disturbance, and uncertainty as well as exact settling time.

A new RL problem method that can identify the most probable diagnoses by optimizing clinical concept mining from a free text case description via assigning related visible indication is introduced by Yuan Ling *et al.* [23]. The agent has learned the optimal policy through iterative search and association of the most appropriate clinical concepts that best define a correct diagnosis during training. A deep Q-network agent is prepared in optimizing the reward which thereby improved the accuracy of the probable diagnoses.

III. Type-II Diabetes

Among all the life-style diseases, Type-II Diabetes (T2D) is a chronic state in which the body failed to produce sufficient insulin and the condition is called insulin resistance. Thus the blood glucose levels uplifted. The T2D results Some important risk factors such as obesity, weight gain [24] and physical inactivity, etc. [25]. Moreover, people having a diet with low fiber and the high glycemic index has been connected with an augmented risk of T2D [26].

When our blood sugar level is higher than normal but not enough to develop diabetes then it occurs prediabetes. This disease happens when our body does not create or routine the hormone insulin appropriately. The prediabetes patients are at much higher risk of developing T2D. They also have the risk of developing other health conditions including heart disease or stroke. Prediabetes can be prevented or delayed the onset of full-blown T2D by changing lifestyle. With the intake of a healthy diet, maintaining a required weight, and exercising regularly, prediabetes patients can be taken good care of.

IV. Machine Learning Classifiers

In this section, some supervised ML classifiers named k-nearest neighbors, Decision Tree and Q-learning classifier have been discussed.

A. k-nearest neighbors

KNN can be used in classifying T2D using diabetic data. In this case, the majority votes of the neighbours are compelled in the classification and a distance function is employed in estimating its K nearest neighbors. Below shows KNN algorithm [27]:

1) The sample size of training data is assumed to be m and $p(x, y)$ be an unknown point that required to be classified.
2) The training samples are kept in 2D array of data points $arr[x, y]$. Every element of this array denotes 2D coordinates of a tuple (x, y).
3) For all training data samples m, Calculate distance from unknown point p.

$$dist(arr[x_{\mathrm{i}}, y_{\mathrm{i}}], p(x, y)) \tag{1}$$

4) Make a set S which includes K number of samples with the smallest distances. Each of these distances resembles an already classified data point.
5) Return the majority label among S.

B. Decision Tree

A DT is a classifier that recursively performs partition of the instance space. It contains nodes that form a tree, a node called "root" that has no incoming edges is the starting point of the tree. All other nodes have one incoming edge. The leaf nodes are known as decision nodes. The child node is nominated by computing Information Gain (IG) and Entropy(E) shown in equations 2 and 3.

$$IG = E(parent) - [weights_{average}] \times E(children) \tag{2}$$

$$E = -P(x_i) \log P(x_i) \tag{3}$$

where $P(x_i)$ is the probability of child node i.

Node with the highest IG will be the parent for next level. This process is continued until it gets a leaf node and completed DT.

The basic steps for generating a DT is as below [28]:

1) Create (T)
2) Calculate frequencies (C_i,T)

3) If all instances belong to the same class, returning leaf
4) For every attribute a test is set for splitting criteria. An attribute that satisfies the test is test node K.
5) Repeating Create (T_i) on each partition T_i. Adding those nodes as children of node K

C. Q-learning

In the Q-learning algorithm, a set of q-values are estimated using an agent utilizing a set of states and actions. The agent will receive positive/negative reward for each (state,action) pair. The agent maximized the positive or negative reward in the long term process through learning optimal policy selection for different unique states.

At the beginning of the Q-learning process value of the q table or q matrix is initialized to zero. After completion of an episode which is to reach the goal state through exploration and exploitation, the q table values are updated and stored for each state (S) and action (A) combination $Q[S_t, A_t]$. The agent may execute a number of episodes to generate more acceptable q-values which becomes a reference table for the agent to make the best decision for a certain circumstance. Agent's action taking decision is done in two manners, one is exploitation and another is exploration. Exploitation indicates to select an action based on the available maximum q-value. On the other hand, exploration is a way to take action randomly instead of selecting based on any policy and discover new states that can be used in exploitation. Both of them are balanced using epsilon (ϵ) that decides how frequently the agent is going to explore or exploit state spaces.

Q value updating is the core iterative process of this algorithm which is a function of current reward (R) and values of the next state action pair Q [S_{t+1}, A_{t+1}]. The equation 4 shows the q value update function:

$$Q[S_t, A_t] = Q[S_t, A_t] + \alpha \times (R_t + \gamma \times max(Q[S_{t+1}, A_{t+1}]) - Q[S_t, A_t]) \quad (4)$$

In the above function the learning rate (α) adjusts step size that defines the acceptance rate between the new value and the old value. Gamma (γ) is a discount factor of the new value that is used to balance immediate and future reward. After completion of several execution the q-table is updated with a stable values and the agent learns an optimal policy to take decision in every individual state.

V. Performance Comparison

A. Process Workflow

The designed framework systematically transforms data into information and aids us to create prediction solution. In this framework, Q-learning algorithm is employed on a specific data set to enhance the accuracy of prediction of T2D and the results have been compared with the other supervised ML algorithms such as KNN and DT.

The workflow of Q-learning based T2D prediction process is shown in Figure 1 where Figure 1(a) explains the Q-learning model training process sequentially and Figure 1 (b) describes the Q-learning model testing process for method validation.

TABLE I
State Space Parameter Levels

State Levels	BMI Range	Glucose Range	Age Range
Level 1	<25.6	<115	<24
Level 2	>25.5 and <27.1	>114 and <130	>23 and <28
Level 3	>27 and <28.1	>129 and <140	>27 and <31
Level 4	>28 and <30.1	>139 and <151	>30 and <35
Level 5	>30 and <32.6	>150	>34 and <46
Level 6	>32.5 and <34.6	NA	>45
Level 7	>34.5 and <36.1	NA	NA
Level 8	>36 and <37.1	NA	NA
Level 9	>37 and <40.1	NA	NA
Level 10	>40 and <45.1	NA	NA
Level 11	>45	NA	NA

1) Feature selection and data cleaning: In this step, PIMA Indian Women Diabetic Dataset has been collected and selected three attributes (Glucose, BMI and Age) after analyzing correlation matrix. These attributes have a significant influence on the outcome of diabetes prediction. The selected dataset contains data of 768 subjects and the modified dataset has been extracted using the three selected attributes.

The modified dataset is then preprocessed to eliminate the noise, then the data is cleaned. Minimum values for many variables are 0. But biological parameters like Glucose and BMI cannot have zero values, looks like null values have been coded as zeros. So removed these noisy data from our dataset and create a new one having 392 subject's data of 3 features without any redundant and misleading data.

2) State and Action Space of Q-learning model: After selecting important features and data cleaning a set of state is defined to describe subjects using their glucose, BMI and age where glucose is divided into 5 different levels {Level 1, Level 2... Level 5}, BMI into 11 different levels {Level 1, Level 2... Level 11} and age into 6 different levels {Level 1, Level 2... Level 6}. Table I presents the state space parameter level range of glucose, BMI and age.

A 2D matrix is plotted to indicate state space of an individual where x-axis denotes the product of glucose level and age (5 × 6 = 30) and the y-axis denotes 11 BMI levels. In this work, the set of health states of a subject is defined by St = (BMI, Glucose, Age). For example, state(111) indicates the subject has BMI <25.6, glucose <115 and age <24.

The Q-agent transitions from state to state by performing an action. In this proposed Q-learning algorithm, the agent can choose any action from a set of four different actions. They are moving Right (X_{i+1}, Y_i), Left (X_{i-1}, Y_i), Up (X_i, Y_{i+1}) and Down (X_i, Y_{i-1}). The agent may select an action based on the maximum q-value of chosen action or randomly. The exploration (when q-agent action randomly) and exploitation (when q-agent select action based on largest q-value) are controlled in our algorithm using epsilon (ϵ) variable. The epsilon is assigned with a value of 0.99. That means the agent explore states 99% times.

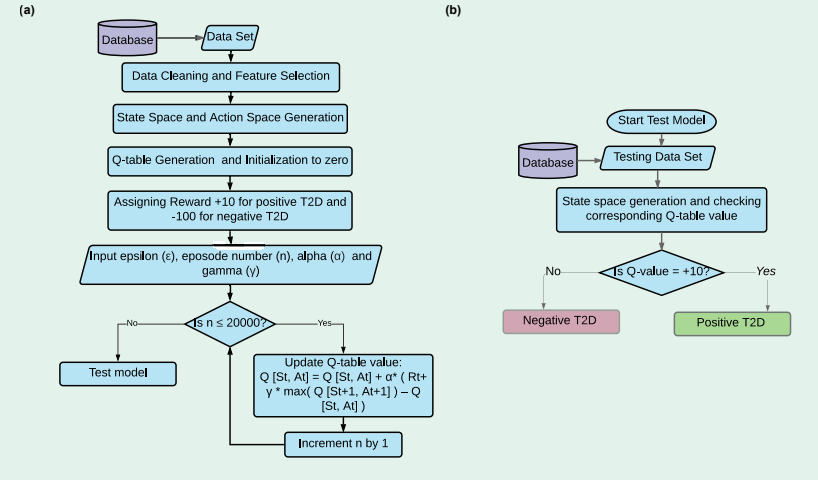

Fig. 1. (a) Q-learning model training flowchart, (b) Q-learning model testing flowchart

3) Q-table generation and initialization: A two-dimensional table having the same rows and columns like state space of q-model and initialized to zero.

4) State Rewards of Q-learning model: In the proposed system, the Q-Agent interacts with the environment moving from starting to the goal state collecting positive or negative rewards according to the state space. Reward +10 is assigned for positive T2D, while all the other states have reward -100.

5) Model Training Process: For predicting T2D based on Q-learning algorithm, we divided our data set into two parts, first one is the training data set (242 rows, 3 columns) and the second one is the testing data set (150 rows, 3 columns).

In the training procedure, the Q-learning agent picks up the optimal T2D prediction policy from the patient's clinical training dataset. In each episode, the Q-agent calculates and updates the q-values table for each combination of state and action. Exploring the state spaces starts from the initial state $S(1,1,1)$ and terminates at $S(11,5,6)$ completing a single episode. We used 20000 complete episodes in our process to draw an acceptable and reliable prediction outcome of T2D while minimizing the error rate and overfitting [29].

6) Model Evaluation: For evaluating the designed model 150 test data from PIMA Indian women diabetes dataset were used. To predict T2D first we calculate the state of the subject and then find q-table value for the corresponding state. If the found q-value is equal to +10 then the subject is considered to have a positive T2D. Otherwise, we level the subject as having a negative T2D.

B. Data Set

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) has created the Pima Indian Women Diabetes data set which contains medical diagnostic data input variables and predicts Type II diabetes (0 or 1) as the output variable [30]. The dataset includes 769 female subjects with 9 variables such as age, insulin level, blood pressure, BMI, skin thickness, blood glucose, diabetes pedigree function, and outcome.

TABLE II
Q-LEARNING RESULTS

BMI	Glucose	Age	Actual*	Q-Agent*	Comment
41.3	198	28	1	1	TP
36.7	197	31	1	1	TP
25.1	195	55	1	-1	FN
25.9	193	24	-1	-1	TN
31.2	189	29	1	1	TP
30.1	189	59	1	1	TP
36.4	187	36	1	1	TP
36.4	187	36	1	1	TP
...					
25	68	25	-1	-1	TN
24.2	56	22	-1	-1	TN

* Outcome

C. Performance Matrics

In an attempt to assess the performance of Q-learning, KNN and DT, confusion matrix, accuracy, precision, recall, f measure, receiver operating characteristic (ROC) curves and the areas under the curve (AUC) for each algorithm have been evaluated for the same dataset with 95% confidence intervals [15], [16], [21], [31].

D. Results and Discussion

The Q-learning model is tested using 150 subject's data from PIMA Indian women diabetes dataset where 59 subjects are labelled with positive T2D and 91 others are with negative T2D. To assess the performance of the process we considered the equivalent lifestyle information of an individual that was used in model building phase; Glucose, BMI and age. Table II shows BMI, Glucose, Age, Actual Outcome and Q-Agent Outcome of several subjects among 150 test data.

Confusion matrix of binary classifier is a two by two table that classifies all the test data into positive or negative class and made by counting the number of the four outcomes usually denoted as True Positives (TP), False Positives (FP),

True Negatives (TN), and False Negatives (FN) describing the accurate positive indicator, inaccurate positive indicator, accurate negative indicator and inaccurate negative indicator respectively. The confusion matrix of this work is presented in Figure 2 which is generated using 150 test data. Here we got zero FP errors including 35 TP, 24 FN and 91 TN classification.

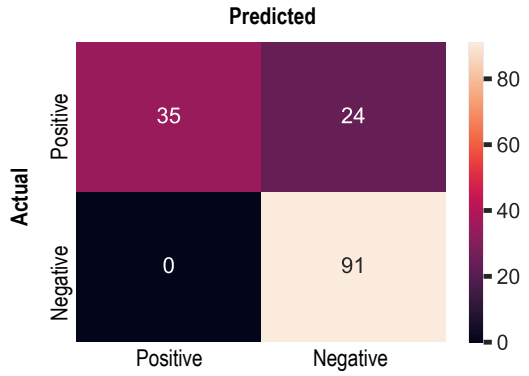

Fig. 2. Q learning classification confusion matrix.

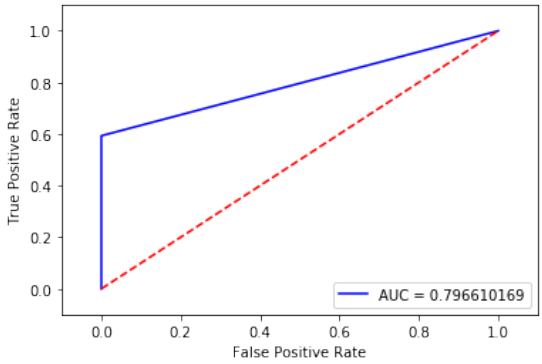

Fig. 3. ROC plot for Q-learning Model.

Figure 3 depicts the ROC curve of the Q-learning model. BMI, glucose level and age of a subject is very crucial for developing an effective T2D prediction model. In this work, we generated an off the policy-based reinforced environment and made our learning agent find an optimal policy to identify patient having T2D considering those 3 crucial factors. We designed the information of a subject in a way that the subject may take place in any of 330 possible states (number of Glucose levels $\times$ number of BMI levels $\times$ number of Age levels = 5 $\times$ 11 $\times$ 6 respectively). Our model identified T2D with an accuracy of 84%, 100% precision, 59% recall 74% f-measure.

The performance of the different ML algorithms are summarized in Table III. Table III indicates that KNN classifier performed below-average with an accuracy of 74.49% having the lowest precision (64.29%). On the other hand, the DT classifier performed with an accuracy rate of 75.51%, recall of 54.55% and f-measure of 1.2121, but with lower precision (66.67%).

TABLE III
SUMMARY OF T2D PREDICTION FOR DIFFERENT ALGORITHMS

Classifier	Accuracy	Precision	Recall	F-measure	AUC
KNN	0.7449	0.6429	0.5455	1.1883	0.7506
DT	0.7551	0.6667	0.5455	1.2121	0.7035
Q-learning	0.84	1.00	0.593	0.7445	0.7966

In contrast to these ML classifiers, the proposed Q-learning model outperforms over its counterpart and obtained an accuracy of 84%, precision of 100% and recall of 59.3%.

The ROC curve of KNN and DT are also shown in Figure 4 and Figure 5 respectively.

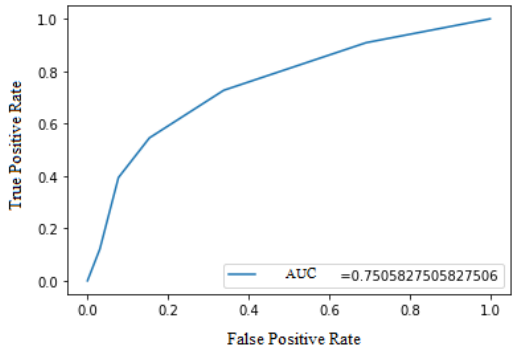

Fig. 4. ROC plot for K-Nearest Neighbors.

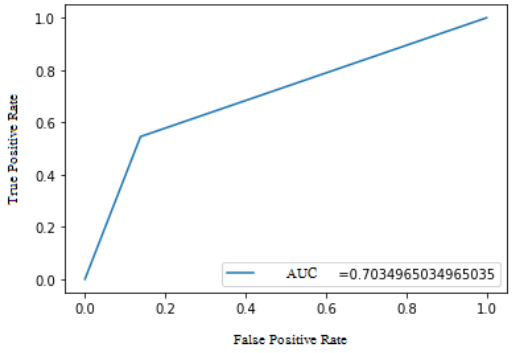

Fig. 5. ROC plot for Decision Tree.

Although in recent years, a rising interest has been noticed in applying machine learning based approaches in analyzing EHRs to identify disease and devise personalized treatment policies. This has enabled a new paradigm of personalised medicine. Towards that goal, this study aims for the the first in predicting T2D using RL model. Firstly, this study considered Glucose, BMI and age as input features that keep a significant rule in T2D outcome.

An effective and easy T2D identification model has been developed using these 3 vital lifestyle factors which was not

considered before in the literature. Secondly, this work can also identify prediabetic patient along with the patient having T2D. In this study, individual q-value for each 330 designed states is generated and considered a subject having T2D if and only if the corresponding q-value of the subject's state is +10. From the calculated q-value table we and also identify prediabetes patient. If a subject gets output of a positive but less than +10 q-value for his inputted state then we can consider the subject has developed prediabetes, which is the preliminary stage of developing T2D. Thirdly, the Q-learning model achieved better accuracy rate (84%) than the supervised ML approaches.Finally, this model predicted T2D with the highest precision (100%) which means the model has accurate positive prediction always. In model validation process there were no single subject's data that was actually negative but the model predicted positive. Minimizing FP rate in clinical prediction is very crucial as it may lead to a wrong diagnosis process.

VI. CONCLUSION

Correct identification of T2D can delay or prevent the complications of the disease like problems with heart and blood vessels, kidneys, eyes, nerves, skin, pregnancy, sleeping, hearing, brain and sex organs. It is extensively presumed that artificial intelligence techniques will profoundly change healthcare industries and reinforcement learning has established human-level or even better output and reliability in the field of medical data processing. This model can aid the medical specialist to reduce the treatment efficiency time as they can identify T2D only using some important lifestyle information of a subject with a zero false-positive error. Along with T2D prediction, this study can also identify the prediabetic condition of patients who are very close to the threshold level of developing positive T2D. Future research can extend this model including other factors responsible for causing diabetes-like insulin, diabetes pedigree function and blood pressure, and considering larger datasets. This model can also be easily used for the detection of other types of the disease using EHRs modifying the state space definition effectively.

REFERENCES

[1] "American diabetes association," (Date last accessed 10-Jan-2020). [Online]. Available: https://www.diabetes.org/resources/statistics/statistics-about-diabetes

[2] W. L and et al., "Prevalence and ethnic pattern of diabetes and prediabetes in china in 2013." *JAMA*, vol. 317, pp. 2515–2523, 2017.

[3] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE trans. neural netw. learn. syst.*, vol. 29, no. 6, pp. 2063–2079, 2018.

[4] H. M. Ali, M. S. Kaiser, and M. Mahmud, "Application of convolutional neural network in segmenting brain regions from mri data," in *Proc. Brain Informatics*. Springer, 2019, pp. 136–146.

[5] O. Orojo, J. Tepper, T. M. McGinnity, and M. Mahmud, "A Multi-recurrent Network for Crude Oil Price Prediction," in *Proc.2019 IEEE SSCI*, 2019, pp. 2953–2958.

[6] S. W. Yahaya, A. Lotfi, and M. Mahmud, "A consensus novelty detection ensemble approach for anomaly detection in activities of daily living," *Appl. Soft Comput.*, vol. 83, p. 105613, 2019.

[7] M. B. T. Noor *et al.*, "Detecting neurodegenerative disease from mri: A brief review on a deep learning perspective," in *Proc. Brain Informatics*. Springer, 2019, pp. 115–125.

[8] Y. Miah, C. N. E. Prima, S. J. Seema, M. Mahmud, and M. S. Kaiser, "Performance comparison of machine learning techniques in identifying dementia from open access clinical datasets," in *Proc. ICACIn*. Springer, Singapore, 2020, pp. 69–78.

[9] G. Rabby, S. Azad, M. Mahmud, K. Z. Zamli, and M. M. Rahman, "Teket: a tree-based unsupervised keyphrase extraction technique," *Cogn. Comput.*, 2020, doi: 10.1007/s12559-019-09706-3, [epub ahead of print].

[10] M. H. Tania *et al.*, "Assay type detection using advanced machine learning algorithms," in *2019 13th International Conference on SKIMA*, 2019, pp. 1–8.

[11] D. Silver and et al., "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, p. 484, 2016.

[12] T. M. N. U. Akhund *et al.*, "Adeptness: Alzheimer's disease patient management system using pervasive sensors-early prototype and preliminary results," in *International Conference on Brain Informatics*. Springer, 2018, pp. 413–422.

[13] A. I. Sumi, M. F. Zohora, M. Mahjabeen, T. J. Faria, M. Mahmud, and M. S. Kaiser, "fassert: A fuzzy assistive system for children with autism using internet of things," in *International Conference on Brain Informatics*. Springer, 2018, pp. 403–412.

[14] M. S. Kaiser *et al.*, "Advances in crowd analysis for urban applications through urban event detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3092–3112, 2018.

[15] V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus — a machine learning approach," in *Proc. 2015 IEEE RAICS*, 2015, pp. 122–127.

[16] S. Hina, A. Shaikh, and S. A. Sattar, "Analyzing diabetes datasets using data mining." *Basic & Applied Sciences*, vol. 13, pp. 466–471, 2017.

[17] R. M. Khalil and A. Al-Jumaily, "Machine learning based prediction of depression among type 2 diabetic patients," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nov 2017, pp. 1–5.

[18] I. Fox and J. Wiens, "Reinforcement learning for blood glucose control: Challenges and opportunities," *ICML 2019 Workshop RL4RealLife*, 2019.

[19] M. Mahmud, M. S. Kaiser, and A. Hussain, "Deep learning in mining biological data," *arXiv*, vol. abs/2003.00108, pp. 1–36, 2020.

[20] O. M. Javad and et al., "Reinforcement learning algorithm for blood glucose control in diabetic patients," in *ASME Conference Publications and Proceedings*, 11 2015, p. V014T06A009.

[21] J. N. Myhre, I. K. Launonen, S. Wei, and F. Godtliebsen, "Controlling blood glucose levels in patients with type 1 diabetes using fitted q-iterations and functional features," in *Proc. 2018 IEEE MLSP*, Sep. 2018, pp. 1–6.

[22] S. G. Nejad, R. Martens, and R. Paranjape, "An agent-based diabetic patient simulation," in *Agent and Multi-Agent Systems: Technologies and Applications*, N. T. Nguyen, G. S. Jo, R. J. Howlett, and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 832–841.

[23] Y. Ling, S. A. Hasan, V. Datla, A. Qadir, K. Lee, J. Liu, and O. Farri, "Diagnostic inferencing via improving clinical concept extraction with deep reinforcement learning: A preliminary study," in *MLHC*, 2017.

[24] G. Colditz *et al.*, "Weight as a risk factor for clinical diabetes in women." *Am J Epidemiol*, vol. 132, pp. 501–13, 1990.

[25] S. Helmrich, D. Ragland, R. Leung, and R. J. Paffenbarger, "Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus." *N Engl J Med*, vol. 325, pp. 147–52, 1991.

[26] L. S, M. JE, and S. M. et al., "A prospective study of wholegrain intake and risk of type 2 diabetes mellitus in us women." *Am J Public Health*, vol. 90, pp. 1409–15, 2000.

[27] "K-nearest neighbours," (Date last accessed 10-Jan-2020). [Online]. Available: https://www.geeksforgeeks.org/k-nearest-neighbours/

[28] D. V. Patil and R. S. Bichkar, "Article: Issues in optimization of decision tree learning: A survey," *Int. J. Appl. Inf. Syst.*, vol. 3, no. 5, pp. 13–29, July 2012.

[29] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering Cambridge, UK, 1994, vol. 37.

[30] NIDDK, "Pima Indians Diabetes Database," 2016, library Catalog: www.kaggle.com. [Online]. Available: https://kaggle.com/uciml/pima-indians-diabetes-database

[31] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation," *J. Mach. Learn. Tech.*, vol. 2, pp. 37–63, 01 2011.