According to Figure 4, after defining the problem, we have to collect the relevant data from the Diagnostic Data Storage. We then pre-process the data to build the prediction model. After that, various machine learning techniques are applied to the training dataset. Finally, the test dataset is used to measure the performance of the techniques for choosing the best classifier to predict diabetes mellitus.

D. Risk Factors Correlation

Progression of diabetes mellitus is strongly correlated to several complications, the leading causes of chronic kidney and blood pressure disease. It is well known that DM covers a wide area of different pathophysiological conditions. The most widely common complications are divided into micro and macrovascular disorders, including diabetic nephropathy, retinopathy, neuropathy, and cardiovascular disease. Because of high DM increase mortality and morbidity. Its related complications need to be prevented. That's why it is essential to eliminate several risk factors related to long term diabetes complications; as a result, longevity can be increase. A correlation measures the relationship between two or more variables indicating the risk factors of DM. To evaluate the correlation between different risk factors of DM, statistical correlation [5] can be applied to the data set attributes that we have considered of chronic kidney disease, blood pressure, hearing loss, and skin problem over the diabetes disease.

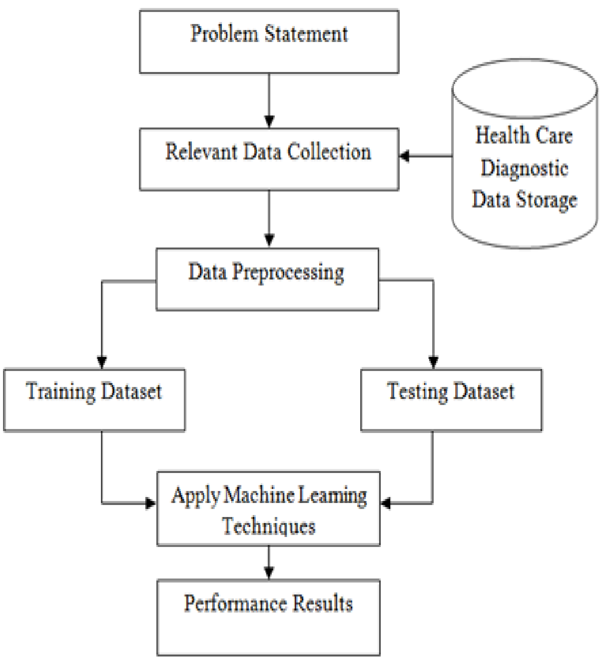


Figure 4. The overall process of our work.

• *Correlation between diabetes and Kidney disease:*

One reason for chronic kidney disease is diabetes mellitus, which is characterized by high blood glucose (sugar) levels. Over time, the high amounts of sugar in the blood vessels that harm millions of small filtering units in the kidney. It is, in the long run, prompts to kidney failure.
Approximately 20 to 30 percent of people with diabetic enlarge to kidney disease (diabetic nephropathy), although not all of these people progress to kidney failure. A person with diabetes is at risk of nephropathy, whether they depend on insulin or not. The risk is correlated to the length of time the person has diabetes. There is no heal for diabetic nephropathy, and therefore the treatment is life-long. Diabetic peoples are also at risk of other kidney problems, including narrowing of the arteries to the kidneys called renal artery renovascular disease.

To build a correlation between chronic kidney disease and diabetic patients, the attributes that we have considered are- age, sex, blood pressure, itching, vomiting, trouble sleeping, chest pain, smoking, heart disease, loss of appetite, too much urine, breath problem, and family history. The sample clinical dataset of diabetic nephropathy patients is shown in Table 9.

Table 9. Sample dataset of diabetic nephropathy patients.

SI	Age	Sex	Itching	Smoking	Vomiting	Sleep_Prob.	Chest Pain	Heart_Dise.	Loss of Appet.	Polyuria	Family Hist.	BP (systole/diastole)	CKD (S.Cret.)	Diabetes (BM/AM)
1	62	Male	Yes	No	No	No	No	No	No	Yes	Yes	128/85	1.4	130/220
2	53	Female	No	No	No	Yes	No	No	No	No	Yes	125/80	0.8	70/150
3	45	Female	No	No	No	Yes	No	No	No	No	No	130/85	1.8	65/155
4	20	Male	No	Yes	No	No	No	No	No	No	No	135/85	1.5	135/250
5	42	Female	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	145/98	2.1	125/230
6	48	Male	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	130/85	1.8	160/270
7	46	Female	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	150/90	2.1	145/245
8	60	Male	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	90/65	1.9	130/210
9	18	Male	No	Yes	No	No	No	No	No	No	No	120/80	2.2	162/245
10	66	Male	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	135/85	1.3	125/224
11	61	Male	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	130/85	1.6	147/250
12	10	Male	No	Yes	No	No	Yes	Yes	No	No	Yes	95/62	1.8	136/256
13	71	Male	Yes	No	No	No	No	No	No	Yes	No	120/80	1.7	150/243
14	69	Male	Yes	No	No	No	No	No	Yes	Yes	No	125/82	1.2	145/240
15	43	Female	Yes	No	No	Yes	No	No	Yes	Yes	No	130/85	2.2	138/230
16	64	Female	No	No	No	No	Yes	Yes	Yes	No	Yes	130/80	0.9	78/135
17	74	Male	No	Yes	No	No	Yes	Yes	Yes	No	Yes	125/80	1.0	80/140
18	49	Female	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	150/90	2.2	150/255

- *Correlation between diabetes and blood pressure disease:*

Many people with diabetes mellitus also have hypertension or blood pressure disease. Blood pressure disease is known as a "silent killer" since it often has no cleared symptoms, and many people are uninformed they have it. According to the 2013 review [36], the American Diabetes Association (ADA) found that a combination of hypertension and diabetes mellitus is particularly deadly and can significantly raise the risk of having a heart attack

or stroke. A person with diabetes must control blood pressure.
In our comparative study to build a correlation between blood pressure disease and diabetic patients the attributes that we have considered are- age, sex, occupation, smoking, blood pressure (both systolic and diastolic), pulse rate, drink, family member, salt in diet, murmur, and cholesterol. The sample clinical dataset of diabetes blood pressure patients is shown in Table 10.

Table 10. Sample dataset of diabetic blood pressure patients.

SI	Age	Sex	Occupation	Smoking	Drink	Salt in Diet	Murmur	Family_Hist.	BP (systole/diastole)	Pulse_Rate	Cholesterol	CKD (S.Cret.)	Diabetes (BM/AM)
1	62	Male	Business	No	No	No	No	Yes	128/85	No	No	1.4	130/220
2	53	Female	Housewife	No	No	Yes	No	Yes	125/80	No	No	0.8	70/150
3	45	Female	Housewife	No	No	Yes	No	No	130/85	No	No	1.8	65/155
4	20	Male	Student	Yes	No	No	No	No	135/85	No	No	1.5	135/250
5	42	Female	Housewife	No	No	Yes	Yes	No	145/98	Yes	Yes	2.1	125/230
6	48	Male	Accountant	Yes	No	Yes	No	Yes	130/85	No	Yes	1.8	160/270
7	46	Female	Housewife	No	No	Yes	Yes	Yes	150/90	Yes	No	2.1	145/245
8	60	Male	Business	Yes	No	No	Yes	No	90/65	Yes	Yes	1.9	130/210
9	18	Male	Student	Yes	No	No	No	No	120/80	No	No	2.2	162/245
10	66	Male	Retired	No	No	No	Yes	Yes	135/85	Yes	Yes	1.3	125/224
11	61	Male	Business	Yes	No	Yes	No	Yes	130/85	No	Yes	1.6	147/250
12	10	Male	Student	Yes	No	No	Yes	Yes	95/62	Yes	No	1.8	136/256
13	71	Male	Retired	No	No	No	No	No	120/80	No	No	1.7	150/243
14	69	Male	Retired	No	No	No	No	No	125/82	No	Yes	1.2	145/240
15	43	Female	Housewife	No	No	Yes	No	No	130/85	No	Yes	2.2	138/230
16	64	Female	Housewife	No	No	No	Yes	Yes	130/80	Yes	Yes	0.9	78/135
17	74	Male	Retired	Yes	No	No	Yes	Yes	125/80	Yes	Yes	1.0	80/140
18	49	Female	Housewife	No	No	Yes	Yes	No	150/90	Yes	Yes	2.2	150/255
19	52	Female	Housewife	No	No	Yes	Yes	Yes	145/95	Yes	No	1.0	166/270
20	63	Male	Business	Yes	No	No	No	No	130/80	No	Yes	0.7	82/130
21	58	Female	Housewife	No	No	No	No	No	125/80	No	No	0.5	73/110
22	61	Male	Service Holder	Yes	No	No	Yes	No	95/65	Yes	No	1.3	86/140

- *Correlation between diabetes and hearing loss disease:*

Diabetes is associated with a risk of hearing loss. Type 2 diabetes may be an independent risk factor for hearing loss. Because high blood sugar effects of hyperglycemia may damage the cochlea. Signs and symptoms that commonly occur in Type 2 diabetes can be related to the immediate effects of hyperglycemia or hypoglycemia (blurred vision and excessive thirst, for example). Many patients may not realize the relation between their hearing impairment and their diabetes condition. According to the National Institutes of Health [39], Hearing loss is common in people with diabetes. To build a correlation between hearing loss and diabetic patients the attributes that we have considered are - age, sex, weight, diet, polyuria, water consumption, excessive thirst, blood pressure, hypertension, tiredness, the problem in vision, kidney problem, hearing loss, skin problem, genetic and diabetic. The sample clinical dataset of diabetes hearing loss patients is shown in Table 11.

- *Correlation between diabetes and skin problem disease:*

Long term Type 2 diabetes with hyperglycemia or high blood glucose, tends to be associated with poor circulation, which decreases blood stream to the skin. It can also affect blood vessels and nerves. The capacity of the white platelets to fend off infections is also decreased in the face of elevated blood sugar. Diminished blood circulation can prompt changes in the skin's collagen. It changes the skin's texture, appearance, and ability to heal.

Table 11. Sample dataset of diabetic hearing loss patients.

SI	Age	Sex	Weight	Diet	Polyuria	Water_Cons	Excessive_Thirst	BP	Hyp_Ten	Tiredness	Problem_in_Vision	Kidney_Problem	Hearing_Loss	Skin_Problem	Genetic	Glucose_Level_PGP	Diabetic
1	62	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	142	Yes
2	53	Female	60	Yes	Yes	No	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	97	No
3	45	Female	55	Yes	Yes	Yes	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	80	No
4	67	Male	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	No	No	Yes	Yes	167	Yes
5	42	Female	52	No	No	No	No	Normal	No	No	No	Yes	No	No	No	172	Yes
6	48	Male	66	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	145	Yes
7	54	Female	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	Yes	No	Yes	Yes	148	Yes
8	60	Male	66	Yes	Yes	Yes	Yes	Low	No	Yes	Yes	Yes	Yes	Yes	No	78	No
9	50	Male	68	No	No	No	No	High	Yes	No	No	Yes	No	No	No	95	No
10	66	Male	62	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	No	Yes	Yes	Yes	156	Yes
11	61	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	141	Yes
12	46	Female	54	No	No	No	No	High	Yes	Yes	No	Yes	No	No	Yes	185	Yes
13	71	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	95	No
14	69	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	88	No
15	43	Female	64	No	No	No	No	Normal	No	No	No	No	No	No	No	158	Yes
16	64	Female	61	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	92	No
17	74	Male	73	Yes	Yes	Yes	Yes	Normal	No	Yes	No	No	Yes	Yes	Yes	88	No

Diabetes can hurt the body skin in two ways:
1. If blood glucose is high, at that point, the body loses liquid or fluid. With less fluid in the body, result skin can get dry. Dry skin can be itchy, causing the body to scratch and make it sore. Cracks cause infections and allow germs. It increases as the blood sugar becomes high. The skin of the legs, feet, elbows, and other places in the body can get dry [52].
2. Nerve damage can decrease the amount of body sweat. Sweating may cause dry skin in the feet and legs [52].

To build a correlation of diabetic patients, the attributes that we have to consider of a skin problem patients are age, sex, weight, diet, polyuria, water consumption, excessive thirst, blood pressure, hypertension, tiredness, the problem in vision, kidney problem, hearing loss, skin problem, genetic and diabetic. The sample clinical dataset of diabetic skin problem patients is shown in Table 12.

4. Experimental Results and Discussion

After the implementation of different Machine Learning models, the next step is to measure the performance of the classification techniques.
It is done by running the models on the test dataset, which was set aside earlier. The test dataset comprised of the original data for diabetic patients. N-fold (N-10) cross-validation technique was done for cardiovascular disease of the original data in the test dataset. To measure the performance N-fold (N=10) cross-validation [53] technique can be used. Cross-validation techniques can follow the following mechanism:

a) The test dataset is separated into N-folds, where each fold is used for classifying the testing and training data to predict the model.
b) Repeats N times until completing the procedure for the testing and training data.

Table 12. Sample dataset of diabetic skin problem patients.

SI	Age	Sex	Weight	Diet	Polyuria	Water_Cons	Excessive_Thirst	BP	Hyp_Ten	Tiredness	Problem_in_Vision	Kidney_Problem	Hearing_Loss	Skin_Problem	Genetic	Glucose_Level_PGP	Diabetic
1	62	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	142	Yes
2	53	Female	60	Yes	Yes	No	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	97	No
3	45	Female	55	Yes	Yes	Yes	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	80	No
4	67	Male	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	No	No	Yes	Yes	167	Yes
5	42	Female	52	No	No	No	No	Normal	No	No	No	Yes	No	No	No	172	Yes
6	48	Male	66	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	145	Yes
7	54	Female	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	Yes	No	Yes	Yes	148	Yes
8	60	Male	66	Yes	Yes	Yes	Yes	Low	No	Yes	Yes	Yes	Yes	Yes	No	78	No
9	50	Male	68	No	No	No	No	High	Yes	No	No	Yes	No	No	No	95	No
10	66	Male	62	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	No	Yes	Yes	Yes	156	Yes
11	61	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	141	Yes
12	46	Female	54	No	No	No	No	High	Yes	Yes	No	Yes	No	No	Yes	185	Yes
13	71	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	95	No
14	69	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	88	No
15	43	Female	64	No	No	No	No	Normal	No	No	No	No	No	No	No	158	Yes
16	64	Female	61	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	92	No
17	74	Male	73	Yes	Yes	Yes	Yes	Normal	No	Yes	No	No	Yes	Yes	Yes	88	No

c) According to N-folds cross-validation, we partition the data into 10-folds, where each fold is nearly the same with other folds in the dataset.
d) Execution of each iteration contains nine folds as a training set to adapt the model, and the remaining 1 fold known as the testing set is used for evaluating performance.
e) In the end, learning scheme techniques performed 10 times on training data sets, and lastly, the prediction accuracy averages for 10 data sets. Various performance metrics, such as precision, recall, F-measure, and accuracy, are described as follow:

A. Evaluation Metric

If TP belongs to true-positive rate and FP belongs to false-positive rate then according to [53] the formal definition of precision is in equation (3),

$$\text{Precision} = \frac{TP}{TP+FP} \text{................(3)}$$

Furthermore, recall is defined as below where FN represents the false-negative rate [53] and represented in equation (4)

$$\text{Recall} = \frac{TP}{TP+FN} \text{..................(4)}$$

The F-measure can be evaluated using the value of precision and recall and defined as below [53] and represented in equation (5)

$$\text{F-measure} = \frac{2 * \mathrm{Re}\,call * \mathrm{Pr}\,ecision}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \text{............(5)}$$

On top of that, we also calculate the accuracy based on the correctly classified instances performed by machine learning techniques. Formally, Accuracy is calculated as below [53] and represented in equation (6)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{.....(6)}$$

B. Comparison Results

The performance of different machine learning techniques has been shown in Table 13, based on precision, recall, and F-measure. The table shows the results of various machine learning techniques such as SVM, NB, KNN, and C4.5.As information gain helps to construct the trees with attributes of the highest gain to lowest in a downward fashion, it is evident that C4.5 achieves better results than other classifiers to predict diabetes mellitus. According to Figure 6, C4.5 achieves 72% precision, 74% recall, and 72% F-measure on this dataset, which is higher than other learning techniques. This experimental result provides evidence that C4.5 Decision Tree performs well on medical datasets to predict diabetes mellitus based on various risk factors discussed in the earlier section.

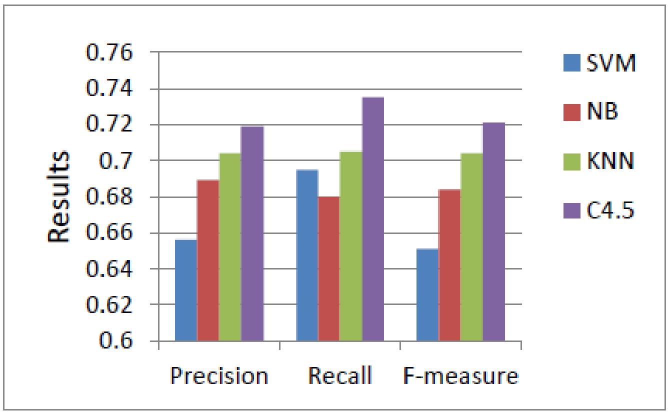


Figure 5. Predictions Results of Various Machine Learning Techniques

In addition to precision, recall, and F-measure, we also calculate the direct accuracy rate in the percentage of all these classifiers shown in Figure 7. If we observe Figure 5, we see that the C4.5 decision tree technique outperforms other techniques to predict Diabetes Mellitus.

Table 13. Comparison of prediction results of various machine learning techniques

Algorithm	Precision	Recall	F-measure
SVM	0.66	0.70	0.65
NB	0.69	0.68	0.68
KNN	0.70	0.71	0.70
C4.5	0.72	0.74	0.72

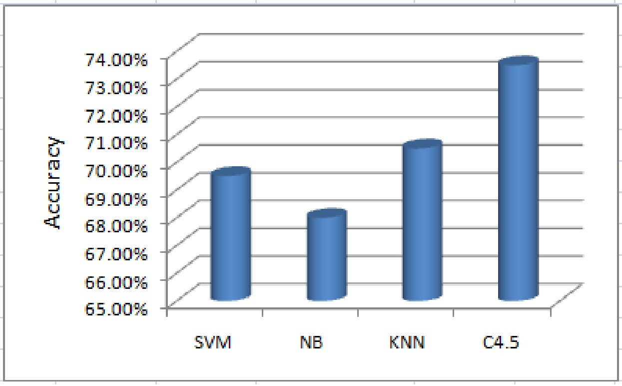


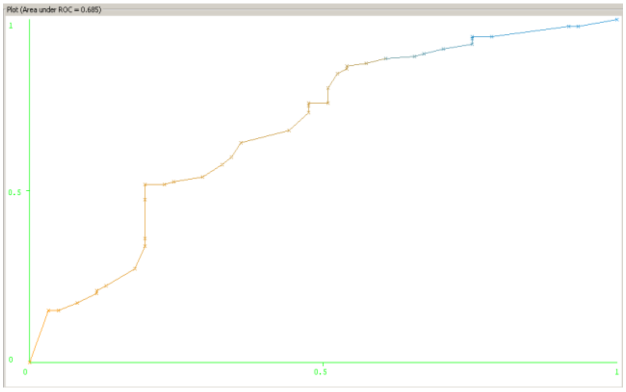
Figure 6. Accuracy Results of Various Machine Learning Techniques

Based on Figure 6, it can be observed that C4.5 Decision Tree achieves a better accuracy of 73.5% to predict diabetes mellitus utilizing a given medical dataset.
In the results, Area under the Receiver Operating Characteristic (ROC) curve of the SVM, NB, KNN, and C4.5 Decision Tree algorithms are 0.56, 0.65, 0.67 and 0.69 respectively, which is shown in Table 14. From Figure 7, the confidence band of the curve is clearly shown for C4.5 Decision Tree rather than other techniques. With the features of the information gain criterion, C4.5 Decision Tree achieves better accuracy for ROC. The experimental results prove that for the diabetic dataset, the area under ROC for the C4.5 algorithm performs best in four learning techniques is 0.69.

Table 14. Comparison of AUC of the four models

Algorithm	The area under the ROC Curve(AUC)
SVM	0.56
NB	0.65
KNN	0.67
C4.5	0.69

To determine the correlation between different risk factors of diabetes mellitus, we collect the dataset consists of various attributes or risk factors of kidney disease and diabetes mellitus of 200 diabetic patients. In the diagnostic dataset results, blood glucose levels (after the meal) are significantly increased for diabetic patients. Serum creatinine levels observed significantly low in non-diabetic patients and high in diabetic patients. A positive



correlation (0.72) is found between serum creatinine and blood glucose level for diabetic patients, which shown in Figure 8.

Figure 7. ROC Curve for C4.5 decision tree

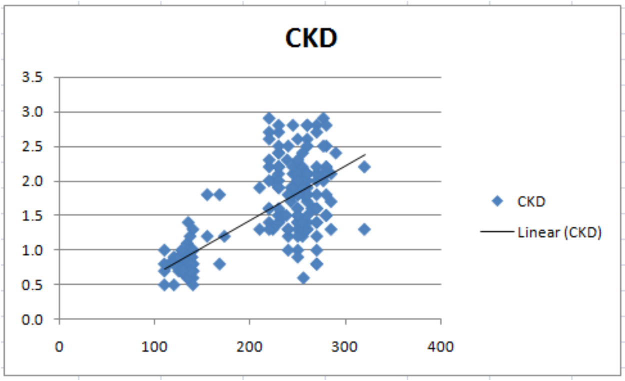


Figure 8. Correlation between Kidney disease and diabetic patients

Diabetes mellitus and blood pressure frequently coexist. Formation o blood glucose (after and before the meal) and blood pressure (systole/diastole) have recently reported as disease markers for diabetes and hypertension, respectively. This study is aimed to find the correlation between diabetes mellitus and blood pressure. An equal number of people of different ages and sex are selected to test. Their blood glucose (after and before the meal), serum creatinine, pulse rate, and cholesterol levels are measured by spectrophotometer. The dataset attributes are correlated by statistical methods. Notably, we see that blood glucose (after and before the meal) levels, as well as blood pressure (systole/diastole) levels, are significantly high for diabetic hypertensive patients. A significant positive correlation (0.81) is found between blood glucose levels and BP levels in diabetic hypertensive patients, shown in Figure 9. These findings suggest that the combination of hypertension and diabetes can be deadly, and together they can enhance the risk of a heart attack or stroke.

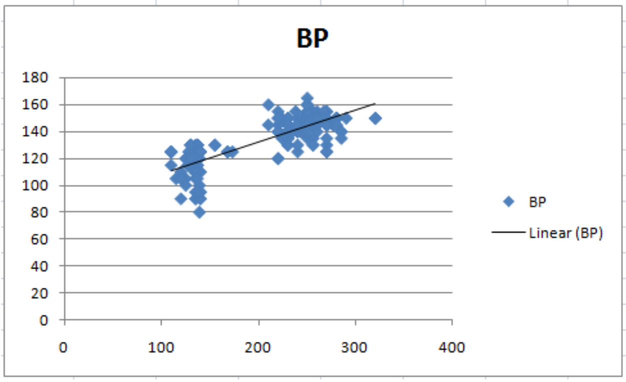


Figure 9. Correlation between blood pressure and diabetic patients

We have analyzed the correlation between diabetes mellitus and hearing loss patients. A negative correlation (-0.72) is found between blood glucose levels and diabetic hearing loss patients, shown in Figure 10. These results suggest that hearing loss and diabetes mellitus are comparatively weak correlated.

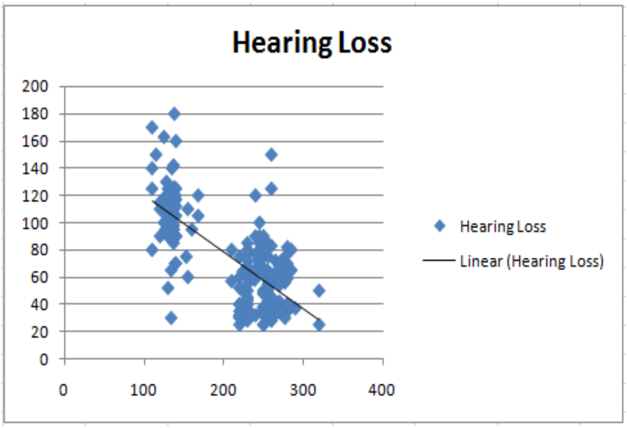


Figure 10. Correlation between hearing loss and diabetic patients

We also evaluate the correlation between diabetes mellitus and skin problems. A negative correlation (-0.76) is found between blood glucose levels and diabetic skin problem patients, shown in Figure 11. These results suggest that skin problem and diabetes mellitus is nearly weak correlated.

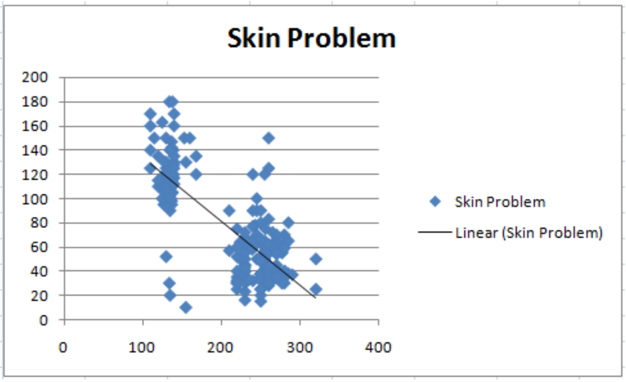


Figure 11. Correlation between skin problem and diabetic patients

In the future, we can collect more data and make decisions based on their correlation with other diseases respective by males and females by considering the concept of recent pattern analysis [57] for building more effective models.

5. Conclusion

In this work, we have explained how Machine Learning can be adopted in clinical diagnostics to predict the probability of diabetes-induced complications. It is done using different Machine Learning algorithms under various circumstances. Knowledge extraction from real health care dataset can be useful to predict diabetes mellitus. To predict diabetes mellitus effectively, we have performed our experiments using four popular machine learning algorithms, such as Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbour (KNN) and C4.5 Decision Tree on the adult population. From the experimental results, we can make the decision that C4.5 Decision Tree is significantly superior to other machine learning techniques on diabetes data. We also find a positive correlation at predicting kidney complications (Nephropathy) and blood pressure (Hypertension) complications and find a negative correlation at predicting hearing loss and skin complications (diabetes dermopathy) from the diabetic patients. For the study, we have collected a diagnostic dataset having 16 attributes diabetic of 200 patients. The experimental results assist the health care centre to make better clinical decisions on diabetes. It is also helpful for individuals to control diabetes.

Acknowledgments

It is our gratitude to give thanks and acknowledge to Dr. Shahed Parvez, Director, Medical Centre Chittagong, and Dr. Mohammad Maruf Faruqi, RMO, ICU, Medical Centre Chittagong, Bangladesh for providing the real diagnostic data and for their cooperation throughout our work.

References

[1] Morteza, M., Franklyn, P., Bharat, S., Linying, D., Karim, K., and Aziz G. 2015. Evaluating the Performance of the Framingham Diabetes Risk Scoring Model in Canadian Electronic Medical Records. Canadian Journal of diabetes 39, 30(April. 2015), 152-156.

[2] V., A. K., and R., C., 2013. Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. 3, (April. 2013), 1797-1801.

[3] Carlo, B G., Valeria, M., and Jesús, D. C., 2011. The impact of diabetes mellitus on healthcare costs in Italy. Expert review of pharmacoeconomics & outcomes research. 11, (Dec. 2011),709-19.

[4] Nahla B., Andrew, P. B., and M., N. B., 2010. Intelligible support vector machines for diagnosis of diabetes mellitus. Information Technology in Biomedicine, IEEE Transactions. 14, (July. 2010), 1114-20.

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

[6] Akinnola N. AKINTUNDE "Path Analysis Step by Step Using Excel."

[7] G. Sparacino, F. Zanderigo, S. Corazza, A Maran, A Facchinetti, and C. Cobelli. "Glucose concentration can be predicted in time from continuous glucose monitoring sensor time-series. Biomedical Engineering", IEEE Transactions on, 54(5): 931{937, May 2007. ISSN 0018-9294. DOI: 10.1109/TBME.2006.889774.

[8] G. Baghdadi and AM. Nasrabadi. "Controlling blood glucose levels in diabetes by neural network predictor. In Engineering in Medicine and Biology Society", 29th Annual International Conference of the IEEE, pages 3216{3219, Aug 2007. DOI: 10.1109/IEMBS.2007.4353014.

[9] C. Marling, M. Wiley, R. Bunescu, J. Shubrook, and F. Schwartz. "Emerging applications for intelligent diabetes management." Artificial Intelligence Magazine, 33(2):67, 2012.

[10] E. Georga, V. Protopappas, D. Polyzos, and D. Fotiadis. "Predictive modeling of glucose metabolism using free-living data of type 1 diabetic patients. In Engineering in Medicine and Biology Society (EMBC)", 2010 Annual International Conference of the IEEE, pages 589{592, Aug 2010. DOI: 10.1109/IEMBS.2010.5626374.

[11] Abdullah A. Aljumah, "Application of data mining: Diabetes health care in young and old patients, Journal of King Saud University" - Computer and Information Sciences, Volume 25, Issue 2, July 2013, Pages 127-136.

[12] Kavakiotis, Ioannis, Olga Tsave, AthanasiosSalifoglou, NicosMaglaveras, IoannisVlahavas, and IoannaChouvarda. "Machine learning and data mining methods in diabetes research." Computational and structural biotechnology journal (2017).

[13] Zheng, Tao et al. "A machine learning-based framework to identify type 2 diabetes through electronic health records." International journal of medical informatics 97 (2017): 120- 127.

[14] Rani, A. Swarupa, and S. Jyothi. "Performance analysis of classification algorithms under different datasets." In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on, pp. 1584-1589. IEEE, 2016.

[15] Kandhasamy, J. Pradeep, and S. Balamurali. "Performance analysis of classifier models to predict diabetes mellitus." Procedia Computer Science 47 (2015): 45-51.

[16] Y. Huang, P. McCullagh, N. Black, R. Harper, "Feature selection and classification model construction on type 2 diabetic patients 'data", Artificial Intelligence in Medicine 41 (3) (2015) 251–262.

[17] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors." The Kaohsiung journal of medical sciences, 29(2), 93-99.

[18] Thirumal, P. C., & Nagarajan, N. "Utilization of data mining techniques for the diagnosis of diabetes mellitus-a case study." ARPN Journal of Engineering and Applied Science, 10(2015).

[19] A. Al Jarullah, "Decision tree discovery for the diagnosis of type II diabetes, in Innovations in Information Technology (IIT)," 2011 International Conference on, 2011, pp. 303–307.

[20] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," Johns Hopkins APL Technical Digest 10 (1988) 262–266.

[21] Komi, Messan, Jun Li, YongxinZhai, and Xianguo Zhang. "Application of data mining methods in diabetes prediction." In Image, Vision, and Computing (ICIVC), 2017 2nd International Conference on, pp. 1006-1010. IEEE, 2017.

[22] Xu, Weifeng, Jianxin Zhang, Qiang Zhang, and Xiaopeng Wei. "Risk prediction of type II diabetes based on the random forest model."In Advances in Electrical, Electronics, Information, Communication, and Bio-Informatics (AEEICB), 2017 Third International Conference on, pp. 382-386. IEEE, 2017.

[23] Song, Yunsheng, Jiye Liang, Jing Lu, and Xingwang Zhao. "An efficient instance selection algorithm for k nearest neighbor regression." Neurocomputing 251 (2017): 26-34.

[24] Perveen, Sajida, Muhammad Shahbaz, Aziz Guergachi, and Karim Keshavjee. "Performance analysis of data mining classification techniques to predict diabetes." Procedia Computer Science 82 (2016): 115-121.

[25] Pradeep, K. R., and N. C. Naveen. "Predictive analysis of diabetes using the J48 algorithm of classification techniques." In Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on, pp. 347-352. IEEE, 2016.

[26] Santhanam, T., and M. S. Padmavathi. "Application of K-means and genetic algorithms for dimension reduction by integrating SVM for a diabetes diagnosis." Procedia Computer Science 47 (2015): 76-83.

[27] Meza-Palacios, Ramiro, Alberto A. Aguilar-Lasserre, Enrique L. Ureña-Bogarín, Carlos F. Vázquez-Rodríguez, Rubén Posada-Gómez, and Armín Trujillo-Mata. "Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus." Expert Systems with Applications 72 (2017): 335-343.

[28] Bashir, Saba, Usman Qamar, Farhan Hassan Khan, and M. YounusJaved. "An Efficient Rule-Based Classification of

Diabetes Using ID3, C4. 5, & CART Ensembles." In Frontiers of Information Technology (FIT), 2014 12th International Conference on, pp. 226-231. IEEE, 2014.

[29] K. Meena, N. Vijayalakshmi, "An Analysis of Risk Factor for Diabetes using Data Mining Approach," Indian Journal of Public Health Research and Development, Vol. 6, Issue No. 2, pp 112-117, April-June 2015.
[30] Varsha Kavi and Divyesh Joshi, "A Survey on Enhancing Data Processing of Positive and Negative Association Rule Mining," International Journal of Computer Sciences and Engineering, Volume-02, Issue-03, Page No (139-143), Mar -2014.
[31] J. Lindstrom and J. Tuomilehto, "The Diabetes Risk Score: A practical tool to predict type 2 diabetes risk," Diabetes Care, 26:3 (2003), 725-731.
[32] Ajay Meshram, Karuna Kachhawa, Vijay Gujar, Pradeep Bokariya-"Correlation Of Dyslipidemia And Type 2 Diabetes Mellitus Amongst The People Of Vidarbha Region Of India". IOSR Journal Of Pharmacy, (e)-ISSN: 2250-3013, (p)-ISSN: 2319-4219 Volume 6, Issue 1 (January 2016), PP. 45-50.
[33] Mustafa Z. Mahmoud, Omer A. Mahmoud, Maram A. Fagiri - "Chronic renal failure secondary to diabetes mellitus." Int. J Case Rep Images 2017;8 (2):124–128.
[34] Dabla PK.- "Renal function in diabetic nephropathy," World Journal Diabetes 2010 May 15;1(2):48–56.
[35] Remuzzi G, Schieppati A, Ruggenenti P. Clinical practice, "Nephropathy in patients with type 2 diabetes", International Journal of Data Mining 2012 Apr 11;346(15):1145–51.
[36] P. Pawelczak, R. Venkatesha –"American Heart Association, Heart diseases, and Stroke Statistics," IEEE International Conference on Machine Learning, Cape Town, South Africa, pp. 1-5, 2010 2009.
[37] Stokes, J., W.B. Kannel, P.A. Wolf, R.B. D'Agostino, and L.A. Cupples, 1989. "Blood pressure as a risk factor for cardiovascular disease: The Framingham Study 30 years of follow-up, Hypertension", IEEE International Conference on hypertension patient 31: I13-I18.
[38] Ersin Elbasi, "Determination of Diabetic Patient's Hearing Sensitivity using Data Mining Techniques," IEEE Transaction on Evolutionary Computation, Special Issue on Artificial Immune System, Volume 6, Issue 3, pp. 239-251, 2012.
[39] Ebrahim Darvishi, "Prediction of diabetes hearing loss patients using machine learning approach," JNCI Journal of National Cancer Inst., Volume 99, Issue 4, pp.268-289, 2017.
[40] Xie, J., & Wang, C., "Classification of Skin Disease using Ensemble Data Mining Techniques," International Research Journal of Engineering and Technology, 2(8), 1544-1547., 2017.
[41] Michelle Duff, "Cutaneous Manifestations of Diabetes Mellitus," In 2017 IEEE Imaging Systems and Techniques (IST) (pp. 1-5).
[42] Platt, John C. "12 fast training of support vector machines using minimal sequential optimization." Advances in kernel methods (1999): 185-208.
[43] Cortes, C., Vapnik, V., "Support-vector networks," Machine Learning, 20(2), pp. 273-297, 1995.
[44] Christopher J.C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery," Springer, 2(2), pp.121-167, 1998.
[45] V. Vapnik, "The Nature of Statistical Learning Theory." NY: Springer- Verlag. 1995.
John, George H., and Pat Langley. "Estimating continuous distributions in Bayesian classifiers." Proceedings of the Eleventh Conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1995.
[46] Zhang, H. (2004). "The optimality of Naive Bayes." International conference artificial intelligence, 1(2), 3.
[47] Aha, David W., Dennis Kibler, and Marc K. Albert. "Instance-based learning algorithms." Machine learning 6.1 (1991): 37-66.
[48] Alpaydin, E. (1997), "Voting over Multiple Condensed Nearest Neighbors," Artificial Intelligence Review, p. 115–132.
[49] Jiang L, Li C, Cai Z. "Learning the decision tree for ranking. Knowledge and Information Systems" in IEEE INFOCOM, pp. 400-406, 2002
[50] Ross Quinlan (1993). "C4.5: Programs for Machine Learning". Morgan Kaufmann Publishers, San Mateo, CA.
[51] Michelle Duff, "Cutaneous Manifestations of Diabetes Mellitus," In 2017 IEEE Imaging Systems and Techniques (IST) (pp. 1-5).
[52] Witten, I. H. et al. (1999). "Weka: Practical machine learning tools and techniques with Java implementations."
[53] K. Ogurtsovaa, J.D. da Rocha Fernandesa, Y. Huanga, U. Linnenkampa, L. Guariguataa, N.H. Choa,b, D.Cavana, J.E. Shawc, L.E. Makaroff "IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040". http://dx.doi.org/10.1016/j.diabres.2017.03.024 0168-8227/2017 Elsevier B.V.
[55] World Health Organization, "Definition, diagnosis, and classification of diabetes mellitus and its complications. Report of a WHO Consultation. Geneva: WHO," World Health Organization, pp. 1-66, 2016.
[56] Sarker, I.H. Context-aware rule learning from smartphone data: survey, challenges, and future directions. *J Big Data* **6,** 95 (2019). https://doi.org/10.1186/s40537-019-0258-4.
[57] Sarker, I.H., Colman, A. & Han, J. RecencyMiner: mining recency-based personalized behavior from contextual smartphone data. *J Big Data* **6,** 49 (2019). https://doi.org/10.1186/s40537-019-0211-6
[58] Sarker, Iqbal H. et al. "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage." *Journal of Big Data* 6 (2019): 1-28.
[59] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, and K. Salah, "BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model," Mobile Networks and Applications, 2019.