Using a fine KNN classifier and a medium KNN classifier, analyze and compare diabetic prediction

Nirupama S[1], Jenila Rani D[2]

[1]Research scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, Tamilnadu, India,Pincode:602105

[2*]Project guide, Corresponding author, Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

ABSTRACT

Aim: The goal of the study is to find out the presence of diabetes using the Fine K-NN (K- Nearest Neighbour) and Medium K-NN (K- Nearest Neighbour) algorithm and comparing the accuracy, specificity and sensitivity. **Materials and Methods:** A compilation of information from Kaggle's website was used in this research. The samples were regarded as (N=25) for Fine KNN and (N= 25) Medium KNN according to clinicalc.com, total sample size calculation was performed by keeping alpha error-threshold value 0.05, enrollment ratio is 0:1, 95% confidence level and power is 80%. The accuracy, specificity, sensitivity was calculated by using Matlab programming software. **Results:** The accuracy (%), specificity (%) and sensitivity (%) is compared using SPSS software using independent sample t tests. There is a statistically insignificant difference, P=0.832, P>0.05 with accuracy (60.4%), P=0.002, P<0.05 with specificity (62.18%) and P=<0.001, p<0.05 with sensitivity (59.18%) and demonstrated a better outcome in comparison to Medium KNN accuracy (47.2%), specificity (48.3%) and sensitivity (34.5%). **Conclusion:** Fine KNN appears to give better accuracy, specificity and sensitivity than Medium KNN to predict diabetes.

KEYWORDS: Diabetic prediction, Novel Fine KNN, Medium KNN, Machine learning, Matlab Programming, Accuracy.

INTRODUCTION

Diabetes is a lethal syndrome located around high society which has a huge hurdle or hitch and it is documented by a large amount of data. This study helps in diabetes diagnosis (Al-Furat,2020) and [Citation error]. The most important of this work is to determine the presence of diabetes [Citation error] using Matlab programming software by reducing the incidence of human error, and therefore regarded the simplest and most cost effective way of forecasting diabetic prediction This research can be carried out at hospitals and endocrine testing facilities [Citation error] and [Citation error] for detection of diabetes at a faster rate.

This study has been linked to 7 Google scholar publications and 3 ScienceDirect articles, being out in recent years to describe the created algorithm and models utilising machine learning methods, such as logistic regression, decision tree, and k closet neighbour (KNN) algorithms to forecast and access their performance in terms of accuracy (%), specificity (%), sensitivity (%). (Sowah et al. 2020) the model has met its objective accurately predicting the labels

of fresh pictures. For particular meal intakes the meal identifications and categorization model obtained the accuracy (95%) [Citation error]. A research work proposed diagnosis of diabetes mellitus using PSO and KNN classifier with the accuracy of (74%) [Citation error] has been encoded with Matlab programming which is the easiest and simplest method to predict diabetes. One among the related work Data mining approach for diabetes prediction using SVM, BPSO, Naive Bayes Classifiers and KNN was found to be more effective with the accuracy of (64.8%) (Niwariya 2020). Diabetes diagnosis based on KNN (Ali et al. 2020) is considered to be more accurate in diabetes prediction compared to other research works.Previously our team has a rich experience in working on various research projects across multiple disciplines(Sathish et al. 2020; Arivazhagan et al. 2020; Pandurangan, Veeraiyan, and Nesappan 2020; Saravanan et al. 2021)

Inefficient early detection of diabetes that eliminates human error rates is the major key point that motivated to work on this project to determine the presence of diabetes at an early stage by figuring out sensitivity (%) and specificity (%). The authors were expertised in the field of machine learning algorithm, and able to conduct studies in comparison to Fine KNN and Medium KNN in the biomedical aspect. The main aim is to analyse and compare diabetic prediction using the novelty Fine KNN classifier and Medium KNN classifier.

MATERIALS AND METHODS

This study was carried out at Microprocessor laboratory, Saveetha School of Engineering, Chennai. The sample size calculation was done using an earlier study results (Ali et al. 2020) using clinicalc.com by subjecting the alpha error- threshold by 0.05, enrollment ratio as 0:1, 95% confidence interval, power at 80%. Group 1 was the novelty Fine KNN algorithm (N=25 study group) in the Medium KNN (N=25 control group). The total sample size was 50.The dataset used in our research were from the kaggle website.

First we used Matlab programming software to train our source dataset,then the data was imported to the classifiers. The Fine KNN and Medium KNN classifiers have to be selected to train the data. A sample dataset of both Novel Fine KNN and Medium KNN was exported to Microsoft Excel document for importing it to the MATLAB as an input. MATLAB 2021 software has to be installed on the PC for training the source data set. (Sowah et al. 2020)The imported data is trained individually for each algorithm by varying k-fold cross-validation. Cross-validation is a model validation method used to estimate the performance of the model A confusion matrix is obtained and true positive, false positive, true negative and false negative values are noted. Accuracy (%), sensitivity (%), and specificity (%) values are calculated from the confusion matrix.

Statistical analysis

The accuracy, specificity and sensitivity percentages comparison of novelty Fine KNN and Medium KNN was done in IBM-SPSS 27.0.1. Since the variables were independent to each other,

an independent sample T-Test was done to collate the accuracy, specificity and sensitivity percentages which are considered as independent variable. There are no dependent variables.

RESULTS

In this research work of predicting diabetes, both the techniques appear to result in the same variable results with accuracy ranging from (47.2%-60.4%), specificity ranging from (34.5%-59.18%) and sensitivity ranging from (48.3%-62.18%). Table 1a depicts the accuracy, specificity and sensitivity of diabetes prediction using Fine KNN classifier. Table 1b depicts the accuracy, specificity and sensitivity of diabetes prediction using the Medium KNN classifier. Table 2 shows the comparison of mean accuracy (60.42%), mean sensitivity (62.18%) and mean specificity (59.18%) using Fine KNN and mean accuracy (47.22%), mean specificity (48.36) and mean sensitivity (34.54) using Medium KNN. Table 3 represents an independent sample t-test that appears to be a statistically significant difference of accuracy (P=60.42%, P<0.05), specificity (P=59.18%, P<0.05) and sensitivity (P=62.18, P<0.05).

Fig. 1 represents the comparison of accuracy between both the classifiers. Confusion matrix shows the true positive rate accounts for 6, while the false positive rate is 6. False negative accounts for 4 and true negative rate is 8. Using the Novel Fine KNN classifier the overall accuracy is determined to be 60.42%. Fig. 2a represents the confusion matrix and also represents the true positive accounts for 9 while the false positive rate is 3. False negative accounts for 4 and true negative accounts for 8. Fig. 2b represents the Medium KNN classifier; the overall accuracy is determined to be 40.42%.

DISCUSSION

In this study project of predicting diabetes, novel Fine KNN had better accuracy P=0.832, P<0.05 (60.42%), P=0.002, P<0.05 specificity (59.18%), and P= <.001, P<0.05 sensitivity (62.18%) in comparison to Medium KNN P=0.832, P<0.05 accuracy (47.22%), P=0.832, P<0.05 specificity (34.54%) and P=0.832, P<0.05 sensitivity (48.36%). The significant difference appears to have increased slightly, although it is not statistically significant (Table 2).

Fine KNN is the simplest and most affordable method of determining the presence of diabetes. [Citation error], have less accuracy (60.7%) than proposed. This involves blood glucose level monitoring, nutritional management, physical activity maintenance, weight and stress management, oral medication monitoring, and if any necessary insulin injection or pump use.[Citation error] ,(Kavakiotis et al. 2017) have accuracy of (98.6%) by using SVM machine learning algorithms. [Citation error] the results of the real world dataset demonstrate that an intelligible SVM method is successful in diabetic prediction that has accuracy of (94%), specificity (93%) and sensitivity of (94%). (Zakaria and Ahmad 2020) have accuracy of (90%).

For validations, the factor which is in limitation with outcomes is the unchanging and smaller sample size. If the sampling size and proportion of the training data set are both increased, the outcomes will grow proportionally as well. It is sensitive to data structure and memory constraints on a local level. It runs slowly

since it is a supervised learning lazy algorithm.

In the near future, we will look at how diabetes detection technology is being used in the healthcare profession and how it can help with more accurate detection of other diseases or illnesses. As a result, this initiative has a promising future as manual forecasting may be quickly converted to computerised output at a minimal cost. A larger dataset of real-time applications combined with other machine learning and deep learning algorithms like SVM, Naive Bayes should yield superior results.

CONCLUSION

In the study of diabetes prediction, the Novel Fine KNN algorithm with the accuracy (60.42%), specificity (62.18%) and sensitivity (59.18%) that operates using Matlab programming appears to give better results when compared to Medium KNN with the accuracy (47.22%) specificity (34.54%) and sensitivity (48.36%). In addition, the performance of the algorithm improved as the amount of data increased, which is not seen in other methods. This model is quite efficient and has a lot of potential in terms of improving the diagnostic efficiency of diabetes, thus it may be used in hospitals and endocrinology centres.

DECLARATIONS

Conflict of Interest

In this manuscript, there are no conflicts of interest.

Authors Contribution

Author SN was involved in data collection, data analysis, and manuscript writing. Author DJR was involved in conceptualisation, data validation and critical review of manuscript.

Acknowledgement

We would like to express our special thanks of gratitude to the management, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing me with necessary infrastructure that was required in completing the project.

Funding: We thank the following organisation for providing financial support that enabled us to complete the study.

1. Axis clinics Pvt.Ltd,Hyderabad.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Ali, Ameer, Mohammed A. T. Alrubei, Laith Falah Mohammed Hassan, Mohannad A. M. Al-Ja'afari, and Saif H. Abdulwahed. 2020. "DIABETES DIAGNOSIS BASED ON KNN." *IIUM Engineering Journal*. https://doi.org/10.31436/iiumej.v21i1.1206.
2. Arivazhagan, R., N. B. Geetha, P. Sivasamy, P. Kumaran, M. Kumara Gnanamithra, S. Sankar, Ganesh Babu Loganathan, and A. Arivarasan. 2020. "Review on Performance Assessment of Phase Change Materials in Buildings for Thermal Management through Passive Approach." *Materials Today: Proceedings* 22 (January): 419–31.
3. Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, NicosMaglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. 2017. "Machine Learning and Data Mining Methods in Diabetes Research." *Computational and*

Structural Biotechnology Journal 15 (January): 104–16.

4. Niwariya, Manoj. 2020. "Data Mining Approach for Diabetes Prediction Using BPSO, SVM, KNN and Naïve Bayes Classifiers." *International Journal of Advanced Trends in Computer Science and Engineering*. https://doi.org/10.30534/ijatcse/2020/4191.52020.
5. Pandurangan, Kiran Kumar, Deepak NallaswamyVeeraiyan, and ThiyaneswaranNesappan. 2020. "In Vitro Evaluation of Fracture Resistance and Cyclic Fatigue Resistance of Computer-Aided Design-on and Hand-Layered Zirconia Crowns Following Cementation on Epoxy Dies." *Journal of Indian Prosthodontic Society* 20 (1): 90–96.
6. Saravanan, A., P. Senthil Kumar, Dai-Viet N. Vo, S. Jeevanantham, S. Karishma, and P. R. Yaashikaa. 2021. "A Review on Catalytic-Enzyme Degradation of Toxic Environmental Pollutants: Microbial Enzymes." *Journal of Hazardous Materials* 419 (October): 126451.
7. Sathish, T., R. V. Sabariraj, K. Muthukumar, and S. Karthick. 2020. "Experimental Investigation of Convective Heat Transfer Coefficient on Nano Particles Mixture Used in Automobile Radiator Based on Mass Flow Rate." *Materials Today: Proceedings* 33 (January): 2524–29.
8. Sowah, Robert A., Adelaide A. Bampoe-Addo, Stephen K. Armoo, Firibu K. Saalia, Francis Gatsi, and Baffour Sarkodie-Mensah. 2020. "Design and Development of Diabetes Management System Using Machine Learning." *International Journal of Telemedicine and Applications* 2020 (July): 8870141.
9. Zakaria, Zahriladha, and Rabiah Ahmad. 2020. *Advances in Electronics Engineering: Proceedings of the ICCEE 2019, Kuala Lumpur, Malaysia*. Springer Nature.

Tables and Figures

Table 1a. Diabetes prediction using Fine KNN classifier.

Sample	Accuracy (%)	Sensitivity(%)	Specificity(%)
1	0.541667	0.555556	0.533333
2	0.666667	0.666667	0.666667
3	0.5833333	0.6	0.571429
4	0.5833333	0.6	0.571429
5	0.666667	0.7	0.642557
6	0.625	0.636364	0.615385
7	0.5833333	0.6	0.571429

8	0.5833333	0.6	0.571429
9	0.625	0.636364	0.571429
10	0.5833333	0.6	0.6
11	0.5833333	0.6	0.533333
12	0.625	0.6666667	0.66666.7
13	0.541667	0.6	0.571429
14	0.666667	0.6	0.571429
15	0.5833333	0.7	0.642857
16	0.5833333	0.636364	0.615385
17	0.666667	0.6	0.571429
18	0.625	0.6	0.571429
19	0.5833333	0.636364	0.615385
20	0.5833333	0.6	0.571429
21	0.625	0.6	0.571429
22	0.5833333	0.6666667	0.615385
23	0.5833333	0.6	0.571429
24	0.625	0.6	0.571429
25	0.5833333	0.7	0.615385

Table 1b. Diabetes prediction using Medium KNN classifier

Sample	**Accuracy (%)**	**Sensitivity(%)**	**Specificity(%)**
1	0.5	0.5	0.5
2	0.458333	0.478261	0
3	0.45833	0.473684	0.4
4	0.45833	0.470588	0.428571
5	0.416667	0.454545	0

6	0.45833	0.473684	0.4
7	0.416667	0.45	0.25
8	0.5	0.5	0.5
9	0.458333	0.478261	0
10	0.5	0.5	0.5
11	0.458333	0.52381	0.666667
12	0.541337	0.5	0.5
13	0.5	0.5	0.5
14	0.458333	0.478261	0
15	0.45833	0.4.73684	0.4
16	0.45833	0.470588	0.428571
17	0.416667	0.454545	0
18	0.45833	0.473684	0.4
19	0.416667	0.45	0.25
20	0.5	0.5	0.5
21	0.458333	0.478261	0
22	0.5	0.5	0.5
23	0.458333	0.52381	0.666667
24	0.45833	0.5	0.5
25	0.416667	0.478261	0.5

Table 2 Comparison of mean, accuracy, specificity and sensitivity using Fine KNN and Medium KNN. Group statistic comparison of accuracy, specificity and sensitivity for diabetes prediction by using Fine KNN classifier and Medium KNN classifier. Fine KNN has better mean compared to Medium KNN. Medium KNN accuracy (25.1167), specificity (11.5000) and sensitivity (8.5833) Medium KNN accuracy (24.0417), specificity(5.2500) and sensitivity (3.2500)

Parameters	**Classifiers**	**N**	**Mean**	**Std.**	**Std. Error**

				Deviation	**Mean**
Accuracy	Fine KNN Classifiers	24	.6042	.03686	.00752
	Medium KNN Classifier	24	.4722	.03617	.00738
Specificity	Fine KNN Classifiers	24	.6218	.03935	.00803
	Medium KNN Classifier	24	.4836	.02104	.00429
Sensitivity	Fine KNN Classifiers	24	.5918	.03647	.00745
	Medium KNN Classifier	24	.3454	.00433	.04579

Table 3. Independent sample t-test in predicting the accuracy, specificity and sensitivity of diabetes using Fine KNN classifier and Medium KNN classifier.

Independent Sample t Test

Parameter	**Equal Variances**	**Levene's Test for Equality of Variances**		**T-test for Equality of Means**							
Accuracy	Assumed	**F**	**Sig**	**t**	**df**	**Significance (one-Sided p)**	**Significance (two-sided p)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence interval (Lower)**	**95% Confidence interval (Upper)**
		.046	.832	12.516	46	<.001	<.001	.13194	.01054	.11073	.15316
	Not assumed			12.516	45.98	<.001	<.001	.13194	.01054	.11072	.15316

Sensitivity	Assumed	10.24	.002	15.178	46	<.001	<.001	.13823	.00911	.11990	.15656
	Not assumed			15.178	35.15	<.001	<.001	.13823	.00911	.11975	.15672
Specificity	Assumed	44.38	<.00	5.312	46	<.001	<.001	.24641	.04639	.15303	.33979
	Not assumed			5.312	24.21	<.001	<.001	.24641	.04639	.15071	.34212

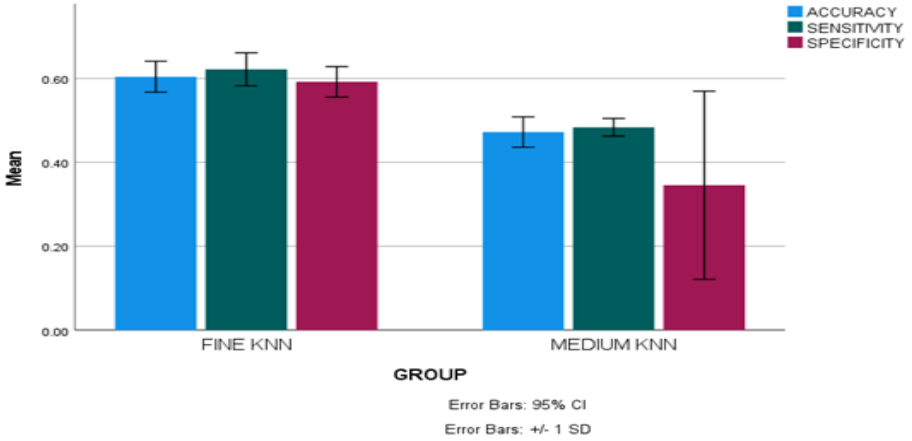

Fig. 1. Simple bar mean of accuracy, mean of specificity and mean of sensitivity using Fine KNN and Medium KNN. Both techniques appear to produce the same variable results with the accuracy ranging from (47.2%-60.42%), specificity from (34.5%-59.18%) and sensitivity ranging from (48.36%- 62.18%). X asis: Fine KNN algorithm vs Medium KNN, algorithm vs Y axis: Mean accuracy of detection +/-1 SD.

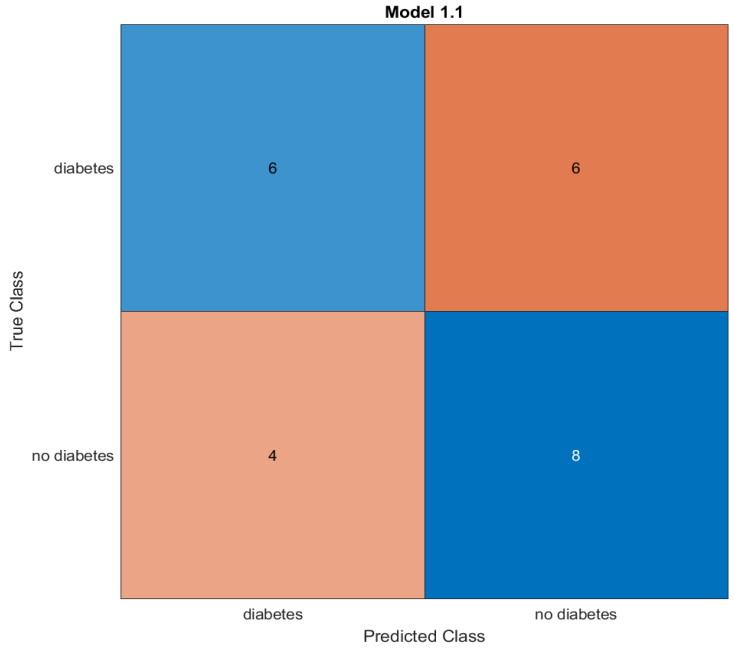

Fig. 2a. Confusion matrix of Fine KNN classifier. True positive rates 6, false positive rates 6 and false negative accounts for 4 and true negative accounts for 8. The total accuracy was found to be 60.42%.

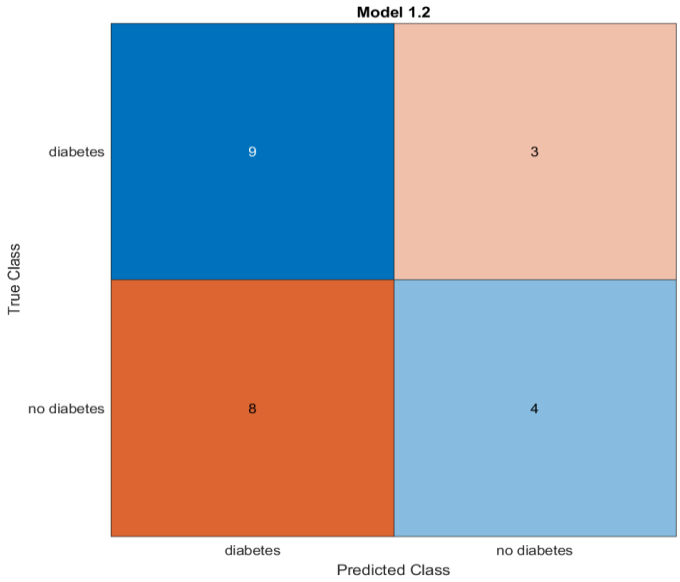

Fig. 2b. Confusion matrix of Medium KNN classifier. True positive accounts for 9, false positive accounts for 3 and false negative rates 4 and true negative rates 8. The total accuracy was found to be 47.2%.