

Table 1. Comparison of Results

Model	Accuracy	F1-Score	Sensitivity	Specificity
KNN (K=3)	0.785	0.804	0.88	0.69
DT	0.875	0.88262	0.94	0.81
RF (d = 10)	0.860	0.87156	0.95	0.77
LR	0.695	0.690	0.68	0.71
NB	0.695	0.687	0.67	0.72
SVM	0.77	0.785	0.84	0.7
XGBoost (d = 6)	0.82	0.829	0.87	0.77
Ensemble (RF + LR)	0.720	0.7021	0.66	0.78

In Table 1, the performance metrics for different ML algorithms have been given and the optimal value for each performance metric is highlighted. Here, authors also implemented an ensemble model that integrates random forest and logistic regression, and it is observed that the proposed ensemble model outperforms all traditional ML models. The graphical comparison of all models for accuracy and fl-score is illustrated in Fig. 3 and Fig. 4 respectively.

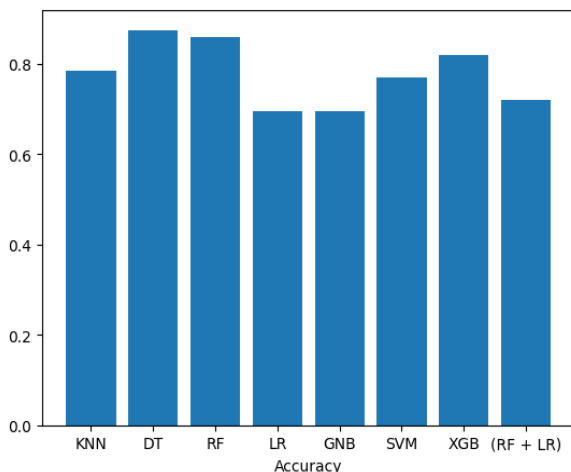


Figure 3. Comparative Analysis of Accuracy

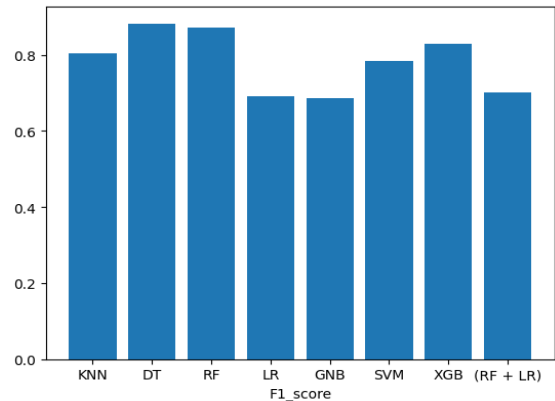


Figure 4. Comparative Analysis of F1_Score

5. Conclusion

This paper discusses the use of various classification algorithms to predict Diabetes using a set of attributes. These models include K- Nearest Neighbor, Decision Tree, Random Forest, Logistic Regression, Gaussian Naive Bayes, Support Vector Machine, XGBoost, and an ensemble of random forest and decision tree. After reviewing the performance of all models, we infer that an ensemble model of Decision Tree and Random Forest would provide us the maximum accuracy of 87.5%. The early prediction of diabetes can help in its treatment and possibly get it under control quickly.

References

- [1] <https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html>
- [2] Mohanty, S.N.; Ghosh, H.; Rahat, I.S.; Reddy, C.V.R. Advanced Deep Learning Models for Corn Leaf Disease Classification: A Field Study in Bangladesh. Eng. Proc. 2023, 59, 69. <https://doi.org/10.3390/engproc2023059069>
- [3] <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>
- [4] Kishan Patel, Manu Nair and Shubham Phansekar. Diabetes Prediction using Machine Learning International Journal of Scientific & Engineering Research Volume 12, Issue 3, March-2021
- [5] P. Tomar, N. Sharma and S. D. Kamble, "Aspect Analysis Based on Statistical Description and Visualization of Data," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 38-42, doi: 10.1109/IC3I56241.2022.10073201.
- [6] Llah, O. and Rista, A., 2021, May. Prediction and Detection of Diabetes using Machine Learning. In RTA-CSIT (pp. 94-102).
- [7] Khadidos A, Khadidos AO, Kannan S, Natarajan Y, Mohanty SN and Tsaramiris G (2020) Analysis of COVID-19 Infections on a CT Image Using DeepSense Model. Front. Public Health 8:599550. doi: 10.3389/fpubh.2020.599550.
- [8] N. Sharma, M. Mangla, M. Ishaque and S. N. Mohanty, "Inferential Statistics and Visualization Techniques for Aspect Analysis," 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), Jeddah,

- Saudi Arabia, 2023, pp. 1-6, doi: 10.1109/ICAISC56366.2023.10085093.
- [9] D. Patel, S. Chopra, N. Sharma and M. Mangla, "Analysis and Visualization of Heart Failure Prediction Dataset," 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST), Delhi, India, 2022, pp. 1-5, doi: 10.1109/AIST55798.2022.10064888.
- [10] S. Gupta, R. Sharma, N. Sharma and M. Mangla, "Aspect Analysis of Dementia Patients," 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2022, pp. 1-5, doi: 10.1109/ICICT55121.2022.10064519.
- [11] N. Sharma, J. Dev, M. Mangla, V. Wadhwa, S. N. Mohanty, S. N., & Kakkar, D. (2021). A heterogeneous ensemble forecasting model for disease prediction. *New Generation Computing*, 1-15.
- [12] M. Mangla, S. K. Shinde, V. Mehta, N. Sharma, & S. N. Mohanty, (Eds.). (2022). *Handbook of Research on Machine Learning: Foundations and Applications*. CRC Press.