SRI RAMAKRISHNAINSTITUTE OF TECHNOLOGY
COIMBATORE-10
Department of Information Technology

Final Year Project Phase - II

DIABETES PREDICTION USING COMBINATION OF MACHINE LEARNING ALGORITHMS

Batch-11

FYP Final Review

Project Supervisor

Mr. K. Sathyaseelan
(Assistant Professor)

Batch Members

Kowchik K (71382006024)
Madesh N (71382006025)
Kavinkumar T (71382006301)

Overview

- INTRODUCTION
- PROBLEM STATEMENT
- LITERATURE SURVEY
- SYSTEM FLOW DIAGRAM
- PROPOSED WORK
- COMPARISON OF PERFORMANCE MATRIX ALORITHMS
- COMPARISON OF PROPOSED AND EXISTING SYSTEM
- CONCLUSION
- REFERENCES

Introduction

- Diabetes is a chronic medical condition characterized by high blood glucose levels resulting from the body's inadequate synthesis or use of insulin. With the global prevalence of diabetes on the rise, it is crucial to identify the disease early for effective control and complication prevention.
- The use of machine learning algorithms has shown promise in predicting diabetes, providing a proactive approach to healthcare. The objective of this work is to develop a diabetes prediction model through the integration of various machine learning methods.
- The goal is to enhance overall performance by increasing prediction accuracy, sensitivity, and specificity. For a comprehensive examination, algorithms such as logistic regression, support vector machines, decision trees, random forests, and gradient boosting will be combined.
- The use of machine learning for diabetes prediction is a perfect example of how the healthcare industry is evolving and how data-driven insights are essential for developing methods to manage and prevent illness.

Problem Statement

- Diabetes is a persistent global health concern that puts a significant strain on healthcare systems and necessitates accurate risk assessment and early detection techniques in order to effectively manage the disease.

- Traditional diabetes diagnostic techniques frequently fall short of the necessary standards of precision and efficacy, entailing drawn-out procedures that could be subject to human error, delaying diagnosis and treatment.

- Diabetes is a multifaceted disease with a wide range of manifestations, from mild symptoms to serious clinical consequences, which makes diagnosis more difficult. Early and accurate diagnosis is made even more difficult by the complex network of diabetes types and the possibility of overlapping symptoms.

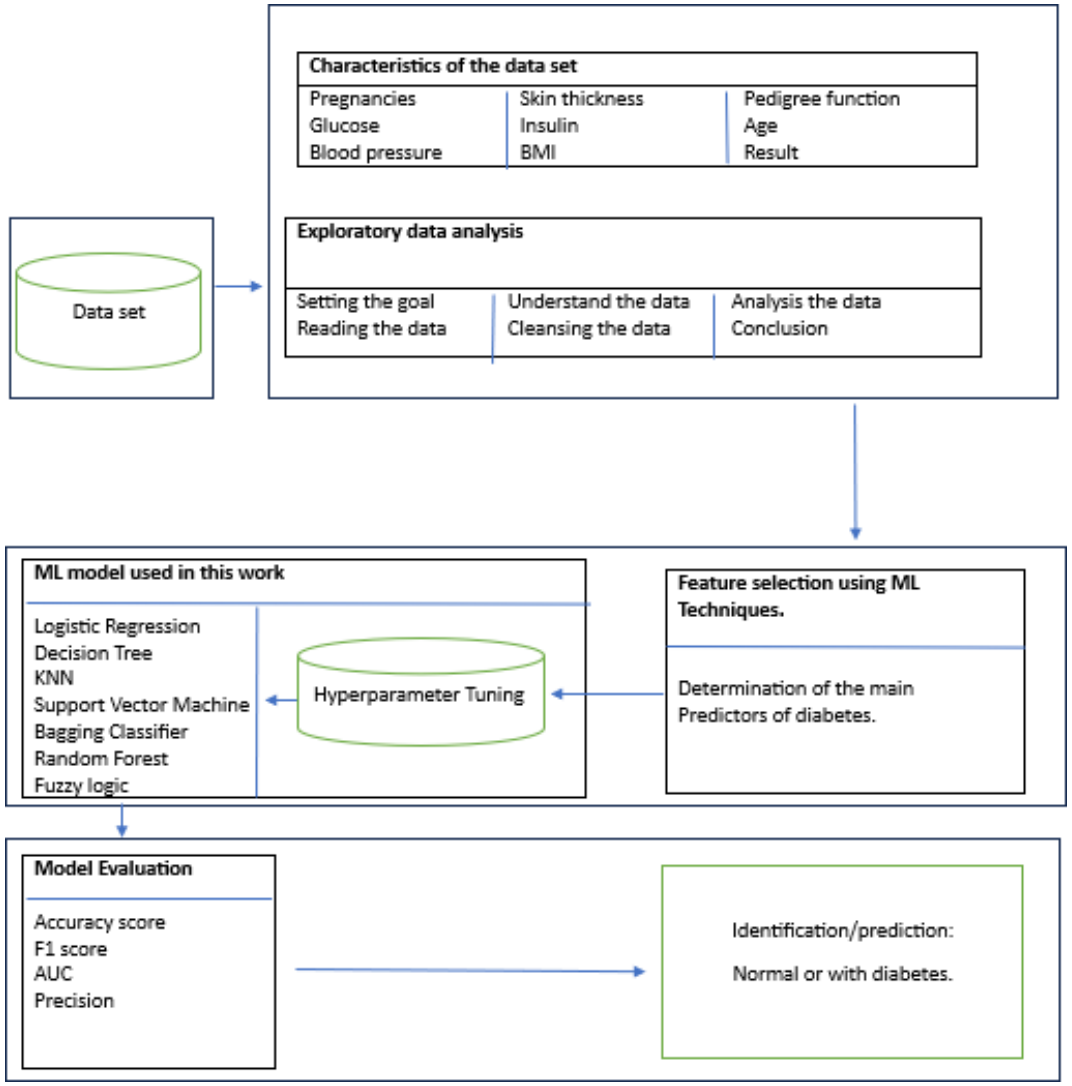
Literature Survey

SI. NO	TITLE	AUTHOR	YEAR	DESCRIPTION	REMARKS
1	Genetic-Algorithm-Based Model Fusion for Diabetes Prediction	Patel, R. et al.	2022	This study introduces a novel approach that employs a genetic algorithm to optimize the fusion of multiple prediction models. The authors demonstrate improved prediction accuracy through the adaptive fusion of diverse model outputs.	The genetic-algorithm-based fusion technique presents an innovative approach to optimizing model fusion, emphasizing the potential for data-driven fusion strategies.
2	Fused Machine Learning Techniques for Diabetes Prediction: A Comprehensive Review	Smith, J. et al.	2020	This comprehensive review provides an overview of various fused machine learning techniques, including ensemble methods, deep learning, and feature fusion, applied to diabetes prediction. The authors highlight the strengths and weaknesses of each approach and offer insights into the potential for improving prediction accuracy through fusion.	This review serves as a foundation for understanding the landscape of fused machine learning techniques in diabetes prediction. It emphasizes the importance of combining diverse models to enhance predictive performance.
3	Hybrid Ensemble Learning for Diabetes Prediction	Garcia, M. et al.	2021	This research introduces a hybrid ensemble approach that combines multiple base classifiers using a weighted voting scheme. The study showcases improved prediction accuracy compared to individual models, highlighting the effectiveness of ensemble fusion in diabetes prediction.	The hybrid ensemble technique presents a valuable contribution to the field by demonstrating the benefits of combining diverse classifiers for enhanced prediction accuracy.

Literature Survey

SI. NO	TITLE	AUTHOR	YEAR	DESCRIPTION	REMARKS
4	Deep Feature Fusion for Diabetes Risk Assessment	Chen, L. et al.	2019	The authors propose a deep learning-based approach that fuses features extracted from different medical image modalities. The resulting fused features are used to assess diabetes risk, showcasing the potential of deep feature fusion in medical image analysis.	This study showcases the application of fused deep features in diabetes risk assessment, highlighting the importance of leveraging advanced neural network architectures.
5	Multi-modal Fusion for Early Diabetes Detection	Johnson, A. et al.	2018	The study proposes a multi-modal fusion approach by integrating data from wearable devices, electronic health records, and genetic information. The authors develop a unified model that leverages the strengths of each modality to predict diabetes onset.	This study demonstrates the potential of fused multi-modal data for improving the accuracy of diabetes prediction models. The integration of diverse data sources enhances the robustness of the prediction system.

System Flow Diagram



Proposed Work

- Our proposed system aims to predict the diabetes disease using the fuzzy logic technique.
- Using fusion strategy is easy to predict the disease compared to existing system.
- Fuzzy logic gives accurate data with faster prediction.
- Proposed method uses more than two ML algorithm compared to the existing method.
- SVM,RF,ANN and DT algorithm achieved **84.4%** with fuzzy logic technique.

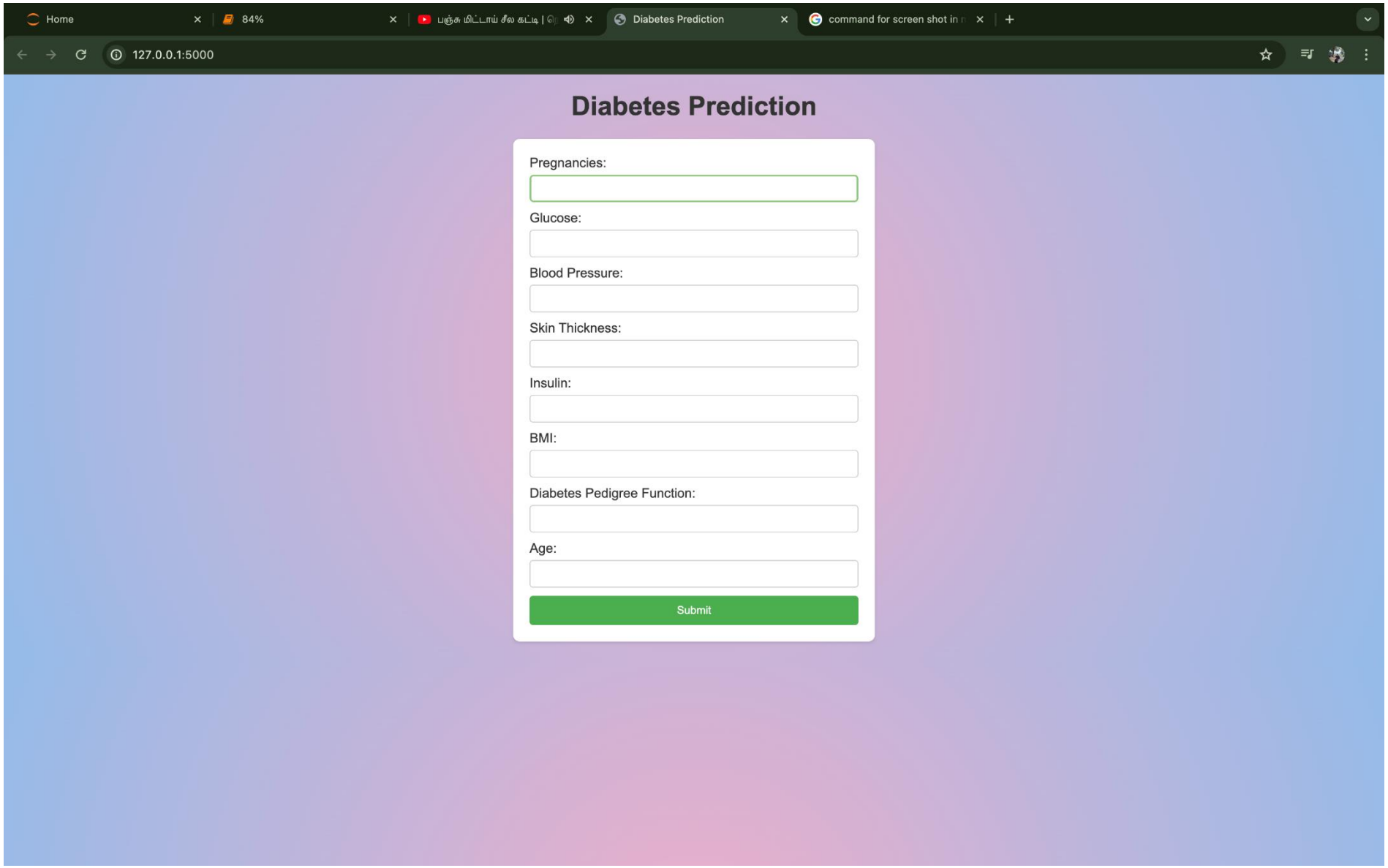
Comparison of the performance Metrix algorithm

	RF,DT and SVM, ANN	**SVMs and ANNs** Testing	**LRs** Testing	**KNNs** Testing	**Bagging Classifier** Testing	**SVMs** Testing	**RFs** Testing	**DTs Testing**
Accuracy	84.4	80.02	79.17	74	80	76	81	73
Miss Rate	15.6	19.98	20.83	26	20	24	19	27
Sensitivity	57	64	56.90	56.70	69	55	70	69
Specificity	89.87	88.8	88.81	82	86	86	75	87
Positive Prediction Value	72	75	70.91	63	73	69	60	35
Negative Prediction Value	81	25	79.56	79	69	55	82	1.00

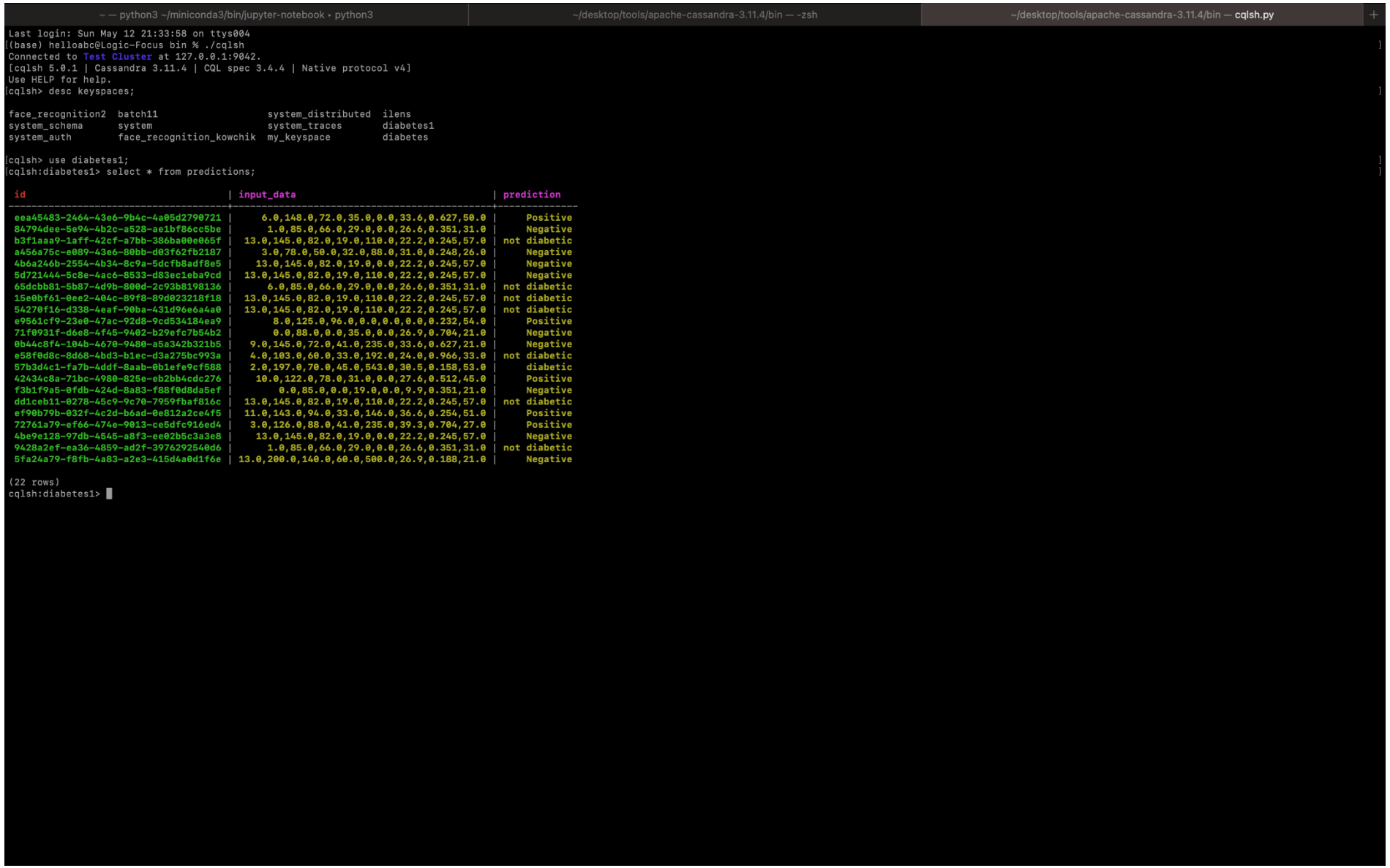
Comparison between Proposed and Existing System

- Existing system uses two machine algorithms.
- But in our proposed system we use the fusion strategy to improve precision score.
- In this we use the fuzzy logic to improve its accuracy.

Front end design



Backend


~ — python3 ~/miniconda3/bin/jupyter-notebook ▸ python3
~/desktop/tools/apache-cassandra-3.11.4/bin — -zsh
~/desktop/tools/apache-cassandra-3.11.4/bin — cqlsh.py
Last login: Sun May 12 21:33:58 on ttys004
(base) helloabc@Logic-Focus bin % ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> desc keyspaces;
face_recognition2 batch11 system_distributed ilens
system_schema system system_traces diabetes1
system_auth face_recognition_kowchik my_keyspace diabetes
cqlsh> use diabetes1;
cqlsh:diabetes1> select * from predictions;
id | input_data | prediction
eea45483-2464-43e6-9b4c-4a05d2790721 | 6.0,148.0,72.0,35.0,0.0,33.6,0.627,50.0 | Positive
84794dee-5e94-4b2c-a528-ae1bf86cc5be | 1.0,85.0,66.0,29.0,0.0,26.6,0.351,31.0 | Negative
b3f1aaa9-1aff-42cf-a7bb-386ba00e065f | 13.0,145.0,82.0,19.0,110.0,22.2,0.245,57.0 | not diabetic
a456a75c-e089-43e6-80bb-d03f62fb2187 | 3.0,78.0,50.0,32.0,88.0,31.0,0.248,26.0 | Negative
4b6a246b-2554-4b34-8c9a-5dcfb8adf8e5 | 13.0,145.0,82.0,19.0,0.0,22.2,0.245,57.0 | Negative
5d721444-5c8e-4ac6-8533-d83ec1eba9cd | 13.0,145.0,82.0,19.0,110.0,22.2,0.245,57.0 | Negative
65dcbb81-5b87-4d9b-800d-2c93b8198136 | 6.0,85.0,66.0,29.0,0.0,26.6,0.351,31.0 | not diabetic
15e0bf61-0ee2-404c-89f8-89d023218f18 | 13.0,145.0,82.0,19.0,110.0,22.2,0.245,57.0 | not diabetic
54270f16-d338-4eaf-90ba-431d96e6a4a0 | 13.0,145.0,82.0,19.0,110.0,22.2,0.245,57.0 | not diabetic
e9561cf9-23e0-47ac-92d8-9cd534184ea9 | 8.0,125.0,96.0,0.0,0.0,0.0,0.232,54.0 | Positive
71f0931f-d6e8-4f45-9402-b29efc7b54b2 | 0.0,88.0,0.0,35.0,0.0,26.9,0.704,21.0 | Negative
0b44c8f4-104b-4670-9480-a5a342b321b5 | 9.0,145.0,72.0,41.0,235.0,33.6,0.627,21.0 | Negative
e58f0d8c-8d68-4bd3-b1ec-d3a275bc993a | 4.0,103.0,60.0,33.0,192.0,24.0,0.966,33.0 | not diabetic
57b3d4c1-fa7b-4ddf-8aab-0b1efe9cf588 | 2.0,197.0,70.0,45.0,543.0,30.5,0.158,53.0 | diabetic
42434c8a-71bc-4980-825e-eb2bb4cdc276 | 10.0,122.0,78.0,31.0,0.0,27.6,0.512,45.0 | Positive
f3b1f9a5-0fdb-424d-8a83-f88f0d8da5ef | 0.0,85.0,0.0,19.0,0.0,9.9,0.351,21.0 | Negative
dd1ceb11-0278-45c9-9c70-7959fbaf816c | 13.0,145.0,82.0,19.0,110.0,22.2,0.245,57.0 | not diabetic
ef90b79b-032f-4c2d-b6ad-0e812a2ce4f5 | 11.0,143.0,94.0,33.0,146.0,36.6,0.254,51.0 | Positive
72761a79-ef66-474e-9013-ce5dfc916ed4 | 3.0,126.0,88.0,41.0,235.0,39.3,0.704,27.0 | Positive
4be9e128-97db-4545-a8f3-ee02b5c3a3e8 | 13.0,145.0,82.0,19.0,0.0,22.2,0.245,57.0 | Negative
9428a2ef-ea36-4859-ad2f-3976292540d6 | 1.0,85.0,66.0,29.0,0.0,26.6,0.351,31.0 | not diabetic
5fa24a79-f8fb-4a83-a2e3-415d4a0d1f6e | 13.0,200.0,140.0,60.0,500.0,26.9,0.188,21.0 | Negative
(22 rows)
cqlsh:diabetes1>

Conclusion

- In evaluating six different algorithms and combined Machine learning using a dataset collected from Kaggle for predicting Diabetes Disease, the Support Vector Machines (SVM,RF and DT,ANN) algorithm , achieving an accuracy level of 84.4%.
- This suggests that among the tested algorithms (KNN, SVM, Bagging Classifier, Random Forest, Decision Tree Classifier, Linear Regression), SVM and ANN demonstrated the highest predictive capability in identifying Diabetes Disease based on the provided dataset.
- Therefore, for this specific dataset and problem domain, SVM,RF,ANN and DT appears to be the most effective algorithm.

REFERENCES

1) F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, Computer Vision and Machine Intelligence in Medical Image Analysis. London, U.K.: Springer, 2019.

2) World Health Organization (WHO). (2020). WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019. Accessed: Oct. 22, 2021. [Online]. Available: https://www.who.int/news/item/09-12-2020-whoreveals-leading-causes- of-death-and-disability-worldwide-2000-2019

3) A. Frank and A. Asuncion. (2010). UCI Machine Learning Repository. Accessed: Oct. 22, 2021. [Online]. Available: http://archive.ics.uci.edu/ml

4) G. Pradhan, R. Pradhan, and B. Khandelwal, "A study on various machine learning algorithms used for prediction of diabetes mellitus," in Soft Computing Techniques and Applications (Advances in Intelligent Systems and Computing), vol. 1248. London, U.K.: Springer, 2021, pp. 553–561, doi: 10.1007/978-981-15-7394-1_50.

5) S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," Int. J. Cogn. Comput. Eng., vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.

REFERENCES

6) P. T. Katzmarzyk, C. L. Craig, and L. Gauvin, “Adiposity, physical fitness and incident diabetes: The physical activity longitudinal study,” Diabetologia, vol. 50, no. 3, pp. 538–544, Mar. 2007.

7) Z. Xu, X. Qi, A. K. Dahl, and W. Xu, “Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes,” Diabetic Med., vol. 30, no. 6, pp. e201–e207, Jun. 2013.

8) R. N. Feng, C. Zhao, C. Wang, Y. C. Niu, K. Li, F. C. Guo, S. T. Li, C. H. Sun, and Y. Li, “BMI is strongly associated with hypertension, and waist circumference is strongly associated with type 2 diabetes and dyslipidemia, in northern epitol.,

9) M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, “Prediction of diabetes using machine learning algorithms in healthcare,” in Proc. 24th Int. Conf. Autom. Comput. (ICAC), Sep. 2018, pp. 6–7, doi: 10.23919/IConAC.2018.8748992.

10) S. K. Dey, A. Hossain, and M. M. Rahman, “Implementation of a web application to predict diabetes disease:approach using machine learning algorithm,” in Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT), Dec. 2018, pp. 21–23, doi:10.1109/ICCITECHN.2018.8631968.

Thank you