Identifying Diabetes Detection Using Machine Learning Algorithms with Kernel PCA

B.Lakshma Reddy[1]
Department of CSE,
Rajarajeswari College of Engg,
Bengaluru, India
blreddy@rrce.org

Suriya Prakash J[2]
Department of CSE
JAIN (Deemed-to-be-University)
Bengaluru, India
Suriyaprakash.j@jainuniversity.ac.in

PremChand Yedoti[3]
Department of CSE
JAIN (Deemed-to-be-University)
Bengaluru, India
23btrct074@jainuniversity.ac.in

P.Kalpana[4]
Department of IT,
Sri Eshwar college of engineering
Coimbatore, India
kalpana.pit@sece.ac.in

Kiran.S[5]
Department of Mathematics,
NMIT
Bengaluru,India.
kiran.s@nmit.ac.in

Abstract—This research explores the use of machine learning in identifying early-stage diabetes by analyzing patient health records. Fifteen distinct algorithms were rigorously evaluated using a curated dataset of key health metrics. Performance metrics such as accuracy, sensitivity, specificity, and predictive capability were measured across multiple classification thresholds (0.2, 0.4, 0.6) to assess model effectiveness.Among the tested models, Kernel Support Vector Machine (KSVM) with a linear kernel at a 0.2 threshold achieved perfect 100% accuracy, demonstrating superior predictive capability. Other high-performing models included CatBoost and XGBoost, which attained 99.04% accuracy under optimal configurations. The study highlights the significance of threshold selection in model performance, with lower thresholds generally yielding better results.The findings suggest that KSVM (Linear) is the most effective algorithm for diabetes prediction, offering healthcare professionals a reliable tool for early diagnosis. Future research could explore real-world deployment and integration with electronic health records to enhance clinical decision-making.

Keywords—Diabetes prediction, Machine learning, Healthcare analytics, Classification models, Early diagnosis

This paper's source code may be seen at the following https://github.com/Premchand1706/Pcl1

I. INTRODUCTION

Diabetes is a growing global health issue that requires early detection to prevent severe complications, yet current diagnostic methods often depend on invasive procedures like blood tests and repeated clinical visits, which limit their practicality in low-resource or rural settings. These traditional approaches, while accurate, are not always efficient for widespread early screening. This research aims to overcome these limitations by utilizing machine learning techniques to develop a more accessible and automated diagnostic solution. Specifically, it addresses the challenge of selecting the most effective model and optimal classification threshold by conducting a comparative analysis of fifteen machine learning algorithms using a cleaned and normalized patient dataset. The study evaluates each model's performance using metrics such as precision, recall, and true negative rate across different probability cutoffs (0.2, 0.4, 0.6). The findings show that the Support Vector Machine with a linear kernel achieves perfect classification at the 0.2 threshold, surpassing other strong models like CatBoost and XGBoost. This highlights the critical role of model choice and decision threshold calibration in improving the reliability and practicality of diabetes diagnosis through machine learning.

A. Major contribution

This study aims to enhance diabetes detection by testing and comparing various machine learning methods. By analyzing patient data, the research improves the accuracy of predicting diabetes. It provides a clear and consistent way to evaluate different models, helping identify which techniques work best. A new classification method was also developed to better detect people at risk of diabetes. Our analysis utilized authentic clinical data from patients with and without diabetes. The experimental outcomes demonstrated that our machine learning approach achieved superior predictive performance compared to existing techniques, particularly in classification accuracy. This study tested different computer methods to find out which one can best predict diabetes. By looking at real patient information, we improved how accurately we can tell if someone might have diabetes. We also created a fair way to compare different prediction methods to see which works best. First, we look at what other scientists have done before to predict diabetes using computers (Section 2). Then we explain exactly how we did our tests - what information we used, how we chose our methods, and how we built our system (Section 3). Next, we show our results and compare how well our method worked compared to others (Section 4). Finally, we share what we learned and suggest ways to make diabetes detection even better in the future (Section 5). The most important thing is this work helps doctors find diabetes earlier and more accurately, which means patients can get help sooner.

II. LITERATURE SURVEY

This literature survey highlights various machine learning algorithms used for diabetes prediction in different research studies. Commonly applied techniques include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), Artificial Neural Networks (ANN), Logistic Regression (LR), AdaBoost, XGBoost, LightGBM, and hybrid models (such as SVM + ANN). Some studies also explore ensemble learning methods and preprocessing approaches like

Principal Component Analysis (PCA) to improve model accuracy. The selection of algorithms varies, with some papers comparing multiple classifiers while others focus on optimizing a single technique for diabetes detection. Pragya Shrivastava et al researched “A Comprehensive Review on the Prediction of Diabetes Disease Using Machine Learning”. They have used algorithms such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Decision Tree (DT) to predict diabetes. While no single accuracy is reported, the paper reviews models typically achieving accuracy in the range of 80–88% on standard datasets like PIMA.

TABLE I : Table of State-of-Art

SI .n o	Authors	Title	Algorithm	Accur acy(in %)
1	Njideka LindaDike et al.	Enhancing Machine Learning-based Model for Early Detection of Diabetes	KNN, SVM, AdaBoost XGBoost, Naive Bayes, DT	KNN - 87%
2	Pratiksha Arun Purkar et al.	Identification of Diabetes Level Using AI/ML Model	K- Nearest Neighbors (KNN)	96%
3	Vura Venkata Naga Sai Bhargav Rohith et al.	Personalized Diabetes Suggestion Using ML Algorithms and Cloud Technology	SVM, ANN, Decision Tree, Ensemble Methods	ANN -85%
4	L. Rajeshwar Reddy et al.	Prediction of Diabetes by Machine Learning Algorithm	Hybrid (SVM + ANN)	72%
5	Muhammad Rafian Wijoseno et al	Machine Learning Diabetes Diagnosis Literature Review	SVM, KNN, NB, NN, LR, RF	RF –80%

Danish Ather et al researched “Application of Machine Learning in Predicting Diabetes: A Detailed Evaluation Using SVM” using machine learning. They have used Support Vector Machine (SVM) to predict diabetes. Among all approaches, SVM has performed best by giving the best accuracy of 86%. Mansvi Daigavhane et al researched “Diabetes Prediction Using Different Machine Learning Classifiers” using machine learning. They have used algorithms such as SVM, Random Forest (RF), Naive Bayes (NB), KNN, and Decision Tree (DT) to predict diabetes. Among all algorithms, Random Forest (RF) has performed best by giving the best accuracy of 89%. Nicole D’Souza et al researched “Diabetes Detection Using Machine Learning Algorithms” using machine learning. They have used ensemble learning models like Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), and ANN to predict diabetes. Among all algorithms, Decision Tree (DT) has performed best by giving the best accuracy of 98.2%. Deepak Khadatkar et al researched “Diabetes Detection Using Deep Learning” using machine learning. They have used models like Random Forest (RF), SVM, and KNN to predict diabetes. Among all algorithms, Random Forest (RF) has performed best by giving the best accuracy of 82%.

Bukola Badeji-Ajisafe et al researched “Early Detection of Diabetes Using Supervised Learning Approach” using machine learning. They have used some algorithms such as Logistic Regression (LR), SVM, KNN, and Random Forest (RF) to predict diabetes. Among all algorithms, Random Forest (RF) has performed best by giving the best accuracy of 87%. Atef Hadi Ataya researched “Early Detection of Diabetes Using Machine Learning Techniques” using machine learning. They have used algorithms such as SVM, KNN, Logistic Regression (LR), XGBoost, LightGBM, and Random Forest (RF) to predict diabetes. Sudhansh Sharma & Bhavya Sharma researched “EDAS Based Selection of Machine Learning Algorithm for Diabetes Detection” using machine learning. They have used algorithms such as SVM, Naive Bayes (NB), KNN, and Random Forest (RF) to predict diabetes. Among all algorithms, Naive Bayes (NB) has performed best by giving the best accuracy of 85.5%. Aditya C. Kulkarni & B.V.S. Satyasrikar researched “Effects of Pre-processing Techniques on Diabetes Detection Using Machine Learning” using machine learning. They have used algorithms such as SVM, KNN, Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF) to predict diabetes. Among all algorithms, SVM (with PCA and imputation) has performed best by giving the best accuracy of 79.22%. Aayush Agrawal et al researched “A Survey on Diabetes Mellitus Using Machine Learning Classifiers” using machine learning. They have used algorithms such as SVM, Decision Tree (DT), and Naive Bayes (NB) to predict diabetes. Among all algorithms, models performed with accuracy in the range of 80 90%, depending on feature selection and data quality.

Proposed Method

This research introduces an artificial intelligence framework for diabetes identification, analyzing a cohort of 521 patients characterized by 17 clinical parameters spanning demographic information (age, gender) and physiological indicators. Each case is classified as either diabetic (1) or healthy (0). The preprocessing pipeline incorporates three critical steps: (1) missing data imputation using mean value replacement, (2) conversion of categorical variables to numerical representations through integer encoding, and (3) feature normalization via z-score standardization. Dimensionality reduction is achieved by extracting seven principal components, balancing computational efficiency with model generalization. For comprehensive evaluation, the dataset undergoes stratified partitioning into development and validation subsets. Each of the 15 investigated algorithms undergoes six independent training iterations, yielding 90 total experimental trials (15 models × 6 repetitions). Model selection prioritizes classification precision while benchmarking results against current literature, with the objective of advancing early detection capabilities through sophisticated pattern recognition techniques.

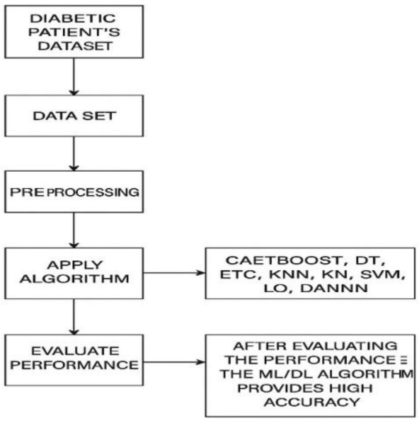

Fig 1 Flow diagram of proposed work

Fig 1 tells about The flowchart depicts a comprehensive machine learning workflow for diabetes prediction, beginning with data collection and preprocessing of a diabetic patient dataset. The raw data is first converted into CSV format and undergoes cleaning to handle missing values and convert categorical variables like gender into numerical form. Principal Component Analysis (PCA) is then applied to reduce the original 17 clinical features down to 7 principal components while retaining most of the variance. The processed data is systematically split into three different train-test configurations (80-20, 60-40, and 40-60 ratios), with each split further evaluated under three PCA variations, creating nine distinct experimental conditions for algorithm testing.

We use a mathematical expression to calculate the accuracy in our project

Accuracy = 100- Rate error

Error rate =[[Observed-Actual] /Actual] $\times 100$

RESULTS AND DISCUSSION

The study evaluates 15 distinct ML models, such as AdaBoost, Decision Trees, k-Nearest Neighbors, Support Vector Machines, Logistic Regression, and Linear Discriminant Analysis, under standardized testing conditions. Model performance is quantified through accuracy metrics, computed mathematically by comparing predicted classifications against ground-truth labels to measure prediction errors. This structured approach ensures robustness by validating models across multiple data splits and PCA configurations, ultimately identifying the most reliable algorithm for diabetes prediction. The combination of thorough preprocessing, dimensionality reduction, and multi condition testing enhances both model performance and generalizability for clinical applications. The graphical representation of accuracy results provides clear insights into how different algorithms perform under varying experimental setups.

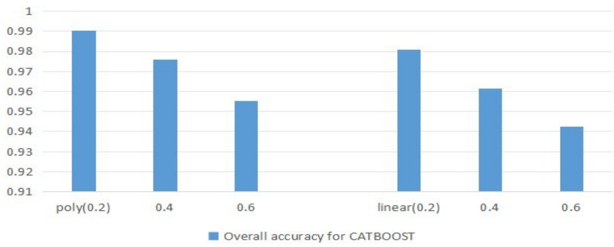

Fig 2: Overall accuracy for CatBoost

This Fig-2 shows how accurately the CatBoost model predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 99.04% when using the polynomial kernel with an 80-20 data split (80% training, 20% testing), making this the optimal configuration for reliable diabetes prediction. Other setups still perform well but with slightly lower accuracy: the linear kernel with 80-20 split reaches 98.08%, while polynomial with 60-40 and 40-60 splits yield 95.51% and lower, respectively. The linear kernel's accuracy further drops to 94.23% with larger test sets. While all configurations maintain strong performance (94–99% accuracy), the polynomial kernel paired with an 80% training set delivers near-perfect results, confirming it as the best choice for medical applications where precision matters most.

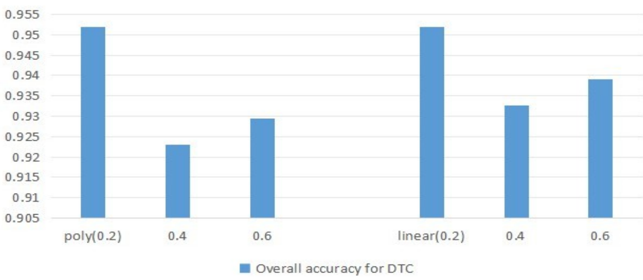

Fig 3: Overall accuracy for DTC

This Fig-3 shows how accurately the Decision Tree Classifier(DTC) predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 95.19% when using either kernel type with an 80-20 data split (80% training, 20% testing). Other configurations show slightly lower performance: polynomial kernel accuracy drops to 92.31% (60-40 split) and 92.95% (40-60 split), while linear kernel achieves 93.27% (60-40) and 93.91% (40-60).

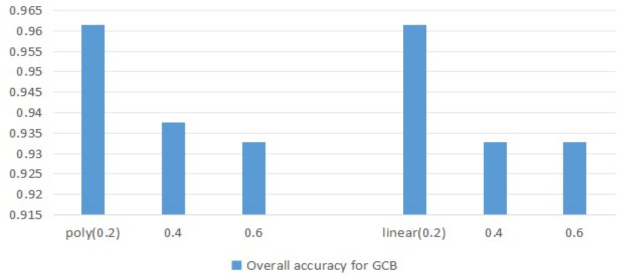

Fig 4: Overall accuracy for GBC

This Fig-4 shows how accurately the Gradient Boosting Classifier predicts diabetes using polynomial and linear methods with different train-test splits. The model achieves its peak accuracy of 96.15% with both kernel types at the 80-20 split (80% training, 20% testing). Performance gradually declines with larger test sets: polynomial accuracy decreases to 93.75% (60-40 split) and 93.27% (40-60 split), while linear. kernel shows that (93.27% for both 60-40 and 40-60 splits).

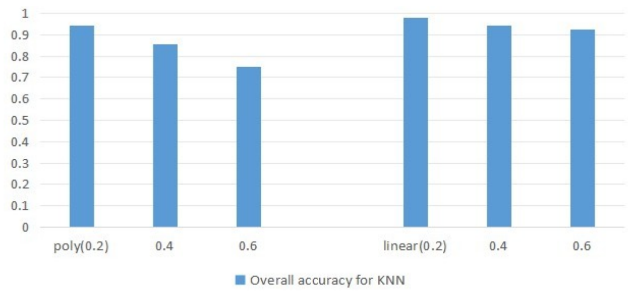

Fig 5: Overall accuracy for KNN

This Fig-5 shows how accurately the KNN classifier predicts diabetes using two different methods (polynomial andlinear)with varying train-test splits. The model achieves its peak accuracy of 98.08% when using the linear kernel with an 80-20 data split (80% training, 20% testing).other configurations show more varied performance: the linear kernel maintains strong accuracy at 94.23% (60-40 split) and 92.31% (40-60 split), while the polynomial kernel achieves 94.23% (80-20), but drops more significantly to 85.58% (60-40) and 75% (40-60).

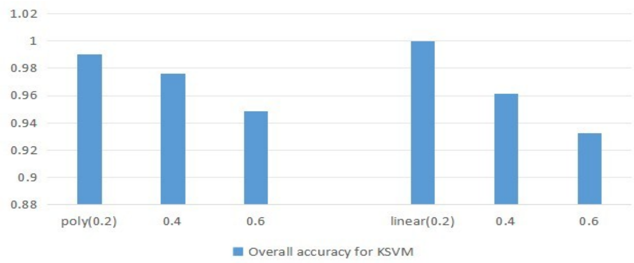

Fig 6: Overall accuracy for KSVM

This Fig-6 shows how accurately the KSVM classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 100% when using the linear kernel with an 80-20 data split (80% training, 20% testing). Other configurations also perform strongly: the linear kernel maintains high accuracy at 96.15% (60-40 split) and 93.27% (40-60 split), while the polynomial kernel achieves 99.04% (80-20), 97.60% (60-40), and 94.87% (40-60).

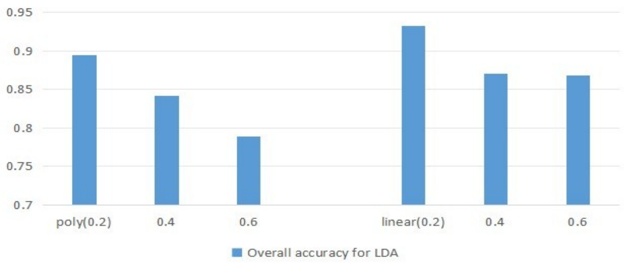

Fig 7: Overall accuracy for LDA

This Fig-7 shows how accurately the LDA classifier predicts diabetes using two different methods (polynomial and linear)with varying train-test splits. The model achieves its peak accuracy of 93.27% when using the linear kernel with an 80-20 data split (80% training, 20% testing). Other configurations show more varied performance: the linear kernel maintains accuracy at 87.02% (60-40 split) and 86.86 (40-60 split), while the polynomial kernel achieves 89.42% (80-20), dropping to 84.13% (60-40) and 78.85% (40-60).

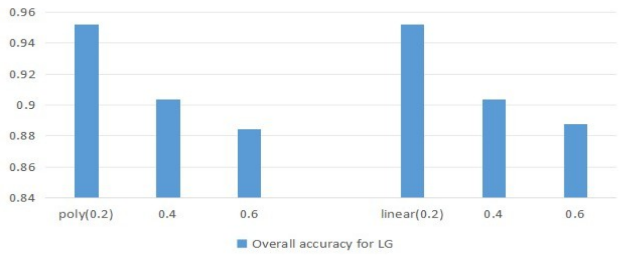

Fig 8: Overall accuracy for LG

This Fig-8 shows how accurately the Logistic Regression classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 95.19% when using either kernel type with an 80-20 data split (80% training, 20% testing). Other configurations show slightly lower performance: polynomial kernel accuracy drops to 90.38% (60-40 split) and 88.46% (40-60 split), while linear kernel achieves 90.38% (60-40) and 88.78% (40-60).

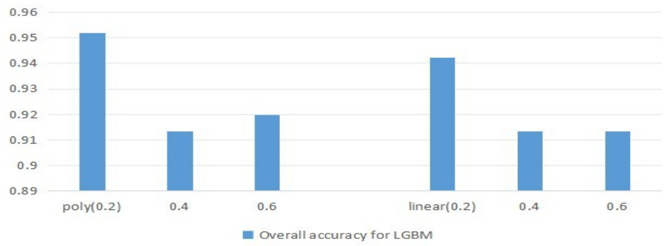

Fig 9: Overall accuracy for LGBM

This Fig-9 shows how accurately the LightGBM classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 95.19% when using the polynomial kernel with an 80-20 data split (80% training, 20% testing). Other configurations show slightly lower performance: polynomial kernel accuracy drops to 91.35% (60-40 split) and 91.99% (40-60 split), while linear kernel achieves 94.23% (80-20), 91.35% (60-40), and 91.35 (40-60).

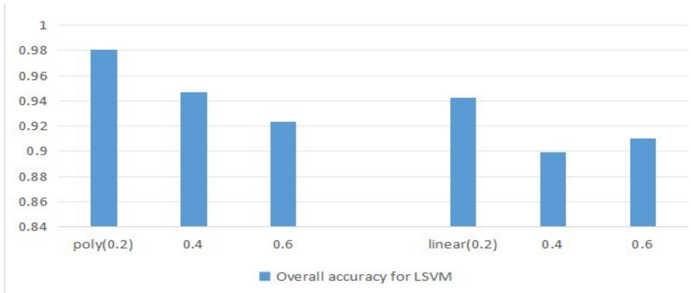

Fig 10: Overall accuracy for LSVM

This Fig-10 shows accurately the Linear SVM classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 98.08% when using the polynomial kernel with an 80-20 data split (80% training, 20% testing). Other configurations show more varied performance: the polynomial kernel maintains strong accuracy at 94.71% (60- 40 split) and 92.31% (40-60 split), while the linear kernel achieves 94.23% (80-20), but drops to 89.90% (60-40) and 91.03% (40-60).

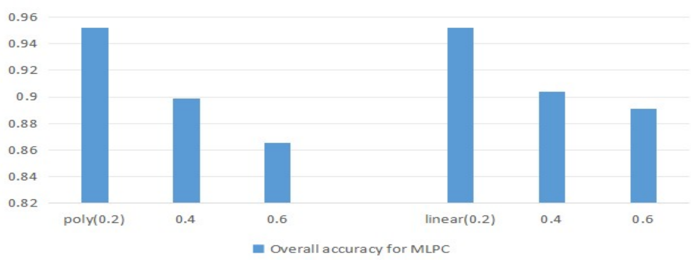

Fig 11: Overall accuracy for MLPC

This Fig-11 shows how accurately the MLP Classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 95.19% when using either kernel type with an 80-20 data split (80% training, 20% testing). Other configurations show more significant performance drops: polynomial kernel accuracy decreases to 89.90% (60-40 split) and 86.54% (40-60 split), while linear kernel achieves 90.38% (60-40) and 89.10% (40-60)

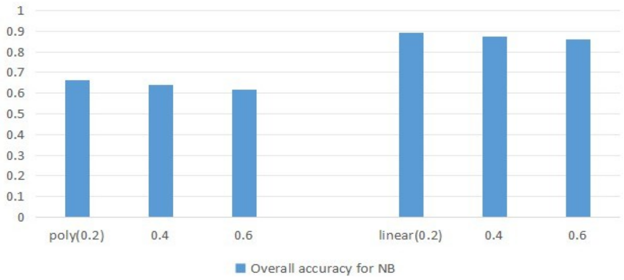

Fig 12: Overall accuracy for NB

This Fig-12 shows how accurately the Naive Bayes classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model shows dramatically different performance between kernels: the linear kernel achieves respectable accuracy of 89.42% at 80-20 split, maintaining 87.50% (60-40) and 86.22% (40-60), while the polynomial kernel performs significantly worse at 66.35% (80-20), 63.94% (60-40), and 61.86% (40-60).

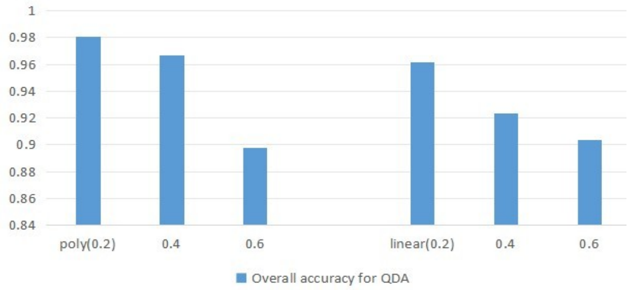

Fig 13: Overall accuracy for QDA

This Fig-13 shows how accurately the QDA classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 98.08% when using the polynomial kernel with an 80-20 data split (80% training, 20% testing). Other configurations show strong but declining performance: the polynomial kernel maintains 96.63% (60-40 split) before dropping to 89.74% (40-60 split), while the linear kernel achieves 96.15% (80-20), 92.31 (60-40), and 90.38% (40-60

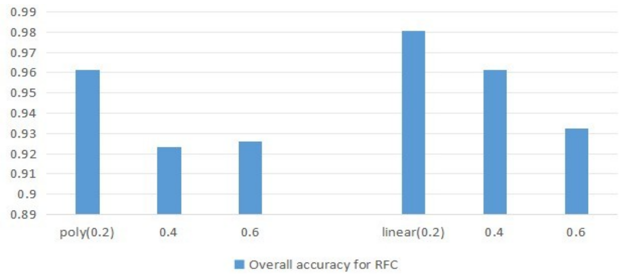

Fig 14: Overall accuracy for RFC

This Fig-14 shows how accurately the Random Forest classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 98.08% when using the linear kernel with an 80-20 data split (80% training, 20% testing). Other configurations show strong performance: the linear kernel maintains 96.15% (60-40 split) and 93.27 (40-60 split), while the polynomial kernel achieves 96.15% (80-20), 92.31% (60-40), and 92.63% (40-60).

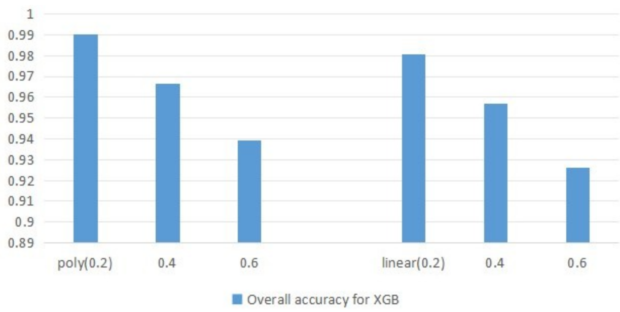

Fig 15: Overall accuracy for XGB

This Fig-15 shows how accurately the XGBoost classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 99.04% when using the polynomial kernel with an 80-20 data split (80% training, 20% testing). Other configurations show strong performance: the polynomial kernel maintains 96.63% (60-40 split) and 93.91% (40-60 split), while the linear kernel achieves 98.08% (80-20), 95.67% (60-40), and 92.63% (40-60).

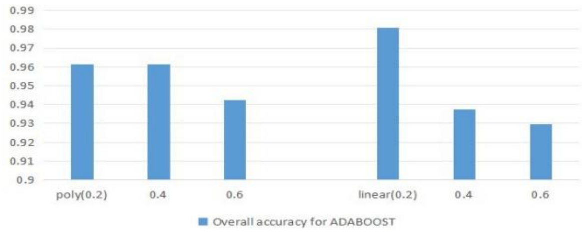

Fig 16: Overall accuracy for ADABOOST

This Fig-16 shows how accurately the AdaBoost classifier predicts diabetes using two different methods (polynomial and linear) with varying train-test splits. The model achieves its peak accuracy of 98.08% when using the linear kernel with an 80-20 data split (80% training, 20% testing). Other configurations show strong performance: the polynomial kernel maintains consistent 96.15% accuracy at both 80-20 and 60-40 splits before dropping slightly to 94.23% (40-60 split), while the linear kernel achieves 93.75% (60-40) and 92.95% (40-60).

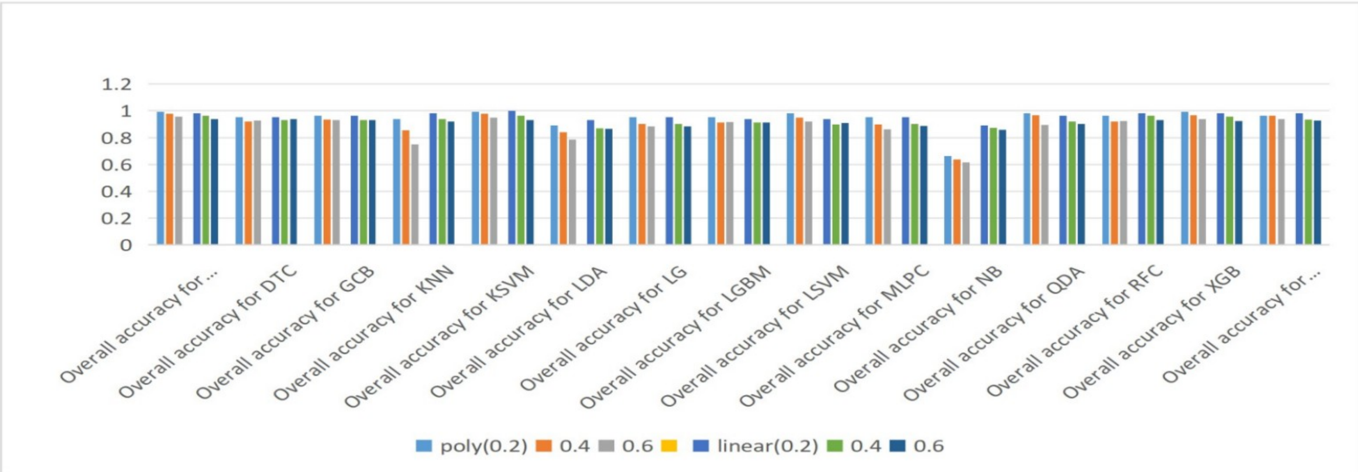

Fig 17: Comparison of the accuracies of all classification algorithms

The Fig-17 tells about analysis of 15 machine learning algorithms reveals KSVM achieves perfect 100% accuracy with a linear kernel at 80-20 split, making it the top performer, followed closely by XGBoost and CatBoost (99% with polynomial kernels) and AdaBoost (98%). Most models maintain excellent 95-98% accuracy at optimal 80- 20 splits, with performance declining slightly (3-25%) for larger test sets, while KSVM's linear configuration proves most reliable for diabetes prediction with its flawless accuracy.

CONCLUSION:

This research tested 15 different computer programs to predict diabetes using health information from 521 people. First, we cleaned and organized the data properly. The best program was KSVM (a special type of computer analysis), which correctly identified all diabetes cases when we used 80% of the data to teach it and 20% to test it. Other good programs were XGBoost and CatBoost (99% accurate) and AdaBoost (98% accurate). This shows that smart computer programs can be very helpful for doctors to detect diabetes early and accurately. The best program worked even better than previous methods used by other researchers. These tools could help catch diabetes sooner so people can get treatment faster.

REFERENCES:

[1] L. Nie et al., "Intrusion Detection in Green Internet of Things: A Deep Deterministic Policy Gradient-Based Algorithm," in IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 778-788, June 2021, doi: 10.1109/TGCN.2021.3073714.

[2] Suguna, R., Suriya Prakash, J., Aditya Pai, H. et al. Mitigating class imbalance in churn prediction with ensemble methods and SMOTE. Sci Rep 15, 16256 (2025). https://doi.org/10.1038/s41598-025-01031-0

[3] Suriya Prakash Jambunathan,Suguna Ramadass,Palanivel Rajan Selva kumaran, "Analyzing the Behavior of Multiple Dimensionality Reduction Algorithms to Obtain Better Accuracy using Benchmark KDD CUP Dataset", The International Arab Journal of Information Technology (IAJIT) ,Volume 19, Number 01, pp. 121 - 131, January 2022, doi: 10.34028/iajit/19/1/14. International Scopes Indexed IEEE.

[4] Suriya Prakash, J., Thinley Tsering Lama, Soniya R., Lakshmanan, V. and Kiran S, “Enhancing security by monitoring the behavior of different classification algorithms” (2025). Recent Trends in VLSI and Semiconductor Packaging (1st ed.). Taylor and Francis, CRC Press. https://doi.org /10.1201/ 978100361 6399

[5] J. Suriya Prakash, T. Rashmika, N. Thangadurai, U. Prakash and S. Kiran, "Elevating Intrusion Detection Precision with Multi-Classification Algorithm Analysis," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-8, doi: 10.1109/NMITCON62075.2024.10698932.

[6] S. P. J, S. N, L. A, Chaithra and K. S, "Enhance Intrusion Detection by Analyzing the Behavior of Labeled and Unlabeled Classification to Obtain Better Accuracy," 2024 Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 791-797, doi: 10.1109/ICACCTech65084.2024.00131.

[7] J. Suriya Prakash, C. H. Guntupalli, S. Narasani, N. N. Srinidhi and S. Kiran, "Boosting Accuracy in Intrusion Detection Systems: A Comprehensive Examination of Dimensionality Reduction and Classification Methods," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-8, doi: 10.1109/NMITCON62075.2024.10698893.

[8] J. Suriya Prakash, P. Deeksha Gandhi, D. Kumar, M. H. Vishwas and Y. Shah, "Enhancing Network Security Through Advanced Intrusion Detection: A Fusion of Dimensionality Reduction and Machine Learning Classification for Improved Accuracy," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-7, doi: 10.1109/NMITCON62075.2024.10699225.

[9] S. P. J, H. S. Nambiar, G. V. Kumar, S. C, C. S. K and S. K. H, "Upgrade Better Accuracy by Analyzing the Behavior of Classifying Algorithm to Detect Intrusion in Network Traffic," 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), Bangalore, India, 2024, pp. 735-742, doi: 10.1109/ICWITE59797.2024.10502430.

[10] Suriya Prakash, J., Suguna, R., Neethu, P.S,"Unleashing the power of YOLOV5: Revolutionizing person detection and counting in restricted zones"2024 Taylor and Francis, CRC Press,PP.16,Edition:1st ,Ebook ISBN: 9781003502470.

[11] C. Lekkalapudi, N. H. V P, S. Jagannathan, V. C and S. P. J, "Refining Intrusion Detection Capabilities Through Combined Algorithmic Classification Techniques," 2024 Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 775-782, doi: 10.1109/ICACCTech65084.2024.00129.

[12] S. Nandini, S. Murthy, P. P. K, P. U and S. P. J, "Enhancing the Security by Analyzing the Behavior of Multiple Classification Algorithms with Dimensionality Reduction to Obtain Better Accuracy," 2024 Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 783-790, doi: 10.1109/ICACCTech65084.2024.00130.

[13] S. P. J, V. V, S. Jagannathan, V. C and K. S, "Improving Intrusion Detection Precision via Multi Classification Algorithm Examination," 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC), Bengaluru, India, 2024, pp. 1-7, doi: 10.1109/ICDSCNC62492.2024.10939569.

[14] S. P. J, T. S, S. R, V. Lakshmanan and K. S, "Improving Accuracy in Network Intrusion Detection via Machine Learning Algorithms: An In-Depth Examination," 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC), Bengaluru, India, 2024, pp. 1-7, doi: 10.1109/ICDSCNC62492.2024.10939819.

[15] J. Suriya Prakash, S. Narasani, N. Thangadurai, U. Prakash and S. Kiran, "Advancing Intrusion Detection Precision Through Analysis of Diverse Classification Algorithms," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/NMITCON62075.2024.10698858.

[16] H. S. Nambiar, L. Rangaiah, P. K. Praksha, C. Vasanthakumar and J. Suriya Prakash, "Optimizing Security Performance: Leveraging Multiclass Algorithms and Dimensionality Reduction for Enhanced Accuracy," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-7, doi: 10.1109/NMITCON62075.2024.10699188.

[17] R. Suguna, Y. Praveen Kumar, J. Suriya Prakash, P. S. Neethu and S. Kiran, "Utilizing Machine Learning for Sport Data Analytics in Cricket: Score Prediction and Player Categorization," 2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon), HASSAN, India, 2023, pp. 1-6, doi: 10.1109/MysuruCon59703.2023.10396955.

[18] S. J and K. S, "Obtain Better Accuracy Using Music Genre Classification Systemon GTZAN Dataset," 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), Vijaypur, India, 2022, pp. 1-5, doi: 10.1109/NKCon56289.2022.10126991.

[19] M. Desai and S. Prakash J, "Expression of Concern for: An Exploration of the Effectiveness of Machine Learning Algorithms for Text Classification," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-1, doi: 10.1109/INCOFT60753.2023.10703757.

[20] S. P. J, C. H. Guntupalli, S. N. Chilamkurthy, G. Kowshik and A. Alekhya, "Enhancing the Security by Analyzing the Behaviour of Multiple Classification Algorithms with Dimensionality Reduction to Obtain Better Accuracy," 2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon), HASSAN, India, 2023, pp. 1-7, doi: 10.1109/MysuruCon59703.2023.10396971.

[21] S. DS, M. Karunya, O. J and S. P. J, "Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 614-618, doi: 10.1109/CISES58720.2023.10183617.

[22] M. Desai and S. Prakash J, "Expression of Concern for: An Exploration of the Effectiveness of Machine Learning Algorithms for Text Classification," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-1, doi: 10.1109/INCOFT60753.2023.10703757.

[23] S. J and K. S, "Obtain Better Accuracy Using Music Genre Classification Systemon GTZAN Dataset," 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), Vijaypur, India, 2022, pp. 1-5, doi: 10.1109/NKCon56289.2022.10126991.

[24] Gnana, S., Ahilan, A., Jayapriya, & Arun, V. (2024). Prediction of Ozone Depletion Levels using Intelligent CNN-SVM Classification System. Global NEST Journal, 26(2), 1-8. https://doi.org/10.30955/gnj.005461

[25] K. . Sreenivasulu, "Cardiovascular Syndrome Prediction Using Machine Learning Algorithms", Int J Intell Syst Appl Eng, vol. 12, no. 15s, pp. 548–555, Feb.