IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

Diabetes Prediction Using Machine Learning

Suvradeep Sahana, Suvankar Biswas, Souhardya Sarkar
Debadrita Aich Sarkar, , Sujata Kundu
Department of Information Technology
Narula Institute of Technology, Kolkata, India
suvradeepsahana10@gmail.com, sb7679250005@gmail.com, sarkarsouhardya67@gmail.com,
debadritaaichsarkar42@gmail.com, sujata.kundu@nit.ac.in

Abstract: *Diabetes is a persistent metabolic condition that impacts millions of individuals globally. Timely identification and intervention are essential for effective management and the prevention of associated complications. This project explores the viability of utilizing machine learning algorithms to forecast the onset of diabetes in individuals. The research employs the well- known Pima Indians Diabetes Dataset, which includes a diverse array of physiological and demographic characteristics. A thorough analysis was carried out using various classification algorithms, such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and K-Nearest Neighbors (KNN). Prior to training the models, extensive data preprocessing methods were applied to handle missing values and standardize features, thereby ensuring optimal performance and generalizability of the models. The evaluation of the models was conducted using a range of metrics, including accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic curve (AUC). The findings indicate that machine learning has significant potential in accurately predicting diabetes, with some algorithms demonstrating superior performance over others. This study underscores the promise of machine learning as an essential resource for healthcare professionals in the early detection of diabetes, facilitating proactive interventions and ultimately enhancing patient outcomes*

Keywords: Diabetes.

I. INTRODUCTION

Diabetes has emerged as a significant global health challenge, affecting millions worldwide. This chronic metabolic disorder disrupts the body's ability to regulate blood sugar levels, leading to a range of complications if left unmanaged. Early detection and intervention are crucial for effective diabetes management and prevention of severe health consequences.

Traditional methods for diabetes diagnosis often rely on clinical examinations and laboratory tests, which can be time-consuming and resource-intensive. This project explores the potential of machine learning to provide an efficient and accurate alternative for early diabetes prediction. By leveraging the power of machine learning algorithms, we aim to develop a predictive model that can effectively identify individuals at high risk of developing diabetes based on their medical and demographic data.

This project will utilize the renowned Pima Indians Diabetes Dataset, a widely used dataset in machine learning research. A variety of classification algorithms, including Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and K-Nearest Neighbors (KNN), will be employed to build and evaluate predictive models. Rigorous data preprocessing techniques, such as handling missing values and feature scaling, will be implemented to ensure optimal model performance and generalizability. The performance of each model will be assessed using a combination of evaluation metrics, including accuracy, precision, recall, F1- score, and the area under the Receiver Operating Characteristic curve (AUC).

The findings of this project have the potential to significantly impact healthcare by assisting healthcare professionals in early diabetes detection. By identifying individuals at high risk, proactive interventions can be implemented, such as lifestyle modifications, medication adjustments, and regular monitoring, leading to improved diabetes management and a reduction in associated health complications.

DOI: 10.48175/IJARSCT-26279

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

II. LITERATURE SURVEY

The application of machine learning (ML) in diabetes prediction has been widely researched, with studies exploring various algorithms, feature selection techniques, and optimization methods to improve predictive accuracy. This literature survey highlights key contributions in this domain, focusing on different ML approaches and their effectiveness in diagnosing diabetes.

1. Traditional Methods for Diabetes Prediction

- Early studies primarily used traditional statistical techniques such as logistic regression (LR) and decision trees (DT) for diabetes prediction.
- Researchers working with the Pima Indians Diabetes Dataset (PIDD) found that logistic regression provided a baseline accuracy but struggled with complex, non-linear patterns in medical data. Decision trees improved interpretability but often suffered from overfitting, limiting their generalization.

2. Supervised Machine Learning Approaches

Several studies have explored supervised learning algorithms, including:

- Support Vector Machines (SVM): Studies have shown that SVM performs well in classifying diabetic and non-diabetic patients, especially when used with kernel functions to capture non-linearity. However, SVM can be computationally expensive with large datasets.
- Random Forest (RF): Research indicates that ensemble models like Random Forest improve diabetes prediction accuracy by reducing bias and variance. Studies comparing RF with other classifiers have found it to be one of the most reliable models due to its ability to handle missing data and feature importance evaluation.
- K-Nearest Neighbors (KNN): KNN has been used for diabetes classification, but its performance varies based on the choice of K- value and dataset size. It is often outperformed by more complex models like neural networks.Deep Learning-Based Approaches

3. Deep Learning-Based Approaches

- Recent studies have applied Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) for diabetes prediction. While ANNs have demonstrated superior accuracy by learning complex feature relationships, they require large amounts of data and careful hyperparameter tuning. CNNs have been explored for analyzing medical images related to diabetes complications, such as diabetic retinopathy detection.

4. Hybrid & Explainable AI Models

- Recent research has explored hybrid models, combining multiple ML techniques to enhance diabetes prediction. For instance, integrating Random Forest with SVM or ANN with Genetic Algorithms has shown promising results. Additionally, Explainable AI (XAI) has gained attention for interpreting ML model decisions, helping healthcare professionals understand predictions and trust AI-based diagnostic tools.

III. AIM & OBJECTIVE

The aim of this project is to develop an accurate and reliable machine learning-based system for predicting the likelihood of diabetes in individuals using clinical and demographic data. To achieve this, the project focuses on several key objectives. First, it involves collecting and preprocessing a relevant dataset that includes medical and lifestyle features associated with diabetes. Suitable machine learning algorithms are then explored and selected to address the binary classification nature of the problem. These models are trained, validated, and evaluated using standard performance metrics such as accuracy, precision, recall, F1- score, and AUC-ROC to ensure effectiveness. Hyperparameter tuning is carried out to optimize the models for improved performance. To enhance generalizability and reduce bias, the system uses cross-validation and diverse data sources. The inclusion of explainable AI (XAI)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

techniques aims to improve the transparency and trustworthiness of the predictions for healthcare professionals. Additionally, the system is designed for ongoing monitoring and maintenance, allowing for periodic updates with new data to maintain performance over time. To encourage user adoption, the final product includes a user-friendly interface and provides appropriate documentation and support for healthcare professionals.

IV. THE MODEL

System Architecture

1. Requirements Analysis:

Functional Requirements-

- Data Input: The system should accept patient data as input, including relevant features like age, BMI, glucose levels, blood pressure, insulin levels, family history, etc. Input methods could include CSV files, manual entry forms, or integration with existing databases.
- Data Preprocessing: The system must perform necessary data preprocessing steps, including:

Handling missing values (imputation, removal). Feature scaling (normalization, standardization). Outlier detection and treatment.

Feature engineering (optional).

- Model Selection and Training: The system should allow for the selection and training of various machine learning models (e.g., Logistic Regression, SVM, Random Forest, Neural Networks). It should support hyperparameter tuning for optimal model performance.

Non-Functional Requirements-

- Performance: The system should provide predictions quickly and efficiently.
- Accuracy: The predictions should be as accurate as possible.
- Scalability: The system should be able to handle large datasets and a high volume of prediction requests.
- Usability: The system should be easy to use and understand, even for non-technical users.
- Security: Patient data should be stored and processed securely, complying with relevant privacy regulations.
- Maintainability: The system should be designed for easy maintenance and updates.
- Portability: The system should be deployable on different platforms.

2. Feasibility Study:

- Technical Feasibility: Python, along with libraries like scikit- learn, pandas, and NumPy, provides a robust platform for developing the system. Cloud computing platforms can be used for handling large datasets and deployment.
- Economic Feasibility: The cost of development depends on the complexity of the system and the chosen deployment method. Open- source tools and cloud computing services can help minimize costs.
- Operational Feasibility: The system can be integrated into existing healthcare workflows with proper training and support for healthcare professionals.
- Legal and Ethical Feasibility: Compliance with data privacy regulations (e.g., HIPAA) is crucial. Ethical considerations regarding the use of AI in healthcare must be addressed.

3. Potential Challenges:

- Data Quality: Inconsistent or incomplete data can negatively impact model performance.
- Model Bias: The training data might contain biases that can lead to unfair or inaccurate predictions.
- Interpretability: Understanding why a model makes a particular prediction can be challenging, especially with complex models.
- Generalizability: Models trained on one population might not perform well on other populations.
- Integration with Existing Systems: Integrating the system with existing healthcare IT infrastructure can be complex.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

- User Acceptance: Healthcare professionals might be hesitant to adopt new technologies.

4. Mitigation Strategies:

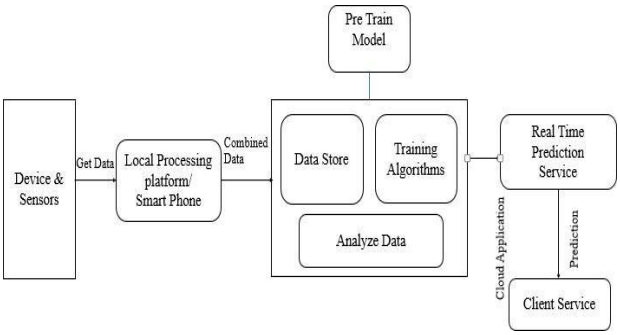
To ensure the responsible and effective use of AI in healthcare, several key factors must be considered. Data quality is foundational and requires the implementation of robust data cleaning and validation procedures to ensure accurate, consistent, and reliable inputs. Addressing model bias is essential and can be achieved by using diverse and representative datasets, along with bias-mitigation techniques, to promote fairness across different patient populations. Enhancing interpretability through explainable AI (XAI) techniques allows clinicians to understand and trust AI-driven decisions, thereby improving transparency and clinical utility. To ensure generalizability, models should be trained and validated on varied datasets using methods like cross- validation, enabling them to perform well across different healthcare settings. Early integration planning is crucial to ensure the AI system aligns seamlessly with existing healthcare infrastructure and workflows. Finally, fostering user acceptance involves providing adequate training and ongoing support to healthcare professionals, empowering them to effectively utilize AI tools in their practice.

5. Success Metrics:

When evaluating the performance and impact of an AI system in healthcare, several key metrics should be considered. Prediction accuracy refers to the percentage of correct predictions made by the model, providing a general sense of its effectiveness. Sensitivity (recall) measures the model's ability to correctly identify individuals with diabetes, while specificity assesses its ability to correctly identify those without the condition—both critical for minimizing false positives and negatives. The F1-score offers a balanced evaluation by combining precision and recall, especially useful when dealing with imbalanced datasets. The AUC- ROC (Area Under the Receiver Operating Characteristic Curve) reflects the model's overall ability to distinguish between classes, offering insight into its discriminative power. Beyond technical performance, the user adoption rate is a vital metric, indicating the extent to which healthcare professionals actively use the AI system in practice—an essential factor for real-world success and clinical impact.

This system analysis provides a comprehensive overview of the requirements, feasibility, and potential challenges associated with developing a diabetes prediction system using machine learning. By carefully considering these factors, developers can create a system that is accurate, reliable, and useful for improving diabetes care.

Framework-



Working of System:-

1. Data Source:

The foundation of the system is the dataset used for training and testing. In this case, it's likely the Pima Indians Diabetes Dataset, but it could be any other relevant dataset containing medical information related to diabetes.This dataset provides the raw data, including features like age, BMI, glucose levels, blood pressure, etc., and the target variable (whether the individual has diabetes or not)

2. Data Preprocessing:

This is a crucial step. Raw data is rarely clean and often needs transformation before it can be used for model training. This stage includes:

DOI: 10.48175/IJARSCT-26279

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

- Handling Missing Values: Strategies like imputation (filling in missing values with the mean, median, or more sophisticated methods) are used.
- Feature Scaling: Normalizing or standardizing the features to a similar range prevents features with larger values from dominating the model.
- Feature Engineering (Optional): Creating new features from existing ones that might improve model performance (e.g., creating a "glucose level category" from raw glucose values).
- Outlier Detection and Treatment: Identifying and handling extreme values that could negatively impact the model.
- Data Cleaning: Removing duplicates, correcting inconsistencies, and ensuring data quality.

3. Model Training:

This stage involves selecting an appropriate machine learning algorithm (or multiple algorithms for comparison) and training it on the preprocessed data. Algorithms like Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, or Neural Networks could be used.The dataset is typically split into training and testing sets. The model learns patterns from the training data.

4. Model Evaluation:

The trained model's performance is evaluated on the testing data (data the model hasn't seen during training). Metrics like accuracy, precision, recall, F1-score, and AUC-ROC are used to assess how well the model predicts diabetes. Techniques like cross-validation might be used to ensure robust evaluation.

5. Model Deployment:

Once a satisfactory model is trained and evaluated, it can be deployed for real-world use. This could involve:

- Creating an API: So that other applications can access the model's predictions.
- Integrating the model into a web or mobile application: To make it accessible to users (e.g., healthcare professionals or individuals).
- Creating a dashboard: To visualize predictions and provide insights.

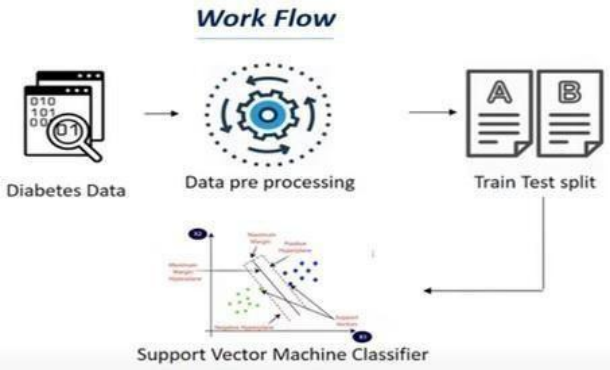
6. Data Flow:

The data flows sequentially from the Data Source through the Preprocessing, Training, and Evaluation stages. The evaluation results inform the model selection and can lead to adjustments in the preprocessing or training phases. Finally, the chosen model is deployed.

Key Considerations:

The development of a diabetes prediction system using machine learning involves several critical steps to ensure optimal performance and long-term reliability. Algorithm selection is a foundational decision, influenced by the characteristics of the data and the specific goals of the project. Once an appropriate algorithm is chosen, hyperparameter tuning is essential to fine-tune the model's performance by adjusting its internal settings. After deployment, continuous monitoring and maintenance are crucial to detect any performance degradation over time, which may require retraining the model with updated data to maintain accuracy and relevance. Overall, this system architecture offers a structured and effective approach to building a robust diabetes prediction system, but it should always be tailored to the unique requirements of the project and the selected algorithms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67



V. KEY ADVANTAGES

Diabetes prediction using Machine Learning with Python offers several key advantages:

1. Early Detection and Prevention: Machine learning models can analyze patient data and identify individuals at high risk of developing diabetes before they exhibit clinical symptoms. This allows for early interventions, lifestyle changes, and preventative measures to delay or even prevent the onset of the disease. This is crucial as early-stage diabetes often has no noticeable symptoms.
2. Improved Accuracy: Compared to traditional risk assessment methods, machine learning models can often achieve higher accuracy in predicting diabetes. They can learn complex patterns and relationships in the data that might be missed by human analysis.
3. Personalized Predictions: Machine learning models can be trained on large datasets that include diverse populations. This allows for the development of more personalized prediction models that take into account individual risk factors and characteristics, leading to more tailored interventions.
4. Cost-Effectiveness: Early detection and prevention can lead to significant cost savings in the long run by reducing the need for expensive treatments and hospitalization due to diabetes complications. Machine learning-based screening can be more efficient than traditional methods, potentially reducing healthcare costs.
5. Automation and Scalability: Once a machine learning model is trained, it can be easily deployed and used to analyze large volumes of patient data quickly and automatically. This makes it a scalable solution for population-wide screening and risk assessment.
6. Identification of Risk Factors: Machine learning models can help identify the most important risk factors for diabetes in specific populations. This information can be used to develop targeted prevention programs and public health initiatives.
7. Data-Driven Insights: Machine learning provides data- driven insights into the factors that contribute to diabetes development. This can be valuable for researchers and healthcare professionals in understanding the disease and developing better prevention and treatment strategies.
8. Continuous Improvement: As more data becomes available, machine learning models can be retrained and updated to improve their accuracy and performance over time. This allows for continuous refinement of the prediction process.
9. Integration with Existing Systems: Machine learning models can be integrated with existing electronic health records (EHRs) and other healthcare IT systems, making it easier to use them in clinical practice.
10. Reduced Burden on Healthcare Professionals: By automating the risk assessment process, machine learning can reduce the burden on healthcare professionals, allowing them to focus on other important tasks.

In summary, diabetes prediction using machine learning with Python has the potential to revolutionize diabetes care by enabling early detection, personalized interventions, and more efficient use of healthcare resources. It empowers both individuals and healthcare providers to take proactive steps in managing and preventing this chronic disease.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

VI. LIMITATIONS

While diabetes prediction using Machine Learning with Python offers significant advantages, it also has limitations that need to be considered:

1. Data Dependency: Machine learning models are heavily reliant on the quality and quantity of data. If the data is incomplete, inconsistent, or biased, the model's performance can be negatively affected. A lack of diverse data can also limit the model's generalizability.
2. Data Privacy and Security: Working with sensitive patient data raises significant privacy and security concerns. Appropriate measures must be taken to protect patient information and comply with regulations like HIPAA.
3. Model Interpretability: Some machine learning models, especially complex ones like deep learning models, can be "black boxes," meaning it's difficult to understand why they make a particular prediction. This lack of interpretability can be a challenge in healthcare settings where understanding the reasoning behind a prediction is important.
4. Overfitting: If a model is too complex and trained on a limited dataset, it might overfit the training data and perform poorly on unseen data. Techniques like cross-validation and regularization can help mitigate overfitting, but it remains a potential issue.
5. Generalizability: A model trained on one population might not perform well on another population with different characteristics. External validation on diverse datasets is crucial to ensure generalizability.
6. Bias in Data: If the training data contains biases, the model can inherit those biases and make unfair or inaccurate predictions for certain groups of people. Addressing bias in data is a complex but essential task.
7. Need for Expertise: Developing and deploying machine learning models requires expertise in data science, machine learning, and software engineering. Healthcare organizations might need to invest in training or hiring specialized personnel.
8. Maintenance and Updates: Machine learning models need to be regularly maintained and updated to ensure their performance remains high as new data becomes available. This requires ongoing effort and resources.
9. Ethical Considerations: The use of AI in healthcare raises ethical questions about responsibility, accountability, and potential biases. Careful consideration of these ethical implications is essential.
10. Integration Challenges: Integrating machine learning models into existing healthcare IT systems can be complex and time-consuming. Interoperability issues might arise.
11. User Acceptance: Healthcare professionals might be hesitant to adopt new technologies like machine learning-based prediction tools. Proper training and communication are needed to ensure user acceptance.
12. Explainability and Trust: If healthcare professionals don't understand how a model makes predictions, they might be less likely to trust it. Explainable AI (XAI) techniques are being developed to address this issue, but it remains a challenge.
13. 13. Cost of Development and Deployment: Developing and deploying a robust machine learning-based prediction system can be expensive, requiring investments in data collection, software development, and infrastructure.

It's important to be aware of these limitations and take appropriate steps to mitigate them when developing and deploying diabetes prediction systems using machine learning. Addressing these challenges will lead to more effective and reliable solutions for improving diabetes care.

VII. FUTURE SCOPE

The future scope of diabetes prediction using machine learning with Python is vast and promising, driven by continuous advancements in healthcare data availability, computing power, and algorithmic efficiency. As more real- time and longitudinal patient data becomes accessible, predictive models can be further refined to detect diabetes at earlier stages, even before clinical symptoms appear.The incorporation of multi-faceted data, such as genomics, lifestyle factors, and real-time information from wearable technology, presents vast opportunities for developing more comprehensive and adaptable risk assessments. Enhanced machine learning methodologies, including deep learning and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

ensemble approaches, will improve prediction precision, while the rise of Explainable AI (XAI) will foster greater transparency and confidence in these models. Tailored prediction and intervention will become feasible through customized risk evaluations and responsive interventions provided through mobile health applications. Cloud computing will facilitate the creation of scalable and accessible systems for immediate data analysis and broader application. A heightened emphasis on predictive analytics for prevention, in combination with focused public health strategies, will support proactive management of diabetes. It is essential to tackle ethical issues such as eliminating bias and protecting data privacy to ensure ethical use of AI. Ultimately, collaborative efforts across disciplines and knowledge exchange will expedite advancements, pushing the field toward enhanced diabetes treatment and ultimately, healthier outcomes.

VIII. CONCLUSION

In summary, this investigation into diabetes prediction utilizing Machine Learning with Python has highlighted the considerable potential of this methodology to transform diabetes management. By harnessing the power of data analysis and advanced algorithms, we can transcend conventional risk assessment approaches to attain earlier and more precise predictions. The capability to pinpoint individuals at risk prior to any symptoms surfacing presents a crucial opportunity for preventive measures and tailored management strategies. Although challenges such as data quality, model interpretability, and ethical issues persist, ongoing research and development in areas like explainable AI, multi-modal data integration, and sophisticated machine learning methods are set to further enhance the effectiveness of these systems. As technology progresses and our insights into diabetes grow, machine learning-based prediction tools will increasingly become essential in addressing this widespread health challenge, ultimately leading to better patient outcomes and a healthier future.

The main aim of this project was to design and implement Diabetes Prediction Using Machine Learning Methods and Performance Analysis of that methods and it has been achieved successfully. The proposed approach uses various classification and ensemble learning method in which SVM, Knn, Random Forest, Decision Tree, Logistic Regression and Gradient Boosting classifiers are used. And 83 % classification accuracy has been achieved. The Experimental results can be asst health care to take early prediction and make early decision to cure diabetes and save humans life.

REFERENCES

[1]. Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University - Computer and Information Sciences 25, 127–136. doi:10.1016/j.jksuci.2012.10.003.

[2]. Arora, R., Suman, 2012. Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. International Journal of Computer Applications 54, 21–25. doi:10.5120/8626-2492.

[3]. Bamnote, M.P., G.R., 2014. Design of Classifier for Detection of Diabetes Mellitus Using Genetic Programming. Advances in Intelligent Systems and Computing 1, 763– 770. doi:10.1007/978-3-319-11933-5.

[4]. Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection, in: Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016), pp. 451–455.

[5]. Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE. pp. 5–10.

[6]. Sharief, A.A., Sheta, A., 2014. Developing a Mathematical Model to Detect Diabetes Using Multigene Genetic Programming. International Journal of Advanced Research in Artificial Intelligence (IJARAI) 3, 54–59. doi:doi:10.14569/IJARAI.2014.031007.

[7]. Sisodia, D., Shrivastava, S.K., Jain, R.C., 2010. ISVM for face recognition. Proceedings - 2010 International Conference on Computational Intelligence and Communication Networks, CICN 2010 , 554–559doi:10.1109/CICN.2010.109.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

[8]. Sisodia, D., Singh, L., Sisodia, S., 2014. Fast and Accurate Face Recognition Using SVM and DCT, in: Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012, Springer. pp. 1027–1038.
[9]. https://www.kaggle.com/johndasilva/diabetes
[10]. Rani, A. S., & Jyothi, S. (2016, March). Performance analysis of classification algorithms under different datasets. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (pp. 1584-1589). IEEE
[10]. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He and Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", Elsevier Informatics in Medicine, vol. 10, pp. 100-107, 2018.
[11]. VeenaVijayan.V, Anjali.C,"Decision Support Systems for Predicting Diabetes Mellitus –A Review", Proceedings of 2015 Global Conference on Communication Technologies(GCCT 2015), 978-1- 4799-8553- 1/15/$31.00
© s2015 IEEE.
[12]. Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine Learning Techniques".Int. Journal of Engineering Research and Application, Vol. 8