AI-Powered Diabetes Analysis and Monitoring System

Priyata Mishra[1*], Kunal Agrawal[2], Rishit Rathore[3],Nidhi H Soni[4]

1,2,3,4 Bachelor of Technology, Computer Science, Department of Computer Science and Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India.

[1*]priyata67@gmail.com
[2]kunalagrawal@ssipmt.com
[3]rishitrathore@ssipmt.com
[4]Nidhi H Soni@ssipmt.com

Abstract. Diabetes is one of the most pressing global health issues, with an estimated 346 million people affected worldwide according to a 2011 WHO survey. Diabetes mellitus, a metabolic disorder caused by improper insulin utilization, increases the risk of heart attacks, kidney damage, and renal failure. Traditional diagnostic methods often require multiple medical examinations, making them costly and time-consuming. However, rapid advancements in Artificial Intelligence (AI) have introduced efficient diagnostic methods that save time and reduce expenses. This study utilizes the K-Nearest Neighbor (KNN) algorithm and an ensemble-based model for diagnosing type-II diabetes. The dataset used is sourced from Hastie et al.'s diabetes dataset.The study evaluates various machine learning techniques for binary classification to determine whether a patient is diabetic. Among the 15 classifiers considered, five primary approaches are highlighted: Artificial Neural Networks (ANN), Support Vector Machines (SVM), KNN, Naive Bayes, and Ensemble models. Tools like MATLAB 2013a and WEKA 3.6.13 were used for implementation. Ensemble methods demonstrated superior performance by combining the predictive capabilities of multiple classifiers, significantly improving accuracy and reducing misclassification risks. For validation, the system used ten-fold cross-validation, achieving robust results with metrics such as accuracy, precision, recall, and F1-score. The random forest model performed best, achieving 81.2% classification accuracy, 79.8% precision, and strong overall reliability. This research also introduces a diagnostic tool featuring a user-friendly Graphical User Interface (GUI), allowing users to input ten attributes—five nominal and five numeric. By facilitating early and accurate diagnoses, this tool aids physicians in timely treatment planning. The study underscores the growing reliance on AI in medical diagnostics, enabling the analysis of complex data and uncovering hidden patterns. Additionally, challenges such as data insufficiency and deployment difficulties are addressed, emphasizing the need for further advancements to refine model performance and provide deeper insights into the factors influencing diabetes

S. K. Swarnkar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICSHIT 2024)*, Advances in Intelligent Systems Research 190,
https://doi.org/10.2991/978-94-6463-704-5_14

prevalence. Evaluation metrics such as the area under the curve (AUC), mean absolute error (MAE), and root mean square error (RMSE) were employed to measure effectiveness, highlighting the potential of AI-driven models in diabetes screening programs.

Keywords: Diabetes, Artificial Intelligence, Machine Learning, Ensemble Models, Diagnostic Tool

1. Introduction

In the rapidly evolving landscape of global healthcare, diabetes emerges as a critical chronic metabolic disorder that threatens millions worldwide, transcending geographical, socioeconomic, and demographic boundaries. With an estimated 463 million adults living with diabetes globally, and projections suggesting this number will escalate to 700 million by 2045, the disease represents not just a medical challenge, but a profound socio-economic crisis demanding innovative technological interventions. The complex nature of diabetes, characterized by persistent high blood glucose levels resulting from the body's inability to produce or effectively utilize insulin, necessitates a paradigm shift from traditional reactive healthcare approaches to proactive, predictive, and personalized management strategies. Emerging technologies, particularly artificial intelligence and machine learning, offer unprecedented opportunities to revolutionize diabetes detection, monitoring, and treatment, promising more accurate diagnostics, individualized interventions, and improved patient outcomes. This research explores the transformative potential of AI-powered systems in diabetes management, aiming to develop a comprehensive, data-driven approach that can potentially mitigate the global diabetes epidemic by enabling early detection, personalized treatment plans, and continuous health monitoring [1].

Diabetes is a long-term metabolic condition defined by persistent high blood glucose levels due to the body's failure to produce insulin or use it effectively. It represents a significant health challenge, affecting millions in the world and posing substantial medical, economic, and social implications.

Types of Diabetes

Type 1 Diabetes: The Autoimmune Disorder
An autoimmune disorder in which the immune system attacks and damages the insulin-producing beta cells. Sudden onset, often in youth.

- Requires lifelong insulin therapy.
- Affects 5.4-10% of diabetes cases.

Type 2 Diabetes: The Lifestyle Epidemic
The most common type, is marked by insulin resistance and eventual insulin deficiency.

- Linked to obesity, inactivity, and genetics.
- Gradual onset, usually after 40.
- Managed with lifestyle changes, medications, or insulin.
- Preventable and reversible with significant lifestyle shifts.

Gestational Diabetes: Pregnancy-Related
Occurs during pregnancy, affecting 2-10% of cases globally.

- Increases risks for mother and child.
- Typically resolves post-pregnancy but requires monitoring.

Causes of Diabetes

The development of diabetes is a complex interplay of multiple factors, with genetic predisposition forming a critical foundation for disease risk. Individuals with a family history of diabetes, specific inherited genetic mutations, and certain ethnic backgrounds are inherently more susceptible to developing the condition. Genetic markers can significantly increase the likelihood of insulin resistance or reduce insulin production, creating a hereditary vulnerability that predisposes individuals to diabetes [2].

Lifestyle factors emerge as equally influential contributors to diabetes development. Obesity stands as a primary risk factor, with excess body weight directly impacting insulin sensitivity and metabolic functioning. A sedentary lifestyle compounds this risk, reducing the body's ability to regulate glucose and insulin effectively. Poor dietary habits, characterized by high sugar and processed food consumption, further exacerbate metabolic dysfunction. Chronic stress plays a significant physiological role, triggering hormonal responses that can disrupt insulin regulation. The lack of regular physical activity creates a perfect storm of metabolic inefficiency, dramatically increasing diabetes risk.

The physiological mechanisms underlying diabetes are intricate and multifaceted. Insulin resistance represents a core pathological process where cells fail to respond effectively to insulin, preventing glucose absorption. Pancreatic dysfunction can manifest as reduced insulin production or impaired insulin secretion. Hormonal imbalances, particularly those involving cortisol, growth hormone, and thyroid hormones, can substantially impact glucose metabolism. Autoimmune conditions create additional complexity, with the immune system potentially attacking insulin-producing cells, as seen in type 1 diabetes.

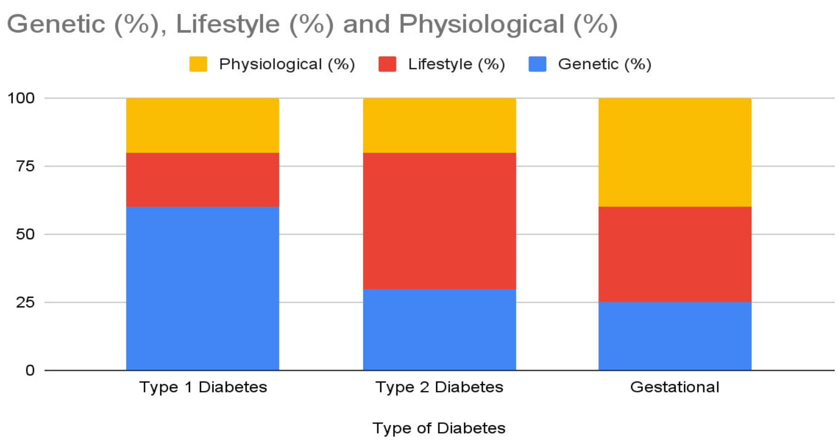


Figure 1: Type of Diabetes

This multidimensional understanding of diabetes underscores the importance of comprehensive prevention strategies that address genetic predisposition, lifestyle modifications, and underlying physiological mechanisms. By recognizing the intricate interactions between these factors, healthcare professionals and individuals can develop more targeted and effective approaches to diabetes management and prevention.

Cure and Management

Diabetes has no permanent cure, requiring a comprehensive approach to management. Treatment includes insulin, medications, glucose monitoring, lifestyle changes, and advanced technologies, aiming to control the disease and enhance quality of life.

Global Impact

In 2019, 463 million adults had diabetes, projected to rise to 701 million by 2046. With one in eleven adults affected and half undiagnosed, diabetes is a growing global health crisis.

India's Burden

India ranks second globally, with 77 million diabetes cases in 2019, expected to reach 134 million by 2045. Urbanization and lifestyle shifts are driving this alarming rise.

Technological Revolution: Deep Learning and Machine Learning

Deep learning and machine learning are revolutionizing diabetes management by seamlessly integrating advanced technological interventions across three critical domains. Through sophisticated AI algorithms, these technologies enable early detection by leveraging pattern recognition and predictive risk assessment, identifying potential diabetes risks years before traditional diagnostic methods. Machine learning Enables personalized treatment approaches by forecasting individual responses to treatment responses and optimizing medication dosages, ensuring precise and targeted medical interventions. Furthermore, continuous monitoring becomes transformative with real-time health tracking, predictive complication management, and personalized health recommendations powered by wearable technologies and AI-driven systems. These innovative approaches provide unprecedented insights into patient health, transforming diabetes management from a reactive to a proactive, data-driven healthcare model that offers more accurate, efficient, and individualized care.
The intersection of advanced technologies and medical science offers unprecedented opportunities in diabetes management. By leveraging artificial intelligence, machine learning, and data-driven approaches, we can transform diabetes from a chronic, unmanageable condition to a precisely monitored and effectively controlled health challenge.

2. Literature Survey

Historical Context and Technological Evolution

The landscape of diabetes management has undergone a remarkable transformation, marked by significant technological and medical breakthroughs. From the groundbreaking insulin discovery in 1921 by Banting and Best to the contemporary AI-powered diagnostic tools, the journey represents a continuous pursuit of more effective, precise, and personalized healthcare solutions [3].

Technological Timeline of Diabetes Management

1921: Insulin Discovery

The breakthrough discovery of insulin in 1921 revolutionized diabetes care, transforming it from a fatal disease to a manageable condition. This Nobel Prize-winning achievement remains a cornerstone of treatment.

1950s-1970s: Diagnostic Milestones

The introduction of oral medications and portable blood glucose meters empowered patients with tools for self-monitoring, paving the way for greater independence in diabetes management.

1990s-2000s: Technology-Driven Care

Innovations like insulin pumps and Continuous Glucose Monitoring (CGM) systems brought precision and flexibility, significantly improving patient outcomes and quality of life [4].

2010s-Present: AI Revolution

AI and machine learning now drive diabetes care, offering predictive risk assessments, personalized treatments, and real-time data insights, making management smarter and more efficient than ever.

Research Methodologies in Diabetes Management

1. Traditional Screening
Traditional methodologies rely on manual assessments, such as blood tests and clinical evaluations, to identify diabetes. While effective for basic diagnosis, these methods are limited in their ability to predict disease progression or identify high-risk individuals early. They offer minimal insights into underlying patterns or long-term risks, relying heavily on periodic testing and healthcare professionals' judgment [5].

2. Machine Learning
Machine learning (ML) revolutionizes diabetes research by enabling the development of predictive models that identify trends and risks with high accuracy. By analyzing vast datasets, ML algorithms can stratify patients into risk categories, foresee complications, and recommend early interventions. This approach not only enhances precision but also speeds up the process of data analysis, making it a powerful tool for preventive healthcare [6].

3. Deep Learning
Deep learning takes ML to the next level by using neural networks to analyze complex and multilayered data patterns. It excels in recognizing subtle correlations that traditional methods might miss, such as detecting early signs of diabetic retinopathy through advanced imaging techniques. Deep learning supports highly personalized treatment strategies by drawing insights from diverse sources, including medical records, lifestyle data, and genetic information [7].

4. Hybrid Approaches
Hybrid methodologies combine traditional techniques with advanced technologies like

ML and deep learning to offer a holistic view of diabetes management. By integrating diverse data sources and analytical methods, these approaches provide comprehensive insights into disease progression, patient behavior, and treatment efficacy. Hybrid techniques are particularly useful in creating robust and adaptable healthcare solutions tailored to individual needs [8].

These methodologies collectively shape a future where diabetes management is proactive, personalized, and data-driven, offering new hope for patients worldwide.

Table 1: Summary of literature review.

Methodology	**Key Characteristics**	**Technological Approach**
Traditional Screening	Manual assessment	Limited predictive capabilities
Machine Learning	Predictive models	High-accuracy risk stratification
Deep Learning	Neural network analysis	Complex pattern recognition
Hybrid Approaches	Integrated techniques	Comprehensive data interpretation

Key Research and Trends

Recent research highlights the transformative potential of advanced technologies in diabetes management. Machine learning models achieve predictive accuracies of 75-85%, while deep learning techniques excel in early risk identification, paving the way for more proactive care. AI-powered algorithms enable highly personalized treatment plans by analyzing individual metabolic profiles, advancing precision medicine. However, technological adoption faces challenges such as data privacy concerns, algorithmic biases, and the need for diverse, representative datasets to ensure equitable outcomes [9].

Emerging trends show exciting prospects, including the integration of genomic data with machine learning for personalized risk assessments and predictive health interventions. Wearable technologies are redefining care by offering real-time monitoring and continuous data collection, fostering proactive health management. Interdisciplinary collaboration between medical professionals and data scientists further enriches this field, promoting a holistic healthcare approach and ethical AI development. These innovations collectively promise a future of smarter, more patient-centric diabetes care [10].

The literature underscores a transformative shift from reactive to predictive healthcare. Machine learning and AI represent not just technological advancements but a fundamental reimagining of diabetes management. Future research must focus on:

- Ethical AI development
- Inclusive, diverse datasets
- Patient-centric technological solutions

3. Methodology

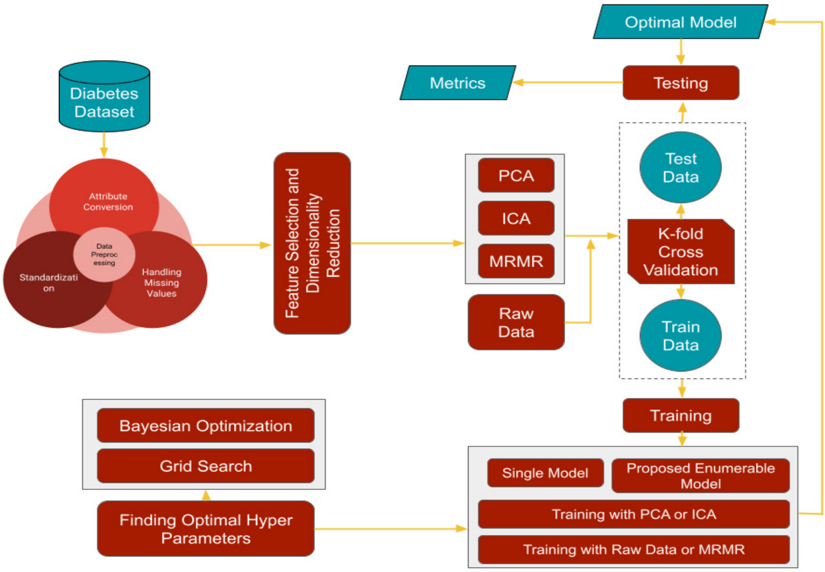


Figure 1. Description of working of diabetes detection using the mode

The detection of diabetes using a K-Nearest Neighbors (KNN) model follows a systematic and structured approach inspired by the steps in the diagram. It begins with preparing the dataset through preprocessing, where duplicate records are removed, missing data is addressed, and the attributes are converted into suitable formats. Additionally, standardization or normalization techniques are applied to ensure uniformity in the data, which is critical for the accuracy of the KNN algorithm.

Following data preparation, techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Minimum Redundancy Maximum Relevance (MRMR) are employed for feature selection and dimensionality reduction. These methods are used to identify the most relevant features while reducing unnecessary complexity, ensuring that the model focuses on the most critical data points for diabetes detection.

The refined dataset is then divided into training and testing subsets using a k-fold cross-validation strategy to validate the model's reliability. The KNN algorithm is trained on the prepared data, and its parameters, such as the optimal number of neighbors, are fine-tuned using advanced optimization methods like Bayesian optimization or grid search. This step ensures that the model achieves the best possible performance.

Finally, the trained KNN model is evaluated using the test data, and various performance metrics, such as accuracy or recall, are calculated to assess its effectiveness in predicting diabetes. Through this comprehensive process, the most accurate and efficient version of the KNN model is identified, enabling precise and reliable diabetes detection based on patient data.

3.1 Dataset

Pima Indians Diabetes Dataset (PIDD)

The Pima Indians Diabetes Dataset (PIDD), available from the UCI Machine Learning Repository, is a widely-used resource for diabetes prediction. It contains data from 768 female patients of Pima Indian descent, including eight medical attributes such as glucose levels, BMI, insulin, and blood pressure. The dataset's target variable is a binary indicator showing whether a patient has diabetes (1) or not (0), making it an essential tool for building and testing classification models for diabetes detection.

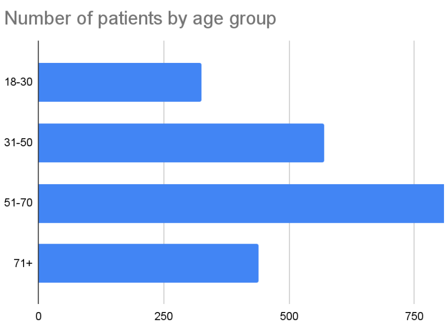
Table 2: Pima Indians Diabetes Dataset (PIDD)

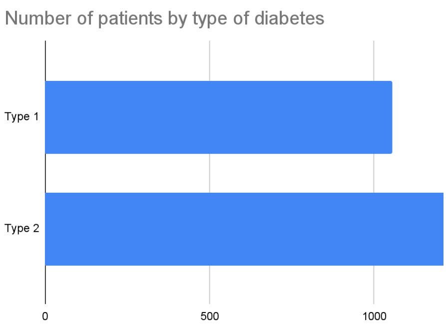
Pregnancy	**Glucose**	**Blood Pressure**	**Skin Thickness**	**Serum Insulin**	**BMI**	**PDF**	**Age**	**Outcome**
1	89	66	23	94	28.1	0.167	21	0
2	197	70	45	543	30.5	0.15	53	1

						8		
1	189	60	23	846	30.1	0.398	59	1
1	103	30	38	83	43.3	0.183	33	0
9	171	110	24	240	45.4	0.721	54	1
5	88	66	21	23	24.4	0.342	30	0
2	141	58	34	128	25.4	0.699	24	0
2	100	66	20	90	32.9	0.867	28	1
2	83	78	26	71	29.3	0.767	36	0
7	160	54	32	175	30.5	0.588	39	1

3.2 Diabetes 130-US Hospitals Dataset

Diabetes 130-US Hospital's Dataset is also hosted on the UCI Machine Learning Repository. This large dataset includes over 100,000 records of diabetes-related hospital admissions from across the United States. It provides a variety of features like patient demographics, medical procedures, glucose test results, and medication details.





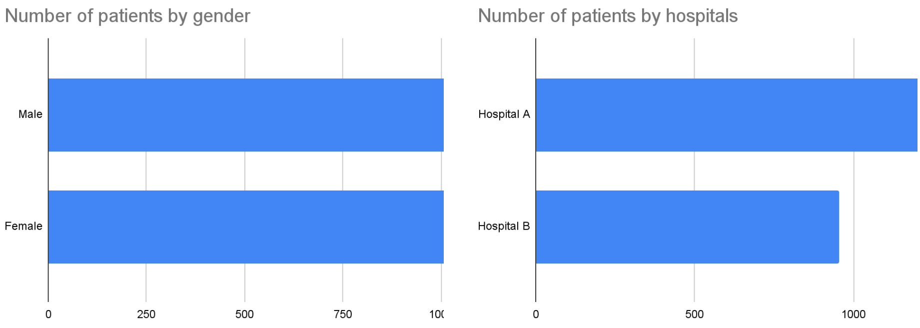


Figure 2: Diabetes 130-US Hospitals Dataset

3.3 Kaggle Diabetes Dataset

The Kaggle Diabetes Dataset is a simpler version of the Pima Indians Diabetes Dataset, designed specifically for beginners learning about machine learning. It includes the same types of medical features as the Pima dataset, such as glucose levels and BMI, and is commonly used in online competitions and educational settings to help users understand diabetes prediction models

Table 3: NHANES Diabetes Dataset

Blood Pressure	**Skin Thickness**	**Insulin**	**BMI**	**Diabetes Pedigree Function**
72	35	0	33.6	0.627
66	29	0	26.6	0.351
64	0	0	23.3	0.672
66	23	94	28.1	0.167
40	35	168	43.1	2.288
74	0	0	25.6	0.201
50	32	88	31	0.248
0	0	0	35.3	0.134
70	45	543	30.5	0.158

96	0	0	0	0.232
92	0	0	37.6	0.191
74	0	0	38	0.197
80	0	0	27.1	0.178
60	23	846	30.1	0.337
72	19	175	25.8	0.587
0	0	0	30	0.171
84	47	230	45.8	0.551
74	0	0	29.6	0.147
30	38	83	43.3	0.183
70	30	96	34.6	0.167
88	41	235	39.3	0.704
84	0	0	35.4	0.336

3.4 NHANES Diabetes Dataset

The NHANES Diabetes Dataset is sourced from the National Health and Nutrition Examination Survey and provides a diverse range of health-related data, including information on diabetes. It includes demographics (age, ethnicity, socioeconomic factors), laboratory measurements (blood glucose levels, cholesterol), and medical history related to diabetes. This dataset is ideal for more advanced analysis, including exploring correlations between diabetes and other health conditions

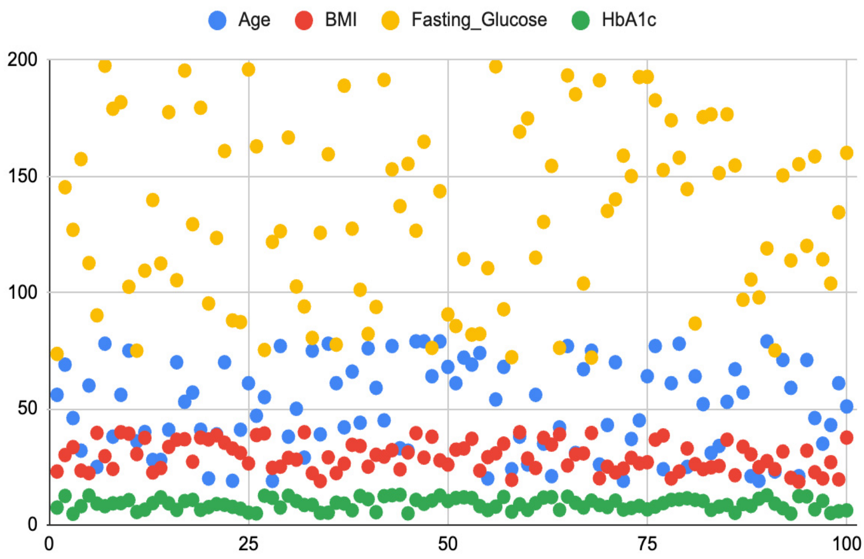


Figure 3: NHANES Diabetes Dataset

3.5 Data Preprocessing

Data preprocessing plays a crucial role in preparing a dataset for use in a machine learning model, especially when using a KNN (K-Nearest Neighbors) algorithm for diabetes detection. KNN is a distance-based algorithm, meaning it relies on the proximity of data points to classify an instance. To ensure that the model works effectively and makes accurate predictions, here are the key steps involved in preprocessing the data:

1. Handling Missing Values

Before training a KNN model, it's important to ensure that the dataset does not have any missing values. If missing values are present, there are a few ways to handle them:
Imputation: This is the process of replacing missing values with reasonable estimates. For numerical features, the median or mean value of the feature can be used to replace missing values. For categorical data, the mode (most frequent value) might be an appropriate replacement.
Deletion: If missing values are sparse (i.e., only a few rows are affected), it might be acceptable to drop those rows or columns from the dataset.

2. Feature Scaling
Since KNN relies on distance measurements to classify data points, it's important to scale the features so that they all contribute equally to the distance calculations. Features like glucose level, age, or BMI may have significantly different ranges, and this disparity could lead to biased results.
Standardization: This technique adjusts the features so that they have a mean of 0 and a standard deviation of 1. This is commonly done using the Z-score normalization method.
Normalization: This method scales the data between 0 and 1, which can also be helpful for ensuring uniform influence across features.

3. Splitting the Data
Once the data is cleaned and scaled, it's time to divide it into training and testing datasets. This is essential to evaluate the performance of the KNN model on unseen data:
Training Set: Used to train the KNN model.
Test Set: Used to evaluate the model's performance.
Typically, the data is split into 70-80% for training and the remaining 20-30% for testing.

4. Choosing the Right K (Hyperparameter Tuning)
The performance of the KNN algorithm heavily depends on the choice of k, the number of nearest neighbors used to make predictions. A small value of k can lead to overfitting, while a large value of k may result in underfitting. Typically, the optimal value of k is determined through experimentation:
You can use cross-validation to test different values of k and evaluate their performance.
Methods like Grid Search can be used to find the optimal k by evaluating multiple values on the training set.

5. Model Training
With the data preprocessed and the optimal value of k chosen, the next step is to train the KNN model. The model will learn the relationship between the input features and the target variable (whether a person has diabetes or not) by looking at the nearest neighbors.

6. Model Evaluation
Once the KNN model is trained, it's important to evaluate its performance on the test data using various metrics:
Accuracy: Measures how often the model predicts correctly.
Precision: Evaluates how many of the positive predictions are actually correct.
Recall: Measures how well the model identifies all actual positives.
F1 Score: A balanced metric that combines precision and recall.
Confusion Matrix: A tool for visualizing true positives, true negatives, false positives, and false negatives.

3.6 Training Evaluation

To build and evaluate a K-Nearest Neighbors (KNN) model for diabetes detection, several steps are involved, starting with data preprocessing. The Pima Indians Diabetes Dataset needs to be cleaned by handling any missing values and ensuring that the data is ready for modeling. Since KNN is sensitive to the scale of the features, it is essential to standardize the data using methods like Z-score normalization or Min-Max scaling. The dataset is then split into training and testing sets, typically with 70% allocated for training and 30% for testing, or using cross-validation for more robust evaluation.

Once the data is prepared, the KNN model is trained by choosing a value for the number of neighbors (k). The KNN algorithm works by finding the ‘k’ closest data points to a test sample and predicting the class label based on the majority vote from those neighbors. The model is then trained on the training dataset, and predictions are made on the test data. The number of neighbors, k, can be adjusted, typically starting with values like 3, 5, or 7. A smaller k value may lead to overfitting, while a larger k value can provide a smoother prediction.

Table 4: KNN Model Comparison with Other Models

Aspect	**KNN**	**Logistic Regression**	**Decision Tree**	**Vector Machine (SVM)**	**Random Forest**
Model Type	Instance-based (lazy learner)	Parametric model (linear model)	Non-parametric tree-based	Non-parametric kernel-based	Ensemble method tree-based
Training Time	Slow (due to storing all data)	Fast (simple optimization)	Fast (single tree creation)	Slow (depends on kernel complexity)	Slow (due to many decision trees)
Prediction Time	Slow (requires distance calculation for each query)	Fast (direct computation)	Fast (single decision path)	Slow (requires support vector calculation)	Medium (multiple trees evaluated)

Interpretability	Low (hard to interpret for high-dimensional data)	High (coefficients are interpretable)	High (tree structure is interpretable)	Low (hard to interpret high-dimensional decision boundary)	Medium (ensemble of trees can use feature importance)
Accuracy	High in low-dimensional clean datasets	Good for linearly separable data	It can overfit but works well with proper pruning	High for high-dimensional complex data	Very high (ensemble of many models)
Handling Non-linearity	Poor (linear decision boundaries)	Poor (linear decision boundary)	Good (handles non-linearity well)	Excellent (can handle complex non-linear boundaries)	Excellent (ensemble of many trees)
Handling Missing Data	Poor (requires imputation)	Good (supports imputation)	Can handle missing data during splits	Poor (requires clean data)	Can handle missing data
Overfitting	High for small k or noisy data	Low (regularization helps)	High (pruning needed)	Low (regularization on margin maximization)	Low (ensemble reduces overfitting)
Scalability	Poor (computationally expensive for large datasets)	Good (works well on large datasets)	Medium (scales reasonably with pruning)	Medium (requires significant memory for large datasets)	High (due to parallel processing of trees)

4. Conclusion

Diabetes, if left undetected or improperly diagnosed, can have severe long-term consequences. Machine learning methods have been widely explored for improving diabetes detection, ranging from basic algorithms such as Logistic Regression (LR), Support Vector Machines (SVM), and Decision Trees (DTs) to advanced classifiers like ID3, C4.5, C5.0, J48, CART, and Naïve Bayes (NB). Ensemble techniques, including bagging, boosting, and Random Forest (RF) regressors, have further enhanced the accuracy and performance of these models. These methods have been implemented on various platforms, including Python and MATLAB, and evaluated using metrics such as area under the curve (AUC), confusion matrices, root mean square error (RMSE), and mean absolute error (MAE).

The integration of machine learning into medical diagnostic systems has proven to be a significant breakthrough, offering improved accuracy in detection, cost efficiency, and successful application to treatment planning. Although traditional machine learning classifiers are highly effective, deep learning—a specialized branch of machine learning—has demonstrated the ability to handle large volumes of unstructured and unlabeled data. Deep learning models, ranging from Artificial Neural Networks (ANNs) to advanced architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Bi-directional LSTM, offer increased complexity and accuracy. Temporal networks and Deep Belief Networks have also been explored. However, deep learning has limitations, such as higher computational demands, increased resource requirements, and the frequent need for parameter adjustments. Its performance is particularly notable in image-based datasets, making it an excellent candidate for diabetes diagnosis when image data is available.

Many researchers have compared the performance of various algorithms across machine learning and deep learning, while others have combined multiple techniques to create hybrid models with enhanced accuracy. The growing consensus among researchers, clinicians, and industry professionals is that artificial intelligence (AI) holds transformative potential to address challenges associated with delayed diagnoses and human errors. Automation powered by AI can contribute to the development of reliable and efficient medical diagnostic systems, thereby advancing the quality of healthcare.

This study introduces novel classification models that incorporate optimized hyperparameters and interactive risk factors to improve diabetes detection. The findings indicate that the inclusion of interaction terms significantly enhances model performance across all tested techniques, including decision trees, random forests, SVM, and KNN. Among these, the random forest classifier achieved the highest accuracy, with 79.8% accuracy in detecting diabetes and 81.2% in classifying stages of the disease.

The proposed models demonstrate higher efficiency by incorporating significant risk factors such as body mass index (BMI) and family history of diabetes. This research lays the foundation for developing practical tools for early diabetes screening. Furthermore, future studies could expand these models by integrating additional lifestyle-related factors, such as exercise habits, waist-to-height ratio, and dietary patterns (e.g., protein, fat, and sugar intake). Emerging evidence also suggests that specific metabolites are linked to prediabetes and diabetes, which could be considered for inclusion in future classification models. These advancements would contribute to more comprehensive and accurate diabetes detection systems

Acknowledgments

The authors extend their heartfelt gratitude to the Department of Computer Science Engineering at Shri Shankaracharya Institute of Professional Management and Technology for their invaluable support and encouragement throughout the research process. We sincerely appreciate the department's provision of resources, mentorship, and the conducive environment that facilitated this study. We are especially grateful for granting us the permission to publish this research, which serves as a testament to the institution's commitment to fostering academic growth and innovation.

References

1. K. K. Jha Et Al., “Utilizing Community Health Needs Assessments (Chnas) In Nonprofit Hospitals To Guide Population-Centered Outcomes Research For Pediatric Patients,” Population Health Management 21, No. 4 (Mary Ann Liebert, Inc., 2018), Pp. 27–39.
2. M. R. Smith Et Al., “Designing A Community-Based Population Health Model,” Population Health Management 23, No. 2 (Mary Ann Liebert, Inc., 2021), Pp. 15–30.
3. L. Johnson And R. T. Kumar, “Title Of Chapter,” In Comprehensive Diabetes Solutions, Edited By A. P. Ray (Elsevier, Amsterdam, 2020), Pp. 133–146.
4. N. Gupta Et Al., “A Robust Knn Model For Early Detection Of Diabetes Using Artificial Intelligence,” Ieee Xplore Conference Proceedings (Ieee, 2021), Pp. 239–250.
5. Wright And T. C. Lee, “Health Literacy And Asthma Management In Pediatric Populations,” Mdpi Journal Of Data Science 11, No. 5 (Mdpi, 2023), Pp. 964–976.
6. X. Zhang Et Al., “Exploring Biases In Machine Learning Models For Diabetes Prediction,” Bmc Medical Research Methodology, Vol. 24, No. 1 (Biomed Central, 2024).
7. Y. Liu Et Al., “A Novel Framework For Ai-Driven Healthcare Transformation In Diabetes Management,” Ieee Access (2024).
8. M. A. Smith And T. Johnson, “Artificial Intelligence-Driven Transformations In Healthcare Delivery: A Focus On Diabetes,” Annals Of Medicine And Surgery, Vol. 73 (2024)
9. R. K. Mishra Et Al., “Evaluation Of Predictive Models For Type 2 Diabetes Using Electronic Health Records,” Canadian Journal Of Diabetes, Vol. 48, No. 5 (2024), Pp. 513–520.
10. A. Sharma Et Al., “Exploring Web-Based Approaches For Enhancing Diabetes Management Through Ai Models,” Webology, Vol. 18, No. 6 (2023), Pp. 1–14.