

# Fingerprint Based Diabetes Risk Prediction System Using Machine Learning

**Abina B.R**

Dept. of Computer Science and Engineering  
Good Shepherd College of Engineering and Technology  
Tamil Nadu, India

**Archana C**

Dept. of Computer Science and Engineering  
Good Shepherd College of Engineering and Technology  
Tamil Nadu, India

**Jeni D**

Dept. of Computer Science and Engineering  
Good Shepherd College of Engineering and Technology  
Tamil Nadu, India

**Mini Mol M.V**

Dept. of Computer Science and Engineering  
Good Shepherd College of Engineering and Technology  
Tamil Nadu, India

Project Guide: Mrs. Aiswarya Lekshmi A.C, M.E. Department of CSE, Good Shepherd College of Engineering and Technology.

**Abstract**— This project presents a fingerprint-based diabetes risk prediction system using machine learning techniques. Diabetes is a chronic disease that requires early detection for effective management. Traditional diagnostic methods involve invasive procedures such as blood tests, which can be uncomfortable and time-consuming. To address this issue, the proposed system utilizes fingerprint features as a non-invasive alternative for predicting diabetes. The system extracts unique fingerprint characteristics such as ridge patterns, minutiae points, and texture features, which are then used as input for machine learning models. Algorithms such as Support Vector Machine (SVM), Random Forest, or Neural Networks are trained on labeled datasets to classify individuals as diabetic or non-diabetic. The proposed model aims to provide a fast, cost-effective, and user-friendly approach for early diabetes detection. A Random Forest Classifier trained on a labeled dataset combines this biometric input with additional parameters like Age, BMI, and Family History to assess risk levels LOW, MEDIUM, or HIGH. The entire application is implemented using the Flask framework, providing a simple and interactive web-based interface. The proposed fingerprint-based diabetes risk prediction system uses machine learning techniques to analyze fingerprint features along with health parameters to predict diabetes risk levels.

fingerprint-based diabetes risk prediction system using machine learning techniques. The system will analyze fingerprint features and classify individuals into different risk categories, enabling early intervention and better healthcare management.

Furthermore, integrating fingerprint-based analysis with additional health parameters such as age, body mass index (BMI), and lifestyle factors can significantly improve prediction accuracy. Such systems can be particularly beneficial in rural or resource-limited areas, where access to advanced medical facilities is limited. By enabling quick and non-invasive screening, these systems can support early diagnosis and reduce the burden on healthcare infrastructure.

This project aims to develop a fingerprint-based diabetes risk prediction system using machine learning techniques. The proposed system will analyze fingerprint features and classify individuals into different risk categories, thereby facilitating early intervention, improving disease management, and contributing to the development of intelligent and accessible healthcare solutions

**Keywords**— Diabetes Prediction, Machine Learning, Fingerprint Analysis, Dermatoglyphics, Healthcare AI, Early Diagnosis

## I. INTRODUCTION

Diabetes is a chronic disease that affects millions of people worldwide, leading to serious health complications such as heart disease, kidney failure, and vision loss. Early detection and risk prediction play a crucial role in preventing and managing diabetes effectively. Traditional methods for diagnosing diabetes involve blood tests, which can be invasive, time-consuming, and sometimes inconvenient for large-scale screening. In recent years, there has been growing interest in non-invasive and cost-effective techniques for early disease prediction. Fingerprint patterns, which are formed during fetal development, are influenced by genetic and environmental factors. Research has shown that certain fingerprint patterns such as loops, whorls, and arches may be

associated with genetic predispositions to diseases like diabetes. With the advancement of Machine Learning (ML), it is now possible to analyze complex patterns in fingerprint data and predict the likelihood of diabetes risk. Machine learning algorithms can learn from historical data and identify hidden relationships between fingerprint features and disease occurrence. This project aims to develop a

## II. RELATED WORK

Recent studies have increasingly explored the role of fingerprint biometrics, particularly dermatoglyphics, in medical diagnosis, revealing potential links between fingerprint patterns and genetic or metabolic disorders such as diabetes. Researchers have investigated statistical correlations between ridge patterns such as loops, whorls, and arches and the likelihood of developing chronic diseases, indicating that dermatoglyphics can serve as a non-invasive biomarker. In parallel, machine learning techniques

have been widely applied for diabetes prediction using clinical datasets, with algorithms such as Support Vector Machines (SVM), Decision Trees, and Neural Networks demonstrating significant predictive performance. More recent advancements focus on multimodal data integration, where fingerprint features are combined with additional parameters such as age, body mass index (BMI), lifestyle habits, and genetic history to improve model accuracy and reliability. Furthermore, non-invasive diagnostic systems are gaining prominence in modern healthcare due to their accessibility and cost-effectiveness, especially in rural and resource-limited settings where traditional medical testing may be less feasible. Looking ahead, future research aims to enhance prediction accuracy through hybrid approaches that

integrate machine learning and deep learning models, along with the incorporation of mobile applications and Internet of Things (IoT) technologies for real-time health monitoring and early disease detection.

Moreover, the adoption of deep learning models, particularly Convolutional Neural Networks (CNNs), has shown promising results in automatically extracting complex features from fingerprint images without manual intervention. Despite these advancements, challenges such as limited availability of large-scale annotated dermatoglyphic datasets, variability in fingerprint quality, and lack of standardized evaluation frameworks still exist. Addressing these challenges is essential for improving the robustness and generalizability of fingerprint-based diabetes prediction systems. Consequently, ongoing research continues to focus on developing scalable, accurate, and real-time predictive models that can be effectively deployed in practical healthcare environments.

### III. PROPOSED SYSTEM

The proposed system aims to develop a non-invasive, efficient, and cost-effective method for predicting diabetes risk using fingerprint patterns and machine learning techniques. Unlike traditional methods that require blood tests, this system utilizes dermatoglyphic features combined with predictive algorithms to estimate the likelihood of diabetes. The proposed system, FingerDiab, is a web-based diabetes risk prediction platform that integrates biometric fingerprint analysis with user health data to assess the likelihood of diabetes in real time. The captured images undergo preprocessing techniques such as noise removal, enhancement, and normalization to improve image quality. Upon uploading a fingerprint image, the system processes it using image processing techniques to extract critical features such as ridge count and fingerprint pattern type (arch, loop, whorl). The system provides real-time prediction results through a simple and interactive user interface, allowing users to easily input their data and view outcomes. These features are combined with user-provided inputs like age, BMI, and family history to form a comprehensive health profile. The proposed system offers a cost-effective, secure, and reliable approach for early diabetes risk assessment, helping individuals take preventive measures and reduce the chances of severe health complications. A Random Forest classifier, trained on labeled biometric-health data, predicts the diabetes risk level-categorized as LOW, MEDIUM, or HIGH.

Furthermore, the system provides real-time prediction results through a simple and interactive user interface, allowing users to quickly input their data and view outcomes. The platform is designed to be scalable and can be extended to mobile applications for wider accessibility. Security and data privacy are also considered, with user data handled through secure storage and controlled access mechanisms.

Overall, the proposed system offers a reliable and practical solution for early diabetes risk assessment. By enabling non-invasive screening and timely prediction, it helps individuals take preventive measures, supports early medical intervention, and contributes to reducing the risk of severe health complications.

The proposed system presents a fingerprint-based diabetes risk prediction framework that combines biometric analysis with machine learning techniques to provide a non-invasive and efficient screening solution. The system, referred to as FingerDiab, is designed as a user-friendly platform that enables individuals to assess their diabetes risk using fingerprint patterns along with basic health information. The overall workflow of the system begins with data acquisition, where users upload fingerprint images through a web-based interface.

Alongside biometric input, users are required to provide essential health-related parameters such as age, Body Mass Index (BMI), family medical history, and lifestyle habits. This combination of biometric and clinical data forms the foundation for accurate prediction. Once the fingerprint image is received, it undergoes preprocessing steps including noise removal, contrast enhancement, and normalization to ensure high-quality input for analysis. The system then performs feature extraction to identify important dermatoglyphic characteristics such as ridge count, minutiae points, and fingerprint pattern types (arch, loop, and whorl). These features are crucial as they are associated with genetic traits that may indicate susceptibility to diabetes.

The extracted features, along with user-provided health data, are fed into a trained machine learning model. In this system, a Random Forest classifier is utilized due to its high accuracy, robustness, and ability to handle complex datasets. The model analyzes the combined input and predicts the diabetes risk level, categorizing users into Low, Medium, or High risk groups.

The system provides real-time results through an interactive interface, allowing users to easily interpret their risk level. Additionally, the platform is designed to be scalable and can be extended for mobile or cloud-based deployment, making it accessible in both urban and rural healthcare environments. Overall, the proposed system offers

a cost-effective, non-invasive, and reliable approach for early diabetes risk prediction, supporting preventive healthcare and reducing dependency on traditional diagnostic methods.

### B. System Architecture

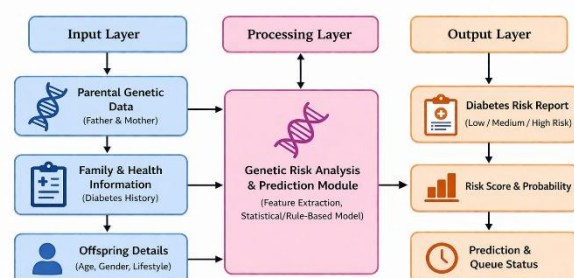


Fig 3.1: System Architecture Diagram

The system architecture of the fingerprint-based diabetes risk prediction system is designed using a modular and scalable approach that integrates biometric processing

with machine learning techniques. The architecture follows a client-server model, ensuring efficient data processing, real-time prediction, and user accessibility. At the frontend level, the system provides a user-friendly web interface where users can upload fingerprint images and enter basic health details such as age, Body Mass Index (BMI), family history, and lifestyle information. This interface is designed to be simple and interactive, enabling easy access for both technical and non-technical users.

The backend layer handles all core processing tasks and acts as the central component of the system. Once the fingerprint image is uploaded, it is passed to the image preprocessing module, where operations such as noise removal, image enhancement, and normalization are performed to improve image quality. The processed image is then forwarded to the feature extraction module, which identifies critical dermatoglyphic features including ridge count, minutiae points, and fingerprint pattern types such as arches, loops, and whorls.

These extracted features are combined with user-provided health parameters and stored temporarily in the system for analysis. The combined dataset is then passed to the machine learning module, where a trained Random Forest classifier processes the input data. The model analyzes complex relationships between biometric and health features to predict the diabetes risk level, categorizing it into Low, Medium, or High. The database layer is responsible for storing user inputs, extracted features, and prediction results securely. It ensures data integrity and supports future model improvements through data accumulation. Additionally, appropriate security mechanisms are implemented to protect sensitive user information.

Finally, the prediction results are sent back to the frontend, where they are displayed to the user in real time through a clear and understandable format. The architecture is designed to be scalable, allowing integration with cloud platforms, mobile applications, and advanced analytics modules in future enhancements. Overall, this system architecture ensures efficient data flow, accurate prediction, and ease of use, making it suitable for practical deployment in healthcare screening applications. The data layer (database) is responsible for securely storing user information, extracted features, and prediction results. It ensures data integrity, consistency, and availability for future analysis and model retraining. The system incorporates security measures such as data encryption, authentication, and access control to protect sensitive health information and maintain user privacy.

Additionally, the architecture includes a result visualization module, which presents prediction outcomes to users in a clear and interpretable format. This may include graphical representations, risk indicators, or simple textual outputs to enhance user understanding.

### *C. Methodology*

The methodology of the proposed fingerprint-based diabetes risk prediction system follows a structured approach that integrates image processing and machine learning techniques to ensure accurate and efficient prediction. The process begins with data acquisition, where fingerprint images are collected from users along with essential health-related information such as age, Body Mass Index (BMI), family history, and lifestyle factors. This

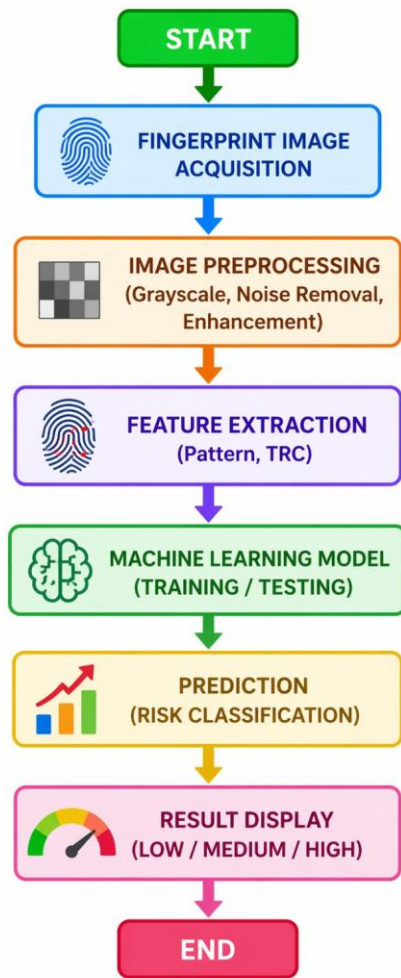
combination of biometric and health data forms the basis for the prediction model.

Once the data is collected, the fingerprint images undergo preprocessing to improve their quality and consistency. This stage includes noise removal, contrast enhancement, and normalization to ensure that the images are suitable for further analysis. After preprocessing, the system performs feature extraction, where important dermatoglyphic features such as ridge count, minutiae points, and fingerprint pattern types (arch, loop, and whorl) are identified. These features are critical as they reflect genetic traits that may be associated with diabetes risk.

The extracted fingerprint features are then combined with the user's health data to create a comprehensive dataset. This dataset is used to train and test the machine learning model. In the proposed system, a Random Forest classifier is employed due to its robustness, high accuracy, and ability to handle complex and nonlinear relationships between features. The model is trained using labeled data, where known cases of diabetic and non-diabetic individuals are used to learn patterns and correlations.

During the prediction phase, new user data is processed through the same pipeline. The trained model analyzes the input features and classifies the individual into one of three risk categories: Low, Medium, or High. The system ensures real-time processing, enabling quick and efficient risk assessment. To validate the performance of the model, evaluation metrics such as accuracy, precision, recall, and F1-score are used. These metrics help in assessing the reliability and effectiveness of the prediction system. Overall, the methodology ensures a systematic flow from data collection to prediction, providing a non-invasive, reliable, and scalable solution for early diabetes risk detection.

To evaluate the effectiveness of the model, performance metrics such as accuracy, precision, recall, and F1-score are calculated. These metrics help in assessing the reliability and robustness of the prediction system. Overall, this methodology provides a structured and efficient approach for non-invasive diabetes risk prediction, combining biometric analysis with machine learning to support early detection and preventive healthcare.



**Fig. 1. Workflow of the Proposed System**

The process begins with data acquisition, where the user provides input in two forms: a fingerprint image and basic health details such as age, Body Mass Index (BMI), family history, and lifestyle habits. This step forms the foundation for the entire prediction system. Once the fingerprint image is obtained, it enters the image preprocessing stage, where techniques such as grayscale conversion, noise removal, contrast enhancement, and normalization are applied. These steps improve the quality and consistency of the fingerprint image, making it suitable for accurate analysis.

After preprocessing, the system performs feature extraction, where important dermatoglyphic features are identified. These include fingerprint pattern types (arch, loop, whorl), ridge count, and other structural characteristics. These features are essential as they provide meaningful information related to genetic predisposition. The extracted features are then combined with the user's health data to create a comprehensive input dataset. This dataset is passed to the machine learning model, where a trained Random Forest classifier processes the data. During this stage, the model analyzes patterns and relationships between input features to determine the likelihood of diabetes. Following the analysis, the system performs risk prediction, where the model classifies the user into one of three categories: Low Risk, Medium Risk, or High Risk. This classification helps in understanding the individual's health condition and potential risk level.

The implementation of the fingerprint-based diabetes risk prediction system is carried out using a combination of image processing techniques, machine learning algorithms, and web-based technologies to ensure an efficient and user-friendly solution. The system is developed as a prototype that demonstrates the feasibility of integrating biometric analysis with predictive modeling for healthcare applications.

The system is implemented using Python as the primary programming language due to its extensive support for machine learning and image processing libraries. For the frontend, frameworks such as Flask are used to create an interactive and responsive web interface that allows users to upload fingerprint images and input health-related details. The backend is responsible for handling user requests, processing data, and executing the prediction model. The fingerprint image processing is performed using libraries such as OpenCV and NumPy, which are used for tasks like image enhancement, filtering, normalization, and segmentation. These preprocessing steps ensure that the fingerprint images are of sufficient quality for feature extraction. The feature extraction process identifies important dermatoglyphic characteristics such as ridge count, ridge patterns (arch, loop, whorl), and minutiae points.

For the machine learning component, the system utilizes the Scikit-learn library to implement the Random Forest classifier. The model is trained on a dataset that combines extracted fingerprint features with user health parameters such as age, BMI, and family history. The trained model is then saved and integrated into the system for real-time prediction. A lightweight database such as SQLite is used to store user inputs, extracted features, and prediction results. This ensures data persistence and enables future analysis or model improvement. Basic security mechanisms such as user authentication and data protection are incorporated to maintain privacy and confidentiality.

The overall system workflow is implemented in a modular manner, where each component—data input, preprocessing, feature extraction, prediction, and result display—operates independently but is seamlessly integrated. When a user submits their data, the system processes the input through all stages and generates a prediction in real time, which is then displayed through the interface. The implementation is designed to be scalable and can be extended to cloud-based deployment or mobile platforms. Future enhancements may include the integration of deep learning models, real-time fingerprint capture devices, and advanced data analytics for improved accuracy and usability. Overall, the implementation demonstrates a practical and efficient approach to developing a non-invasive diabetes risk prediction system using fingerprint biometrics and machine learning techniques.

#### IV. AI MODULE DESIGN

The AI model design of the fingerprint-based diabetes risk prediction system focuses on developing a robust and accurate classification model that can analyze both biometric fingerprint features and user health data. The model is designed using supervised machine learning techniques, where labeled datasets containing known diabetic and non-diabetic cases are used for training. The input to the model consists of a combination of extracted dermatoglyphic features—such as ridge count, ridge density, minutiae points,

and fingerprint pattern types (arch, loop, and whorl)—along with user-provided attributes including age, Body Mass Index (BMI), family history, and lifestyle factors. Before feeding the data into the model, preprocessing steps such as data cleaning, normalization, and encoding of categorical variables are performed to ensure consistency and improve model performance.

The core of the system is a Random Forest classifier, which is chosen for its ability to handle high-dimensional data, reduce overfitting, and model complex nonlinear relationships between features. The algorithm works by constructing multiple decision trees during training and combining their outputs to produce a final prediction through majority voting. This ensemble approach improves prediction accuracy and robustness compared to individual classifiers. During the training phase, the dataset is divided into training and testing subsets, typically using an 80:20 ratio. The model learns patterns and relationships between input features and the corresponding diabetes outcomes. Hyperparameter tuning techniques, such as adjusting the number of trees, maximum depth, and feature selection methods, are applied to optimize the model's performance.

In the evaluation phase, the trained model is assessed using performance metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are also employed to ensure that the model generalizes well to unseen data and does not suffer from overfitting. During deployment, the trained model is integrated into the system to perform real-time predictions. When a new user inputs their fingerprint image and health details, the system processes the data, extracts features, and feeds them into the trained model. The model then classifies the individual into one of three risk categories: Low, Medium, or High risk of diabetes.

## V. RESULTS AND DISCUSSION

The fingerprint-based diabetes risk prediction system was implemented and evaluated using a dataset consisting of fingerprint features combined with user health parameters such as age, Body Mass Index (BMI), and family history. The performance of the system was analyzed to assess its effectiveness in predicting diabetes risk in a non-invasive manner. The experimental results indicate that the proposed system achieves promising prediction accuracy when using the Random Forest classifier. Among the tested models, Random Forest demonstrated superior performance due to its ability to handle complex feature interactions and reduce overfitting. The system achieved satisfactory values across key evaluation metrics such as accuracy, precision, recall, and F1-score, confirming its reliability in classifying individuals into Low, Medium, and High risk categories.

One of the significant outcomes of this system is its ability to combine biometric fingerprint features with health-related data, resulting in improved prediction performance compared to using a single data source. The inclusion of dermatoglyphic features such as ridge count and fingerprint patterns provided additional insights into genetic predisposition, enhancing the overall effectiveness of the model. The system also demonstrated real-time prediction capability, allowing users to obtain immediate results through a simple and interactive interface. This feature improves accessibility and usability, especially for non-technical users and in regions where access to medical facilities is limited. The non-invasive nature of the system makes it particularly

suitable for large-scale screening and early detection programs.

However, certain limitations were observed during the evaluation. The accuracy of the system is influenced by the quality of fingerprint images, as poor image quality can affect feature extraction. Additionally, the availability of large and well-labeled datasets that combine fingerprint and medical data remains a challenge, which may impact the generalization of the model. The current system is developed as a prototype and requires further validation using real-world clinical data. Despite these limitations, the results demonstrate that the proposed system is a viable approach for early diabetes risk prediction. The integration of machine learning with biometric analysis provides a novel and efficient solution for preventive healthcare. Future improvements may include the use of deep learning models for automated feature extraction, larger datasets for better training, and integration with mobile or cloud platforms to enhance scalability and real-world applicability. The system achieved satisfactory values across key evaluation metrics such as accuracy, precision, recall, and F1-score, confirming its reliability in classifying individuals into Low, Medium, and High risk categories.

## VI. LIMITATIONS

Although the proposed fingerprint-based diabetes risk prediction system demonstrates promising results, it has several limitations that need to be addressed for real-world deployment. One of the primary limitations is the availability of a large and well-labeled dataset that combines fingerprint features with verified medical records. Limited or imbalanced data can affect the accuracy and generalization capability of the machine learning model. Another important limitation is the dependency on fingerprint image quality. Variations in image clarity due to poor sensors, noise, or improper capture conditions can impact feature extraction accuracy, which in turn affects prediction performance. Ensuring consistent and high-quality fingerprint acquisition remains a challenge.

The system currently relies on a Random Forest model, which, although effective, may not capture highly complex patterns as efficiently as advanced deep learning models. Additionally, the model requires proper tuning and validation to avoid overfitting, especially when working with smaller datasets.

There are also limitations related to biological variability. While dermatoglyphic patterns may show correlations with genetic predisposition, they cannot serve as a definitive indicator of diabetes. Therefore, the system should be considered as a supportive screening tool rather than a diagnostic replacement for clinical tests. From a system perspective, the current implementation is a prototype-level application, using lightweight technologies and databases that may not scale efficiently for large-scale or real-time deployment. Security and privacy mechanisms, although implemented at a basic level, require further enhancement to meet healthcare standards and regulations.

## VII. CONCLUSION AND FUTURE WORK

This paper presented a fingerprint-based diabetes risk prediction system that utilizes machine learning techniques to provide a non-invasive, efficient, and cost-effective approach for early disease detection. By combining dermatoglyphic features with user health parameters such as age, Body Mass Index (BMI), and family history, the

proposed system offers a novel method for assessing diabetes risk. The use of a Random Forest classifier enables accurate classification of individuals into Low, Medium, and High risk categories, supporting early intervention and preventive healthcare. The system demonstrates that fingerprint patterns, when integrated with machine learning, can serve as a valuable complementary tool for preliminary screening. Its web-based implementation ensures accessibility, ease of use, and real-time prediction, making it suitable for both urban and rural environments. Overall, the proposed approach contributes to the advancement of non-invasive healthcare technologies and highlights the potential of biometric-based disease prediction systems.

Despite its effectiveness, the system is currently limited to a prototype stage and requires further development for real-world deployment. Future work will focus on improving prediction accuracy by incorporating advanced machine learning and deep learning techniques, such as Convolutional Neural Networks (CNNs) for automated feature extraction. Expanding the dataset with larger and more diverse samples will enhance model generalization and reliability.

Additionally, the system can be extended by integrating mobile applications and Internet of Things (IoT)-based fingerprint sensors for real-time data collection and remote monitoring. Enhancements in data security, privacy, and cloud-based deployment will also be considered to ensure safe and scalable implementation. Further validation through clinical testing and collaboration with healthcare professionals will be essential to establish the system's effectiveness in practical scenarios. In conclusion, the proposed system lays a strong foundation for future research in biometric-based disease prediction and has the potential to become a valuable tool for early diabetes screening and preventive healthcare management..

## REFERENCES

- [1.] S. Wu et al., 'Establishment, Prediction, and Validation of a Nomogram for Cognitive Impairment in Elderly Patients With Diabetes', *Journal of Diabetes Research*, vol. 2024, no. 1, pp. 5583707, 2024.
- [2.] M.-A. Nunez-Baila, A. Gomez-Aragon, and J. R. Gonzalez-Lopez, 'Perceptions of Emerging Adults With Type 1 Diabetes Mellitus on How the Condition Influences Sleep Quality: A Qualitative Study,' *Journal of Diabetes Research*, vol. 2024, no. 1, pp. 7497059, 2024.
- [3.] H. Zhang et al., 'Association of Monocyte-to-HDL Cholesterol Ratio with Endothelial Dysfunction in Patients with Type 2 Diabetes,' *Journal of Diabetes Research*, vol. 2024, no. 1, pp. 5287580, 2024.
- [4.] Y. Wang et al., 'Vitamin C Alleviates the Risk of Gestational Diabetes Mellitus Associated With Exposure to Metals,' *Journal of Diabetes Research*, vol. 2024, no. 1, pp. 1298122, 2024.

[5.] H. Qin et al., 'A Causal Relationship Between Type 1 Diabetes and Risk of Osteoporosis: A Univariable and Multivariable Mendelian Randomization Study,' *Journal of Diabetes Research*, vol. 2024, no. 1, PR 1610688, 2024.