CS4175 - SENIOR PROJECT 2

Tayaqn: An AI-Driven Diabetes Prediction System Using Federated Learning Architecture

Araa AlMarhbi, Arwa Abu Rub, Duaa Suroor

Supervisor : Prof. Passent M. ElKafrawy

Computer Science Department,
College of Engineering,
Effat University, Jeddah, Saudi Arabia

AGENDA

01	Introduction & Background
02	Literature Review
03	Competitor Analysis
04	Methodology
05	Business Model
06	System Analysis & Design
07	Experiment details
08	Future Directions
09	Roadmap
10	Conclusion

01 INTRODUCTION & BACKGROUND

INTRODUCTION & BACKGROUND

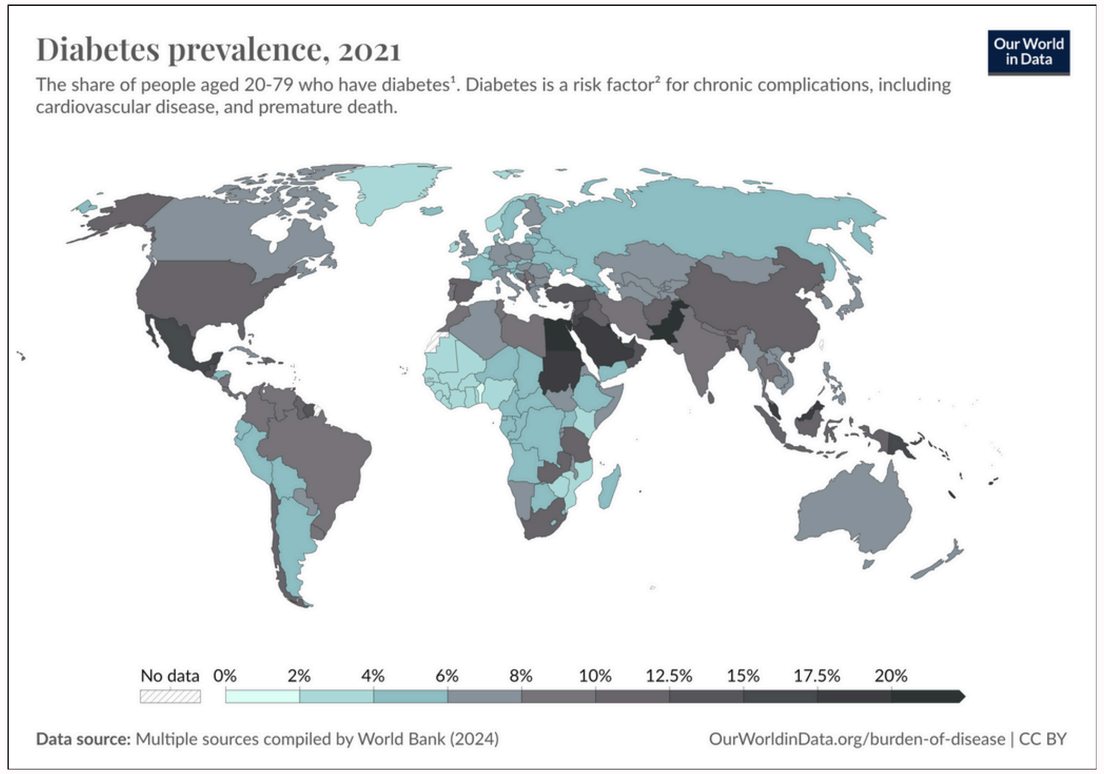
Motivation fo Tayaqn

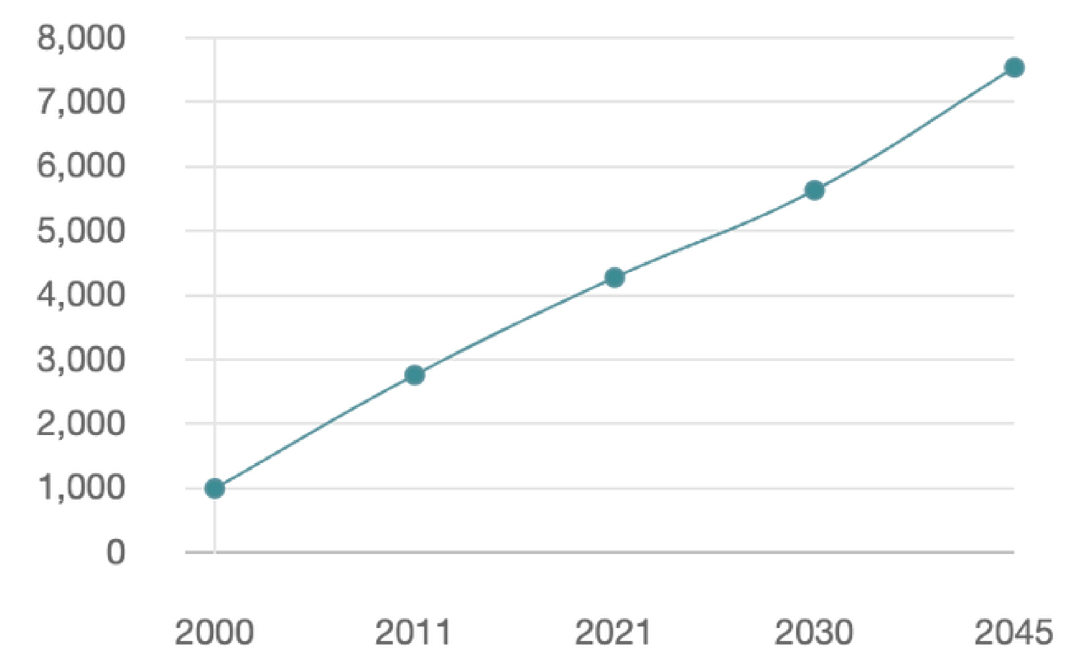
537M
Adults with Diabetes
1 in 10 people worldwide (2021)

783M
Projected by 2045
Dramatic increase expected

6.7M
Annual Deaths
One death every five seconds

7.5M
Saudi Arabia by 2045
Up from 997,000 in 2000





Saudi Arabia Diabetes report 2000 — 2045

INTRODUCTION & BACKGROUND

Problem statement, Research Questionm, Aim

PROBLEM STATEMENT

Diabetes requires early and accurate prediction, but traditional models risk patient privacy. A secure and scalable solution is needed to protect data while enabling effective prediction.

RESEARCH QUESTION

How can federated learning be effectively utilized to develop a scalable and privacy-preserving system for early diabetes prediction?

OUR AIM

Build a secure and accurate diabetes prediction system using modern technologies, ensuring high-precision risk prediction while protecting user data privacy.

INTRODUCTION & BACKGROUND

Objectives

Accurate Predictions

Design, train, and validate machine learning models across distributed datasets.

Data Privacy and Security

Implement federated learning to keep sensitive data local to individual hospitals.

Interactive Dashboards

Develop intuitive visualizations for real-time diabetes risk monitoring.

INTRODUCTION & BACKGROUND
Significance

VISION رؤية
2030
المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

3 GOOD HEALTH AND WELL-BEING

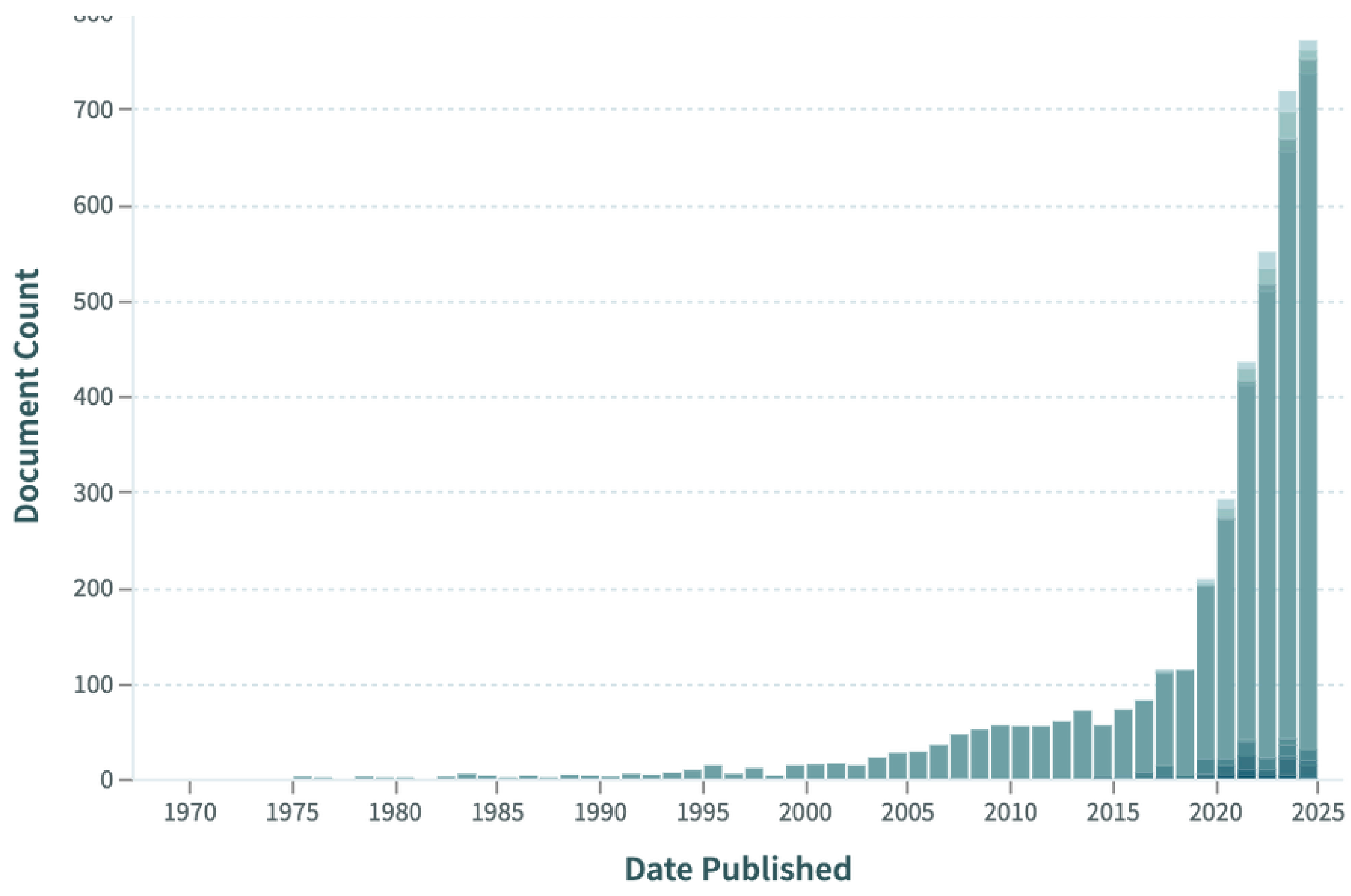
SUSTAINABLE DEVELOPMENT GOALS

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

02 LITERATURE REVIEW

LITERATURE REVIEW
Publications & Resources

Main Research Databases



This figure illustrates the exponential increase in scholarly publications on diabetes prediction using artificial intelligence from 1970 to 2025, with a notable surge after 2015. This trend reflects growing research interest and advancements in AI applications for healthcare.

LITERATURE REVIEW

Summary of Existing Work

AUTHOR	TITLE	DATASET	APPROACH	KEY FINDINGS	LIMITATIONS
A. Mujumdar and V. Vaidehi, 2019	Diabetes Prediction using Machine Learning Algorithms	Diabetes dataset (800 records, 10 attributes)	Logistic Regression, SVM, Random Forest, AdaBoost, Gradient Boost	High accuracy (98.8%) using AdaBoost with pipelining	Limited dataset size (800 records)
A. Nurdin et al., 2023	Using Machine Learning for the Prediction of Diabetes with Emphasis on Blood Content	LMCH Diabetes Dataset (Mendeley)	Random Forest, MLP, SVM	Random Forest achieved perfect accuracy, precision, recall, and F1-score	Dataset imbalance, potential information leakage
A. Z. Woldaregay et al., 2019	Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes	Various studies (2000–2018)	Feed-forward and Recurrent Neural Networks, SVM, Genetic Programming, Gaussian Processes	Systematic review of multiple models, hybrid approaches proposed	Challenges in generalizability and handling external factors
I. Tasin et al., 2022	Diabetes prediction using machine learning and explainable AI techniques	PIMA Indian and private dataset (203 samples)	Decision Tree, SVM, Random Forest, Logistic Regression, KNN, XGBoost	XGBoost achieved highest performance, explainable AI adds transparency	Moderate accuracy (81%), requires further optimization

LITERATURE REVIEW

Summary of Existing Work

AUTHOR	TITLE	DATASET	APPROACH	KEY FINDINGS	LIMITATIONS
H. Naz and S. Ahuja, 2020	Deep learning approach for diabetes prediction using PIMA Indian dataset	PIMA Indian Dataset	Decision Tree, Naive Bayes, ANN, Deep Learning	DL outperformed traditional ML methods	Focused on one dataset (PIMA), lacks generalizability
K. Abnoosian et al., 2023	Prediction of diabetes disease using an ensemble of machine learning multi-classifier models	Iraqi Patient Dataset for Diabetes	k-NN, SVM, Decision Trees, Random Forest, AdaBoost, Gaussian Naive Bayes	Robust handling of dataset imbalance, ensemble methods improve performance	Requires advanced preprocessing and computational resources
A. El-Bashbishy Sayyid et al., 2024	Pediatric diabetes prediction using deep learning	MUCHD Dataset (548 samples)	Deep Neural Network (DNN)	99.8% accuracy, effective for early diagnosis in children	Focused on a single dataset, lacks external validation
Z. Zhang et al., 2021	Machine Learning Prediction Models for Gestational Diabetes	Multiple studies (meta-analysis)	Ensemble-based methods, Logistic Regression	High AUROC for ensemble-based methods	Interstudy heterogeneity, data biases

LITERATURE REVIEW
Gaps

Dataset Challenges
Lack of standardization and underrepresentation of certain demographic groups.

Privacy Concerns
Limited research on privacy-preserving methods like differential privacy.

Real-World Application
There is a pressing need to enhance secure, AI-driven systems for effective and privacy-preserving diabetes detection.

Bias Issues
Data imbalance leads to biased predictions.

Scalability Limitations
Traditional models struggle with generalization and scalability.

03 COMPETITOR ANALYSIS

WHAT MAKES OUR SYSTEM DIFFERENT?

Privacy-Preserving Architecture

Uses federated learning to train models without sharing raw patient data.

Advanced AI Models

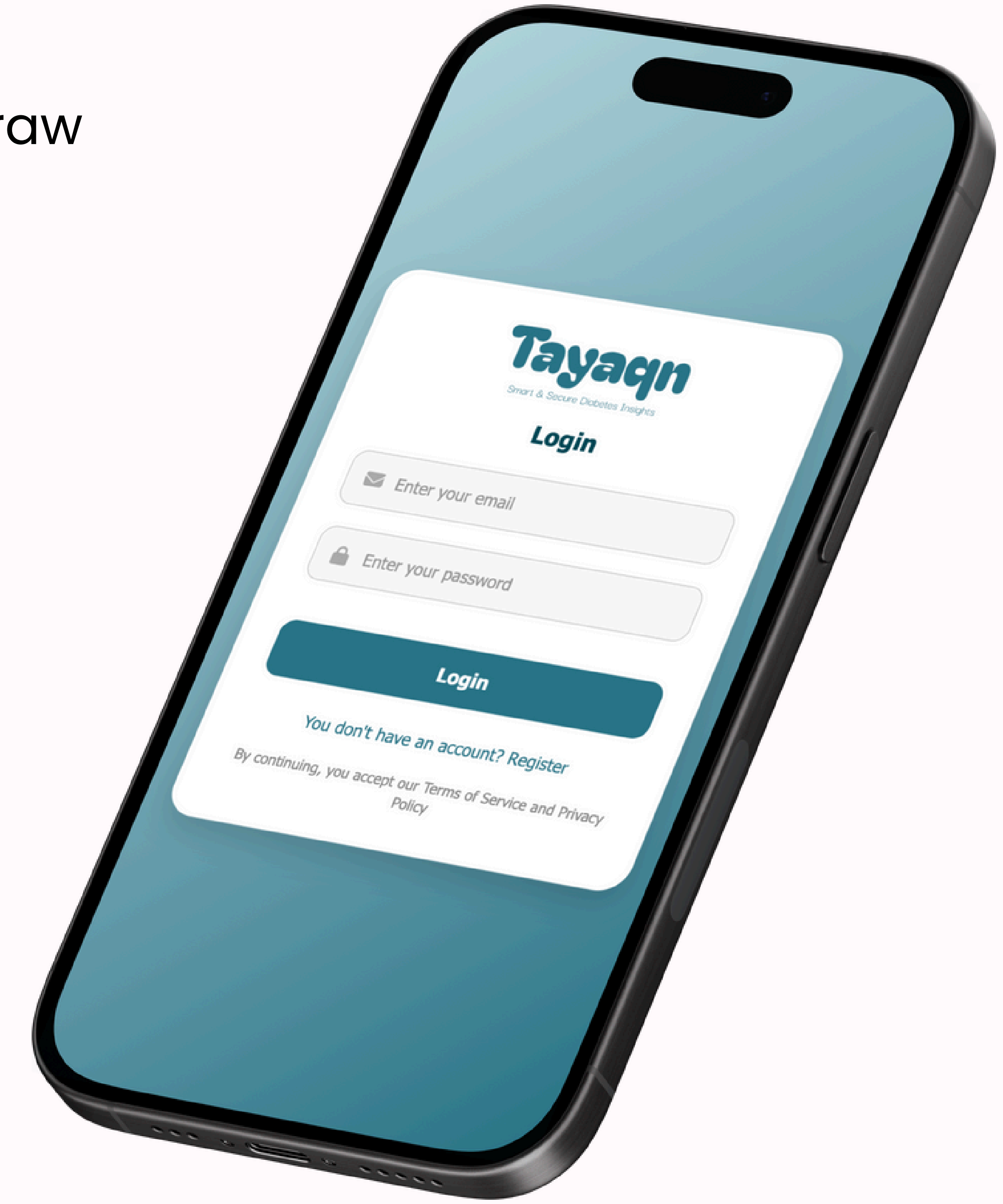
Employs ML/DL techniques for accurate, early-stage diabetes prediction.

Enhanced Security Measures

Incorporates encryption, hashing, and SQL injection protection.

Interactive Dashboards

User-friendly visualizations for monitoring risk levels and trends.



04 METHODOLOGY

METHODOLOGY

1 Data Collection

Gathering relevant diabetes datasets

2 Data Preprocessing

Cleaning and preparing data

3 Model Development

Model Training & Model Validation

4 Federated Learning

Local Training & Global Aggregation

5 System Design

Architectural Planning & API integration

6 Implemntation

Develop backend & frontend components

7 System Testing

Functional , Integration , and Performance

8 Deployment

Launching the system in production

05 BUSINESS MODEL

Business Model Canvas

Designed For: Senior Project

Designed By: Araa , Arwa , and Duaa

Date: 15/9/2024

Version: #1

Key Partners

- **Healthcare providers:** Hospitals, clinics, diagnostic centers for data validation.
- **Technology partners:** Cloud service providers for hosting, FL frameworks, and database services.
- **Research institutions:** Collaboration on Federated Learning advancements and medical data insights.
- **Regulatory bodies:** Compliance with GDPR, HIPAA, etc.
- **Community health organizations:** Partnerships for underserved regions.

Key Activities

- **Development:** Building and integrating frontend, backend, FL server, ML models.
- **Data Management:** Collecting, processing, and securing user health data.
- **Model Training:** Implementing Federated Learning workflows for model updates.
- **User Support:** Providing training, tutorials, technical assistance.
- **Marketing:** Promoting platform through campaigns and partnerships.

Key Resources

- **Team:** Developers, data scientists, healthcare experts, and regulatory advisors.
- **Technology:** Angular, FastAPI, PyTorch, SQLite, and cloud hosting.
- **Data:** Diabetes-related health indicator datasets.
- **Funding:** Capital for development, operations, and marketing.

Value Propositions

- **Privacy-preserving predictions via Federated Learning.**
- **Advanced ML models for reliable diabetes predictions.**
- **User-friendly dashboards for patients and clinicians.**
- **Scalable system for diverse user bases.**
- **Improved healthcare decision-making through actionable insights.**

VISION رؤيـــة 2030
المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

Customer Relationship

Direct Support: Dedicated customer service for technical issues.
User Engagement Tools: Tutorials, FAQs, help documentation.
Community Building: Online forums and user groups.
Feedback Mechanisms: Regular updates and improvements.

Channels

- **Digital Platforms:** Official website, mobile applications, and email campaigns.
- **Healthcare Institutions:** Partner clinics and hospitals to onboard clinicians.
- **Conferences and Seminars:** Health-tech events to demonstrate system capabilities.
- **Social Media Campaigns:** Targeted outreach to patients and healthcare professionals.
- **Publications:** Case studies and articles in healthcare and technology journals.

Customer Segments

- **Primary:** Patients seeking diabetes risk insights and early diagnosis tools.
- **Secondary:** Clinicians and healthcare professionals requiring decision-support tools.
- **Tertiary:** Public health organizations and researchers interested in privacy-preserving predictive models.

Cost Structure

- **Development Costs:** Salaries for developers and researchers, technology tools, and infrastructure setup.
- **Operational Costs:** Cloud hosting, maintenance, and customer support.
- **Marketing Costs:** Advertising, educational campaigns, and partnerships.
- **Compliance Costs:** Ensuring regulatory adherence to privacy and security standards.

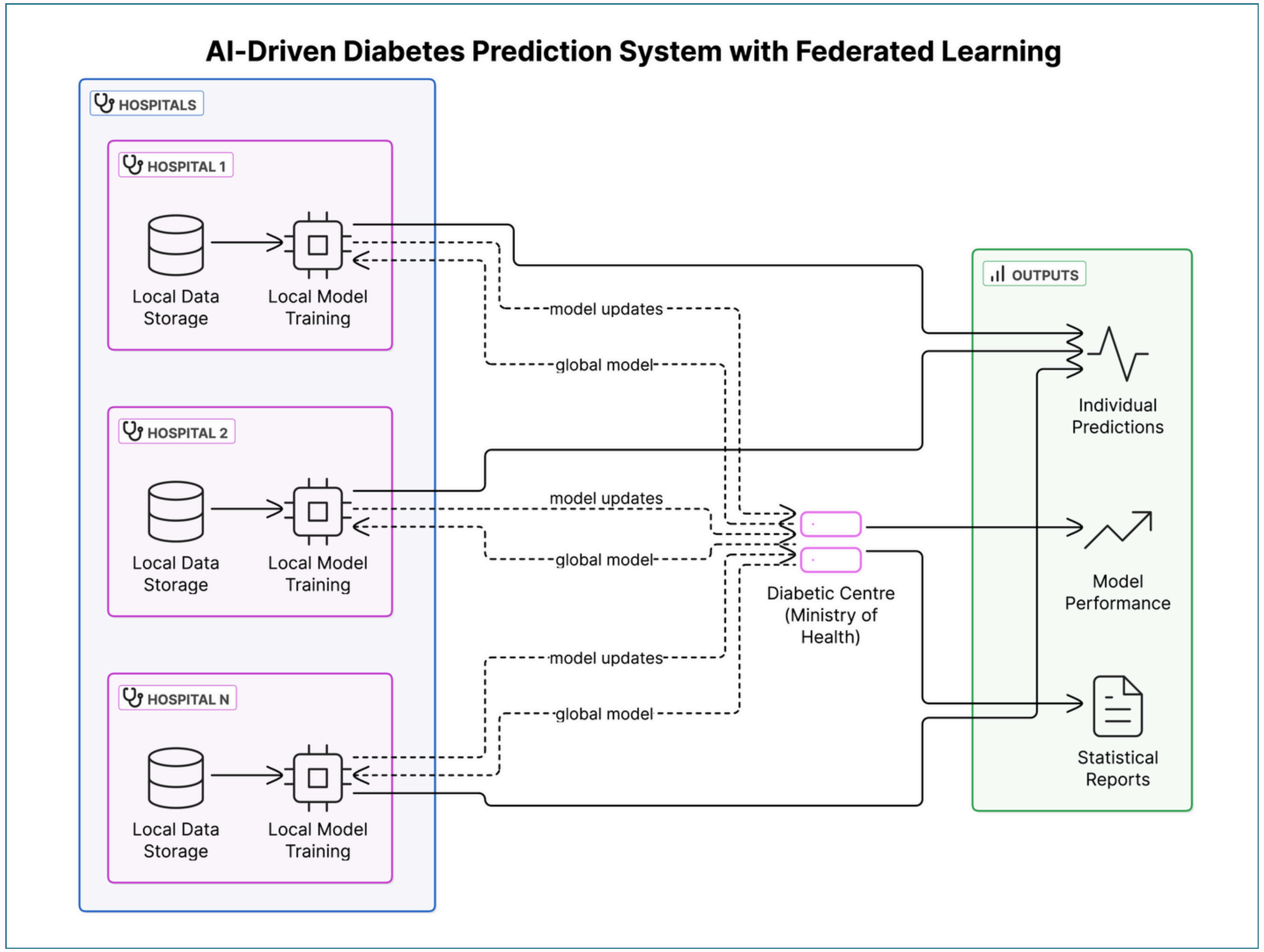
Revenue Stream

- **Subscription Model:** Monthly or annual fees for advanced analytics and insights.
- **Freemium Model:** Free basic predictions with premium features for in-depth analysis.
- **Institutional Partnerships:** Licensing the platform to healthcare providers and organizations.
- **Grants and Funding:** Research grants for privacy-preserving AI in healthcare.
- **Data Insights Services**: Aggregated, anonymized data insights (with user consent) for research purposes.

06 SYSTEM ANALYSIS & DESIGN

SYSTEM ANALYSIS & DESIGN

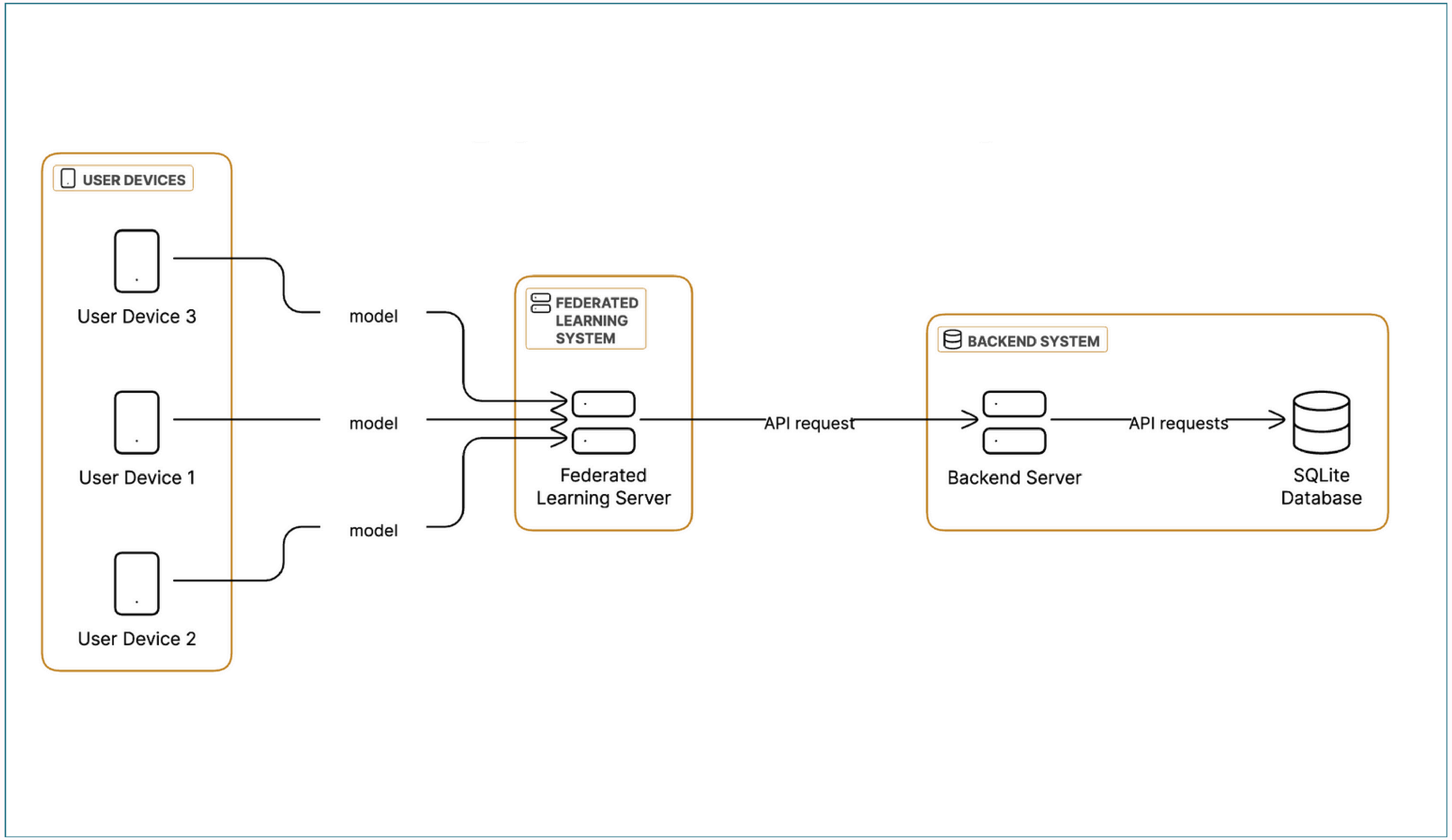
Architectural designs



Federated Learning Architecture Across Multiple Hospitals for Diabetes Prediction

SYSTEM ANALYSIS & DESIGN

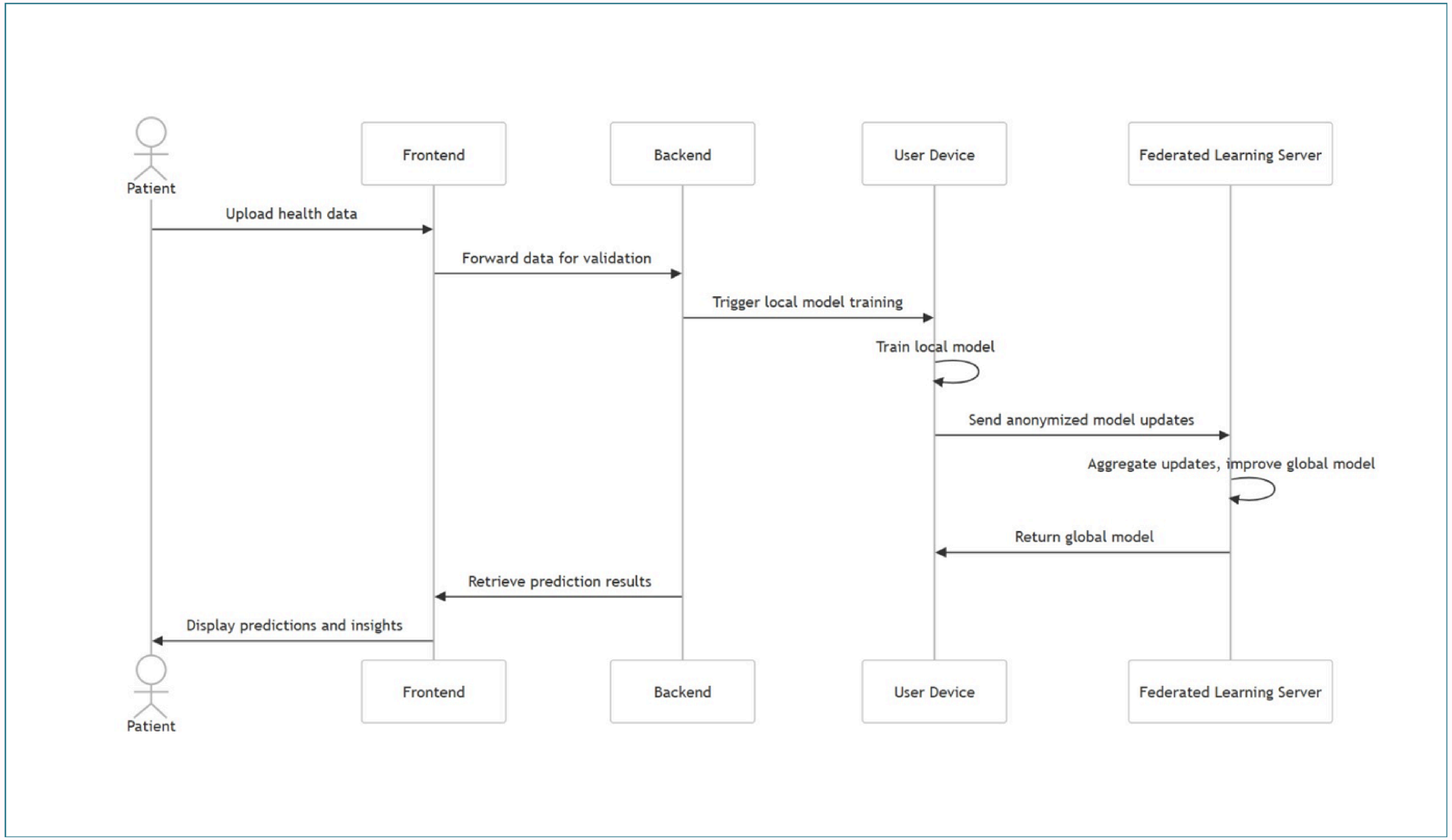
Architectural designs



System Architecture for Tayaqn Web Application

SYSTEM ANALYSIS & DESIGN

Architectural designs



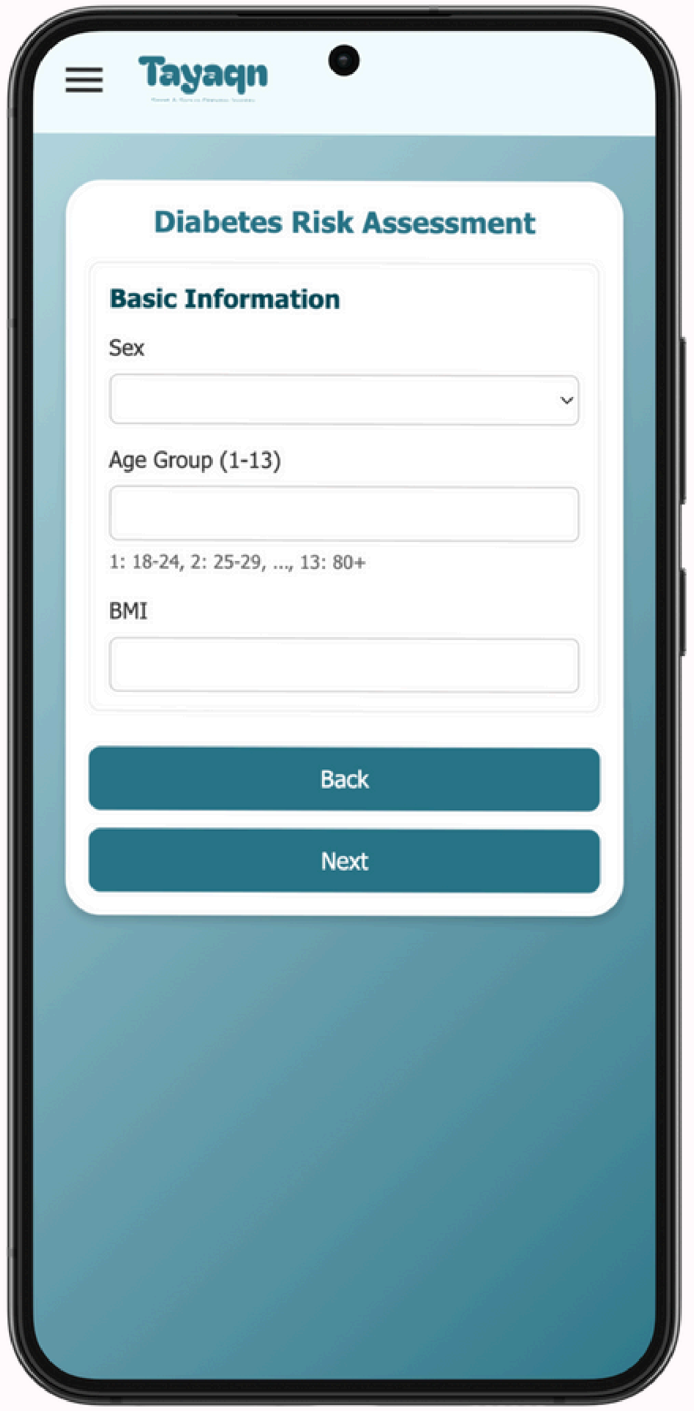
Sequence Diagram

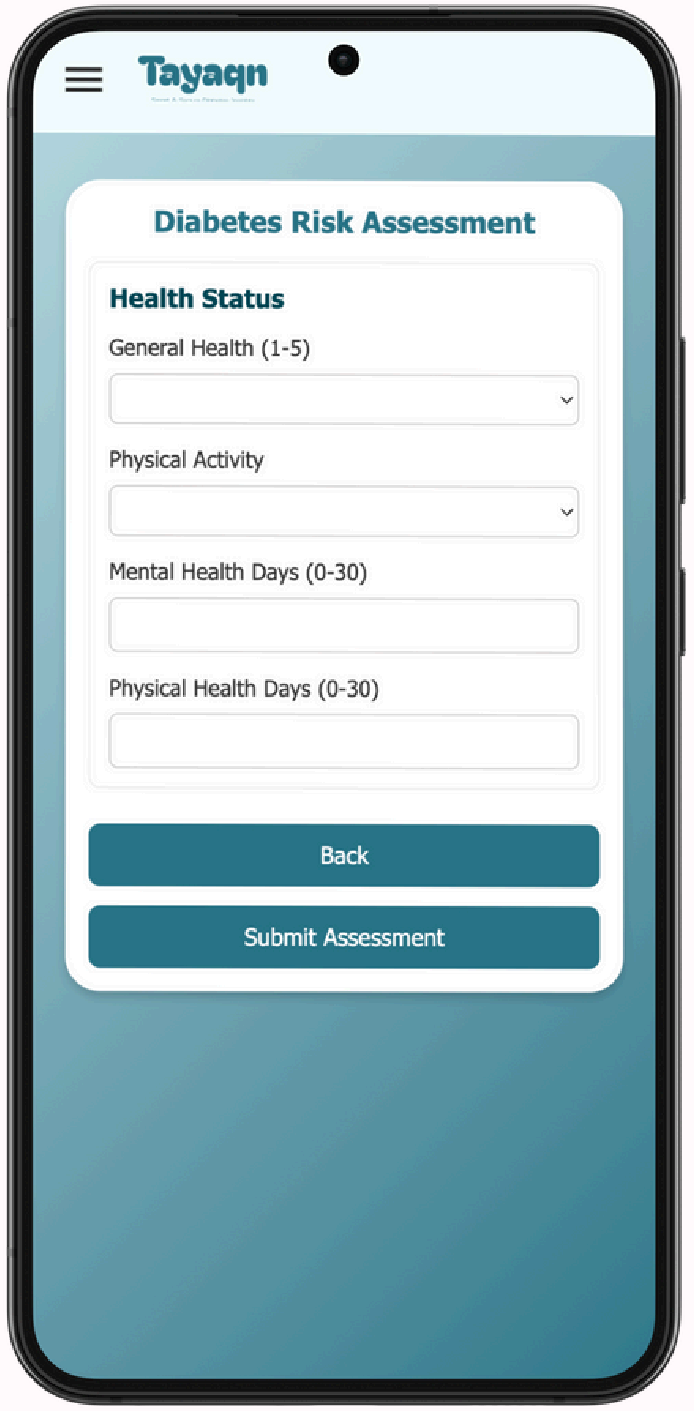
SYSTEM ANALYSIS & DESIGN

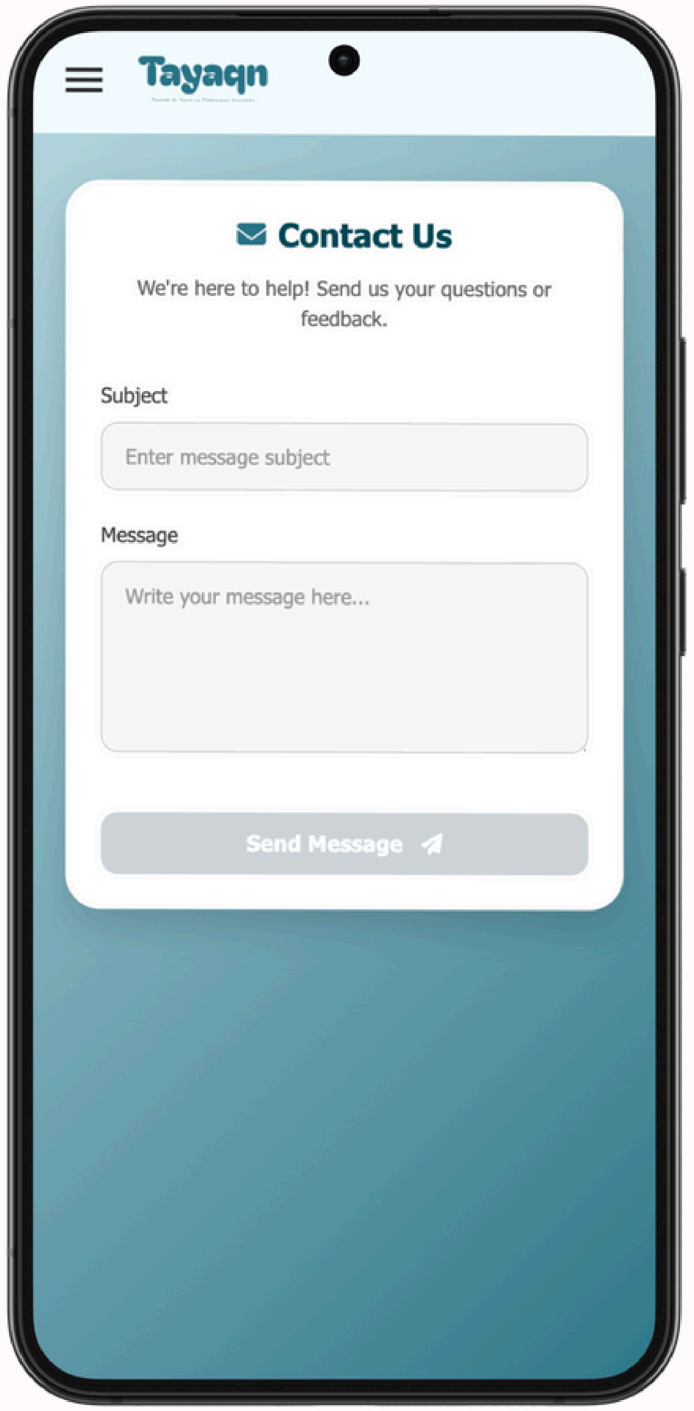
User Interfaces (UI)

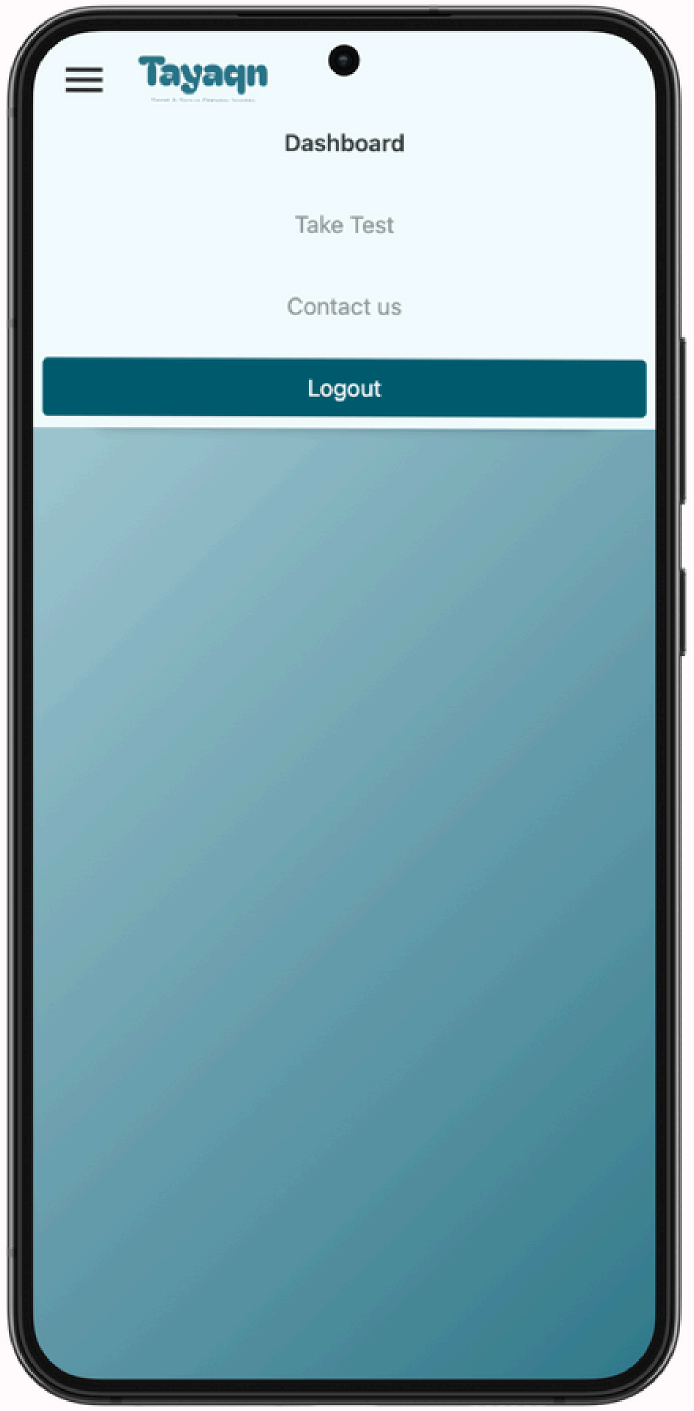
SYSTEM ANALYSIS & DESIGN

User Interfaces (UI)









07 EXPERIMENT DETAILS

01 AI MODEL DEVELOPMENT

1.1 Dataset Details

Dataset Name	Diabetes Health Indicators Dataset (BRFSS 2015)
Size	253,680 survey responses
Target Variable	Diabetes_binary
Features	21 (demographic, health status, lifestyle factors)
Feature Types	Categorical and Integer
Challenges	Imbalanced dataset requiring preprocessing

01 AI MODEL DEVELOPMENT

1.2 AI Model Development Steps

Data Loading & Exploration
Import and analyze diabetes indicators dataset
01

Data Preprocessing
Clean, normalize, and prepare features
02

DL Model Development
Build and train deep learning architecture
03

Evaluation & Visualization
Assess performance and create insights
04

Federated Training
Implement distributed learning protocol
05

01 AI MODEL DEVELOPMENT

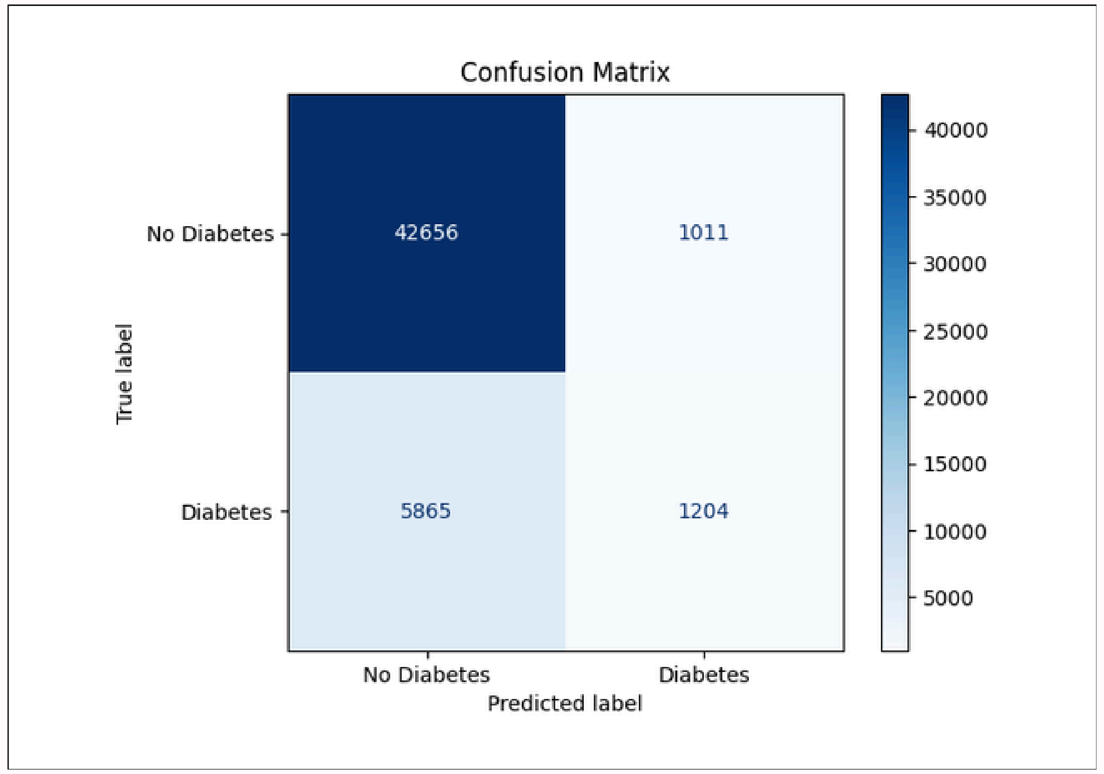
1.3 Evaluation & Results

Classification Report:

	precision	recall	f1-score	support
No Diabetes	0.88	0.98	0.93	43667
Diabetes	0.54	0.17	0.26	7069
accuracy			0.86	50736
macro avg	0.71	0.57	0.59	50736
weighted avg	0.83	0.86	0.83	50736

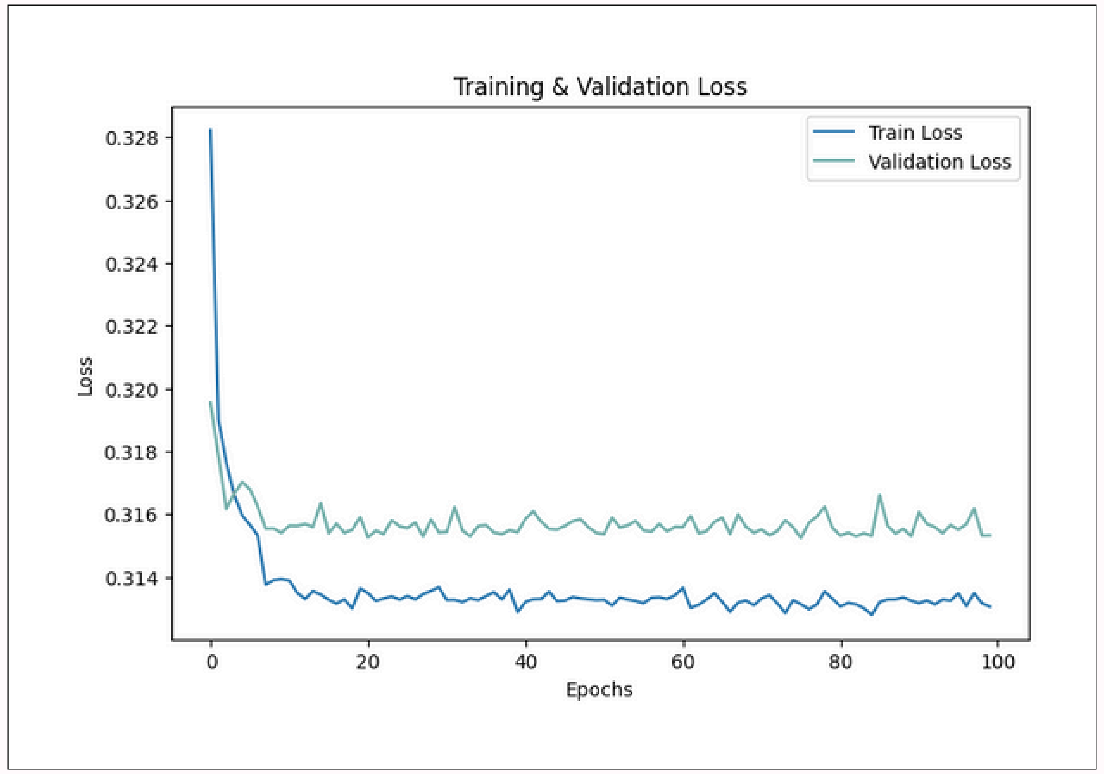
Classification Report

Precision, Recall, and F1-score



Confusion Matrix

Performance visualization

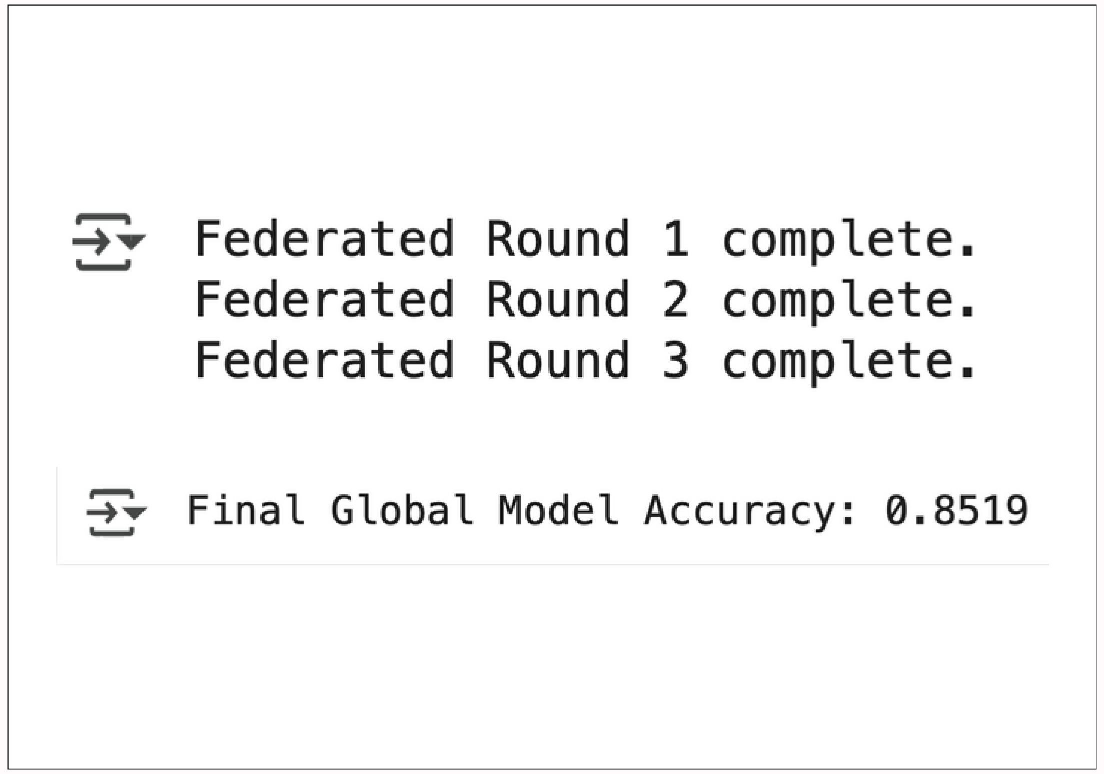


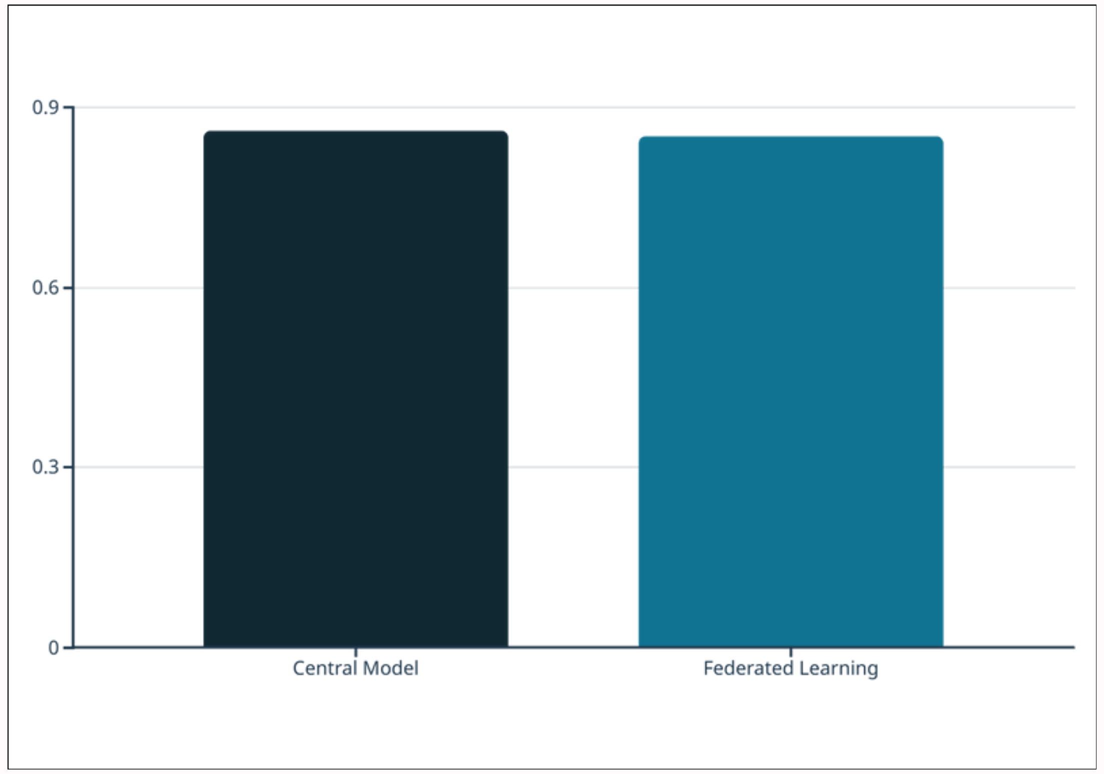
Training and Validation Loss Curves

Model convergence monitoring

01 AI MODEL DEVELOPMENT

1.4 Evaluation & Results





The federated learning approach achieves comparable accuracy (0.8519) to the central model (0.86), demonstrating that privacy preservation comes with minimal performance loss.

02 SECURITY MECHANISMS

01

Sql Injection Prevention

SQLAlchemy ORM is used for protecting the database from sql injections attacks, using parameterized queries.

02

Password Hashing

Bcrypt is used for password hashing to protect the passwords from brute force attacks and rainbow table attacks by using salting.

03

Environment Configuration Management

The file defines and manages environment-specific settings to ensure the application behaves appropriately in different environments (development, staging, production) and keeps the sensitive data out of the main code base.

04

Token Handeling

Token handling protects against unauthorized access and session hijacking by securely storing, validating, and encrypting authentication tokens.

05

Sanitizer Function

Sanitizer function prevents Cross-Site Scripting Security bugs (XSS) attacks by sanitizing any HTML content before being processed.

06

Email Validation

The email validation protects against invalid email formats by using a regular expression (regex) to ensure the email follows a proper structure before processing it further.

08 FUTURE DIRECTIONS

FUTURE DIRECTIONS

01 Mobile Accessibility

Expand the system to a mobile app, making it more accessible to users on the go.

02 Arabic Language Support

Support Arabic language to reach a broader and more diverse user base.

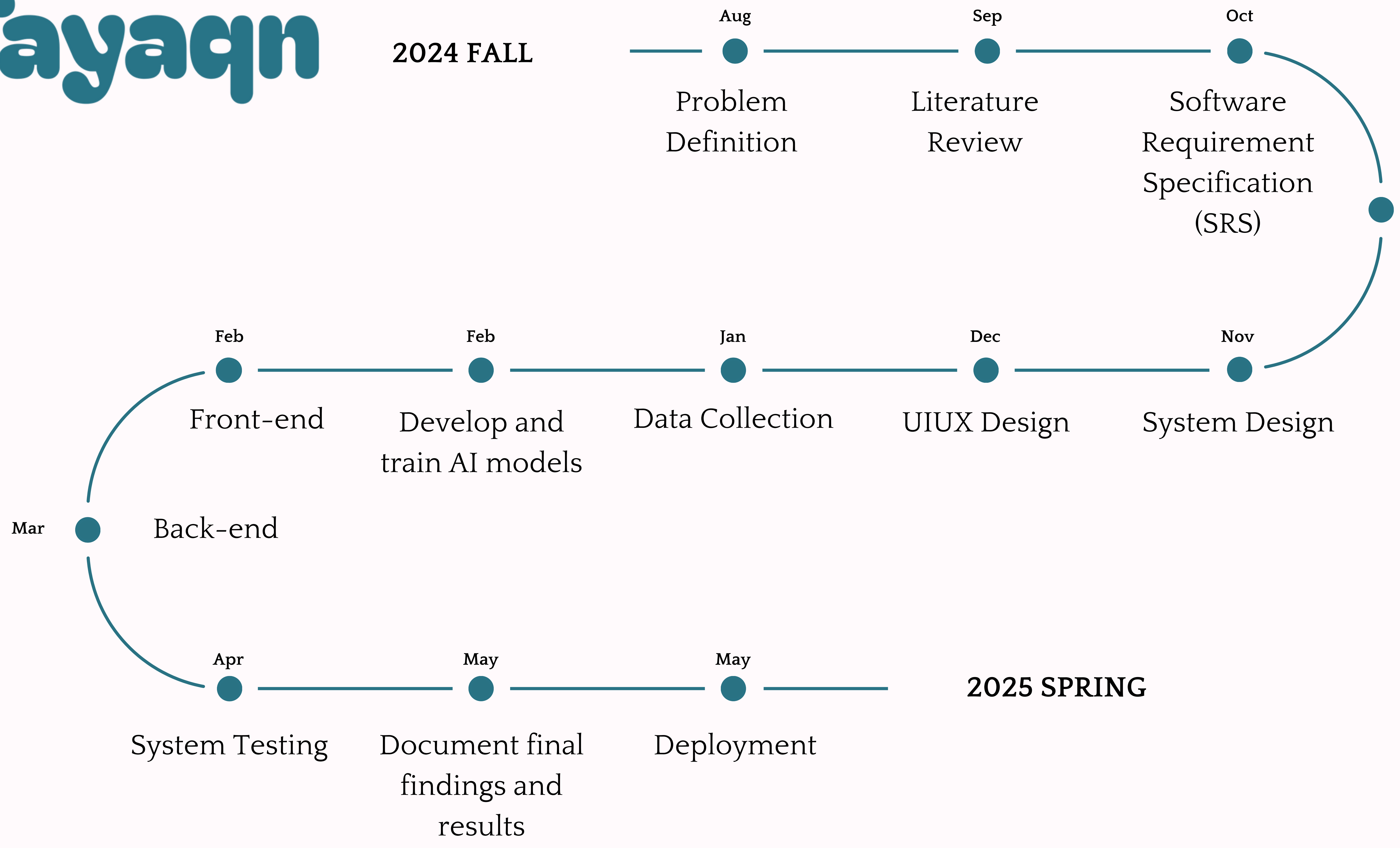
03 Explainable AI

Incorporate Explainable AI techniques to provide transparent and understandable predictions.

09 ROADMAP

PROJECT TIMELINE

Tayaqn



10 CONCLUSION

CONCLUSION

- The **AI-driven Diabetes Detection System** uses **Federated Learning** to ensure data privacy while providing accurate and timely diabetes predictions.
- This innovation **enhances healthcare outcomes** by empowering patients and clinicians with secure, actionable insights for early diagnosis.

THANK YOU

ANY QUESTIONS?

REFERENCES

- A. Mujumdar and V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms,” Procedia Computer Science, vol. 165, pp. 292–299, Jan. 2019, doi: 10.1016/j.procs.2020.01.047.
- A. Nurdin, M. M. Tane, R. W. T. Tumewu, K. M. Suryaningrum, and H. A. Saputri, “Using Machine Learning for the Prediction of Diabetes with Emphasis on Blood Content,” *Procedia Computer Science*, vol. 227, pp. 990–1001, Jan. 2023, doi: 10.1016/j.procs.2023.10.608.
- A. Z. Woldaregay a *et al.*, “Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes,” Elsevier B.V., journal-article, Jul. 2019. [Online]. Available: https://doi.org/10.1016/j.artmed.2019.07.007
- H. Naz and S. Ahuja, “Deep learning approach for diabetes prediction using PIMA Indian dataset,” *Journal of Diabetes & Metabolic Disorders*, vol. 19, no. 1, pp. 391–403, Apr. 2020, doi: 10.1007/s40200-020-00520-5.
- K. Abnoosian, R. Farnoosh, and M. H. Behzadi, “Prediction of diabetes disease using an ensemble of machine learning multi-classifier models,” *BMC Bioinformatics*, vol. 24, no. 1, Sep. 2023, doi: 10.1186/s12859-023-05465-z.
- A. El-Bashbishy Sayyid, H. M. El-Bakry, Information Systems Department, Faculty of Computer and Information Sciences, Mansoura University, Mansoura, Egypt, and Head of Information Systems Department, Faculty of Computer and Information Sciences, Mansoura University, Mansoura, Egypt, “Pediatric diabetes prediction using deep learning,” journal-article, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-51438-4
- Z. Zhang *et al.*, “Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis,” *Journal of Medical Internet Research*, vol. 24, no. 3, p. e26634, Dec. 2021, doi: 10.2196/26634.
- Nassiwa, F., & Zeng, J. (2024). Traditional machine learning model performance testing in predicting the onset of diabetes using the Pima Indians dataset. https://doi.org/10.13140/RG.2.2.22126.22081
- Gkillas, A. (2024). Optimization-based models for explainable deep learning architectures in multidimensional signal processing applications. PhD Thesis, University of Patras. DOI: 10.12681/eadd/57630.
- Ali, M. S., Ahsan, M. M., Tasnim, L., Afrin, S., Biswas, K., Hossain, M. M., Ahmed, M. M., & Islam, M. K. (2024). Federated Learning in Healthcare: Model Misconducts, Security, Challenges, Applications, and Future Research Directions — A Systematic Review. arXiv preprint arXiv:2405.13832.
- Nadella, G. S., Satish, S., Meduri, K., & Meduri, S. S. (n.d.-a). A systematic literature review of advancements, challenges and future directions of AI and ML in Healthcare. International Journal of Machine Learning for Sustainable Development. https://www.ijsdcs.com/index.php/IJMLSD/article/view/519/214
- Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2020, November 12). Federated Learning for Healthcare Informatics - Journal of Healthcare Informatics Research. SpringerLink. https://link.springer.com/article/10.1007/s41666-020-00082-4
- Gu, X., Sabrina, F., Fan, Z., & Sohail, S. (2023, August 7). A review of Privacy Enhancement Methods for Federated Learning in healthcare systems. MDPI. https://www.mdpi.com/1660-4601/20/15/6539
- Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., & Raad, A. (2023, February 13). Reviewing Federated Machine Learning and its use in disease prediction. MDPI. https://www.mdpi.com/1424-8220/23/4/2112
- Loftus, T. J., Ruppert, M. M., Shickel, B., et al. (2022). Federated learning for preserving data privacy in collaborative healthcare research. *Digital Health, 8*, 20552076221134455. https://doi.org/10.1177/20552076221134455