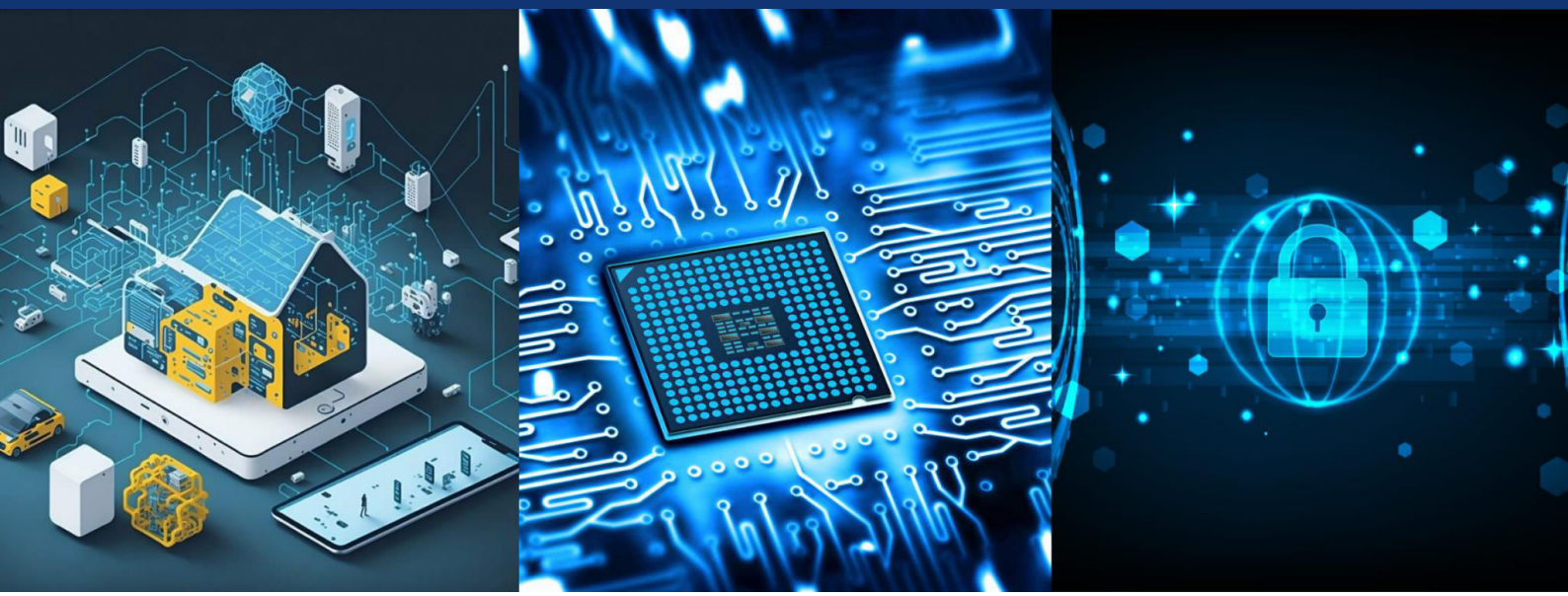
IJIRCCE

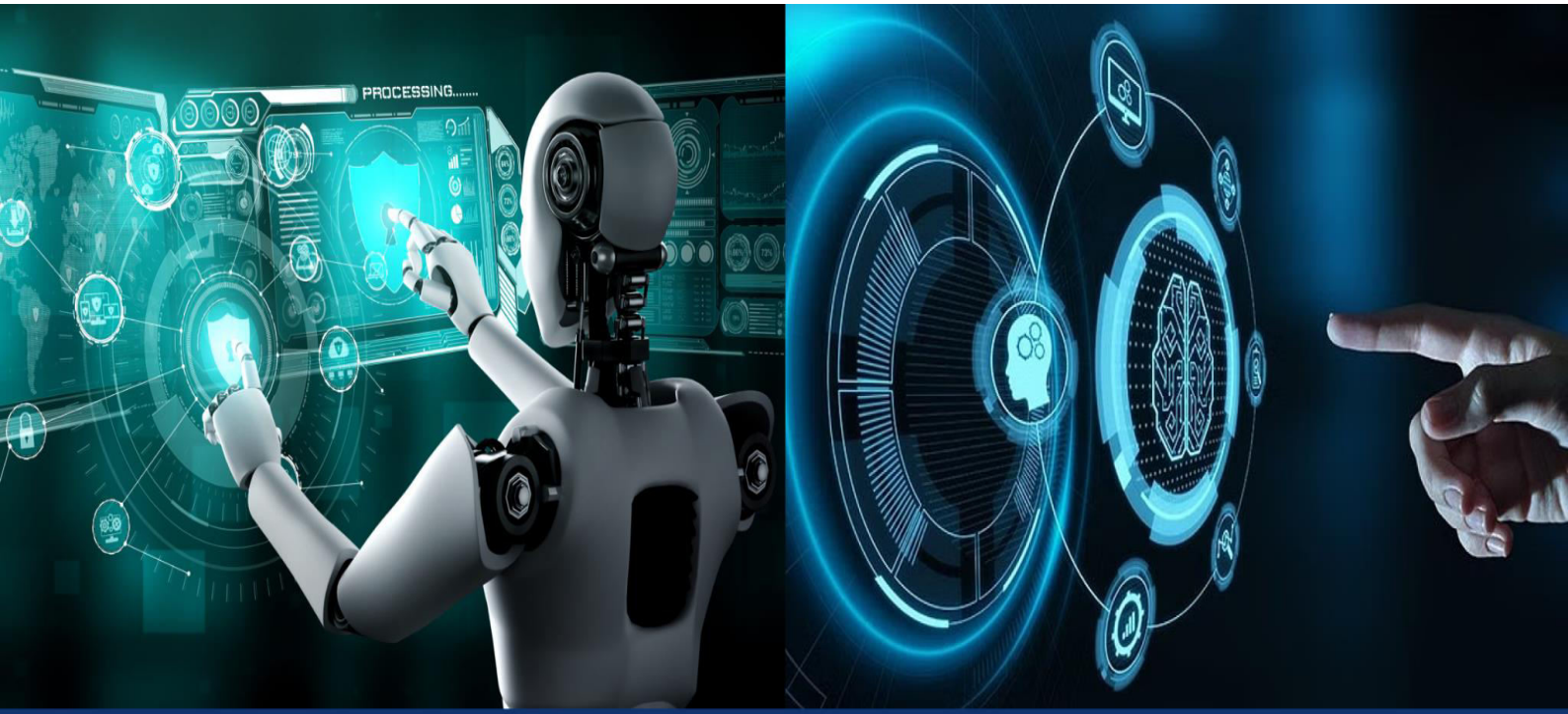
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 5, May 2026

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405034

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Diabetes Prediction Using Machine Learning Approaches

Ramya K[1], Heera Shiny V[1], Mariammal M[1], Lega M[1], J. Revathy[2]

Department of Artificial Intelligence and Data Science, Christ the King Engineering College, Coimbatore, Tamil Nadu, India[1]

Project Guide, Department of Artificial Intelligence and Data Science, Christ the King Engineering College, Coimbatore, Tamil Nadu, India[2]

ABSTRACT: Diabetes mellitus is a chronic metabolic disorder affecting over 537 million adults worldwide. Timely detection of diabetes risk is critical for early medical intervention. This paper presents an integrated diabetes prediction and dietary monitoring system deployed as a Flask web application. A Random Forest Classifier trained on the PIMA Indian Diabetes Dataset (768 records, 8 features) achieves 82.03% accuracy and an AUC-ROC of 0.862. A MobileNetV2 CNN trained on 200 food images across four categories achieves 84.38% training accuracy with a confidence-based fallback using a pretrained ImageNet MobileNetV2 for predictions below 0.70. The application provides secure authentication, automated BMI computation, personalized health tips, dietary recommendations, and PDF report generation. All nine functional test cases passed end-to-end validation, confirming system integrity.

KEYWORDS: Diabetes Prediction; Random Forest; MobileNetV2; CNN; Flask; Food Classification; Machine Learning; Deep Learning; PIMA Dataset; Dietary Monitoring.

I. INTRODUCTION

Diabetes mellitus affects approximately 537 million adults globally, projected to rise to 783 million by 2045 [1]. A large proportion of cases remain undiagnosed until complications—cardiovascular disease, nephropathy, neuropathy, and retinopathy—arise, due to the asymptomatic nature of early-stage diabetes and limited access to specialist care.

Ensemble ML methods such as Random Forest provide robust predictions by aggregating multiple decision trees [2]. Simultaneously, dietary habits critically influence diabetes management; MobileNetV2 CNNs offer lightweight, accurate food classification suitable for web deployment [3]. No existing deployed application integrates both capabilities in a unified, patient-facing interface. This paper bridges that gap, combining ML-based prediction, CNN-based dietary monitoring, secure authentication, automated BMI computation, personalized health guidance, and PDF reporting in a single Flask application.

II. LITERATURE SURVEY

A. Machine Learning for Diabetes Prediction

Sneha and Gangil [4] reported Logistic Regression achieving 77.6% on the PIMA dataset; Sisodia and Sisodia [5] evaluated Naive Bayes (76.3%) and SVM (65.1%), noting class-imbalance issues. Priya and Aruna [2] showed Random Forest achieving 81.97%; Mujumdar and Vaidehi [6] applied Gradient Boosting with SMOTE to reach 83.5%; Kaur and Kumari [7] achieved 78.4% with SVM-RBF. Ayon and Islam [8] obtained 82.1% via a three-layer ANN. Zhu et al. [9] achieved 87.3% with a hybrid LSTM-RF model on longitudinal EHR data, demonstrating advantages of sequential modelling.

B. CNN-Based Food Image Classification

Bossard et al. [10] introduced Food-101 as a benchmark; Liu et al. [11] achieved 82.48% by fine-tuning VGG-16 with augmentation. Subramanian et al. [3] applied MobileNetV2 for South Indian food classification across 30 categories, achieving 88.5% accuracy—directly motivating its adoption in the proposed system.

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405034

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Web-Based Healthcare Applications & Research Gaps

Dey et al. [12] and Ahuja et al. [13] confirmed Flask's suitability for serving ML models via HTTP endpoints. Key gaps remain: (i) isolated prediction models lack patient-facing interfaces; (ii) dietary monitoring and diabetes prediction are rarely unified; (iii) Indian food contexts are underrepresented; and (iv) no deployed application combines both in a single Flask system.

III. SYSTEM DESIGN AND ARCHITECTURE

A. Overall Architecture

The system follows a three-tier architecture: (1) Presentation Layer—HTML/Bootstrap templates via Flask's Jinja2, covering login, registration, profile, prediction, food scan, health tips, and PDF export; (2) Application Logic Layer—Flask handles routing, session management, and inference for both models; (3) Data Layer—SQLite stores user profiles; food images are saved to a static uploads directory.

TABLE I: Technology Stack of the Proposed System

Component	Technology	Purpose
Web Framework	Flask 2.x	HTTP routing, sessions, template rendering
ML Prediction	scikit-learn RandomForest	Diabetes classification from 8 clinical features
DL Food Scan	TensorFlow/Keras MobileNetV2	Food image classification (224×224 input)
Database	SQLite3	User profile and account persistent storage
Auth Security	Werkzeug PBKDF2	Secure password hashing and verification
Frontend	HTML5, CSS3, Bootstrap	Responsive user interface

B. Functional Module Design

The Authentication Module hashes passwords with Werkzeug PBKDF2-HMAC-SHA256 and verifies all protected routes via session. The Profile Module auto-computes BMI = weight / (height/100)² from stored SQLite values, enforcing completeness before prediction access. The Prediction Module assembles an 8-feature vector (6 user-entered + auto-computed BMI + profile age) and calls the Random Forest classifier. The Food Scan Module resizes uploads to 224×224, normalizes to [0,1], runs the custom MobileNetV2, and falls back to ImageNet MobileNetV2 if confidence < 0.70. The Health Tips Module delivers personalized guidance by BMI category, gender, and prediction outcome, with a ReportLab PDF export.

TABLE II: Input Features for Diabetes Prediction Model

#	Feature	Source	Description
1	Pregnancies	User Input	Number of times pregnant
2	Glucose	User Input	Plasma glucose concentration (mg/dL)
3	Blood Pressure	User Input	Diastolic blood pressure (mm Hg)
4	Skin Thickness	User Input	Triceps skin fold thickness (mm)
5	Insulin	User Input	2-hour serum insulin (mu U/ml)
6	DPF	User Input	Diabetes pedigree function score
7	BMI	Auto-computed	Derived from stored height and weight
8	Age	User Profile	Retrieved from SQLite user profile

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405034

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Security Design

Plaintext passwords are never stored. The Flask secret key is loaded from an environment variable. Werkzeug's secure_filename prevents directory traversal. The database enforces a UNIQUE constraint on email, and all sensitive routes redirect unauthenticated requests to login.

IV. IMPLEMENTATION

A. Datasets

Diabetes: PIMA Indian Diabetes Dataset (UCI), 768 records, 8 features, binary label. Food: 200 images across 4 classes (apple, banana, burger, pizza)—160 training (40/class), 40 validation (10/class), loaded via Keras ImageDataGenerator at 224×224.

B. Diabetes Prediction Model

A Random Forest Classifier (n_estimators=100, max_depth=5, random_state=42) is trained on the full PIMA dataset. At inference, BMI is computed from stored profile data and the 8-feature vector is passed to model.predict() to return 1 (Diabetic) or 0 (Non-Diabetic).

C. CNN Food Classification & Flask Routes

MobileNetV2 (ImageNet weights) serves as a feature extractor; custom dense layers add 4-class softmax output. The model trains for 30 epochs (Adam optimizer, categorical cross-entropy). If max softmax probability < 0.70, the pretrained ImageNet MobileNetV2 is invoked with keyword-based vocabulary mapping. Key routes: /predict (RF inference), /scan (CNN inference), /download_pdf (ReportLab PDF via send_file()). Main implementation challenges—CNN overfitting, BMI inconsistency, incomplete profiles, and session food-list persistence—were resolved via the confidence fallback, SQLite auto-compute, null-checks, and pre-logout list extraction respectively.

TABLE III: CNN Training Progress (Selected Epochs)

Epoch	Train Accuracy	Train Loss	Val. Accuracy
1	40.00%	1.9632	27.50%
10	67.50%	0.8921	30.00%
20	78.75%	0.5634	32.50%
30	84.38%	0.3973	30.00%

V. RESULTS AND ANALYSIS

A. Diabetes Prediction Model Results

The Random Forest achieved 82.03% accuracy and AUC-ROC of 0.862, indicating strong class separability. Glucose, BMI, and age were the most influential features. Lower recall (74.63%) vs. precision (79.41%) reflects PIMA's class imbalance (65% non-diabetic, 35% diabetic).

TABLE IV: Random Forest Classifier Performance Metrics

Metric	Value	Parameter	Setting
Accuracy	82.03%	n_estimators	100
Precision	79.41%	max_depth	5
Recall	74.63%	CV Folds	5
F1-Score	76.94%	random_state	42
AUC-ROC	0.862	Dataset Size	768

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405034

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. CNN Food Classification Results

Training accuracy reached 84.38% at Epoch 30; validation accuracy remained at 30–35%, indicating overfitting from the small dataset. Apple achieved the highest F1-score (0.40) due to distinctive texture. The confidence-based fallback ensures practical usability. Expanding to 500+ images per class with augmentation is expected to substantially close this gap.

C. Comparative Analysis

TABLE V: Comparative Summary of Diabetes Prediction Methods

Author (Year)	Algorithm	Dataset	Accuracy
Sneha & Gangil (2019)	Logistic Regression	PIMA	77.60%
Sisodia & Sisodia (2018)	Naive Bayes	PIMA	76.30%
Kaur & Kumari (2018)	SVM (RBF)	PIMA	78.40%
Priya & Aruna (2013)	Random Forest	PIMA	81.97%
Mujumdar & Vaidehi (2019)	Gradient Boosting	PIMA	83.50%
Ayon & Islam (2019)	ANN (MLP)	PIMA	82.10%
Zhu et al. (2021)	LSTM + RF	EHR	87.30%
Proposed System	Random Forest (n=100)	PIMA	82.03%

D. Functional Test Results

All nine test cases passed, confirming end-to-end system integrity across authentication, prediction, food scanning, PDF generation, and session management (Table VI).

TABLE VI: Functional Test Results

Module	Test Case	Result
Authentication	Registration with hashed password storage	Pass
Authentication	Login with valid and invalid credentials	Pass
Profile	BMI auto-computation from stored height and weight	Pass
Prediction	Diabetic result for high-risk input values	Pass
Prediction	Non-diabetic result for normal clinical values	Pass
Food Scan	Correct food identification (apple, banana)	Pass
Food Scan	Fallback activation on low-confidence prediction	Pass
PDF Generation	Saved food list exported as downloadable PDF	Pass
Session Mgmt	Saved food list retained after logout	Pass

VI. CONCLUSION AND FUTURE WORK

This paper presented an integrated diabetes prediction and dietary monitoring system deployed as a Flask web application. The Random Forest Classifier achieved 82.03% accuracy and AUC-ROC of 0.862, with glucose, BMI, and age as the most influential features. The MobileNetV2 CNN achieved 84.38% training accuracy with a confidence-based fallback ensuring reliable food classification. All nine functional modules passed testing, bridging the gap between isolated predictive models and accessible patient-facing healthcare tools.

Future work includes: (i) retraining on larger, multi-demographic datasets with SMOTE oversampling; (ii) expanding food data to 50+ categories with augmentation; (iii) integrating YOLO for real-time portion-size estimation; (iv) cloud

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

deployment (AWS/GCP) with Gunicorn and HTTPS; (v) native Android/iOS mobile application; (vi) wearable and CGM data integration for longitudinal monitoring; and (vii) regional Indian language support (Tamil, Hindi, Telugu).

REFERENCES

[1] IDF, "Diabetes Atlas," 10th ed., International Diabetes Federation, Brussels, Belgium, 2021.
[2] R. Priya and P. Aruna, "SVM and Neural Network Based Diagnosis of Diabetic Retinopathy," Int. J. Computer Applications, vol. 41, no. 1, pp. 6–12, 2013.
[3] V. Subramanian, S. Sriram and R. Lavanya, "Transfer Learning Based South Indian Food Classification Using MobileNetV2," Int. J. Advanced Computer Science and Applications, vol. 11, no. 9, pp. 423–430, 2020.
[4] N. Sneha and T. Gangil, "Analysis of Diabetes Mellitus for Early Prediction using Optimal Features Selection," J. Big Data, vol. 6, no. 1, pp. 1–19, 2019.
[5] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," Procedia Computer Science, vol. 132, pp. 1578–1585, 2018.
[6] A. Mujumdar and V. Vaidehi, "Diabetes Disease Prediction using ML on Big Data with Apache Spark," Procedia Computer Science, vol. 165, pp. 23–38, 2019.
[7] H. Kaur and V. Kumari, "Predictive Modelling and Analytics for Diabetes using a ML Approach," Applied Computing and Informatics, vol. 18, no. 1/2, pp. 90–100, 2018.
[8] S. I. Ayon and Md. M. Islam, "Diabetes Prediction: A Deep Learning Approach," Int. J. Information Engineering and Electronic Business, vol. 11, no. 2, pp. 21–27, 2019.
[9] T. Zhu, K. Li, P. Herrero and P. Georgiou, "Deep Learning for Diabetes: A Systematic Review," IEEE J. Biomedical and Health Informatics, vol. 25, no. 7, pp. 2744–2757, 2021.
[10] L. Bossard, M. Guillaumin and L. Van Gool, "Food-101 — Mining Discriminative Components with Random Forests," in Proc. ECCV, Zurich, pp. 446–461, 2014.
[11] C. Liu et al., "DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment," in Proc. Int. Conf. Inclusive Smart Cities and Digital Health, Springer, pp. 37–48, 2016.
[12] S. Dey, C. Bhatt and A. S. Ashour, "Big Data for Remote Sensing: Visualization, Analysis and Interpretation," Springer, Cham, pp. 13–35, 2021.
[13] R. Ahuja et al., "Classification and Clustering Algorithms of ML with their Applications," Machine Learning and Big Data Analysis, Springer, Singapore, pp. 225–248, 2019.
[14] Q. Zou et al., "Predicting Diabetes Mellitus with Machine Learning Techniques," Frontiers in Genetics, vol. 9, Article 515, pp. 1–10, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details