Mesopotamian journal of Big Data
Vol. **(2025)**, 2025, **pp.** 211–221
DOI: https://doi.org/10.58496/MJBD/2025/014 , ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

A Hybrid Machine Learning Approach for Enhanced Diabetes Prediction: Integrating Image and Numerical Data

Ahmad Shaker Abdalrada[1,*], Ali Fahem Neamah [2], Huda Lafta Majeed [3], Ahmed Raad Al-Sudani [4]

[1,2,4] *Department of Software, College of Computer Science and Information Technology, Wasit University, Iraq*
[3] *Department of Computer Science, College of Computer Science and Information Technology, Wasit University, Iraq*

ARTICLE INFO

Article History

Received 23 May 2025
Revised 24 Jul 2025
Accepted 10 Aug 2025
Published 19 Aug 2025

Keywords

Diabetes Mellitus (DM)
Machine Learning (ML)
Deep Learning (DL)
Multimodal Data Integration.

ABSTRACT

Diabetes mellitus (DM) continues to escalate as a worldwide health emergency issue, with approximately 537 million adults currently diagnosed and forecasts estimating a further increase to 643 million by 2030. Early and precise foretelling of DM remains a decisive factor for timely intervention, thereby mitigating severe downstream sequelae such as cardiovascular disease, peripheral neuropathy, and diabetic retinopathy (DR). Conventional prognostic frameworks typically depend on exclusively either structured tabular measurements or visual medical imagery, which constrains comprehensive diagnostic capacity. This contribution confronts such limitation by advancing a hybrid machine learning (ML) methodology that synergistically combines deep learning—specifically, convolutional neural networks (CNNs) dedicated to retinal photograph scrutiny—with gradient-boosting machines (GBMs) that ingest structured demographic and clinical variables. Two publicly accessible repositories supplied training material: the Pima Indians Diabetes Database for tabular covariates and the Asia Pacific Tele-Ophthalmology Society (APTOS 2019) Blindness Detection corpus for fundus imagery. Retinal studies underwent standardised pre-processing re-scaling, pixel normalisation, Gaussian denoising, and multiplicative augmentation while tabular patient records underwent rigorous feature ranking. Outcome representations from both data strata were concatenated into a consolidated tensor, thereby rendering simultaneous latent-space learning achievable. The experimental results demonstrate that the hybrid model outperforms single-modality models, achieving an accuracy of 96%, a macro average F1 score of 0.96, and an area under the receiver operating characteristic curve (AUC-ROC) of 0.994. The proposed approach offers a comprehensive diagnostic framework by combining systemic and localized disease indicators, thereby enhancing robustness, reducing variance, and supporting more informed clinical decision-making. This work highlights the potential of multimodal ML integration for complex disease prediction and sets the stage for future extensions to other chronic conditions.

1. INTRODUCTION

Diabetes mellitus (DM) is a chronic metabolic disorder that presents with chronically elevated blood glucose levels; if left untreated, DM can lead to serious complications such as cardiovascular disease, peripheral neuropathy, and diabetic retinopathy (DR)[1] . As estimated by the International Diabetes Federation (IDF), approximately 537 million adults are currently living with DM, and this number is expected to reach 643 million in 2030, representing an increasing public health and economic burden worldwide [2]. Hence, early recognition of the disease and its risk stratification are necessary to avoid advancement of the condition and provide better outcomes [3, 4].

Standard diagnostic practice is largely dependent on laboratory markers such as plasma glucose and HbA1c, which still play key roles in clinical decision-making [5]. Simultaneously, improvements in retinal imaging are also allowing earlier detection of DR-associated microvascular changes, often long before recognizable symptoms are detected, or potential intervention is necessary[6]. Specifically, retinal fundus photography can detect changes in the microvasculature that herald the development of the disease process of diabetes even when patients are asymptomatic [7]. These observations motivate the inclusion of both systemic and localized indicators as well as mechanisms in predictive frameworks for DM.

Moreover, extracting reliable knowledge from heterogeneous data poses nontrivial challenges. Machine learning (ML) methods have demonstrated strong performance in medical prediction tasks and screening workflows [8-11], yet many models are designed around a single modality—either structured numerical variables or medical images—thereby capturing

*Corresponding author. Email: aabdalra@uowasit.edu.iq

only part of the underlying pathophysiology. Integrating disparate modalities introduces issues of feature fusion, preprocessing compatibility, and generalizability across populations. Prior investigations suggest that combining complementary information sources can increase robustness and accuracy when properly unified within a single learning pipeline [12].

In this context, the present study proposes a hybrid ML approach that processes retinal images and structured numerical patient variables concurrently. Convolutional neural networks (CNNs) are employed to learn discriminative visual representations from retinal images, whereas gradient boosting machines (GBMs) model nonlinear relationships among demographic and clinical measurements. By concatenating modality-specific feature vectors into a unified representation, the framework aims to uncover patterns that may remain latent when either modality is considered in isolation. The ensuing evaluation examines predictive performance via accuracy, precision, recall, the F1 score, and the area under the receiver operating characteristic curve (AUC-ROC) and analyses the contribution of each modality within the integrated setting.

This work addresses a key gap in DM prediction by bridging systemic and localized indicators within a single, multimodal model. Beyond performance gains, such integration aspires to support clinical decision-making with richer, more contextualized evidence, potentially enabling earlier intervention and improved outcomes. The following sections detail related studies, the proposed methodology for data preprocessing, feature fusion and model construction, and a comprehensive analysis of the results and implications.

The aims of this study are as follows: (1) A hybrid model that employs a powerful and effective feature fusion of image and numerical features to perform prediction on diabetes images; (2) a performance evaluation of the model is conducted through several statistical and classification assessment metrics; and (3) a comparative analysis involving several experiments of the proposed model against various unimodal and multimodal models establishes the relative benefits of combining image and numerical features via the proposed level of modelling. Additionally, this study will attempt to address two further research questions: What is the effect of multimodal integration on predictive accuracy compared with unimodal approaches? Which feature combination strategies lead to the greatest performance improvements?

This work is important because it contributes to the development of predictive healthcare analytics. This finding not only highlights the importance of multimodal learning in practical medical applications but also shows that hybrid architectures can provide a useful alternative to conventional single-modality models. In addition, the results may help shape the basis of clinical decision-support systems to support clinicians in early diagnosis, personalized planning of treatment, and long-term management of the disease.

2. RELATED WORKS

Over the past few years, the capabilities of a variety of machine learning (ML) and deep learning (DL) tools have facilitated the creation of data fusion-based predictive models of diabetes that incorporate several data modalities and algorithms. Recent interest in hybrid architectures arises from the inability of single-modality methods to capture a comprehensive picture for diagnosis. The use of image-derived features in combination with structured patient data has been investigated in various studies to facilitate diabetes risk prediction. For example, Yao et al. developed the ‘SynthA1c’ method in [13], which integrates the CT scan phenotype and physical examination data via neural networks and decision tree models to predict bloodless HbA1c levels, with up to 87.6% sensitivity. Although this highlights the potential of multimodal synergies, the fusion of retinal images with systemic variables was not adequately evaluated as part of an ophthalmic-based predictive framework.

In addition to multimodal fusion, a variety of hybrid ML methods have been developed for structured datasets utilizing both ensemble and dimensionality reduction strategies. Sampath, et al. [14] used the synthetic minority oversampling technique (SMOTE) along with ensemble classifiers, which gave them a balanced class, and they were able to achieve 94.12% accuracy on the NHANES data and 89.47% accuracy on the PIMA Indian data. Similarly, Bülbül [15] combined genetic algorithms with stacked autoencoders and Softmax classifiers, reaching an accuracy of 98.72%. Poornima [16] applied random projection for dimensionality reduction, followed by a hybrid classification pipeline, resulting in improved sensitivity (0.95), specificity (0.98), accuracy (0.97), and AUC (0.97). Abnoosian, et al. [17] reported that bagging, boosting, and stacking outperform single classifiers, achieving an AUC of 0.999, whereas Liu, et al. [18] integrated unsupervised clustering with supervised classifiers to enhance pattern discovery, reporting accuracy above 99%. Further evidence from Hasan, et al. [19] and Dutta, et al. [20] confirmed that ensemble methods yield more robust predictions, with Dutta et al. demonstrating notable improvements through weighted ensemble learning. Despite these advances, such methods have been largely confined to structured data, overlooking the potential of visual biomarkers such as retinal images.

Concurrently, DL-based methods have achieved significant success in diabetic retinopathy (DR) detection from retinal images. Bhimavarapu and Battineni [21] improved CNN efficiency through optimized activation functions, delivering superior DR classification results. Gulshan, et al. [22] trained a deep CNN on a large, multiethnic dataset, achieving 90.3% sensitivity and 98.1% specificity, whereas Pratt, et al. [23] reported 75% accuracy in DR severity classification. The Kaggle DR competition [24] further demonstrated the capability of CNN-based solutions, with AUC-ROC scores ranging from 0.85--0.9. Moustari, et al. [25] proposed a two-branch attention-guided CNN architecture, reaching an AUC of 0.998, and Lam, et al. [26] combined multiple CNN architectures to improve DR classification, with an AUC of 0.97. Additionally, Rajalakshmi, et al. [27] developed a mobile-based DR detection system that achieved high sensitivity (95.8%) and moderate specificity (80.2%). These studies highlight the impressive performance of image-based models, but they generally lack integration with systemic variables, which could improve diagnostic completeness.

Overall, existing research has demonstrated strong predictive ability when focusing on either image-based or numerical-data-based approaches. However, multimodal learning—particularly the combination of DR-related visual indicators with systemic patient data—remains underutilized. Most ensemble and hybrid methods emphasize tabular datasets, whereas image-based DL methods omit complementary systemic risk factors. This gap underscores the need for predictive frameworks that jointly leverage both modalities, aiming to increase accuracy, robustness, and clinical relevance.

To provide a concise comparative overview, Table 1 summarizes the main characteristics, methodologies, datasets, performance metrics, and limitations of the reviewed studies, highlighting the distinctions and research gaps that motivate the present work.

TABLE I. SUMMARY AND COMPARISON OF REVIEWED STUDIES

Study	Data type(s)	Method(s)	Dataset(s)	Best metric(s)	Limitations
Yao, et al. [13]	CT images + physical data	Neural networks + decision trees	Clinical CT + exam data	Sensitivity = 87.6%	No retinal images, limited modality fusion
Sampath, et al. [14]	Numerical	SMOTE + ensemble ML	NHANES, PIMA	Acc. 94.12%, 89.47%	No image data
Bülbül [15]	Numerical	Genetic algorithm + stacked AE + Softmax	Clinical tabular	Acc. 98.72%	No multimodal inputs
Poornima [16]	Numerical	RP + hybrid classifiers	UCI diabetes	Sens. 0.95, Spec. 0.98, Acc. 0.97, AUC 0.97	No image features
Abnoosian, et al. [17]	Numerical	Ensemble (bagging, boosting, stacking)	Multiple diabetes datasets	AUC 0.999	No visual biomarkers
Liu, et al. [18]	Numerical	Unsupervised clustering + supervised learning	Clinical tabular	Acc. >99%	No multimodal fusion
Hasan, et al. [19]	Numerical	Ensemble ML	Multiple datasets	Improved AUC	No image features
Dutta, et al. [20]	Numerical	Weighted ensemble	Clinical tabular	Significant metric gain	No visual biomarkers
Bhimavarapu and Battineni [21]	Retinal images	CNN with enhanced activation	Kaggle DR	High DR classification perf.	No systemic data
Gulshan, et al. [22]	Retinal images	Deep CNN	Large multiethnic	Sens. 90.3%, Spec. 98.1%	No multimodal features
Pratt, et al. [23]	Retinal images	CNN	DR datasets	Acc. 75%	Limited performance
Kaggle comp. [24]	Retinal images	CNN variants	APTOS DR	AUC 0.85–0.9	No systemic data
Moustari, et al. [25]	Retinal images	AG-CNN	APTOS DR	Acc. 0.9848, AUC 0.998	No systemic data
Lam, et al. [26]	Retinal images	CNN ensemble	DR datasets	AUC 0.97	No multimodal fusion
Rajalakshmi, et al. [27]	Retinal images	Mobile DL	Fundus photos	Sens. 95.8%, Spec. 80.2%	Limited computational depth

3. PROPOSED MODEL

This section illustrates the proposed model for the prediction of diabetes disease incidence. Figure 1 illustrates the overall architecture of the proposed hybrid machine learning framework, highlighting the sequential stages from dataset description through preprocessing and feature concatenation to the final hybrid model and performance evaluation metrics.

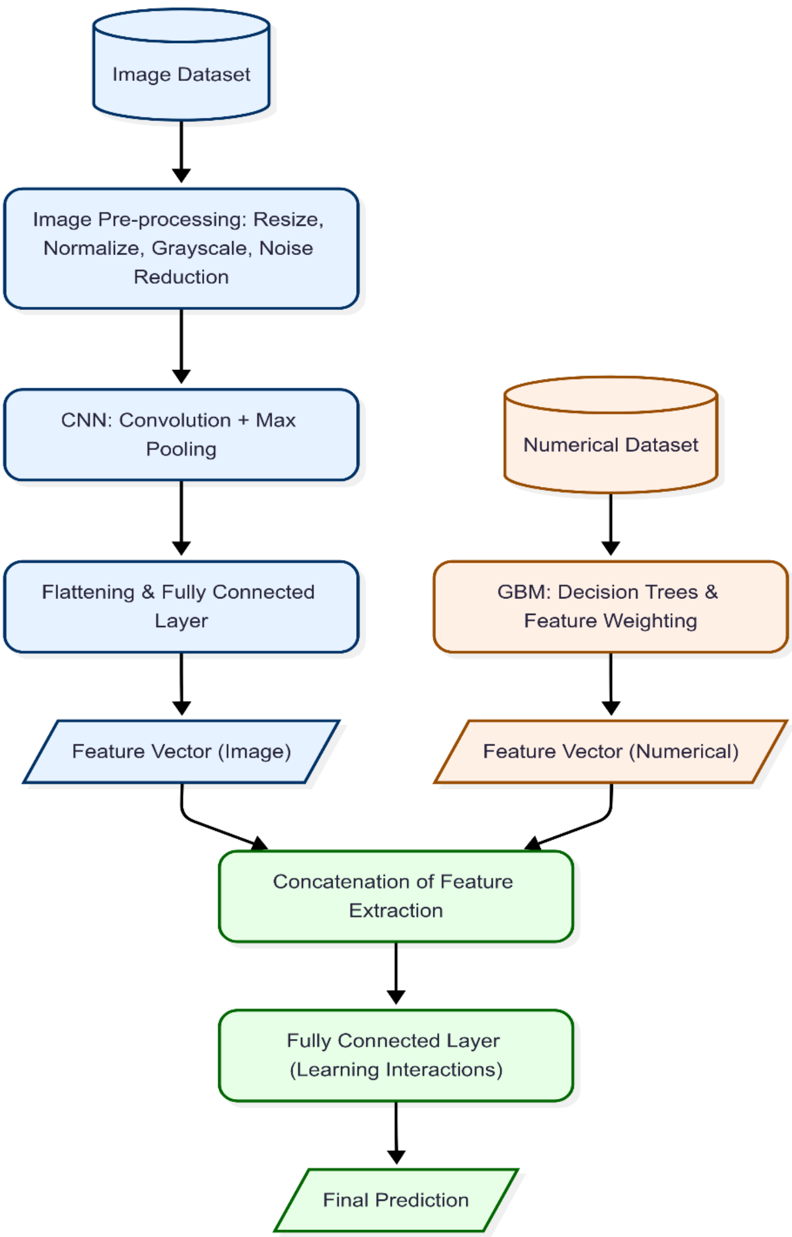

Figure 1. Workflow of the proposed hybrid machine learning framework for diabetes prediction, which integrates retinal image features with numerical patient data.

The proposed model is a hybrid machine learning framework designed to predict diabetes by integrating both image and numerical data. This model leverages the strengths of convolutional neural networks (CNNs) for image analysis and gradient boosting machines (GBMs) for numerical data processing. The integration aims to provide a more comprehensive understanding of the factors contributing to diabetes prediction.

3.1 Dataset

Owing to the lack of a comprehensive diabetes dataset, for this study, two different datasets were employed. The Pima Indians Diabetes Database was chosen for the numerical dataset, which has information regarding the population of Pima Indians, which is a group of Native Indians in a specific region of Arizona, USA, and their potential risk of developing diabetes[28]. The dataset includes several medical and demographic factors, such as age, BMI, blood pressure, and glucose level, along with the determination of whether a person developed diabetes within 5 years of the initial examination. As shown in Table 2.

The dataset generated by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) is expected to be used within machine learning and data mining fields; indeed, this database is mostly used as a benchmark dataset for diabetes predictive models. It contains 768 observations with 8 variables.

TABLE II. PRESENT THE DATASET

variables	Description
Pregnancies (preg)	"number of times pregnant"
Glucose(plas)	"plasma glucose concentration after 2 hours in an oral glucose tolerance test"
BloodPressure(pres)	"diastolic blood pressure (mm Hg)"
SkinThickness(skin)	"triceps skin fold thickness (mm)"
Insulin(insu)	"2-hour serum insulin (mu U/ml)"
BMI(mass)	"body mass index (weight in kg/(height in m)^2)"
DiabetesPedigreeFunction(pedi)	"diabetes pedigree function (a function which scores the likelihood of diabetes based on family history)"
Age(age)	"age in years"
The target variable is Outcome, which indicates whether or not an individual developed diabetes within 5 years of the initial examination (0 = no diabetes, 1 = diabetes).	

For the image data in this model, we used datasets such as retinal images for diabetic retinopathy detection from the APTOS 2019 Blindness Detection Dataset [29]. This dataset is part of a Kaggle competition and contains high-resolution images of retinas taken under a variety of imaging conditions. The goal is to predict the presence and severity of diabetic retinopathy. Over 3500 images labelled with different levels of diabetic retinopathy severity, No DR (0), Mild DR (1), Moderate DR (2), Severe DR (3), and Proliferative DR (4), are included, as shown in Figure 2. In this study, only 10000 images were used after augmentation.


(No DR) (Mild) (Moderate) (Severe) (Proliferative DR)

Fig. 2. Levels of diabetic retinopathy

3.2 Preprocessing Steps

In the integrated model that uses both image and numerical data for diabetic disease classification, preprocessing image data is crucial for ensuring that the images are in a suitable format and quality for the model to effectively learn patterns and make accurate predictions. The preprocessing steps help standardize the images, enhance important features, and reduce computational complexity. Image resizing is applied to ensure that all the input images have the same dimensions, which is necessary for batch processing in deep learning models. In this model, images are resized to a fixed size (e.g., 128x128 pixels). Normalization was also employed to scale the pixel values to a range that is suitable for the neural network. Typically, pixel values are normalized to a range of [0, 1] or [-1, 1]. This is achieved by dividing the pixel values by 255 (if the range is [0, 255]) to normalize them between [0, 1]. Data augmentation is a process of applying random heat transformations to artificially increase the diversity of training dataset. This mitigates the overfitting and increases the generalizability of the model. Some of the images were converted to grayscale through a process called grayscale conversion. Noise, which was identified as the potential noise in image that effects the important features of the image generating poor performance of model, was reduced by implementing a Gaussian filter that smooths the image. Gaussian filter was used to attenuate the high-frequency noise and highlights the features appearing in the retinal images. The implementation of the filter was done using OpenCV in Python with the cv2 function. GaussianBlur(). A kernel size of (5, 5) and a sigma of 1.0 were used as the parameters. The setting you have chosen made the filter to gently smooth the images but keeping the important structures like blood vessels and lesions [1]. The Gaussian filter increased the robustness of CNN by maintaining the uniformity of input.

3.3 Concatenating Feature Vectors

Feature vectors are numeric representations of data, and concatenating them involves combining these vectors into a single unified representation. This unified vector can then be fed into subsequent layers of the hybrid model for prediction. The process works as follows:

3.3.1 Feature Vector Extraction

- **From Numeric Data:**
  After processing the numeric data via RF feature selection, the output is typically a 1D feature vector. For example, if there are 10 numeric features (e.g., age, BMI, blood pressure), the numeric model might output a feature vector of size n, such as [f1,f2...,fn]

- **From image data:**
  The convolutional neural network (CNN) processes the image and produces a high-dimensional feature map (e.g., a 3D tensor). This feature map is flattened into a 1D vector via techniques such as global average pooling (GAP), max pooling, or fully connected layers. For example, a CNN might output a vector of size m, such as [g1,g2...,gm]

3.3.2 Aligning feature dimensions

Feature dimension alignment is a crucial task whenever we have models working on different types of data (e.g., numerical and image data). It ensures the input data has uniformity and is compatible with the model for processes. Mismatched feature dimensions can lead to errors, non-optimal training or convergence failures.

3.3.3 Concatenation process

Once the feature vectors are prepared, they are concatenated along the feature axis. Suppose that the numeric feature vector is Fnum= [f1, f2...,fn] of size n and that the image feature vector is Fimg=[g1,g2...,gm] of size m. The concatenated feature vector FconcatF is formed as Fconcat=[f1,f2...,fn,g1,g2...,gm]. The resulting vector has a size of n+ m.

3.4 Model Construction

The CNN model is designed to extract features from retinal images. It consists of several convolutional layers followed by max pooling layers to capture spatial hierarchies in the image data. The architecture of the CNN employed in this model consists of four convolutional layers, and each layer has a 3×3 kernel and ReLU activation functions. After each two convolutional layers, MaxPooling layers with a 2×2 window were applied. This is done to reduce the spatial dimensions. All the features are extracted after passing through fully connected layers. Those layers are 128 and 64 neurons, respectively, followed by ReLU activation. A sigmoid activation function was used at the end to the output layer for classification. Moreover, a dropout rate of 0.3 and L2 regularization ($\lambda = 0.0001$) were also used to avoid overfitting and improve the generalizability. The CNN was also configured with the Adam optimizer with a learning rate of 0.001. For the loss function, binary cross-entropy was used, which is appropriate for the binary classification task of diabetes prediction. In the convolutional and dense layers, ReLU activation functions were applied, whereas in the final output layer, sigmoid activation was applied. Dropout with a rate of 0.3 is used after dense layers, as well as L2 regularization with a coefficient of 0.0001. Those configurations were applied to avoid overfitting. The model was trained for 60 epochs with a batch size of 32, using early stopping with a patience of 7 epochs to halt training when the validation performance stopped improving. The extracted features are flattened and passed through fully connected layers to generate a feature vector that represents the image data. The GBM model is used to analyse the numerical data. It applies an ensemble learning technique that builds a sequence of decision trees, where each tree corrects the errors of the previous tree. The model assigns different weights to numerical features to optimize the predictive performance.

The feature vectors from both the CNN and GBM models are concatenated to form a unified feature representation. A fully connected layer processes the combined features, enabling the model to learn interactions between image and numerical data. The proposed model is designed to integrate both image and numerical data concurrently for optimal performance. However, it can still function if only one type of data (either numerical or image) is provided; however, with some limitations, if only image data are loaded, the model relies solely on the image processing branch (the CNN). The numerical data processing branch will essentially be bypassed, or its input will be set to default or null values. The classification accuracy may decrease compared with that of the integrated model since the model cannot leverage additional insights from numerical data. It performs similarly to a traditional CNN-based model that only processes images.

When only numerical data are provided, the model uses the fully connected neural network (FCNN) branch to process these inputs. The image processing branch will be inactive, or its input will be omitted. The model functions like a typical neural network designed for tabular data. The prediction accuracy may be inferior to that of the integrated model since the pairwise integrated model has a visual context with retinal images. To keep the model constitutional for only one typology of data, some adaptations can be made — for example, the architecture can be adjusted dynamically on the basis of the typology of data available through conditional logic. One should also note when training the model that it should be able to function when only one part of the data is provided, possibly with some submodels or imputation for missing data. This flexibility can be useful in practice when not all the data may be absorbed. In some instances, for instance, we may only get numerical data, and no imaging data, as we do not have an imaging apparatus, and the other way around.

3.5 Performance Metrics

The quantitative evaluation of the proposed hybrid model is based on some relevant metrics widely used for assessing the performance of machine learning classifiers, in particular in the field of medical image analysis and diagnostic prediction systems. These metrics give complementary insights around the prediction capabilities of the model (accuracy, precision, recall or sensitivity, specificity, and F1 score) so that we do not define ourselves using a unique metric. These metrics tell you how well the classifier worked overall and how well it does on a per-class basis, etc.

Accuracy is the first metric that determines how correct the model is overall by measuring the ratio of correctly classified instances (both positive and negative) to the total number of predictions. Although the most obvious performance measure is usually accuracy, in the case of medical data, where a highly unequal distribution of classes may be observed (i.e., one diagnostic class is very frequent, while the other is very rare), it may not be a good indicator. As a result, accuracy is evaluated in conjunction with various other discriminative metrics.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1).$$

where \(TP \) denotes true positives; (TN \) denotes true negatives; (FP \) denotes false positives; and \(FN \) denotes false negatives.

Precision quantifies the proportion of predicted positive cases that are truly positive, providing a measure of the model's reliability when it asserts the presence of a condition (e.g., diabetic retinopathy). High precision indicates a low rate of false positives, which is critical in healthcare settings to avoid unnecessary interventions or follow-up procedures.

$$Precision = \frac{TP}{TP+FP} \qquad (2).$$

Recall (sensitivity) measures the proportion of actual positive cases that are correctly identified by the model. In clinical diagnostics, a high recall value is essential because it reflects the system's ability to capture as many true cases as possible, thereby reducing the likelihood of missing patients who genuinely require treatment.

$$Recall = \frac{TP}{TP+FN} \qquad (3).$$

The F1 score harmonizes precision and recall into a single measure by computing their harmonic mean. This metric is particularly valuable when there is an uneven distribution of classes, as it balances the trade-off between capturing true positives and avoiding false positives.

$$\mathrm{F1-score} = \frac{Precision * Recall}{Precision+Recall} * 2 \qquad (4).$$

Finally, the performance of the model on the test dataset is evaluated via the receiver operating characteristic curve and the area under the curve. The ROC curve is used to illustrate sensitivity versus specificity, whereas the AUC score provides a single descriptive statistic of the discriminatory performance of the model at different classification threshold levels. The statistic approaches a value of 1.0 when the entire range of classification thresholds is used, whereas below-chance performance is represented by an AUC = 0.5. The test dataset is not used in either the training or the validation sets, making this evaluation unbiased, and cross-validation is used to evaluate the predictive stability and generalizability of the developed model. The evaluation approaches ensure that the final assessment of the hybrid model's performance is based on a range of performance measures, including the hybrid model's classification power, which takes into account its clinical usefulness, reliability, and robustness to data splitting.

The experiments performed in this research were performed via a personal computer with Windows 8 alongside a 2.5 GHz Intel Core i7 processor and 12 GB of RAM, running python language via Kaggle. As delineated in section 4.1, various metrics are used to evaluate the model's predictive capabilities.

To test the model, the holdout technique was employed to obtain an accurate estimate of the generalization error. The whole dataset was split such that 70% of these parts were used for training, whereas the remaining parts were used for validation (15%) and testing (15%).

4. RESULTS AND DISCUSSION

The evaluation of the proposed hybrid machine learning model was carried out with a rigorous, multistage analysis designed to capture its true predictive potential and clinical relevance. Beginning with the preprocessing stage, both retinal image data and numerical patient data were subjected to specialized transformations to maximize the quality of the extracted features. Images were enhanced through resizing, normalization, and noise reduction, ensuring that spatial patterns critical for diabetic detection were preserved without distortion. Numerical features underwent z score normalization to eliminate scale disparities between variables such as glucose level, BMI, and blood pressure. This standardization not only prevented high-magnitude attributes from dominating the learning process but also ensured that the gradient updates during training were more stable and convergent.

Following feature preparation, the dual-branch architecture—a CNN for image features and a GBM-inspired fully connected network for numerical features—was trained independently before their outputs were concatenated into a unified representation. The fusion layer allows the model to learn complex interdependencies between systemic health indicators and localized retinal biomarkers. For only image data, the model achieves lower accuracy, indicating that while image data alone are informative, it benefits significantly from the addition of numerical data. With respect to only numerical data, the average correctness is acceptable. This finding indicates that images are valuable because they facilitate a more precise diagnosis. The performance of the model is consistent, albeit poorer than when both types of data are used. The overall accuracy reached 96%, reflecting that a high proportion of correct predictions across both classes achieved 96% accuracy and an AUC of 0.994, demonstrating that multimodal integration leads to a richer decision boundary and greater discriminatory power.

The classification results for the integrated dataset are shown in Table 3, where both the precision and recall scores remained high across classes, indicating balanced predictive behavior. The nondiabetic class achieved a precision of 0.98 and a recall of 0.96, whereas the diabetic class scored 0.92 precision and 0.97 recall. These results translate into a macroaverage F1 score of 0.96, confirming the model's capacity to reduce both false positives and false negatives—a critical attribute in a clinical screening context where errors carry significant medical consequences.

TABLE III. PRESENT THE PERFORMANCE OF THE MODEL.

Class	Precision	Recall	F1-Score	Support
Not Diabetic	0.98	0.96	0.97	976
Diabetic	0.92	0.97	0.95	524
Accuracy			0.96	1500
macro avg	0.95	0.96	0.96	1500
weighted avg	0.96	0.96	0.96	1500

The training and validation accuracy curves (Figure 3) revealed a steady upwards trajectory with minimal divergence, indicating strong generalization without overfitting. Correspondingly, the training and validation loss curves (Figure 4) both displayed consistent downwards trends, with a narrow gap that reflects an optimal bias–variance trade-off. The confusion matrix (Figure 5) further confirmed this performance: of the 976 nondiabetic individuals, only 42 were misclassified, and of the 524 diabetic patients, 17 were missed. Such a low false-negative count is particularly vital in medical applications, as it minimizes the risk of undetected disease progression.

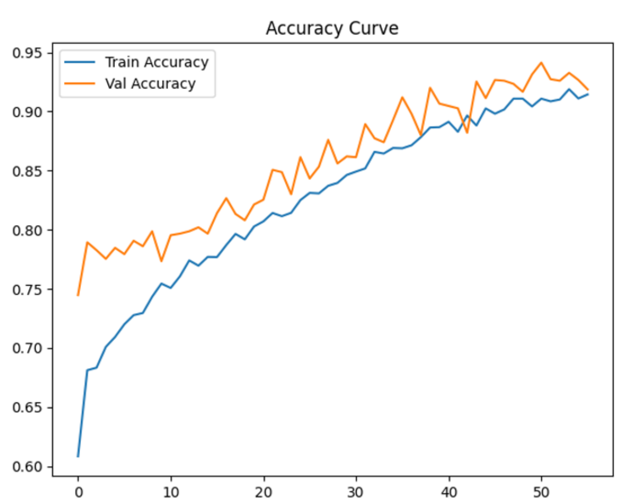

Fig. 3. Training and validation accuracy

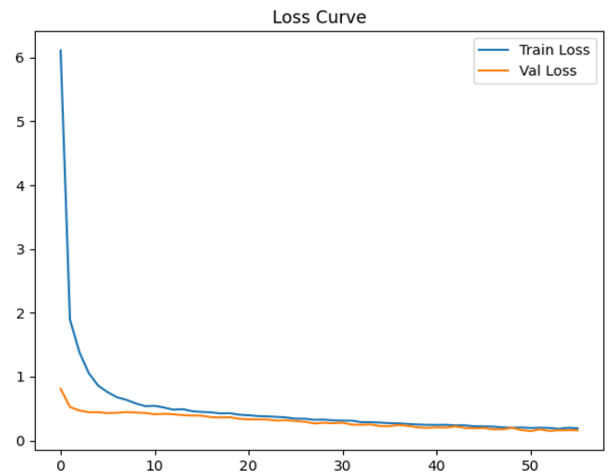

Fig.4. Training and Validation of Loss

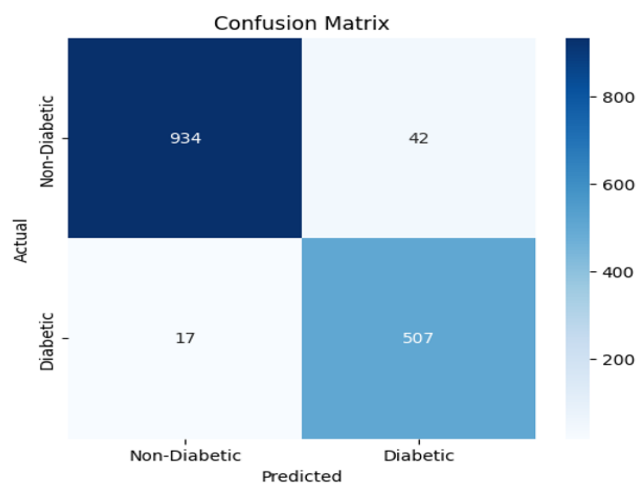

Fig.5: Confusion matrix of the model

The superiority of the hybrid model stems from its ability to capture orthogonal yet complementary data representations. Retinal images reveal microvascular changes, hemorrhages, and exudates—subtle yet powerful predictors of diabetic status—while numerical features quantify systemic factors that may not manifest visually in the early stages. The fusion of these modalities creates a decision space where complex patterns, inaccessible to single-modality models, become discernible. Additionally, the adaptive architecture lends itself to where the model can continue to work (with a minor accuracy penalty) in the absence of one or more modalities, improving the model's robustness to data availability challenges encountered in practice.

These metrics, together with low rates of misclassification and a high area under the curve (AUC), demonstrate the clinical feasibility of the proposed method. We believe its performance regardless of the presence of signs, along with its strong generalization ability, make it a strong candidate for the missing tool in large scale diabetic screening and late disease detection programs, where it could contribute to better patient outcomes through the provision of timely diagnoses.

Compared with existing methods, the performance of our proposed model is competitive with that of state-of-the-art systems. Although only numerical data are used, Bülbül [15]achieves 98.72% accuracy, and Moustari, et al. [25] achieve an AUC of 0.998 when only image data are used, which fails to account for the potential gain from multimodal integration. The proposed framework bridges this gap by achieving state-of-the-art performance while improving the completeness of the reported diagnoses, providing a tool that is both accurate and clinically meaningful. These results confirm that the hybrid model creates an exciting synthesis of the visual and structured data that enables it to detect patterns that would likely be missed in any uni-modality system. The achievements of balanced metric performance, lowest misclassification rates, and high AUC support the promising real-world deployment potential of this novel method as a diabetes screening and early detection tool to promote better-informed clinical decision making for improved diabetes-related patient outcomes.

6. Limitations

Despite the encouraging predictive performance and robustness of the proposed hybrid machine learning framework in providing additional insights into the combination of retinal image features with other numerical patient data towards predicting diabetes, more limitations remain to be overcome.

Finally, this study uses openly available datasets, the Pima Indians Diabetes Database and the APTOS 2019 Blindness Detection dataset, which are well used for research purposes, but are possibly not representative of true clinical populations. This may even vary considerably across health-care systems, given the diversity in demographics, imaging protocols and measurement standards.

Second, the dataset we used for retinal images in this study was preprocessed and augmented for better model learning. These steps introduce greater robustness against overfitting but do not optimally achieve the representation of image quality and variability within real-world clinical imaging environments, where the imaging devices and operator skill vary greatly.

Third, the integration process in this study involves late fusion via concatenation of feature vectors from the CNN and GBM branches. This is a very good approach, but it does not explore more sophisticated fusion strategies, such as attention or transformer-based multimodal frameworks, which may promote more rich interdependencies between modalities.

Fourth, patients were available for evaluation of the model in a static setting (including numerical and image data from all patients). However, in the real world, some data that are either missing or incomplete are common, and an evaluation of the model in such situations should be conducted to understand how robust the model is.

Finally, this study used a binary classification approach (diabetic vs. nondiabetic) without analysing the differences across stages or degrees of disease severity. Leveraging the model for multiclass classification would help gain insights into its clinical utility, specifically in early intervention planning and disease management.

Future research efforts to overcome these limitations may consider testing the model on larger and more diverse patient populations, exploring novel multimodal fusion capabilities with advanced deep learning architectures, and assessing performance under more realistic clinical constraints.

7. CONCLUSION

This study illustrates the accuracy of a hybrid model that uses numerical data alongside images for diabetes prediction. The model was able to leverage both image and numeric data together to produce a mean accuracy of 96% and an AUC-ROC of 0.994, which can be interpreted as an indication of the predictive power and trustworthiness of the model. The model is trained on the image data that zoomed the region of interest which is eye here to check for any anomalies along with blood sugar levels and BMI, and the numerical data helps to deeper analyse the system. The combined analysis of the two data modalities allowed for a more integrated model that can make better predictions compared to single modality models, and improve outcomes also for patients with early or late diabetes.

Further iterations of this model could be developed through advanced feature engineering, attention mechanisms, or the application of additional data modalities including historical or lifestyle information from patients. This is an example of a hybrid model which broadly enhances the positive diagnosis which is one of the biggest challenges in the area of diagnostic error documentation and medical error documentation.

REFERENCES

[1] D. M. Nathan, "Diabetes: advances in diagnosis and treatment," Jama, vol. 314, no. 10, pp. 1052-1062, 2015.

[2] D. J. Magliano and E. J. Boyko, "IDF diabetes atlas," 2022.

[3] R. Deepa and A. Sivasamy, "Advancements in early detection of diabetes and diabetic retinopathy screening using artificial intelligence," AIP Advances, vol. 13, no. 11, 2023.

[4] A. S. Abdalrada, J. Abawajy, T. Al-Quraishi, and S. M. S. Islam, "Prediction of cardiac autonomic neuropathy using a machine learning model in patients with diabetes," Therapeutic advances in endocrinology and metabolism, vol. 13, p. 20420188221086693, 2022.

[5] A. D. A. P. P. Committee, "1. Improving Care and Promoting Health in Populations: Standards of Medical Care in Diabetes—2022," Diabetes Care, vol. 45, no. Supplement_1, pp. S8-S16, 2021, doi: 10.2337/dc22-S001.

[6] Z. Zhang, C. Deng, and Y. M. Paulus, "Advances in structural and functional retinal imaging and biomarkers for early detection of diabetic retinopathy," Biomedicines, vol. 12, no. 7, p. 1405, 2024.

[7] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," NPJ digital medicine, vol. 1, no. 1, p. 39, 2018.

[8] D. A. Kadhim and M. A. Mohammed, "Advanced machine learning models for accurate kidney cancer classification using CT images," Mesopotamian Journal of Big Data, vol. 2025, pp. 1-25, 2025.

[9] A. Abdalrada, A. F. Neamah, and H. Murad, "Predicting diabetes disease occurrence using logistic regression: An early detection approach," Iraqi Journal For Computer Science and Mathematics, vol. 5, no. 1, pp. 160-167, 2024.
[10] M. S. Qadir and G. BİLGİN, "Active learning with Bayesian CNN using the BALD method for hyperspectral image classification," Mesopotamian Journal of Big Data, vol. 2023, pp. 53-60, 2023.
[11] A. S. Abdalrada, "The role of various risk factors in the prevalence of cardiac autonomic neuropathy and associated diseases," Deakin University, 2018.
[12] R. C. Deo, "Machine learning in medicine," Circulation, vol. 132, no. 20, pp. 1920-1930, 2015.
[13] M. S. Yao et al., "SynthA1c: Towards Clinically Interpretable Patient Representations for Diabetes Risk Stratification," in International Workshop on PRedictive Intelligence In MEdicine, 2023: Springer, pp. 46-57.
[14] P. Sampath et al., "Robust diabetic prediction using ensemble machine learning models with synthetic minority over-sampling technique," Scientific Reports, vol. 14, no. 1, p. 28984, 2024.
[15] M. A. Bülbül, "A novel hybrid deep learning model for early stage diabetes risk prediction," The Journal of Supercomputing, pp. 1-23, 2024.
[16] V. Poornima, "A Hybrid Model for Prediction of Diabetes Using Machine Learning Classification Algorithms and Random Projection," Wireless Personal Communications, pp. 1-13, 2024.
[17] 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, p. 337, 2023.
[18] J. Liu, B. Peng, and Z. Yin, "A Hybrid Machine Learning Method for Diabetes Detection based on Unsupervised Clustering," in Proceedings of the 2023 7th International Conference on Machine Learning and Soft Computing, 2023, pp. 144-149.
[19] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," IEEE Access, vol. 8, pp. 76516-76531, 2020.
[20] A. Dutta et al., "Early prediction of diabetes using an ensemble of machine learning models," International Journal of Environmental Research and Public Health, vol. 19, no. 19, p. 12378, 2022.
[21] U. Bhimavarapu and G. Battineni, "Deep learning for the detection and classification of diabetic retinopathy with an improved activation function," in Healthcare, 2022, vol. 11, no. 1: MDPI, p. 97.
[22] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," jama, vol. 316, no. 22, pp. 2402-2410, 2016.
[23] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," Procedia computer science, vol. 90, pp. 200-205, 2016.
[24] B. Graham, "Kaggle diabetic retinopathy detection competition report," University of Warwick, vol. 22, no. 9, 2015.
[25] A. M. Moustari, Y. Brik, B. Attallah, and R. Bouaouina, "Two-stage deep learning classification for diabetic retinopathy using gradient weighted class activation mapping," Automatika, vol. 65, no. 3, pp. 1284-1299, 2024/07/02 2024, doi: 10.1080/00051144.2024.2363692.
[26] C. Lam, D. Yi, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," AMIA summits on translational science proceedings, vol. 2018, p. 147, 2018.
[27] R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," Eye, vol. 32, no. 6, pp. 1138-1144, 2018.
[28] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in Proceedings of the annual symposium on computer application in medical care, 1988, p. 261.
[29] A. Kaggle, "Blindness Detection," ed, 2019.