INSPIRE HEALTH

Inspire Health Journal
ISSN 3110-9438, Vol. 1, No. 3, 2026
doi.org/10.65718/inspireHealth.2026.2012

InspireQuill.org
Inspire Academic Press (Canada)

Research Paper OPEN ACCESS

Calibrated Source-Aware Stacking Ensemble for Early Diabetes Risk Prediction Across Diverse Medical Datasets

Manivir Kaur [1], Jaspreet Kaur [1, *] and Mridula Sharma [1]

[1] Department of Computing Science, Thompson Rivers University, Canada
[*] Corresponding Author: jaspkaur@tru.ca

Abstract

Diabetes mellitus is a prevalent disease that can cause severe complications if undetected or poorly managed, affecting overall health. Diabetes prediction using machine learning is the second most prominent research focus in the field. Early prediction is important for preventive interventions and personalized healthcare. Traditional predictive models typically rely on a single data source, such as clinical, imaging, or genomics, limiting their ability to capture the multifactorial nature of disease while existing ensemble approaches assume homogeneous patient populations. This paper presents a source-aware stacking ensemble pipeline for early diabetes risk prediction that effectively integrates heterogeneous medical datasets and addresses missing patient-level data across modalities. The meta-model incorporates five base models, trained on clinical and genetic factors, retinal images, thermal foot images, and wearable-based stress predictions, one-hot encoded source identifiers and combining calibrated probabilities for final prediction. This enables it to learn from base model outputs while adapting to data source ensuring interpretability and robustness. Among five evaluated meta-classifiers, Logistic Regression performs best. This modular approach allows each model to be optimally trained on its respective dataset while enabling integrated, multimodal predictions, explicitly accounting for dataset heterogeneity. Beyond diabetes, this framework offers scalable solutions for combining diverse factors across unrelated datasets.

Keywords: Diabetes Prediction; Multimodal Data; Diverse Datasets; Ensemble Methods; Source Aware Learning.

1. Introduction

Various research has been conducted for prevention, detection as well as diabetes management. However, there is currently no definitive cure [1]. Early signs of diabetes might go unnoticed for years before diabetes is diagnosed [2]. The disease can be prevented through timely interventions before onset or managed effectively once diagnosed. Previous studies have applied various machine learning techniques using diverse risk factors to predict diabetes. However, single modality approaches alone may not fully capture the complex, multifactorial nature of diabetes [3].

Integrating multiple input factors is therefore necessary to improve prediction. However, to achieve this, it requires per patient datasets that have all the relevant factors which are unlikely to be found easily unless collected specifically for the purpose stated. The key challenge lies in integrating different datasets and different data

Academic Editor:
Jaafar Alghazo

Received: 12/01/2026
Revised: 17/03/2026
Accepted: 07/05/2026
Published: 05/07/2026

Citation
Kaur M., Kaur J., and Sharma M. (2026). Calibrated Source-Aware Stacking Ensemble for Early Diabetes Risk Prediction Across Diverse Medical Datasets. *Inspire health Journal*, 1(3), 130-142.

modalities with no shared patient base where sample level fusion is not feasible. Data fragmentation is a real hurdle, and the traditional multimodal fusion methods cannot be directly used because of the assumption that all data modalities come from the same sample or patient.

Despite these challenges, it is important to allow different types of data to leverage together for better coverage and holistic predictions for diabetes. This paper proposes an approach to combine different factors into a single prediction model when not all patient level factors are available. While the initial goal was to evaluate how the inclusion of diverse risk factors in a prediction model affects the accuracy and robustness of Type 2 diabetes risk evaluation for real-world clinical applications and timely prevention? Along the way, the research question has evolved considering the challenges and the research gap of the use of diverse datasets and data modalities without feature level fusion. The focus shifted on how to combine the heterogeneous data that is not per patient.

This model proposes an approach that uses diverse datasets for prediction of diabetes where each base model is trained on different data modality rather than just using multiple models for the same dataset or multiple models trained on varied datasets of single modality. The primary objective of this study is to demonstrate a source aware meta model pipeline that can integrate heterogeneous data sources even when per patient alignment is not feasible. This is relevant in real world medical applications where data availability is often fragmented across sources.

2. Literature Review

Various models have been developed to predict the risk of Type 2 diabetes (T2D) by incorporating different factors. For example, [4] proposed a prediction model that combines retinal images with traditional risk factors such as HbA1c, age, and gender using deep learning algorithms. Their study, utilizing data from the UK Biobank, demonstrated improvements in risk prediction by including retinal images. However, the factors used in this model do not encompass a wide variety of potential predictors.

Another study, [5] explored the association between C-reactive protein (hs-CRP), an inflammation marker, and the increased risk of T2D. They applied Multivariable Cox proportional hazard regression to data from the China Health and Retirement Longitudinal Study. Another work by [6] has used thermal foot images for diabetes prediction. Similarly, [7] used a logistic regression model and decision tree to predict T2D risk among Pima Indian women, achieving a prediction accuracy of 78.26%. Their model included factors such as glucose, BMI, age, diabetes pedigree function, and pregnancy, but did not incorporate additional risk factors like genetic data, inflammation markers, or stress levels.

[8] incorporated a polygenic risk score alongside factors like age, BMI, and gender for T2D prediction using UK Biobank data. Moreover, [9] and [10] investigated the role of gestational diabetes and postpartum circulating miRNAs in predicting T2D. They found that including these factors alongside traditional ones improved model accuracy. However, like other studies, their models did not incorporate factors like genomic data, stress levels, or retinal images. Another study by [11] emphasized the impact of stress levels on glucose level prediction, which can then be used for predicting T2D by combining it with other factors. It has used data from wearable devices and from clinical experiments to apply machine learning techniques.

Research work done by [12]and by [13] has both used traditional factors such as blood pressure and glucose, among others. The former study employed both parallel and sequential ensemble machine learning methods and achieved 100% classification accuracy, significantly outperforming base algorithms. The latter has compared the performance of the Naive Bayes and k-Nearest Neighbor (KNN) algorithms for predicting diabetes using the Pima Indians dataset. However, these have not considered other major factors like genetics and stress levels.

Although a prediction model by [14], that is using Random Forests based on data from the Pima Indians Diabetes Database, includes blood pressure, glucose, insulin, body mass index, HbA1c, and numerically encoded the risk of retinopathy, neuropathy, and nephropathy, however, this study has not used factors like retinal images, pregnancy, genetic factor, inflammations markers, and stress levels.

While existing models show reasonable accuracy due to the multifactorial complexity of diabetes, they are also limited by a narrow set of predictors and are missing relevant factors influencing T2D risk. The proposed model will fill this gap by incorporating a wide variety of factors.

3. Methodology

This section introduces a source-aware stacking ensemble meta model that takes clinical and genetic factors, retinal images, thermal foot images, and stress prediction from wearables to predict the risk of diabetes. Considering heterogeneous datasets without per patient data available for all the factors, the figure 1 below is the overview diagram that shows how the meta model combines the five base models and the steps in the entire pipeline of the meta model.

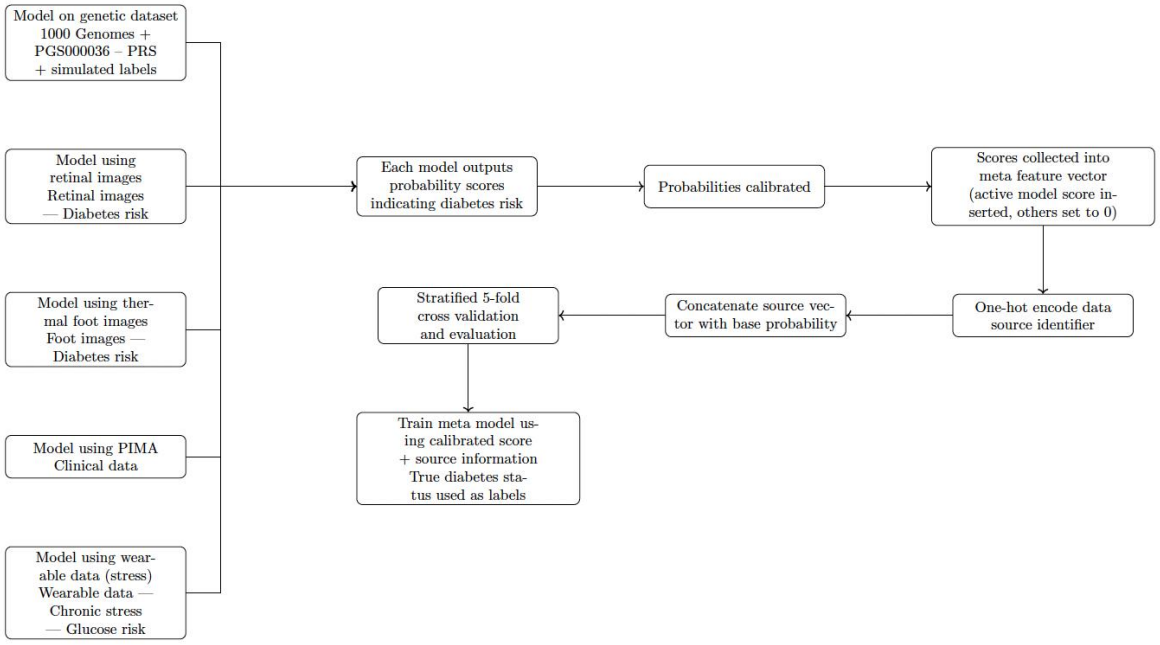

Figure 1. Multi-model diabetes risk prediction pipeline using stacked meta-learning

3.1. Base Models Development

This section includes five base models each trained independently per dataset and the best model per source was chosen. This is done because feature level fusion was not feasible since the five datasets originate from entirely different patient populations, making it impossible to create a unified feature vector. The datasets used in each of the base models are:

3.1.1 Retinal Fundus Image Analysis

Retinal fundus images (RFMiD) [15] dataset was used, preprocessing images to 224×224 pixels for Vision Transformer (ViT) feature extraction. Figure 2(a) below shows a diabetic retinopathy case while Figure 2(b) shows a non-diabetic case. The inherent class imbalance was addressed through computed class weights, while feature standardization ensured consistent input distributions. Logistic Regression, MLP and XGBoost were evaluated, and the best model was chosen based on accuracy as primary metric and F1 score as secondary. After evaluating,

MLP emerged as the optimal performer with 88.44% accuracy and 92.70% F1-score, forming one of the base models for the final meta model.

Figure 2(a). Diabetic retinal image

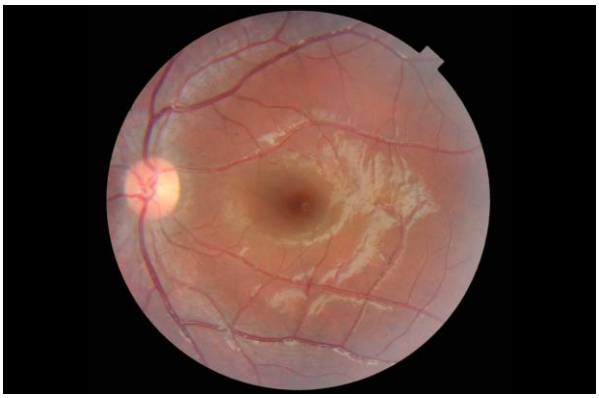

Figure 2(b). Not-diabetic retinal image

3.1.2 Clinical Data

The PIMA Indians Diabetes Database [16] provided traditional clinical markers including glucose levels, BMI, and demographic factors. The missing values were imputed with the median value of each respective feature. Z-score normalization standardized feature scales, and SMOTE addressed class imbalance concerns. The MLP classifier achieved 79.50% accuracy with an 80.00% F1-score, demonstrating reliable performance on established clinical predictors.

3.1.3 Wearable derived Stress Indicators

HUPA-UCM Diabetes dataset [17] presented unique challenges in temporal data processing. Glucose readings were averaged in 15-minute intervals, while heart rate and step data were averaged daily. Glucose and stress features were extracted on a weekly basis and rolling average was calculated for RMSSD, SDNN, resting heart rate and a custom stress index combining HRV, heart rate and physical activity. A week was labeled high risk if the glucose was above 180 mg/dl for more than 25% of the time. SMOTE was used to handle class imbalance. Logistic Regression, Random Forest, MLP, XGBoost, AdaBoost were evaluated using a 5-fold cross validation and the best model was chosen based on harmonic mean of accuracy and F1 score. Despite these, Logistic Regression achieved a harmonic mean of only 41.38%, highlighting the complexity of stress-glucose relationships in wearable data.

3.1.4 Thermal Foot Images

Thermography Images of Diabetic Foot [18] were used and like retinal model, the images were resized to 224*224 pixels and Vision Transformer Model (ViT) was used to extract features. Figure 3(a) below shows the diabetic foot while figure 3(b) is from the control group. Thereafter, features were standardized using feature normalization and the class weights were computed to handle the imbalanced data. Logistic Regression, MLP and XGBoost were evaluated, and the best model was chosen based on accuracy as primary metric and F1 score as secondary. XGBoost emerged as the best model which formed one of the base models for the final meta model. XGBoost accuracy was 87.06% and F1 score was 89.11%.

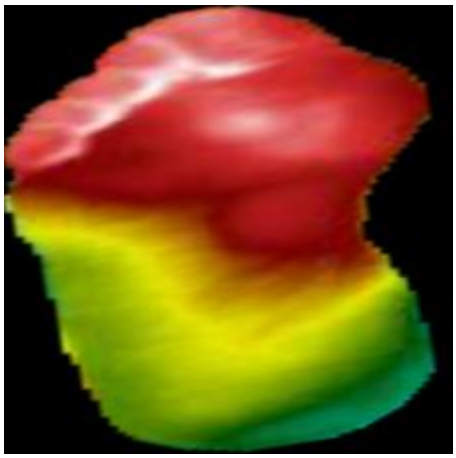

Figure 3(a). Diabetic Foot

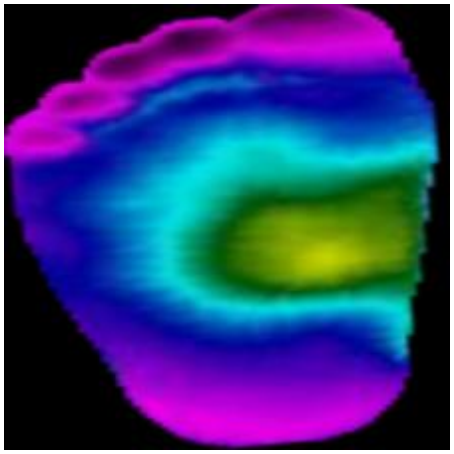

Figure 3(b). Control Group (non-diabetic)

3.1.5 Genetic Risk Assessment

Polygenic risk score (PRS) computation combined 1000 Genomes Project data [19] with PGS000036 [20] to quantify genetic predisposition. Single nucleotide polymorphisms (SNPs) from the genomic dataset were matched against PGS catalog entries to calculate weighted risk scores. Given the absence of corresponding diabetes labels, we generated simulated outcomes using logistic transformation with 10% disease prevalence, reflecting population-level diabetes rates. Logistic Regression, Random Forest, MLP, XGBoost and AdaBoost were evaluated, and the best model was chosen based on harmonic mean of accuracy and F1 score. Logistic Regression achieved a harmonic mean of 33.03%, indicating the challenges in genetic risk prediction without phenotypic validation data.

All five models were included in the meta model ensemble regardless of their individual performance scores. This was done to allow the meta model to learn from the calibrated probabilistic outputs of each source and treat each as a potentially informative signal. These models, though individually weaker, may capture rare or edge-case signals that could be missed by stronger models, and so the probabilistic outputs of weaker models can still provide complementary value. The meta model neither uses manual filtering nor weighting based on base model performance before training the meta model but was expected to learn how to weigh and interpret these signals through calibration and source encoding. The meta-model is not re-learning diabetes prediction from raw data. Instead, it is learning patterns in the confidence levels of expert base models and using source identity to weigh them.

3.2. Calibration

Base model predicted probabilities were calibrated using isotonic calibration and the test dataset for the respective model. The meta model relies on probabilistic outcomes from different base models and if one model is overconfident and the other underconfident, it can mislead the meta model. Therefore, calibration enables the meta model to better learn and interpret those probabilities. For instance, if a model predicts 0.85 risk of diabetes, calibration makes sure that the specific model is confident 85% of the time. This is achieved by calibrating the outputs of the already/pretrained base models and it also makes them comparable across base models. It makes each model's predicted probability a true, reliable estimate of diabetes risk within its own data type. Once each base model's prediction was calibrated, the next step was to inform the meta model about the origin of each prediction. This was achieved by including one-hot encoding as described below.

3.3. Source-aware Encoding

One-hot encoding of data sources explicitly informs the meta-model about prediction origin. This architectural choice eliminates guesswork about data source while enabling the meta-model to learn source specific reliability patterns. For each prediction, we constructed feature vectors where calibrated probabilities occupied positions corresponding to their source models, with remaining slots set to zero. A retinal prediction with 85% confidence would generate the vector [0.0, 0.85, 0.0, 0.0, 0.0], followed by one-hot source encoding [0.0, 1.0, 0.0, 0.0, 0.0]. Even though each sample has only one non-zero-base probability in this model, source-aware one-hot encoding explicitly states that, helping it learns source-specific reliability and adjust its decision accordingly. For instance, the retinal predictions at 75% confidence may prove more reliable than genomic predictions at equivalent confidence levels. This allows the model to generalize better when faced with new data from known sources, improving practical applicability in multimodal diagnoses.

3.4. Meta-Model

This meta model is trained in calibrated outputs of five independently trained base models, each suitable for its specific data modality. The meta dataset that contains feature vectors, including calibrated probabilities, along with the source encoding is split using five stratified folds for cross validation to ensure balanced diabetic/non-diabetic cases, ensuring reduced overfitting and providing a reliable estimate of the meta-model's ability to generalize across all data sources. Thereafter, the meta model is tested using the original labels of the respective base models. While the base models were calibrated using their own test datasets, these datasets were not used to train the base models again or train the meta model directly. However, the meta model is trained in predicted probabilities from the base models that makes it learn from model outputs and not the raw input features. The meta model does not re-predict diabetes but rather learns the reliability patterns of base model outputs and adapts its final prediction based on both the predicted probability and its source. Logistic Regression, Random Forest, Gradient Boosting, XGBoost, and MLP were then tested using the original test labels for the respective datasets.

Model Pipeline: Each of the five saved models were loaded along with their test datasets. The predicted probabilities from the test set from each of the base models were calibrated using isotonic calibration with 'cv = prefit', since all the base models were already trained. The test sets for each of the base models were as follows:

Foot: 85 samples
Retina: 640 samples
Stress: 20 samples
PIMA: 200 samples
Genomics: 501 samples

Test sets from all five base models together were used to form the dataset for the meta model. For each prediction from a model, a feature vector was created where the calibrated probability was placed in the position corresponding to its source model with all other slots set to zero. For instance, for a sample from the PIMA dataset and the predicted probability of 0.84, the initial vector would be [0.0, 0.0, 0.0, 0.84, 0.0]. One-hot encoded vector was appended to the feature vector to explicitly indicate the data source. For the same PIMA sample, the final feature vector would become [0.0, 0.0, 0.0, 0.84, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0] where the last five values indicate the source was PIMA. These feature vectors from all the five datasets together formed the meta dataset for the meta model. The meta dataset was split into train and test sets in the 80:20 ratio. Stratified 5-fold cross-validation was performed on the training set to maintain the balance between diabetic and non-diabetic samples. The held-out test set from the meta dataset was used to evaluate the model using accuracy, F1 score, precision, and recall. The best meta model was chosen based on the harmonic mean of accuracy and F1 score.

4. Experimental Results

Calibration and source-aware encoding are critical components in improving the performance of the meta model. Calibration aligns the predicted probabilities with actual outcome likelihoods, while source-aware encoding informs the meta-model about the origin of each prediction, helping it learns source-specific reliability patterns. To quantify calibration quality, Expected Calibration Error (ECE) was used. ECE measures how well the predicted probabilities from a model reflect the true likelihood of an event. A lower ECE indicates more reliable probability estimates. Below is Table 1 that shows ECE for each of the datasets with and without calibration.

Table 1. ECE (Expected Calibration Error) for datasets

Datasets	**Expected Calibration Error Without Calibration**	**Expected Calibration Error With Calibration**
Foot	0.1618	0.0000
Retina	0.1143	0.0000
Stress	0.1464	0.0000
PIMA	0.1690	0.0000
Genomics	0.3861	0.0000

Meta Model ECE on the test set without calibration was 0.0552 while with calibration it improved to 0.0384. Logistic Regression emerges as the best model based on harmonic means of accuracy and F1 score. This ensures balanced prediction performance and minimizes false negatives which are critical for healthcare. Table 2 below shows the evaluation metrics for the meta model Logistic Regression.

Table 2. Meta Model Logistic Regression Metrics

Meta model – Logistic Regression Metrics	**Value**
Accuracy	84.44%
F1 score	85.19%
Recall	89.61%
Precision	81.22%

This reflects a well-balanced performance, particularly with high recall, which is critical for identifying diabetic cases effectively. Below are the tables and corresponding graphs for meta model comparisons with and without calibration using metrics accuracy, f1 score, recall, and precision. Table 3 and Table 4 shows accuracy and f1-score comparisons using all the five base models respectively.

Table 3. Accuracy comparison across meta models with and without calibration

Model	**Accuracy Without Calibration**	**Accuracy With Calibration**
Meta_GB	0.8195	0.8410
Meta_Logistic	0.8375	0.8444
Meta_MLP	0.8403	0.8396
Meta_RF	0.7905	0.8389
Meta_XGB	0.8161	0.8430

Table 4. F1 Score comparison across meta models with and without calibration

Model	**F1 Score Without Calibration**	**F1 Score With Calibration**
Meta_GB	0.8241	0.8486
Meta_Logistic	0.8449	0.8519
Meta_MLP	0.8473	0.8455
Meta_RF	0.7854	0.8467
Meta_XGB	0.8154	0.8508

The following figure 4 shows the above values for accuracy and f1 score in a graph.

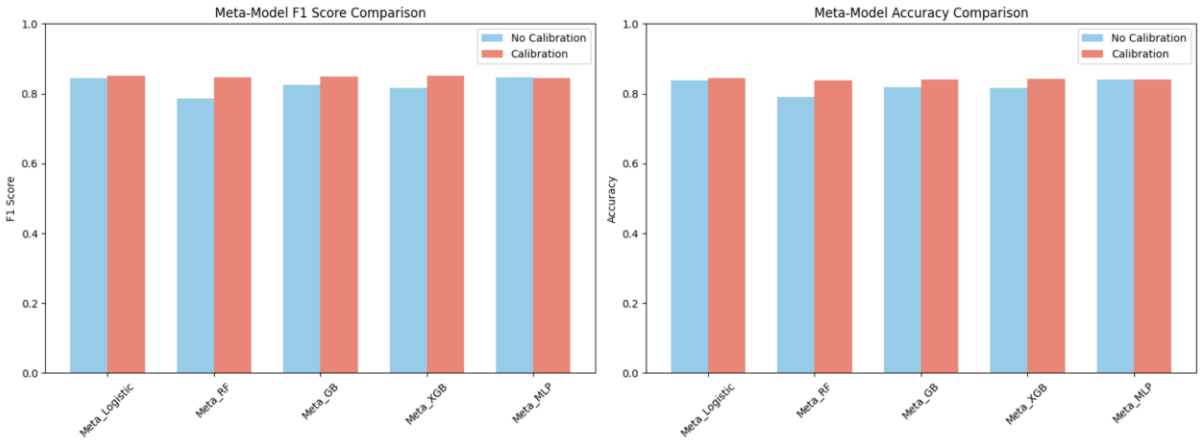

Figure 4. Accuracy and F1 Score comparison across meta models with and without calibration

Precision and Recall Comparisons using all the five base models is shown in Table 5 and Table 6 below:

Table 5. Precision score comparison across meta models with and without calibration

Model	Precision Without Calibration	Precision With Calibration
Meta_GB	0.8036	0.8095
Meta_Logistic	0.8075	0.8122
Meta_MLP	0.8112	0.8146
Meta_RF	0.8034	0.8073
Meta_XGB	0.8180	0.8104

Table 6. Recall comparison across meta models with and without calibration

Model	Recall Without Calibration	Recall With Calibration
Meta_GB	0.8463	0.8920
Meta_Logistic	0.8864	0.8961
Meta_MLP	0.8878	0.8809
Meta_RF	0.7687	0.8906
Meta_XGB	0.8144	0.8961

The following figure 5 shows the above values for precision and recall in a graph.

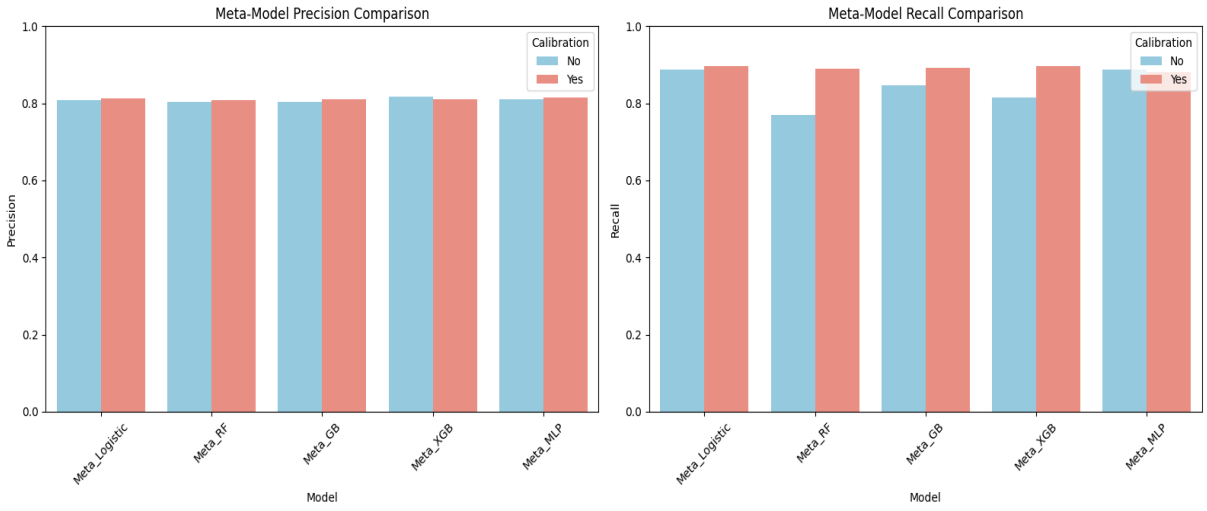

Figure 5. Precision and recall score comparison across meta models with and without calibration

The confusion matrix for Logistic regression (with five base models) is shown below in figure 6.

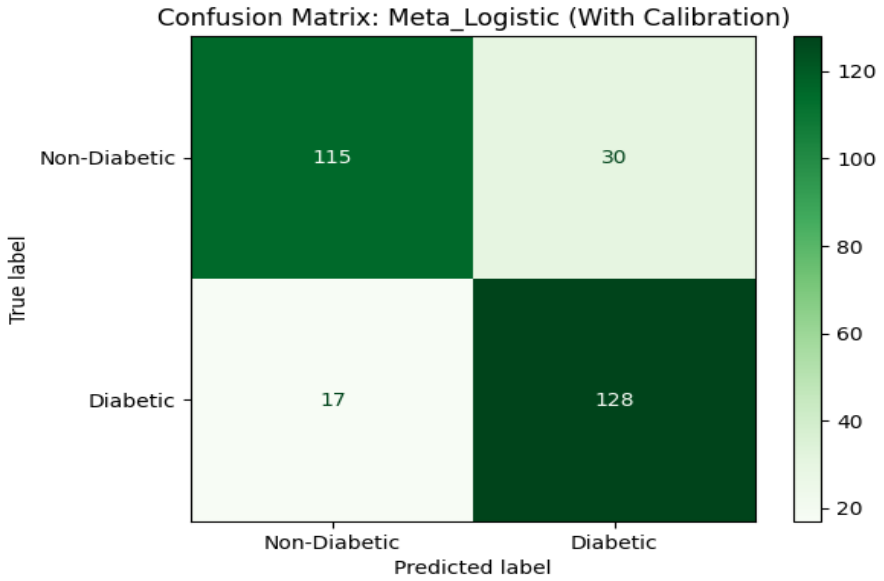

Figure 6. Confusion matrix for best performing meta model (Logistic Regression)

Other insightful results include how the meta model handles the weaker base models when their inclusion to the meta model neither hampers the overall performance nor introduces redundancy but can be considered rare, extreme and meaningful cases. The harmonic mean of accuracy and F1 score for the meta model with all five base models and the meta model with three base models removing the weaker Stress and Genetics model is illustrated below. This shows that even with the inclusion of weaker models, the meta models have slightly higher overall performance across the different ML models. Table 7 below shows the harmonic mean comparison using all five base models and using three base models removing the weaker base models on Stress and Genetics.

Table 7. Harmonic Mean Comparison

Model	**Harmonic Mean (5 Models)**	**Harmonic Mean (3 Models)**
Meta_Logistic	0.8481	0.7788
Meta_RF	0.8428	0.8334
Meta_GB	0.8447	0.8378
Meta_XGB	0.8469	0.8387
Meta_MLP	0.8425	0.8479

Accuracy and F1 Score Comparison graphs using all five base models and using three base models removing the weaker base models on Stress and Genetics are illustrated below in figure 7:

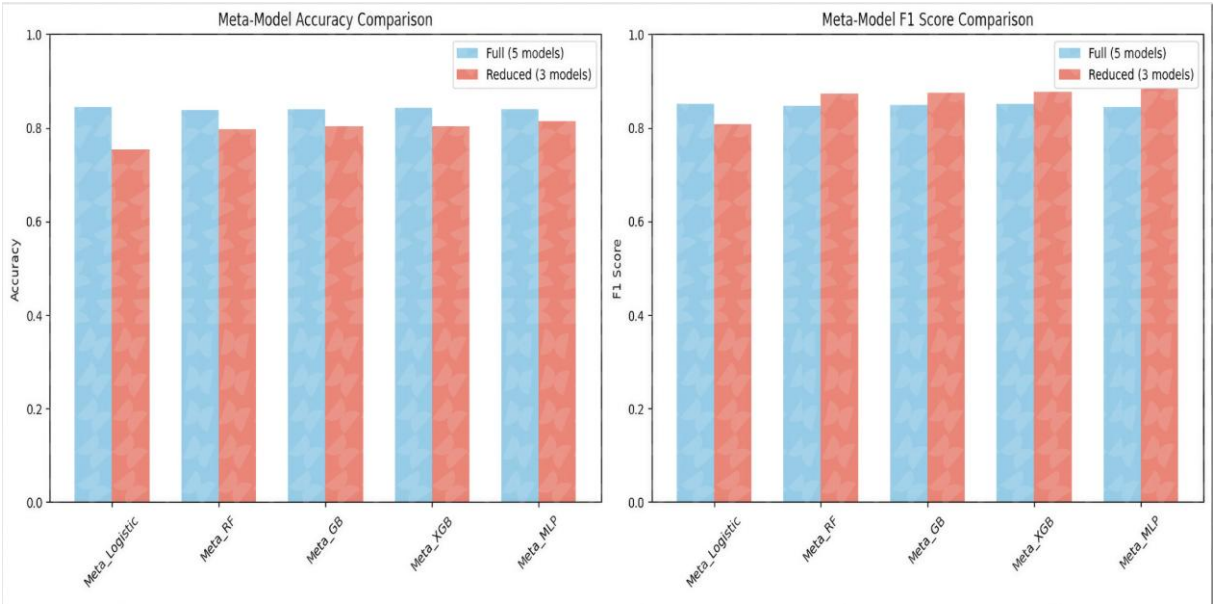

Figure 7. Accuracy and F1 Score comparison for meta models using three base models and five base models

This demonstrates that inclusion of all five models leads to more robust and consistent performance across architectures, validating the idea that even less accurate models can contribute meaningfully when their outputs are interpreted probabilistically.

5. Discussion

This meta-model does not learn from raw input features but from the calibrated confidence scores output by each base model. Crucially, it does not treat all base models as equally reliable. Instead, it learns to trust each source differently based on source-specific prediction patterns, enabling the system to dynamically adjust for varying data quality and signal strength. Even when individual base models perform poorly in isolation, their inclusion does not degrade the overall system. On the contrary, these models can contribute valuable signals in rare or edge cases, improving the overall robustness of the ensemble.

As evident from the results, removing weaker models like those based on stress and genomic data did not lead to any major improvements, and in some cases, slightly reduced performance. This highlights the benefit of including even the low performing models that can provide useful information. This flexibility makes the meta-model more reliable and better suited for real world clinical settings, where data quality and availability often vary across sources. The system is designed to operate seamlessly, with no additional burden on the end user. The inference process is fully modular and source aware. The user only needs to provide input from a known source. This input is processed by the corresponding base model generating a calibrated input. This probability along with source encoding creates the feature vector which the meta model evaluates and produces the final diabetes risk prediction. This pipeline ensures modularity and interpretability, making the system highly adaptable for both current clinical usage and future integration with additional data modalities.

6. Conclusion

This modular approach handles heterogeneous data modalities from diverse sources without requiring feature level fusion. This flexibility allows for easy addition or removal of data sources or data models, making the system adaptable and scalable. By using calibrated probabilities and source-aware encoding, the meta-model learns which

base models to trust and how to interpret their outputs, rather than relying on raw features. This enhances interpretability and robustness, especially in real world healthcare settings where data availability and quality vary across modalities. However, a limitation of this study is the reliance on test sets for calibration, training, and evaluation of the meta-model. Future work could address this by employing proper training/validation splits specifically for calibration and meta-model training, to avoid potential data leakage and to ensure a more rigorous evaluation. This approach has broad applicability beyond diabetes. It could be adapted to predict other complex, multifactorial diseases such as cancer, cardiovascular conditions, and neurodegenerative disorders and particularly in contexts where centralized data sharing is restricted due to privacy or regulatory constraints. Since the model operates on calibrated confidence scores rather than raw data, it respects data boundaries while still enabling integration. Further improvements could include incorporating multiple base models per modality like using different architectures or datasets, enhancing base model performance, especially for lower-performing sources, extending to other domains like remote sensing, public health, and wearable analytics where multimodal, decentralized data is common. Moreover, collecting the per patient data for all the factors could lead to a more powerful and personalized predictive model. This could open the opportunities for comparison between the traditional neural networks using a single unified feature vector and the proposed modular meta model.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The author declares no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1]. Adam, J. (2025, May 9). *The future of diabetes treatment: Is a cure possible?* Labiotech. https://www.labiotech.eu/in-depth/diabetes-treatment-cure-review/

[2]. Diaz-Santana, M. V., O'Brien, K. M., Park, Y. M., Sandler, D. P., & Weinberg, C. R. (2022). Persistence of risk for type 2 diabetes after gestational diabetes mellitus. *Diabetes Care, 45*(4), 864–870.

[3]. Febrian, M. E., Ferdinan, F. X., Sendani, G. P., Suryanigrum, K. M., & Yunanda, R. (2023). Diabetes prediction using supervised machine learning. *Procedia Computer Science, 216*, 21–30.

[4]. Gulshan, N., & Arora, A. S. (2024). Automated prediction of diabetes mellitus using infrared thermal foot images: Recurrent neural network approach. *Biomedical Physics & Engineering Express, 10*(2), 025025.

[5]. HUPA-UCM Dataset. (n.d.). *IEEE DataPort*. https://ieee-dataport.org/documents/hupa-ucm-dataset

[6]. Index of /gbdb/hg19/1000Genomes/phase3. (n.d.). *UCSC Genome Browser*. https://hgdownload.cse.ucsc.edu/gbdb/hg19/1000Genomes/phase3/

[7]. Joglekar, M. V., Wong, W. K. M., Ema, F. K., Georgiou, H. M., Shub, A., Hardikar, A. A., & Lappas, M. (2021). Postpartum circulating microRNA enhances prediction of future type 2 diabetes in women with previous gestational diabetes. *Diabetologia, 64*(7), 1516–1526.

[8]. Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health, 18*(14), 7346.

[9]. Lara-Abelenda, F. J., Chushig-Muzo, D., Wägner, A. M., Tayefi, M., & Soguero-Ruiz, C. (2025). Interpretable and multimodal fusion methodology to predict severe hypoglycemia in adults with type 1 diabetes. *Engineering Applications of Artificial Intelligence, 144*, 110142.

[10]. Liu, W., Zhuang, Z., Wang, W., Huang, T., & Liu, Z. (2021). An improved genome-wide polygenic score model for predicting the risk of type 2 diabetes. *Frontiers in Genetics, 12*.

[11]. Olorunfemi, B. O., Ogunde, A. O., Almogren, A., Adeniyi, A. E., Ajagbe, S. A., Bharany, S., Altameem, A., Rehman, A. U., Mehmood, A., & Hamam, H. (2025). Efficient diagnosis of diabetes mellitus using an improved ensemble method. *Scientific Reports, 15*(1).

[12]. Perry, M. (2001). Early detection of type 2 diabetes: The role of the community nurse. *British Journal of Community Nursing, 6*(4), 176–179.

[13]. PGS Catalog. (n.d.). *Browse polygenic scores (PGS).* https://www.pgscatalog.org/browse/scores/

[14]. PIMA Indians Diabetes Database. (2016, October 6). *Kaggle*. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[15]. Ranganathan, C. S., Nandhini, R., Nisha, K., Sumathi, G., Sudhakar, G., & Srinivasan, C. (2024). *Random forests for predicting diabetes progression and complications* (pp. 1–6). IEEE. https://ieeexplore.ieee.org/xpl/conhome/10774571/proceeding

[16]. Retinal Disease Classification. (2021, August 16). *Kaggle*. https://www.kaggle.com/datasets/andrewmvd/retinal-disease-classification

[17]. Sevil, M., Rashid, M., Hajizadeh, I., Park, M., Quinn, L., & Cinar, A. (2021). Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management. *IEEE Transactions on Biomedical Engineering, 68*(7), 2251–2260.

[18]. Thermography images of diabetic foot. (2022, November 10). *Kaggle*. https://www.kaggle.com/datasets/vuppalaadithyasairam/thermography-images-of-diabetic-foot

[19]. Yang, X., Tao, S., Peng, J., Zhao, J., Li, S., Wu, N., Wen, Y., Xue, Q., Yang, C., & Pan, X. (2021). High sensitivity C-reactive protein and risk of type 2 diabetes: A nationwide cohort study and updated meta-analysis. *Diabetes/Metabolism Research and Reviews, 37*(8).

[20]. Yun, J., Kim, J., Jung, S., Cha, S., Ko, S., Ahn, Y., Won, H., Sohn, K., & Kim, D. (2022). A deep learning model for screening type 2 diabetes from retinal photographs. *Nutrition, Metabolism and Cardiovascular Diseases, 32*(5), 1218–1226.