



Research Article

Prediction of Diabetes Status Using Binary Logistic Regression with Artificial Neural Network

Zaharadeen Abdullahi, Adamu Abubakar and A. U. Shelleng

Department of Mathematical Sciences, Gombe State University, Gombe, Nigeria

*Corresponding author's Email: abumalala@gsu.edu.ng, doi.org/10.55639/607.020100108

ARTICLE INFO:

Keywords:

Ensemble,
Binary-Logistic
Regression,
Artificial-Neural
Network.

ABSTRACT

Diabetes is among the most prevalent non communicable diseases globally, contributing significantly to morbidity, mortality, and healthcare costs, particularly in low- and middle-income countries. Early detection of this condition is essential for effective prevention and management; however, traditional diagnostic approaches are often costly, time consuming, and inaccessible to many populations. This study developed and evaluated predictive models for diabetes using Binary Logistic Regression (BLR), Artificial Neural Networks (ANN), and a weighted Ensemble model that integrates both approaches. The study uses retrospective patient data obtained from Federal Teaching Hospital Gombe and Specialist Hospital Gombe, Nigeria. A total of 3,146 patient records containing demographic, clinical, lifestyle, behavioral and environmental variables were analyzed. Data preprocessing included cleaning, encoding of categorical variables, normalization of continuous variables, and outlier handling. Model performance was evaluated using accuracy, sensitivity, specificity, precision, F1 score, and Area Under the Receiver Operating Characteristic Curve (AUC ROC). Model stability was assessed using fivefold cross validation. Results showed that ANN model achieved the highest accuracy (74.72%) among individual models, while the Ensemble model produced the best overall discrimination with an AUC ROC of 0.792 and an F1 score of 79.08% on the test set. Logistic regression models provided interpretable insights into significant risk factors but were outperformed by ANN and Ensemble models in predictive accuracy. The findings demonstrate that integrating traditional statistical methods with machine learning techniques improves predictive performance and model robustness. The Ensemble modeling approach offers a balanced and effective framework for early detection of diabetes.

Corresponding author: Zaharadeen Abdullahi, **Email:** abumalala@gsu.edu.ng
Department of Mathematical Sciences, Gombe State University, Gombe, Nigeria

INTRODUCTION

Diabetes is among the most common chronic diseases affecting millions of people worldwide. In 2017, approximately 450 million people were diagnosed with diabetes, resulting in around 1.37 million deaths worldwide (Cho *et al.*,

2018). It does not just impact health it also strains healthcare systems and reduce quality of life for those affected. Early prediction is important for effective prevention. If we can predict who is at risk, we can take steps to prevent or manage the condition before it

escalate. Traditionally doctors rely on tests and clinical evaluations to diagnose it, but these methods can be expensive, time consuming, and sometimes inaccessible in regions or community that lack adequate access to healthcare (Algandi *et al.*, 2017). It also strain healthcare systems and reduce quality of life for those affected. Early detection can make a difference, if we can predict who is at risk, we can take steps to prevent or manage the condition before it escalate. Despite the advancements in medical technology, predicting diabetes remains a significant challenge. Traditional diagnostic methods, while effective, are often inaccessible to large portions of the population, particularly in rural areas. This creates a gap in early detection and prevention, leading to unnecessary suffering and healthcare costs, (Fitriyani *et al.*, 2019). By analyzing patient data, we can build models that can predict the likelihood of someone developing diabetes, however, not all models are created equal. Some, like logistic regression, are simple and interpretable but may lack the flexibility to capture complex relationships in the data. Others, like artificial neural networks, are highly accurate but can be difficult to interpret.

By combining binary logistic regression and artificial neural networks to improve prediction accuracy of diabetes, this study will bridge the gap between traditional statistical methods and modern machine learning techniques. The study aim to build predictive models using binary logistic regression, artificial neural networks (MLPNN) and combining both models to improve accuracy in predicting the presence or absence of diabetes by analyzing variables related to demographics, clinical data, lifestyle, behavioral and environmental factors. The objectives was (i) Identify key risk factors using Wald test to determine which variable significantly contribute to the risk of developing diabetics, (ii) Assess and evaluate model performance using performance metrics (accuracy, sensitivity, specificity, F1score, precision and AUC), (iii) Compare model effectiveness by differentiating between individuals (with & without) diabetics, (iv) Evaluate the stability of model predictions using fivefold cross validation and (v) Combine the predicted probabilities of both models using a

weighted average approach to improve prediction accuracy.

Diabetes mellitus, particularly type 2 diabetes (T2D), is a major global health concern, contributing to significant morbidity, mortality, and economic costs. Early prediction and diagnosis of diabetes is critical for implementing preventive measures and reducing complications such as cardiovascular diseases, kidney failure, and vision loss (Joshi & Dhakal, 2021).

Machine learning (ML) and statistical models, have been widely employed to predict diabetes risk factors (Alshammari, 2016). Logistic regression is a traditional yet robust method for binary classification, often performing comparably to more complex ML algorithms (Christodoulou *et al.*, 2019).

Joshi and Dhakal (2021) compared binary logistic regression and classification tree models to predict type 2 diabetes in Pima Indian women. Using the Pima Indian Diabetes dataset, they identified glucose levels, pregnancy frequency, BMI, age, and diabetes pedigree function as the most significant predictors. Despite advancements, challenges remain in diabetes prediction models, including dataset limitations and variability in risk factors across populations (Bashir, Qamar and Khan, 2017). For instance, the Pima Indian dataset used by Joshi and Dhakal (2021) may not generalize to other ethnic groups due to genetic and lifestyle differences. Additionally, missing data handling, such as median imputation, can introduce bias (Joshi & Dhakal, 2021). Future research should incorporate broader datasets with variables like genetic markers, physical activity, and socioeconomic status to enhance model generalizability (Wilson *et al.*, 2007). Another research by Edlitz and Segal (2022) developed simple, interpretable predictive models for Type 2 Diabetes Mellitus (T2D) using logistic regression-based scorecards. Using data from 44,709 non-diabetic participants from the UK Biobank, the authors created two types of models: first a non-laboratory model based on anthropometric features (e.g., BMI, waist to hip ratio), and second a laboratory-based model using four common blood tests (HbA1c, HDL, triglycerides, and GGT). Although previous studies have successfully developed models for predicting diabetes status, but majority of the

models mostly focused on clinical and demographic variables. they often neglect lifestyle, behavioral, and environmental factors that play a significant role in disease risk. Therefore, this study seeks to address this limitation by integrating lifestyle, behavioral, and environmental related variables and ensemble the logistic regression with neural network models to improve the prediction accuracy of diabetes status.

METHODOLOGY

2.1 Source of Data

The data used for this study was, obtained from Federal Teaching Hospital (FTH) Gombe and Specialist Hospital Gombe. The target population include adult with 18 years and above who received screening at both hospitals. The inclusion criteria ensured that only complete record relevant to Diabetes risk factors was taken. Independent Variables are Demographic: Age, sex, marital status, region. Clinical: Fasting blood glucose(FBG), Random blood glucose, Body mass index (BMI), insulin level, cholesterol, family history. LifeStyle and behavioral: Smoking, alcohol, physical activity, diet, sugar intake, salt intake, sleep duration. Environmental: Exposure to environmental risks such as arsenic and pesticide. while Dependent Variables is the presence or absence of the disease (Diabetes Status : 1 = Disease present and 0 = Disease Not Present).

Data Preprocessing

- Data Cleaning: Handle missing values (imputation and removal), used median (numerical) and mode (categorical).
- Encoding: Categorical variables converted using dummy encoding.
- Normalization: Continuous variables (e.g., glucose, BMI and etc.) scaled for ANN input.
- Outliers Detection: Removed and adjusted outliers using interquartile range for preventing skewed model training.

2.2 Binary Logistic Regression

The logistic regression model estimate the probability $P(Y=1)$ (Presence of Diabetes) is given as:

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_kx_k)}}$$

Where,

P = is the probability of disease presence

Y = Binary Outcome variable 0 or 1, (1 = Disease present and 0 = Disease Not Present)

x_1, x_2, \dots, x_k = Predictor variables

β_0 = Intercept

$\beta_1, \beta_2, \dots, \beta_k$ = Coefficient for each predictor (regression Coefficients)

Model Statistics:

- i. Wald Test for significance of coefficient
It purpose is to test the null hypothesis that a particular parameter or set of parameters is equal to zero or some other specified value.

For coefficient $\beta_1, \beta_2, \dots, \beta_k$.

Null hypothesis (H_0) $\beta_1, \beta_2, \dots, \beta_k = 0$

Alternative hypothesis (H_1) $\beta_1, \beta_2, \dots, \beta_k \neq 0$

$$W = \left(\frac{\beta}{SE(\beta)} \right)^2$$

Where,

W = Wald Test

β = the estimate coefficient

$SE(\beta)$ = Standard error of the coefficient

If the Wald Test gives a small p value (any p value < 0.05) we reject the null hypothesis and conclude that the coefficient is statistically significance

- ii. Variance Inflation Factor (VIF) for multicollinearity

Variance Inflation Factor (VIF) is a statistical measure used to detect multicollinearity in a multiple regression model.

Multicollineraity occurs when two or more predictor variables are highly correlated, leading to unreliable or unstable estimates of regression coefficients.

$$VIF = \frac{1}{1 - R_i^2}$$

Where,

R^2_i is the coefficient of determination from regressing the i th independent variable on all the other predictors.

(VIF) Interpretation:

1 = No multicollinearity

1 to 5 = Moderate multicollinearity (Acceptable)

> 5 Potentially Problematic multicollinearity

> 10 Serious multicollinearity

2.3 Artificial Neural Network

A multilayer perceptron neural network (MLPNN) was build.

The model includes:

- Input layer: Number of neurons (nodes) = Number of predictors. Each nodes represent an input variables (e.g., Age, sex, ...)
- Hidden layers: one layer with 10 neurons (nodes), Optimized by trial
- Output layer: Single neuron with sigmoid activation for binary output (1 = Disease present and 0 = Disease Not Present)

2.3.1 Artificial Neural Network (ANN) Diagram

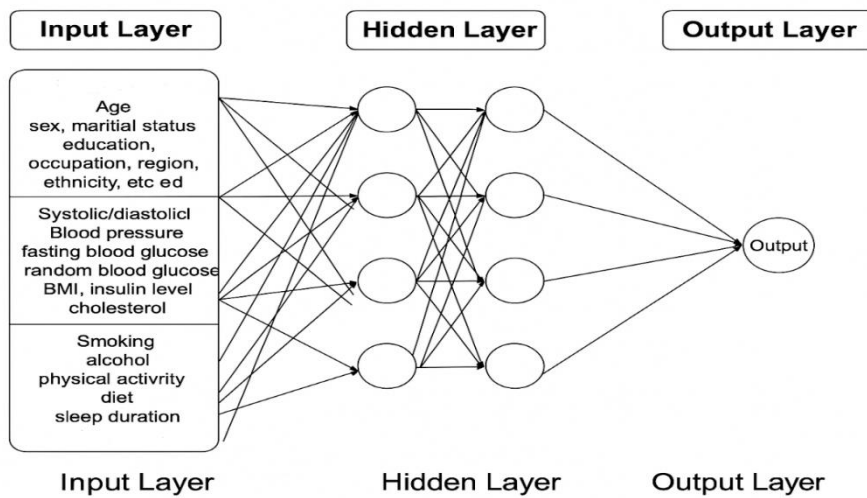


Figure 1: ANN Diagram

2.4 Weighted Average Model Combination

A weighted average method was used by combining the predicted probabilities from the binary logistic regression and artificial neural network models. Predicted probabilities of

- $W_{(BLR)} = \frac{\text{Accuracy}(BLR)}{\text{Accuracy}(BLR) + \text{Accuracy}(ANN)}$
- $W_{(ANN)} = \frac{\text{Accuracy}(ANN)}{\text{Accuracy}(ANN) + \text{Accuracy}(BLR)}$

$$P_{(Ensemble)} = W_{(BLR)} \times P_{(BLR)} + W_{(ANN)} \times P_{(ANN)}$$

Where,

$P_{(Ensemble)}$ is Final probability from the ensemble model

$P_{(BLR)}$ is Probability from binary logistic regression

$W_{(BLR)}$ is Weight assigned to binary logistic regression

$P_{(ANN)}$ is Probability from artificial neural network

diabetes were generated for the testing dataset from both models. These probabilities were combine in to a single, integrated prediction using a weighted average formula. The weights were determined based on the individual model performances on the validation data.

Weight Calculation,

$W_{(ANN)}$ is Weight assigned to artificial neural network

2.5 Model Evaluation

2.5.1 Data Splitting Strategy

To ensure robust model evaluation and prevent overfitting, the dataset was partitioned into training (60%) validation (20%) and testing (20%). using a stratified random sampling approach that maintained the original class distribution in each subset and both Artificial neural network and Binary logistic regression models were evaluated using the performance metrics:

$$\text{Accuracy: } \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity (recall): } \frac{TP}{TP+FN}$$

$$\text{Specificity: } \frac{TN}{TN+FP}$$

$$\text{Precision: } \frac{TP}{TP+FP}$$

$$\text{F1score: } \frac{2 \times (\text{Precision} \times \text{Sensitivity})}{(\text{Precision} + \text{Sensitivity})}$$

AUCROC: Area under receiver operating characteristics curve

2.5.2 Confusion matrix

A confusion matrix is a table used to evaluate the performance of a binary classification model by showing how well the predicted classes match the actual classes.

Table 1: Table of Confusion matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

2.5.2.1 Definitions:

True Positive (TP): Model predicted yes and it was yes (the number of cases where the model correctly predicted the positive class)

True Negative (TN): Model predicted no and it was no (the number of cases where the model correctly predicted the negative class)

False Positive (FP): Model predicted yes but it was no (the number of cases where the model incorrectly predicted the positive class when it was actually negative)

False Negative (FN): Model predicted no but it was yes (the number of cases where the model

incorrectly predicted the negative class when it was actually positive)

2.5.3 Model Comparison

Artificial neural network and Binary logistic regression models were compare using the above metrics

2.5.4 Model Stability Assessment

To evaluate the stability of model prediction the following strategies were used

- Cross Validation: was performed using Five Fold cross validation the data sets were divided into five multiple subsets

where the models were trained and evaluated five times.

the performance of the models and ensemble models for diabetes prediction is presented below;

III. RESULTS AND DISCUSSIONS

The results of the predictive modeling analysis conducted to identify key risk factors, compare

Table 2: Correlation Matrix For Selected Numeric Predictors (Top 5×5)

Variable	Age	Systolic BP	Diastolic BP	Fasting Glucose	Life Style
Age	1.00	0.08	0.03	0.07	0.07
Systolic BP	0.08	1.00	0.01	0.05	0.07
Diastolic BP	0.03	0.01	1.00	0.09	0.00
Fasting Glucose	0.07	0.05	0.09	1.00	0.03
Life Style	0.07	0.07	0.00	0.03	1.00

The correlation analysis revealed generally weak to moderate relationships among the numeric predictors, suggesting minimal multicollinearity concerns. Age showed weak positive correlations with systolic blood pressure ($r = 0.08$) and fasting glucose ($r = 0.07$), consistent

with the known age related increases in these clinical parameters. The absence of strong correlations ($r > 0.7$) between predictors suggests that each variable contributes unique information to the prediction models, which is favorable for model development

Table 3: Data Partitioning for Model Development and Evaluation

Dataset	Training Set	Validation Set	Test Set	Total
Sample Size	1,888 (60%)	629 (20%)	629 (20%)	3,146
Positive Cases	1,102	368	367	1,837

The 60 20 20 split was employed to allocate sufficient data for model training while reserving adequate samples for validation (hyper parameter tuning and model selection) and final testing (unbiased performance assessment). The

stratified sampling ensured that each subset maintained approximately the same proportion of positive and negative cases as the original dataset, which is critical for developing models that generalize well to new da

Table 4: Significant Predictors for Diabetes (Logistic Regression Model)

Predictor	Coefficient (β)	Std. Error	z value	p value	Odds Ratio	95% CI
Intercept	5.359	0.743	7.213	<0.001***	0.005	
Age	0.023	0.004	6.172	<0.001***	1.023	[1.016, 1.031]
Fasting Blood Glucose	0.024	0.002	10.826	<0.001***	1.024	[1.020, 1.028]
Family History (Diabetes)	0.571	0.101	5.657	<0.001***	1.770	[1.453, 2.155]

Note: *** $p < 0.001$

The logistic regression analysis identified three highly significant predictors of diabetes risk. Age demonstrated a significant positive association with diabetes ($\beta = 0.023$, $p < 0.001$), with an odds ratio of 1.023, indicating that each additional year of age increases the odds of diabetes by approximately 2.3%. Fasting blood glucose turned out to a predictor ($\beta = 0.024$, $z = 10.826$, $p < 0.001$), with an odds ratio of 1.024, meaning each 1 mg/dL increase in fasting glucose increases diabetes odds by 2.4%. Family

history of diabetes showed a strong effect ($\beta = 0.571$, $p < 0.001$), with individuals having a positive family history being 1.77 times more likely to have diabetes compared to those without such history.

The model's Variance Inflation Factor (VIF) values ranged from 1.014 to 1.074, all well below the threshold of 5, confirming the absence of problematic multicollinearity. The model achieved an AIC of 2,395, indicating reasonable fit while reducing model complexity.

Table 5: Logistic Regression Performance Metrics (Validation Set)

Metric	Value
Accuracy	64.71%
Sensitivity (Recall)	77.93%
Specificity	45.06%
Precision	67.82%
F1 Score	72.52%
AUC ROC	0.658

On the validation set, the logistic regression model achieved an overall accuracy of 64.71%. The model demonstrated high sensitivity (77.93%), indicating it correctly identified approximately 78% of actual diabetes cases. However, specificity was moderate at 45.06%, suggesting the model had more difficulty correctly classifying non diabetic individuals. The F1 score of 72.52% reflects a reasonable

balance between precision and recall. The AUC ROC of 0.658 indicates fair discriminative ability.

3.1 Artificial Neural Network Model

The neural network architecture consisted of one hidden layer with 10 neurons, using a decay parameter of 0.01 for regularization and trained for 500 iterations.

Table 6: Neural Network Performance Metrics (Validation Set)

Metric	Value
Accuracy	74.72%
Sensitivity (Recall)	76.60%
Specificity	71.94%
Precision	80.22%
F1 Score	78.37%
AUC ROC	0.762

The neural network model outperformed logistic regression across all metrics. With an accuracy of 74.72%, the model showed a 10.01 percentage point improvement over logistic regression. The specificity increased greatly to 71.94% (from 45.06%), indicating better ability to correctly classify negative cases while maintaining high sensitivity (76.60%). The

precision of 80.22% demonstrates that when the model predicted diabetes, it was correct approximately 80% of the time. The AUC ROC of 0.762 represents acceptable discriminative performance.

3.2 Ensemble Model

The ensemble model combines predictions from both logistic regression and neural network

- $W_{(BLR)} = \frac{64.71}{64.71 + 74.72} = 0.464$
- $W_{(ANN)} = \frac{74.72}{74.72 + 64.71} = 0.536$

models using weighted averaging based on their validation performance.

Weight Calculation,

Table 7: Ensemble Model Configuration and Performance

Component	Weight	Individual Accuracy
Logistic Regression	0.464	64.71%
Neural Network	0.536	74.72%

Ensemble Performance Metrics Validation Set	
Accuracy	73.45%
Sensitivity (Recall)	77.66%
Specificity	67.19%
Precision	77.87%
F1 Score	77.76%
AUC ROC	0.767

The ensemble model achieved an accuracy of 73.45%, which exceeded logistic regression (64.71%) but was slightly below the neural network (74.72%). However, the ensemble demonstrated a more balanced performance profile with sensitivity of 77.66% and specificity of 67.19%, indicating better balance between

identifying positive and negative cases. The AUC ROC of 0.767 was marginally higher than the neural network's 0.762, suggesting slightly better overall discrimination ability. The ensemble approach effectively used the interpretability of logistic regression while capturing the complex patterns identified by the neural network.

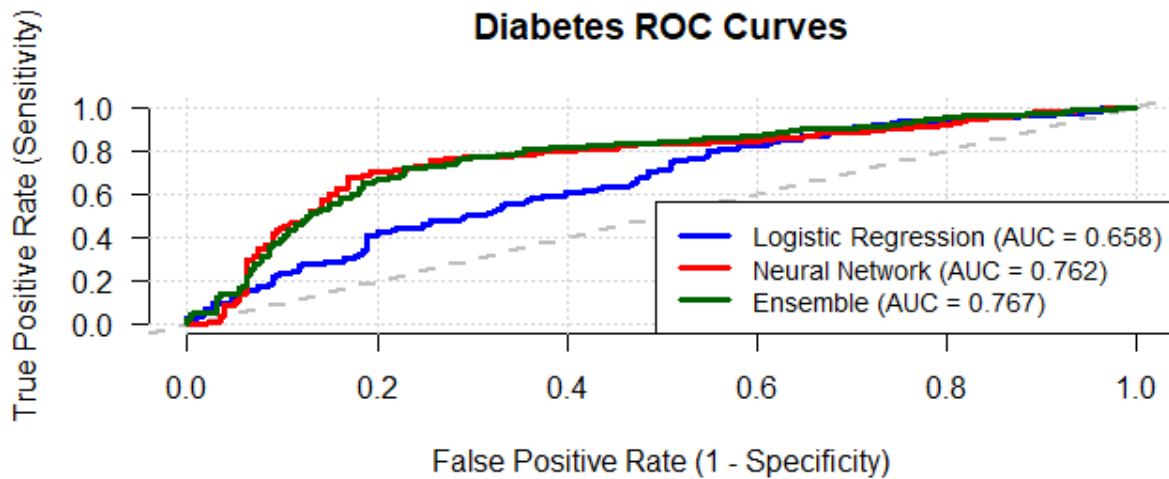


Figure 2: Receiver Operating Characteristic (ROC) Curve Diabetes

Figure 2 Illustrates the balanced between sensitivity and specificity for the diabetes prediction models across different classification thresholds. The curve lying well above the

diagonal reference line indicates strong discriminatory power, while the Area Under the Curve (AUC) confirms that the models perform better than random classification.

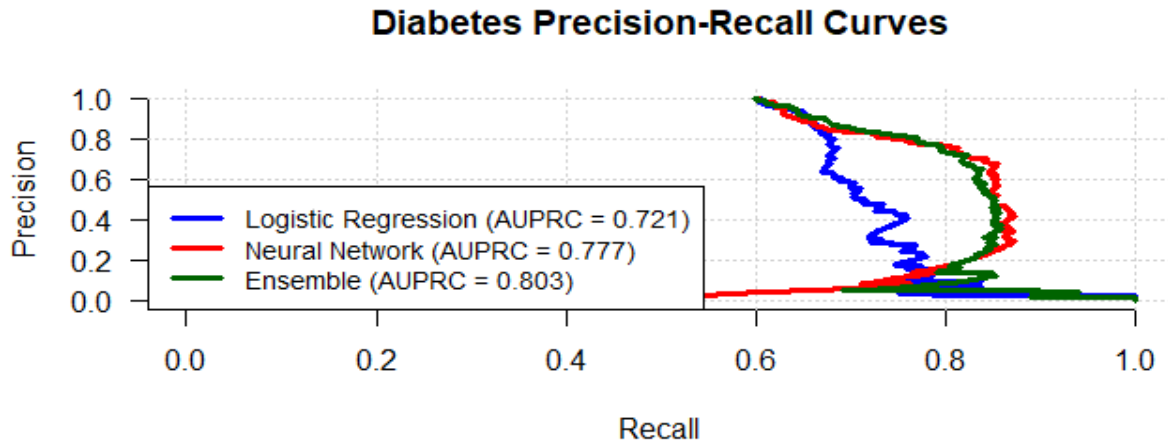


Figure 3: Precision Recall Curve Diabetes

Figure 3 highlights the model’s ability to correctly identify diabetes cases under class imbalance. High precision across increasing recall levels demonstrates that the model

maintains low false positive rates while successfully detecting true positive cases, indicating robustness for screening purposes.

3.3 Model Stability Assessment through Cross Validation

Table 8: Detailed Cross Validation Performance across Five Folds

Fold	Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
1	LR	63.02%	73.38%	46.67%	68.48%	70.85%	0.667
1	NN	71.77%	75.65%	65.64%	77.67%	76.64%	0.749
1	Ensemble	74.35%	79.22%	66.67%	78.96%	79.09%	0.763
2	LR	63.89%	80.14%	41.51%	65.36%	72.00%	0.687
2	NN	71.83%	77.40%	64.15%	74.83%	76.09%	0.763
2	Ensemble	69.25%	79.11%	55.66%	71.08%	74.88%	0.775
3	LR	65.08%	83.62%	40.55%	65.04%	73.17%	0.660
3	NN	70.04%	77.35%	60.37%	72.08%	74.62%	0.736
3	Ensemble	69.84%	81.18%	54.84%	70.39%	75.40%	0.745
4	LR	63.62%	79.78%	43.81%	63.51%	70.72%	0.678
4	NN	69.58%	74.73%	63.27%	71.38%	73.02%	0.710
4	Ensemble	68.39%	76.90%	57.96%	69.16%	72.82%	0.730
5	LR	67.99%	83.93%	47.98%	66.95%	74.48%	0.674
5	NN	72.37%	79.64%	63.23%	73.11%	76.24%	0.745
5	Ensemble	69.38%	81.07%	54.71%	69.21%	74.67%	0.751

The five-fold cross validation results demonstrate stable performance patterns across all folds. Logistic regression showed stable but lower accuracy (63.02 to 67.99%), while neural

networks outperformed with accuracy ranging from 69.58% to 72.37%.

The ensemble model achieved intermediate accuracy (68.39 to 74.35%) with the highest first fold performance (74.35%).

Table 9: Cross Validation Summary Statistics

Metric	Logistic Regression (LR)	Neural Network (NN)	Ensemble
Mean Accuracy (%)	64.72	71.12	70.24
SD Accuracy (%)	1.98	1.23	2.36
Mean Sensitivity (%)	80.17	76.95	79.50
SD Sensitivity (%)	4.25	1.89	1.75
Mean Specificity (%)	44.10	63.33	57.97
SD Specificity (%)	3.20	1.92	5.04
Mean F1 (%)	72.24	75.32	75.37
SD F1 (%)	1.60	1.50	2.30
Mean AUC	0.673	0.741	0.753
SD AUC	0.010	0.020	0.017
CV (%)	3.06	1.73	3.36

The cross validation results demonstrate that all three models shows good stability across the five folds, with generally low standard deviations. The neural network stands out for its superior consistency, achieving the lowest coefficient of variation in accuracy (CV = 1.73%), indicating minimal fluctuation in performance across different data splits. This stability, ensemble with the highest mean specificity (63.33%), suggests that the neural network reliably avoids

false positives while maintaining robust overall performance.

The ensemble model emerges as the strongest performer in terms of discriminative ability, recording the highest mean AUC (0.753) and the highest mean sensitivity (79.50%). These results highlight its stable capability to correctly identify true diabetes cases across folds, making it particularly valuable in clinical contexts where missing positive cases carries high cost.

Table 10: Comprehensive Model Comparison For Diabetes Prediction (Test Set)

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC ROC
Logistic Regression	68.04%	81.17%	46.19%	71.52%	76.04%	0.704
Neural Network	74.72%	76.34%	72.03%	81.97%	79.05%	0.774
Ensemble	73.93%	78.88%	65.68%	79.28%	79.08%	0.792

On the independent test set, the ensemble model achieved the highest AUC ROC (0.792), indicating superior overall discriminative ability. While the neural network had marginally higher accuracy (74.72% vs. 73.93%), the ensemble demonstrated more balanced performance with improved specificity (65.68%) compared to logistic regression (46.19%). The ensemble model F1 score of 79.08% was highest among all models, suggesting perfect balance between precision and recall. The ensemble approach successfully combines the strengths of both

models while reducing their individual weaknesses.

IV. CONCLUSION

The findings of this research demonstrate the practical value of statistical and machine learning approaches for predicting diabetes within Gombe metropolis. Across all models evaluated, meaningful discriminatory performance was observed, as evidenced by Receiver Operating Characteristic (ROC) curves consistently above the line of no discrimination and corresponding Area Under the Curve (AUC)

values exceeding random classification. These results confirm that the joint consideration of demographic, clinical, lifestyle, behavioral and environmental risk factors provides a robust basis for Diabetes prediction. Fasting blood glucose, family history of diabetes, and age emerged as the most influential predictors. The strong association between fasting blood glucose and diabetes risk reinforces its central role in both diagnosis and early risk identification, with incremental increases translating into clinically meaningful rises in disease odds. Similarly, the elevated risk associated with family history shows the genetic and shared environmental components of diabetes, supporting recommendations for targeted screening among high risk households. The observed age effect reflects the progressive nature of insulin resistance and beta cell dysfunction, aligning with global epidemiological trends that show increasing diabetes prevalence with advancing age.

The ensemble modeling approach delivered the balanced and stable performance. By integrating predictions from logistic regression and neural networks, the ensemble model effectively reduced variance and compensated for individual model weaknesses, resulting in superior AUC values and more stable cross validation performance. Evaluation using Precision Recall (PR) curves further strengthened these conclusions. While ROC analysis confirmed overall discrimination, PR curves provided a more informative assessment of model performance under class imbalance, which is common in population health data, therefore ensemble modeling approaches are recommended for deployment in population level diabetes risk assessment, as they provide balanced and stable predictions while reducing misclassification errors. Predictive models such as ANN and Ensemble models should be integrated into routine screening programs to identify individuals at high risk of diabetes. The research limitation was that the models may not be generalized to other ethnic groups due to genetics and lifestyle differences, regularization and cross validation were applied but the possibility of overfitting particularly in the neural network may not be completely ruled out.

REFERENCE

- Alghamdi, M., Al Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., & Sakr, S. (2017). Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. *PLoS ONE*, *12*(7), e0179805. <https://doi.org/10.1371/journal.pone.0179805>
- Alshammari, M. H. S. (2016). *Comparison of artificial neural network and logistic regression models for prediction of diabetes type II with complications* (Doctoral dissertation, Rutgers, The State University of New Jersey).
- Bashir, S., Qamar, U., & Khan, F. H. (2017). A multiclassifier based disease prediction system. *Journal of Medical Systems*, *41*(5), 1–10. <https://doi.org/10.1007/s10916-017-0723-6>
- Centers for Disease Control and Prevention. (2020). *National Diabetes Statistics Report*. U.S. Department of Health and Human Services. <https://www.cdc.gov/diabetes/library/reports/report.html>
- Chen, Q., Hu, H., She, Y., He, Q., Huang, X., Shi, H., Cao, X., Zhang, X., & Xu, Y. (2024). An artificial neural network model for evaluating the risk of hyperuricemia in type 2 diabetes mellitus. *Scientific Reports*, *14*(1), 2197. <https://doi.org/10.1038/s41598-024-52550-1>
- Cho, N., Shaw, J., Karuranga, S., Huang, Y., da Rocha Fernandes, J., Ohlrogge, A., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, *138*, 271–281. <https://doi.org/10.1016/j.diabres.2018.02.023>
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic

- regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22.
<https://doi.org/10.1016/j.jclinepi.2019.02.004>
- Edlitz, Y., & Segal, E. (2022). Prediction of type 2 diabetes mellitus onset using logistic regression based scorecards. *eLife*, *11*, e71862.
<https://doi.org/10.7554/eLife.71862>
- Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2019). Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access*, *7*, 144777–144789.
<https://doi.org/10.1109/ACCESS.2019.2945129>
- 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.
<https://doi.org/10.3390/ijerph18147346>
- Wilson, P. W., Meigs, J. B., Sullivan, L., Fox, C. S., Nathan, D. M., & D'Agostino, R. B. (2007). Prediction of incident diabetes mellitus in middle aged adults: The Framingham Offspring Study. *Archives of Internal Medicine*, *167*(10), 1068–1074.
<https://doi.org/10.1001/archinte.167.10.1068>