DOI: https://doi.org/10.47391/JPMA.41436

Research Article

Predicting diabetic foot amputation using machine learning and statistical models: a prospective cohort study

Syed Muhammad Adnan[1], Samreen Fatima[2]

1 Department of Community Medicine, Dow International Medical College, Dow University of health Sciences, Karachi, Pakistan.

2 Department of Statistics, University of Karachi, Karachi, Pakistan.

Correspondence: Syed Muhammad Adnan. **Email**: syed.adnan@duhs.edu.pk

ORCID ID: 0000-0002-0680-3109

Abstract

Objective: To compare the predictive performance of an artificial neural network model with statistical logistic regression model for predicting the risk of diabetic foot amputation.

Method: The prospective cohort study was conducted at the Dow University Hospital, Karachi, between October 2024 and June 2025 and adult comprised patients with diabetic foot ulcers. The patients were monitored for the occurrence of both minor and major diabetic foot amputations. Demographic, clinical and laboratory data was collected at baseline. Artificial neural network and logistic regression models were created to evaluate the likelihood of diabetic foot amputation. Data were analysed using R version 4.5.0, and model performance was evaluated using appropriate performance measures.

Results: Of the 496 patients, 93(18.75%) underwent diabetic foot amputation; 63(67.74%) males and 30(32.26%) females with mean age 67.36±0.36 years. The remaining 403(81.25%) patients did not undergo diabetic foot amputation; 243(60.30%) males and 160(39.70%) females with mean age 51.29±2.19 years.

Provisionally Accepted For Publication

Among those who underwent amputation, 82(88.17%) had minor amputations and 11(11.83%) had major amputations. Amputation was more common in male patients 63(67.74%) than in female patients 30(32.26%). The artificial neural network ANN model outperformed the logistic regression model with an area under the receiver operating characteristic curve 0.862, accuracy 86.0%, and 95% confidence interval 0.791-0.892 ($p<0.001$). The corresponding values for the logistic regression model were 0.726, 71.2% and 0.651-0.756 ($p=0.042$).

Conclusion: Based on empirical findings, the artificial neural network ANN model was found to be a useful tool for predicting the risk of diabetic foot amputation, outperforming the classic logistic regression model.

Key Words: Diabetic foot, Amputation, Machine learning, Prognosis, Logistic models.

Introduction

Diabetic foot ulcer (DFU) is one of the most severe and debilitating consequences of diabetes mellitus (DM), caused mostly by peripheral neuropathy and vascular damage. DFU, defined by an open wound on the foot, is linked to infection, delayed wound healing, lower-extremity amputation and premature death.[1] Aside from its clinical repercussions, DFU places a significant economic burden on healthcare systems and significantly reduces the quality of life (QoL) of patients and families. The annual healthcare costs associated with DFU have been estimated to range between US$3,368 and US$30,131 per patient, depending on disease severity.[2] DFU develops and progresses by intricate interactions among metabolic, vascular, neurological and viral factors.[3]

Diabetic foot amputation (DFA) remains one of the most serious consequences of DFU. Individuals with long-standing DM account for around 75% of all lower-extremity amputations. A recent systematic review and meta-analysis found that the global yearly incidences of minor and major amputations were 139.97 and 94.82 incidents per 100,000 diabetics, respectively, from 2010 to 2020.[4] Another meta-

Provisionally Accepted For Publication

analysis estimated that 31% of individuals with DFUs required amputation.[5] In patients with healed DFUs, the 5-year cumulative ulcer recurrence and amputation rates are 66% and 12%, respectively.[6] Mortality after amputation is very high. A study in New Zealand found that more than 11% of individuals receiving significant DM-related amputation died within 30 days, and nearly 18% died within 90 days.[7] Subsequently, DFA is associated not only with profound functional disability and reduced QoL, but also with significantly increased mortality.[8]

A number of clinical and demographic characteristics, such as male gender, smoking, elevated body mass index (BMI), hypertension (HTN), cardiovascular disease (CVD), chronic kidney disease (CKD), elevated white blood cell (WBC) count, low haemoglobin (Hb) concentration and hypoalbuminemia, have been repeatedly linked to an increased risk of lower-extremity amputation in patients with DFUs.[9]

In Pakistan, DFA is a significant public health concern. Despite an incidence of DFU similar to that reported globally, reported amputation rates vary from 21% to 48%, with one study reporting that 25% patients with DFUs got lower-limb amputation, and 63% eventually needed limb amputation.[10] Equally concerning is the mortality rate after amputation, which is about 30% in the first year, 50% in three years, and 70% in five years.[11] Poor glycaemic control, long-term DM, neuropathy, retinopathy and nephropathy have all been established as important predictors of amputation.[12] Early diagnosis of patients at high risk of amputation is crucial for prompt intervention, improved limb salvage, and lower healthcare costs.[13]

Conventional statistical approaches, particularly logistic regression (LR), have been widely used to identify risk factors and predict amputation among patients with DFUs. However, these methods assume linear relationships between predictor variables and outcomes, and may therefore fail to capture the complex, non-linear interactions frequently observed in clinical datasets. As a result, their predictive performance may be limited when modelling heterogeneous patient populations.[14]

Recent advances in machine learning (ML) have enabled the development of more sophisticated predictive models capable of learning complex patterns from large

Provisionally Accepted For Publication

clinical datasets. Unlike traditional statistical models, ML algorithms primarily aim at optimising predictive performance without requiring strict distributional assumptions.[15] Growing evidence suggests that ML can substantially improve diagnostic accuracy, risk stratification and prognostic prediction across a wide range of medical conditions. In diabetic foot care, ML techniques have demonstrated promising performance in detecting DFUs, predicting wound healing, identifying patients at high risk of amputation, and supporting personalised clinical decision-making.[16]

Among the ML approaches, artificial neural network (ANN) has shown considerable promise because of its ability to model highly complex and non-linear relationships between clinical variables and outcomes. ANNs have been successfully applied to disease diagnosis, outcome prediction and medical image analysis by identifying intricate patterns within structured and unstructured clinical data.[17] Despite these advances, few studies have directly compared the predictive performance of ANN models with conventional statistical approaches for predicting DFA, particularly in low- and middle-income countries (LMICs).[15-17]

The current study was planned to compare the predictive performance of an ANN model with statistical LR model for predicting DFA risk.

Patients and Methods

The prospective, cohort study was conducted at the Dow University Hospital, Karachi, between October 30, 2024, and June 30, 2025. After approval form the institutional ethics review committee, the sample comprised adult DFU patients of either gender. Patients with pre-existing DFA (minor or major), type 1 diabetes mellitus, pregnancy or other endocrine disorders were excluded. For each participant enrolled after taking informed consent, follow-up time was calculated from the date of enrolment until the earliest occurrence of the study outcome (amputation), the last follow-up visit, loss to follow-up, transfer to another healthcare facility, or the administrative end of the study. Those who did not need to undergo DFA till the end of study were

Provisionally Accepted For Publication

administratively censored on the last day of the study. Individual follow-up times were summed up across all the participants to obtain the total person-time of observation, which was expressed in person-years by dividing the accumulated follow-up time in days by 365.

Baseline demographic, clinical and biochemical data included gender, age, educational level, marital status, monthly income, BMI, glycated haemoglobin (HbA1c), random blood glucose (RBG), fasting blood glucose (FBG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TGs), total cholesterol (TC), blood pressure (BP), serum creatinine (Cr), CVD, HTN, neuropathy, nephropathy, retinopathy, smoking status and physical activity. The participants were prospectively followed until DFA or completion of the study. Cases with missing data were excluded. Minor and major DFAs were defined as amputations below and above the ankle, respectively.

LR and ANN were selected a priori to compare a conventional statistical approach with a non-linear ML approach. LR served as the interpretable statistical benchmark, whereas ANN was selected for its ability to capture complex non-linear relationships and interactions. A uniform set of 13 predefined baseline predictors was entered into both the models to ensure a consistent comparison.

Data was analysed using the R version 4.5.0 focusing on the caret, tidyverse, neuralnet and receiver operating characteristic curves in R (ROCR) packages. Data normality was assessed using the Shapiro-Wilk test. Continuous variables were presented as mean ± standard deviation (SD) or median with interquartile range (IQR), as appropriate, while categorical variables were presented as frequencies and percentages. Comparisons between patients with and without DFA were done using the independent-samples t-test and Pearson's chi-square test, respectively. Least Absolute Shrinkage and Selection Operator Linear Regression (LASSO LR) was used for variable selection before multivariable LR. Subsequently, the pre-specified DFA predictors were entered into the model, with the penalty parameter (λ) selected using 10-fold cross-validation. Predictors with non-zero coefficients were retained for

Provisionally Accepted For Publication

multivariable LR. Multicollinearity was assessed using the variance inflation factor (VIF), and model goodness-of-fit was evaluated using the Hosmer-Lemeshow test. Adjusted odds ratios (AORs) and 95% confidence intervals (CIs) were calculated. P<0.05 was considered significant.

The ANN model the dataset was randomly divided (70:30%) into a training dataset and an independent testing dataset. Continuous variables were normalised using minimum and maximum scaling based on the training dataset, and the similar scaling parameters were used to the testing dataset to prevent data leakage. Categorical variables were binary encoded. The outcome variable was DFA (present=1, absent=0). Predictor variables included age, gender, BMI, HDL-C, HbA1c, serum Cr, CVD, HTN, physical activity, smoking status, neuropathy, nephropathy and retinopathy. The ANN comprised an input layer (13 neurons), single hidden layer and output neuron. The optimal number of hidden neurons was determined using a grid-search strategy, in which candidate network architectures were systematically evaluated using repeated 10-fold cross-validation with two repetitions on the training dataset. The architecture with the best cross-validated predictive performance was selected as the final model. Model parameters were optimised using the back-propagation algorithm with a sigmoid activation function and a classification threshold of 0.50 (Figure 1). The independent testing set was used only for final model evaluation. Model performance was evaluated using accuracy, sensitivity, specificity, area under the ROC curve (AUC), F1-score, positive predictive value (PPV) and negative predictive value (NPV). Variable importance was estimated using the varImp function.[18]

Results

Of the 604 patients screened, 496(82.12%) were included. Of them, 93(18.75%) patients underwent DFA; 63(67.74%) males and 30(32.26%) females with mean age 67.36±0.36 years. The remaining 403(81.25%) patients did not undergo DFA; 243(60.30%) males and 160(39.70%) females with mean age 51.29±2.19 years. Among those who underwent DFA, 82(88.17%) had minor amputations and

Provisionally Accepted For Publication

11(11.83%) had major amputations. Amputation was more common in male patients 63(67.74%) than in female patients 30(32.26%). The median follow-up duration was 6.3 months (IQR: 3.5-7.8 months), with a cumulative follow-up of approximately 307.75 person-years. LASSO LR retained all the 13 candidate predictors with non-zero coefficients; therefore, no variable was excluded before multivariable LR. No substantial multicollinearity was observed, with VIF values ranging 1.08-2.41. The goodness-of-fit indicated adequate model fit (p=0.554). Neuropathy was the strongest independent predictor of DFA (AOR=4.18, 95% CI: 2.37-7.35; p<0.001), followed by HbA1c ≥7%, physical inactivity, nephropathy, serum Cr ≥1.5mg/dL, CVD, BMI ≥25kg/m², retinopathy, low HDL-C, HTN, age, active smoking, and male gender and low average monthly income (p<0.05) (Table 1).

After controlling for potential confounding variables, multivariable LR analysis revealed independent predictors of DFA, which included neuropathy, HbA1c ≥7%, physical inactivity, nephropathy, serum Cr ≥1.5mg/dL, CVD, BMI ≥25kg/m², retinopathy, low HDL-C, HTN, active smoking, male gender and age 50-59 years (Table 2).

According to variable importance analysis, neuropathy was the strong predictor in the ANN model (score=1.00), followed by HbA1c (0.86), peripheral retinopathy (0.78), nephropathy (0.71) and serum Cr (0.62) (Figure 2). Although their relative importance decreased over time, age (0.55), BMI (0.48), HTN (0.43), CVD (0.37), HDL-C (0.29), physical inactivity (0.22), active smoking (0.15) and male gender (0.08) had an impact on model performance, indicating that the main factors influencing the ANN model's predictive performance were poor glycaemic management and diabetic microvascular problems. In every performance indicator, the ANN consistently performed better than the traditional statistical model. Compared to LR, the ANN's prediction accuracy was much higher, and the ANN outperformed LR in terms of sensitivity, specificity, PPV, NPV and F1-score (Table 3). The superior discriminative performance of the ANN model was further validated by AUC. While the LR model achieved an AUC of 0.726,

Provisionally Accepted For Publication

indicating merely acceptable discrimination, the ANN model achieved an AUC of 0.862, showing excellent discrimination (Figure 3).

Discussion

The current study showed that when it came to predicting DFA in patients with DFUs, the ANN model performed noticeably better than the traditional statistical model.[19] In terms of predictive accuracy, discriminative capacity, specificity, sensitivity, PPV, NPV and F1-score, the ANN performed better than LR. These findings suggest that ANNs are better at mimicking the complex, nonlinear interactions between clinical risk variables that lead to DFAs.[20] The improved effectiveness of the ANN model is consistent with prior study on ML techniques for DF care. A recent study found that an ANN model beat conventional statistical approaches in prediction, with an AUC of 0.840 and an overall accuracy of 77.8%.[21] Another study employed ANN to predict both minor and major lower-extremity amputations, with an AUC of 0.820 and 74.9% accuracy.[13] The current findings add to the existing literature by demonstrating even greater predictive accuracy in a prospective cohort of Pakistani patients, underscoring ANN's potential as a clinically useful decision-support tool.

Because of its ability to learn complex patterns from high-dimensional clinical data, ANNs are increasingly being employed in a wide range of healthcare settings, including disease diagnosis, prognosis, survival prediction, medical imaging and cardiovascular risk assessment.[22] Unlike LR, which assumes a linear relationship between predictors and outcomes, ANN models can uncover complex nonlinear interactions without relying on established statistical assumptions. As a result, ANN models usually outperform other prediction models when examining diverse clinical datasets with multiple interacting risk factors.[23] The current study also found several independent predictors of DFA. Neuropathy was the biggest predictor, followed by poor glycaemic control (HbA1c ≥7%), nephropathy, high serum Cr, physical inactivity, retinopathy, CVD, HTN, obesity, active smoking, increasing age and male gender. These findings are biologically plausible and consistent with earlier research,

Provisionally Accepted For Publication

which has shown that diabetic microvascular problems, decreased renal function, cardiovascular comorbidity and poor metabolic management significantly enhance the risk of lower-extremity amputation.[24] Furthermore, the ANN variable importance analysis identified neuropathy, HbA1c, peripheral retinopathy, nephropathy and serum Cr as the most influential variables driving model prediction, highlighting the clinical importance of these established risk factors.[25]

In clinical practice, early risk classification and individualised therapy depend on precise DFA prediction. Patients who have been recognised as high-risk may benefit from increased foot surveillance, interdisciplinary treatment, glycaemic control optimisation, smoking cessation programmes and prompt vascular or surgical procedures. The observed predictive performance of the ANN model suggests its potential future utility for early risk stratification of patients with DFUs. Repeated 10-fold cross-validation with two repetitions within the training set strengthened the architecture-selection process, while the independent 30% testing set provided an unbiased assessment of predictive performance. However, whether its use can improve clinical decision-making, reduce amputations, enhance limb preservation or reduce healthcare costs remains to be established through external validation and prospective clinical utility studies.

In Pakistan, DF disease is still widespread, with high rates of ulcer-related amputations and consequent death. Early identification of high-risk patients is still a major clinical challenge, particularly in healthcare environments with limited resources. According to the current study, ML models may provide a helpful addition to conventional clinical evaluation by improving prediction accuracy without requiring more invasive testing. The strengths of the current study include its prospective cohort design, comprehensive evaluation of clinical, biochemical and demographic risk factors, direct comparison of traditional statistical and ML techniques, and rigorous model evaluation using an independent testing dataset and repeated 10-fold cross-validation. Additionally, variable significance analysis enhanced the interpretability of the ANN model by identifying the most important factors on DFA.

Provisionally Accepted For Publication

The current study also has its limitations. First, the study was conducted at a single tertiary care centre, and, as such, the findings may not be generalizable to other healthcare settings. A single random 70:30 training-testing split may result in performance estimates that are sensitive to sampling variability, although repeated 10-fold cross-validation with two repetitions was used within the training set for architecture selection. Further, the sample size was still too small for ML applications even if subsequent follow-up decreased the likelihood of recollection bias. Also, further assessment of the ANN model's resilience and generalisability is necessary because separate multicentre datasets were not utilised for external validation. Even though variable importance analysis made the model easier to understand, ANN remains a more complicated "black-box" method than LR. Finally, the study only focused on predicting DFA rather than evaluating other clinically significant outcomes such as ulcer healing, ulcer recurrence, mortality, QoL, healthcare utilisation or cost-effectiveness. Future multicentre prospective studies with bigger and more ethnically diverse populations are needed to externally validate the ANN model and assess its therapeutic value. Further research should examine explainable AI methods to improve model transparency and clinician confidence when evaluating the integration of ANN models into traditional DF care pathways.

Conclusion

The ANN model demonstrated significantly superior predictive performance compared to the conventional statistical model for predicting DFA among patients with DFUs. The ANN achieved higher discrimination, predictive accuracy, sensitivity, specificity and overall classification performance, indicating its greater ability to capture the complex interactions among clinical risk factors associated with DFA. Neuropathy, poor glycaemic control, retinopathy, nephropathy and impaired renal function were the most influential predictors identified by the model. The findings demonstrated the potential of ANN-based modelling for early risk stratification of

Provisionally Accepted For Publication

patients with DFUs. However, the model needs external validation before routine implementation in clinical practice can be considered.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: None.

References

1. Wang X, Yuan CX, Xu B, Yu Z. Diabetic foot ulcers: Classification, risk factors and management. World J Diabetes. 2022 Dec 15;13(12):1049-1065. doi: 10.4239/wjd.v13.i12.1049.
2. Lo ZJ, Surendra NK, Saxena A, Car J. Clinical and economic burden of diabetic foot ulcers: A 5-year longitudinal multi-ethnic cohort study from the tropics. Int Wound J. 2021 Jun;18(3):375-386. doi: 10.1111/iwj.13540.
3. Akkus G, Sert M. Diabetic foot ulcers: A devastating complication of diabetes mellitus continues non-stop in spite of new medical treatment modalities. World J Diabetes. 2022 Dec 15;13(12):1106-1121. doi: 10.4239/wjd.v13.i12.1106.
4. Ezzatvar Y, García-Hermoso A. Global estimates of diabetes-related amputations incidence in 2010-2020: A systematic review and meta-analysis. Diabetes Res Clin Pract. 2023 Jan;195:110194. doi: 10.1016/j.diabres.2022.110194
5. Luo Y, Liu C, Li C, Jin M, Pi L, Jin Z. The incidence of lower extremity amputation and its associated risk factors in patients with diabetic foot ulcers: A meta-analysis. Int Wound J. 2024;21(7):e14931. doi: 10.1111/iwj.14931.
6. Hunt DL. Diabetes: foot ulcers and amputations. BMJ Clin Evid. 2011 Aug 26;2011:0602.
7. Walicka M, Raczynska M, Marcinkowska K, Lisicka I, Czaicki A, Wierzba W, et al. Amputations of Lower Limb in Subjects with Diabetes Mellitus: Reasons

Provisionally Accepted For Publication

and 30-Day Mortality. J Diabetes Res. 2021 Jul 24;2021:8866126. doi: 10.1155/2021/8866126

8. Popa AD, Gavril RS, Popa IV, Mihalache L, Gherasim A, Nita G, et al. Survival Prediction in Diabetic Foot Ulcers: A Machine Learning Approach. J Clin Med. 2023 Sep 7;12(18):5816. doi: 10.3390/jcm12185816.
9. Beulens JWJ, Yauw JS, Elders PJM, Feenstra T, Herings R, Slieker RC, et al. Prognostic models for predicting the risk of foot ulcer or amputation in people with type 2 diabetes: a systematic review and external validation study. Diabetologia. 2021 Jul;64(7):1550-1562. doi: 10.1007/s00125-021-05448-w.
10. Akhtar S, Ali A, Ahmad S, Khan MI, Shah S, Hassan F. The prevalence of foot ulcers in diabetic patients in Pakistan: A systematic review and meta-analysis. Front Public Health. 2022 Oct 25;10:1017201. doi: 10.3389/fpubh.2022.1017201.
11. Riaz M, Miyan Z, Waris N, Zaidi SIH, Tahir B, Fawwad A, et al. Impact of multidisciplinary foot care team on outcome of diabetic foot ulcer in term of lower extremity amputation at a tertiary care unit in Karachi, Pakistan. Int Wound J. 2019 Jun;16(3):768-772. doi: 10.1111/iwj.13095.
12. Rout P, Jialal I, Doerr C. Diabetic Nephropathy (Nursing). StatPearls. Treasure Island (FL)2026.
13. Oei CW, Chan YM, Zhang X, Leo KH, Yong E, Chong RC, et al. Risk Prediction of Diabetic Foot Amputation Using Machine Learning and Explainable Artificial Intelligence. J Diabetes Sci Technol. 2025 Jul;19(4):1008-1022. doi: 10.1177/19322968241228606
14. Won SH, Chung CY, Park MS, Lee T, Sung KH, Lee SY, et al. Risk factors associated with amputation-free survival in patient with diabetic foot ulcers. Yonsei Med J. 2014 Sep;55(5):1373-8. doi: 10.3349/ymj.2014.55.5.1373.
15. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of Conventional Statistical Methods with Machine Learning in Medicine:

Provisionally Accepted For Publication

Diagnosis, Drug Development, and Treatment. Medicina (Kaunas). 2020 Sep 8;56(9):455. doi: 10.3390/medicina56090455.

16. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. JAMA Netw Open. 2020 Jan 3;3(1):e1918962. doi: 10.1001/jamanetworkopen.2019.18962.

17. Abhari S, Niakan Kalhori SR, Ebrahimi M, Hasannejadasl H, Garavand A. Artificial Intelligence Applications in Type 2 Diabetes Mellitus Care: Focus on Machine Learning Methods. Healthc Inform Res. 2019 Oct;25(4):248-261. doi: 10.4258/hir.2019.25.4.248.

18. Altmann A, Tolosi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010 May 15;26(10):1340-7. doi: 10.1093/bioinformatics/btq134.

19. Stefanopoulos S, Qiu Q, Ren G, Ahmed A, Osman M, Brunicardi FC, et al. A Machine Learning Model for Prediction of Amputation in Diabetics. J Diabetes Sci Technol. 2024 Jul;18(4):874-881. doi: 10.1177/19322968221142899.

20. Guan H, Wang Y, Niu P, Zhang Y, Zhang Y, Miao R, et al. The role of machine learning in advancing diabetic foot: a review. Front Endocrinol (Lausanne). 2024 Apr 29;15:1325434. doi: 10.3389/fendo.2024.1325434.

21. Tekkesin AI. Artificial Intelligence in Healthcare: Past, Present and Future. Anatol J Cardiol. 2019 Oct;22(Suppl 2):8-9. doi: 10.14744/AnatolJCardiol.2019.28661.

22. Wang L, Li Q, Chen X, Wang Z. Clinical characteristics and risk factors of lower extremity amputation in patients with diabetic foot. Pak J Med Sci. 2022 Nov-Dec;38(8):2253-2258. doi: 10.12669/pjms.38.8.5635.

23. Panesar SS, D'Souza RN, Yeh FC, Fernandez-Miranda JC. Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a

Provisionally Accepted For Publication

Small, Heterogeneous Glioma Database. World Neurosurg X. 2019 Jan 24;2:100012. doi: 10.1016/j.wnsx.2019.100012.

24. Huang S, Xu Y, Yue L, Wei S, Liu L, Gan X, et al. Evaluating the risk of hypertension using an artificial neural network method in rural residents over the age of 35 years in a Chinese area. Hypertens Res. 2010 Jul;33(7):722-6. doi: 10.1038/hr.2010.73.

25. Albagmi FM, Hussain M, Kamal K, Sheikh MF, AlNujaidi HY, Bah S, et al. Predicting Multimorbidity Using Saudi Health Indicators (Sharik) Nationwide Data: Statistical and Machine Learning Approach. Healthcare (Basel). 2023 Jul 31;11(15):2176. doi: 10.3390/healthcare11152176.

Table 1: Patient characteristics (n=496).

Variables	Diabetic Foot Amputation		P value
	Yes	No	
	n=93	n= 403	
Gender Male: Female	63:30	243:160	<0.001
Age (years)	67.36±0.36	51.29±2.19	0.019
BMI (kg/m^2)	31.09±1.12	24.93±2.09	0.036
BP Systolic (mmHg)	165.98±1.36	135.2±4.23	<0.001
BP Diastolic (mmHg)	94.56±1.66	72.23±1.65	0.043
Total Cholesterol (mg/dL)	322.23±2.08	226.29±3.01	0.002
HDL (mg/dL)	28.23±1.39	61.12±2.68	0.013
LDL (mg/dL)	205.56±1.65	146.26±1.89	0.569
Triglyceride (mg/dL)	498.12±1.98	220.05±6.65	0.009
HbA1c (%)	12.98±2.76	6.40±2.56	<0.001
Creatinine (mg/dL)	1.99±1.36	1.07±1.11	0.005
Income (Rupees)	52,562	97,129	0.026
Amputation (minor: major)	82:11	--	--
Exercise (Yes :No)	27:66	311:92	0.045
Active Smokers (Yes: No)	35:58	82:321	0.005
Neuropathy Yes	74 (79.57%)	162 (40.20%)	0.009

Provisionally Accepted For Publication

No	19 (20.43%)	241 (59.80%)	
Nephropathy Yes No	 63 (67.74%) 30 (32.26%)	 96 (23.82%) 307 (76.18%)	 <0.001
Retinopathy Yes No	 52 (55.91%) 41 (44.09%)	 118 (29.28%) 384 (70.72%)	 0.026
CVD Yes No	 39 (41.94%) 54 (58.06%)	 82 (20.35%) 321 (79.65%)	 <0.001

BMI, Body mass index, BP: Blood ressure, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, HbA1c: Glycated haemoglobin, CVD: Cardiovascular disease.

Table 2: Multivariable logistic regression analysis of predictors of diabetic foot amputation (DFA).

Characteristics	Yes (n=93)	No (n=403)	AOR (95% CI)	p-value
Gender				
Male	63 (67.74%)	243 (60.30%)	1.76 (1.08–2.89)	0.024
Female (Ref.)	30 (32.26%)	160 (39.70%)	1	
Age (years)				
30–39	9 (9.68%)	78 (19.35%)	0.84 (0.36–1.96)	0.683
40–49	11 (11.83%)	101 (25.06%)	1.19 (0.58–2.42)	0.632
50–59	31 (33.33%)	131 (32.51%)	2.08 (1.18–3.67)	0.011
≥60 (Ref.)	42 (45.16%)	93 (23.08%)	1	
BMI				
≥25 kg/m²	54 (58.06%)	129 (32.01%)	2.47 (1.49–4.10)	<0.001
<25 kg/m² (Ref.)	39 (41.94%)	274 (67.99%)	1	
HDL				
Low HDL	61 (65.59%)	182 (45.16%)	2.21 (1.34–3.65)	0.002
Normal HDL (Ref.)	32 (34.41%)	221 (54.84%)	1	
HbA1c				
≥7.0%	69 (74.19%)	181 (44.91%)	3.18 (1.88–5.39)	<0.001
<7.0% (Ref.)	24 (25.81%)	222 (55.09%)	1	
Serum Creatinine				
≥1.5 mg/dL	66 (70.97%)	86 (21.34%)	2.69 (1.60–4.52)	<0.001
<1.5 mg/dL (Ref.)	27 (29.03%)	317 (78.66%)	1	
Cardiovascular Disease				
Yes	39 (41.94%)	82 (20.35%)	2.58 (1.53–4.36)	<0.001

Provisionally Accepted For Publication

No (Ref.)	54 (58.06%)	321 (79.65%)	1	
Hypertension				
Yes	58 (62.37%)	193 (47.89%)	2.12 (1.28–3.50)	0.003
No (Ref.)	35 (37.63%)	210 (52.11%)	1	
Physical Activity				
Inactive	66 (70.97%)	177 (43.92%)	2.96 (1.75–5.00)	<0.001
Active (Ref.)	27 (29.03%)	226 (56.08%)	1	
Active Smoking				
Yes	35 (37.63%)	82 (20.35%)	2.04 (1.21–3.45)	0.007
No (Ref.)	58 (62.37%)	321 (79.65%)	1	
Neuropathy				
Yes	74 (79.57%)	162 (40.20%)	4.18 (2.37–7.35)	<0.001
No (Ref.)	19 (20.43%)	241 (59.80%)	1	
Nephropathy				
Yes	63 (67.74%)	96 (23.82%)	2.81 (1.68–4.69)	<0.001
No (Ref.)	30 (32.26%)	307 (76.18%)	1	
Retinopathy				
Yes	52 (55.91%)	118 (29.28%)	2.43 (1.47–4.02)	<0.001
No (Ref.)	41 (44.09%)	384 (70.72%)	1	

BMI, Body mass index, HDL: High-density lipoprotein, HbA1c: Glycated haemoglobin, AOR: Adjusted odds ratio, CI: Confidence interval.

Table 3: Comparison of predictive accuracy of machine learning (ML) model with logistic regression (LR) model.

Model Performance Indicators	Predictive Models	
	ANN	LR
Accuracy (95 % CI, p- value)	0.860 (0.791-0.892), <0.001)	0.712 (0.651-0.756, 0.042)
Sensitivity (95 % CI)	0.825 (0.780-0.859)	0.732 (0.682-0.759)
Specificity (95 % CI)	0.813 (0.772-0.826)	0.719 (0.659-0.739)
PPV (95 % CI)	0.822 (0.786- 0.865)	0.706 (0.651- 0.732)
NPV (95 % CI)	0.809 (0.761- 0.839)	0.669 (0.671- 0.689)
F1 score (95 % CI)	0.852 (0.829-0.891)	0.722 (0.681- 0.742)
AUC	0.862	0.726

ANN: Artificial neural network, AUC: Area under the receiver operating characteristic curve, PPV: Positive predictive value, NPV: Negative predictive value, CI: Confidence interval.

Provisionally Accepted For Publication

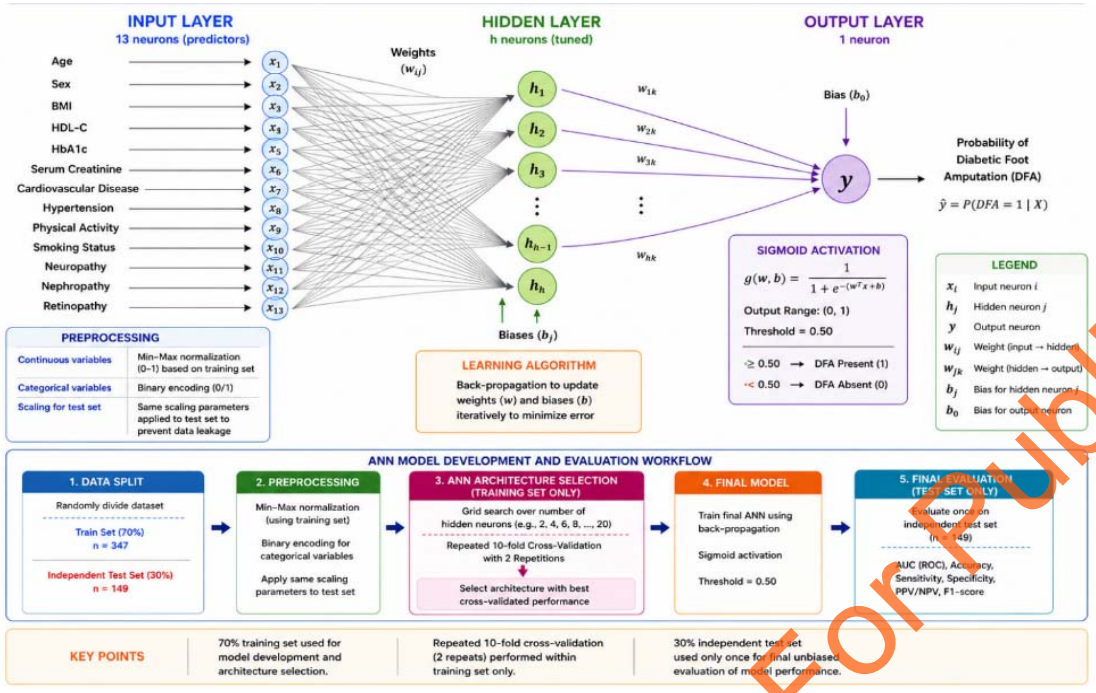

Figure 1: Summarisation of the architecture of artificial neural network (ANN) model.

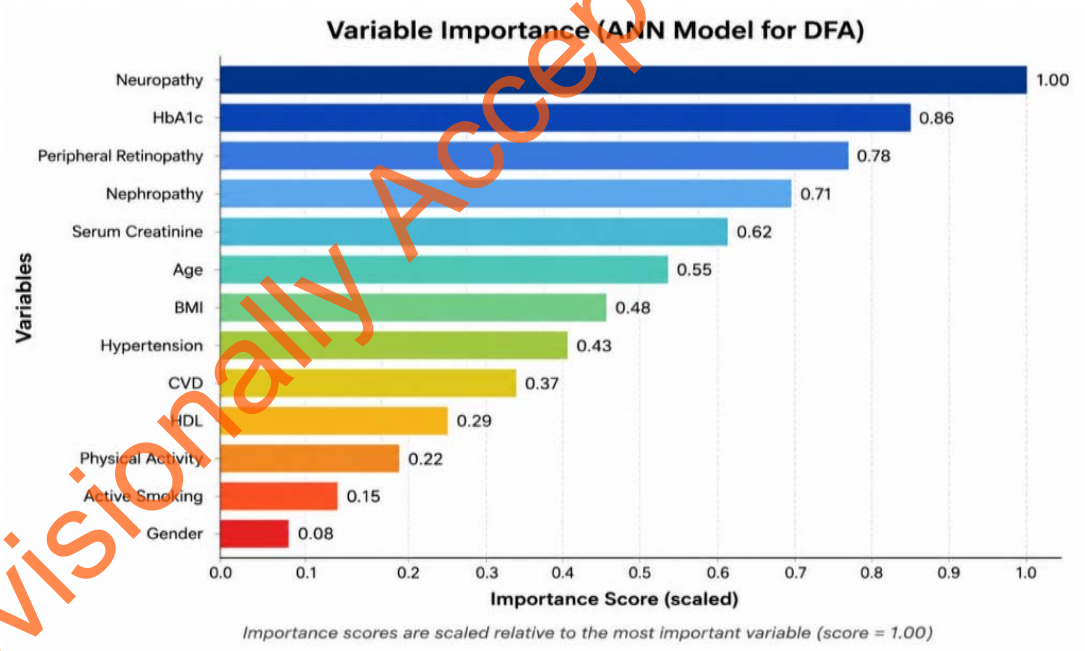

Figure 2: Variable importance for predicting machine learning (ML) artificial neural network (ANN) model.

Provisionally Accepted For Publication

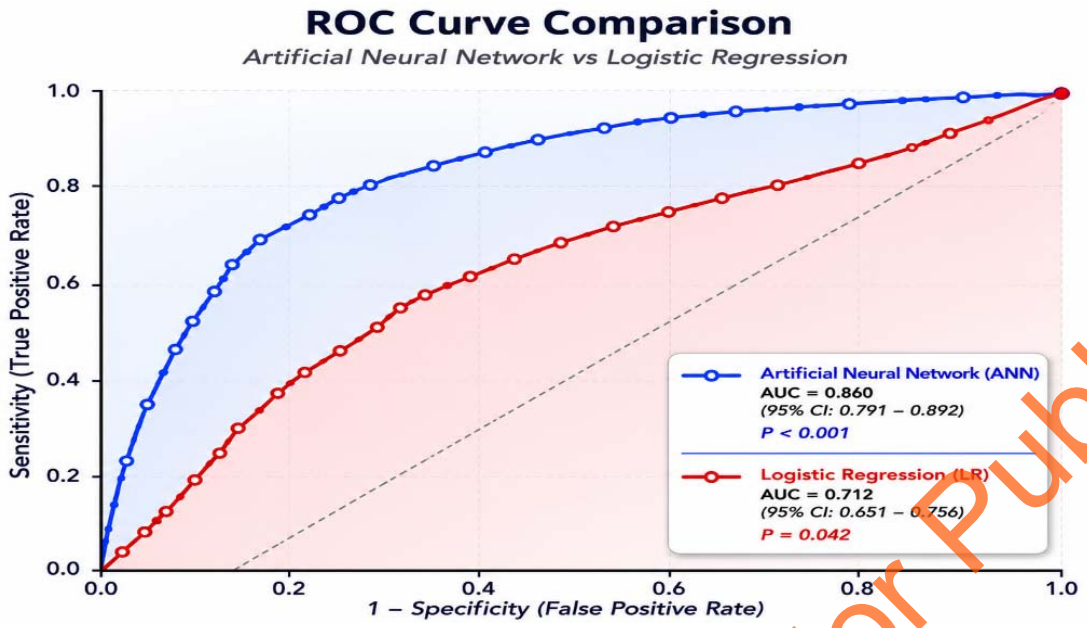

Figure 3: Comparison of accuracy between artificial neural network (ANN) and logistic regression (LR) models.

Submission completed	09-05-2026
1st Revision received	21-07-2026
Last Revision received	20-10-2026
Acceptance	21-10-2026

AUTHORS' CONTRIBUTIONS:

SMA: Concept, design, data acquisition, analysis, interpretation, drafting, revision, final approval and agreement to be accountable for all aspects of the work.

SF: Drafting, revision and final approval.

Provisionally Accepted For Publication