Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng. Vol. 2, No. 4 (2025) 275–285.

Paper Type: Original Article

# A Machine Learning Model for Multi-Level Classification of Diabetic Peripheral Neuropathy Using Clinical, Lifestyle, and Familial Factors

**Mohammad Hosein Amouei[1,*], Mohammad Mehdi Lotfi[1], Nasim Namiranian[2]**

[1] Department of Industrial Engineering, Yazd University, Yazd, Iran; amouie@stu.yazd.ac.ir; lotfi@yazd.ac.ir.

[2] Department of Social Medicine, Faculty of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran; namiranian.nasim@gmail.com.

**Citation:**

| | |
|---|---|
| *Received: 16 May 2025*<br>*Revised: 15 August 2025*<br>*Accepted: 06 October 2025* | Amouei, M. H., Lotfi, M. M., & Namiranian, N. (2025). A machine learning model for multi-level classification of diabetic peripheral neuropathy using clinical, lifestyle, and familial factors. *Annals of healthcare systems engineering, 2*(4), 275-285. |

## Abstract

This study aims to develop a robust Machine Learning (ML) framework for multi-level classification of Diabetic Peripheral Neuropathy (DPN) severity by integrating clinical indicators, lifestyle factors, and familial history in patients with type 2 diabetes. The dataset, collected from the Diabetes Research and Treatment Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran, underwent comprehensive preprocessing including normalization via MinMaxScaler and class rebalancing using the Synthetic Minority Over-sampling Technique (SMOTE). Several ML algorithms — Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Voting Classifier — were implemented and systematically compared. Among these, the RF model achieved the best performance with an accuracy of 86.1%, demonstrating superior stability and interpretability, closely followed by XGBoost. Feature engineering and the incorporation of clinically meaningful composite indices significantly enhanced model performance by capturing complex relationships among physiological and lifestyle variables. Model evaluation based on accuracy, sensitivity, specificity, F1-score, and ROC-AUC confirmed both predictive reliability and clinical applicability. To further enhance interpretability, SHapley Additive exPlanations (SHAP) analysis was conducted using the XGBoost framework due to its higher compatibility with gradient-based explanation methods. The SHAP results confirmed the consistency of feature importance observed in RF, revealing that lower Mean Reflex values, reduced vibration sensitivity (Tuning Fork Test), and higher Body Mass Index (BMI) were strongly associated with severe neuropathy levels. These findings highlight that combining predictive modeling with explainable Artificial Intelligence (AI) approaches can provide transparent, clinically interpretable insights — paving the way for intelligent, explainable decision-support systems in diabetic care.

**Keywords:** Diabetes, Diabetic peripheral neuropathy, Machine learning, Clinical indicators, Lifestyle factors, Family history.

## 1 | Introduction

Diabetic Peripheral Neuropathy (DPN) is one of the most prevalent and debilitating chronic complications of diabetes, significantly impairing patients' quality of life. It results from progressive damage to peripheral nerves, manifesting as pain, numbness, tingling, and muscle weakness—symptoms that typically begin in the

Corresponding Author: amouie@stu.yazd.ac.ir

https://doi.org/10.22105/ahse.v2i4.56

lower extremities. The condition is often asymptomatic in its early stages, making timely detection and intervention particularly challenging. With the global prevalence of diabetes continuing to rise, DPN has emerged as a major public health concern, contributing substantially to disability, hospitalization, and healthcare costs [1].

Early and accurate assessment of DPN severity is therefore essential for preventing disease progression and improving long-term outcomes. However, traditional diagnostic methods such as clinical examinations, reflex testing, and electrophysiological assessments are often subjective, time-consuming, and limited in their ability to detect subclinical neuropathic changes [2]. These limitations underscore the urgent need for data-driven and automated approaches capable of capturing complex, non-linear relationships among diverse clinical and behavioral factors.

In this context, Machine Learning (ML) has emerged as a transformative tool in medical data analysis. ML algorithms can uncover hidden patterns in large, heterogeneous datasets, enabling more precise prediction, classification, and risk stratification compared to conventional statistical models [3]. Recent advances in supervised learning methods, particularly ensemble-based algorithms such as Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), have demonstrated remarkable performance in diagnosing and predicting diabetes-related complications [4].

Moreover, growing evidence suggests that combining clinical indicators (e.g., reflex tests, Body Mass Index (BMI)), lifestyle behaviors (e.g., physical activity, education), and familial risk factors can significantly improve predictive accuracy for DPN severity [5]. Yet, the optimal integration of these diverse data sources within an interpretable and clinically applicable framework remains an open research challenge.

The present study seeks to address this gap by developing and evaluating a multi-level classification framework for predicting DPN severity using a combination of clinical, lifestyle, and familial data. By systematically comparing multiple ML models namely RF, XGBoost, LightGBM, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Voting Classifier—this research aims to identify the most effective and generalizable algorithm for DPN classification. Furthermore, the study introduces feature engineering strategies to create composite indices that better represent the physiological and behavioral dimensions of diabetic neuropathy.

Ultimately, this study aims to contribute to the development of intelligent Clinical Decision Support Systems (CDSS) for early identification and monitoring of patients at risk for DPN, bridging the gap between data science and practical diabetes management.

## 1.1|Problem Statement and Motivation

DPN poses a major challenge in diabetes management, affecting a substantial proportion of individuals with long-standing diabetes. Although early detection can prevent irreversible nerve damage, current diagnostic approaches often fail to identify the disease during its subclinical or early symptomatic stages [6]. Traditional assessments rely heavily on subjective clinical judgment, which limits reproducibility and scalability across patient populations [7].

ML offers a promising alternative by leveraging complex datasets encompassing clinical, physiological, lifestyle, and genetic dimensions. These algorithms can learn intricate, non-linear relationships and interactions that are beyond the reach of traditional regression-based analyses [8]. By incorporating diverse patient information—including reflex assessments, vibration tests, BMI, educational level, and family history—ML models can provide a more holistic and accurate prediction of DPN severity.

This research, therefore, investigates how ensemble-based ML algorithms (specifically RF and XGBoost) can improve the precision and robustness of DPN classification. The motivation stems from the need to translate data-driven insights into clinically interpretable models that can assist healthcare professionals in making timely, evidence-based decisions. Achieving this goal would mark a significant step toward early diagnosis, personalized intervention, and prevention of diabetic neuropathy and its associated complications.

# 2|Previous Studies on Machine Learning Approaches for Diabetic Peripheral Neuropathy

DPN is among the most prevalent and disabling complications of diabetes mellitus, affecting a substantial proportion of patients worldwide. It manifests through symptoms such as pain, numbness, muscle weakness, and sensory loss; in advanced stages, it may progress to foot ulcers and lower-limb amputations, imposing severe clinical and socioeconomic burdens. With the global rise in diabetes prevalence, the early identification, classification, and management of DPN have become critical priorities in diabetic care.

In recent years, ML techniques have gained growing attention for their potential to enhance diagnostic accuracy and support clinical decision-making in DPN management. These models can analyze complex medical datasets to uncover subtle, non-linear relationships between variables, enabling earlier and more precise detection of neuropathic complications.

For instance, Yu et al. [9] demonstrated that ML-based systems could accurately predict Diabetic Foot Ulcers (DFUs)—one of the most severe manifestations of DPN—thus supporting preventive interventions. Similarly, Haque et al. [10] reported that traditional ML algorithms such as SVM and RF markedly improved diagnostic precision compared to conventional statistical approaches, underscoring the practical value of ML models in clinical contexts.

## 2.1|Early Detection and Prediction of Diabetic Neuropathy

Accurate early prediction of DPN onset and progression remains a key challenge in diabetic care, as many patients remain asymptomatic until substantial nerve damage occurs. ML approaches have shown great promise in addressing this issue by integrating clinical, sensory, and biometric indicators to identify patients at risk before irreversible damage develops.

Ma et al. [7] conducted a systematic review of ML-based methods for DPN prediction and emphasized that incorporating biometric and clinical measurements—such as reflex tests, vibration assessments, and metabolic markers—significantly enhances model performance. Similarly, Sheikh et al. [11] utilized plantar pressure distribution data in combination with advanced ML algorithms to predict peripheral neuropathy with high accuracy, illustrating the value of integrating physiological signals with clinical data.

Collectively, these studies highlight the capability of ML techniques to detect neuropathic abnormalities earlier than traditional screening methods, enabling timely intervention and improved patient outcomes.

## 2.2|Machine Learning for Predicting Diabetic Foot Ulcers

Predicting and preventing DFUs represents one of the most critical applications of ML in DPN management, as DFUs often lead to infection, hospitalization, and amputation. Ensemble and deep learning models have demonstrated particular strength in improving DFU prediction accuracy by processing high-dimensional and heterogeneous data.

Ma et al. [12] evaluated several ML algorithms—such as RF and SVM—for DFU classification and achieved notable improvements in diagnostic accuracy and computational efficiency. Their findings showed that ML-assisted diagnostic systems could reduce diagnostic time, facilitate early intervention, and improve treatment outcomes, particularly in high-risk diabetic populations.

## 2.3|Existing Challenges and Research Gaps

Despite remarkable progress, several methodological and practical challenges persist in applying ML to DPN prediction and classification. A major limitation involves data quality and heterogeneity, especially in developing regions where access to comprehensive, high-quality medical data is limited. Incomplete or noisy data can reduce model reliability and generalizability, particularly in real-world healthcare settings.

Almutairi et al. [13] highlighted that data incompleteness and imbalance remain key barriers to implementing ML-based prediction systems in clinical practice. Moreover, many existing models focus solely on biochemical or demographic variables, overlooking the potential contribution of lifestyle and familial factors—two dimensions that are crucial for personalized prediction. The lack of multi-modal data integration and insufficient model interpretability also restrict their translation into routine clinical workflows.

## 2.4 | Recent Advances and Future Directions

Recent advances in ML and data science have opened new avenues for developing more comprehensive and interpretable DPN prediction models. Emerging approaches such as transfer learning and federated learning allow model training on decentralized datasets while addressing data privacy and scarcity concerns. These frameworks enhance model generalizability across institutions and patient populations.

Liu et al. [1] developed a hybrid model combining clinical and biometric data, achieving high predictive accuracy while maintaining interpretability. Similarly, multi-modal ML frameworks that integrate structured clinical data, imaging, and lifestyle indicators are gaining traction as promising solutions for personalized neuropathy prediction.

However, despite these advances, there remains a clear need for comprehensive, population-specific models that integrate clinical, lifestyle, and familial data within an interpretable ML framework. This study aims to contribute to this growing body of research by designing and validating such a model for multi-level classification of DPN severity.

# 3 | Methodology

## 3.1 | Dataset Description

This study used electronic medical records of 1,204 patients with type 2 diabetes, collected between 2019–2022 at the Yazd Diabetes Research and Treatment Center, Shahid Sadoughi University of Medical Sciences, Iran. The initial dataset contained 70 variables encompassing clinical, neurological, metabolic, lifestyle, and familial factors related to DPN.

Data sources included:

I. Clinical features: BMI, blood pressure, and vascular assessments.

II. Neurological tests: achilles reflex, vibration perception (tuning fork), and muscle cramp sensation.

III. Lifestyle indicators: education, occupation, smoking, and alcohol use.

IV. Familial history: diabetes, neuropathy, and cardiovascular disease in first-degree relatives.

V. Neuropathy assessment: michigan Neuropathy Screening Instrument (MNSI), integrating symptom-based and physical examination scores.

The MNSI score was used as the target variable, categorized into four severity levels: 1) 0: No neuropathy, 2) 1: Mild, 3) 2: Moderate, and 4) 3: Severe.

## 3.2 | Data Preprocessing

To ensure data reliability, the following preprocessing steps were applied:

I. Missing data handling: records with >20% missing values were excluded; remaining gaps were imputed (mean for numerical, mode for categorical).

II. Outlier removal: the Interquartile Range (IQR) method was applied.

III. Normalization: continuous features were scaled to the [0,1] range using MinMaxScaler.

IV. Class balancing: class imbalance was corrected using Synthetic Minority Over-sampling Technique (SMOTE) to ensure proportional representation across severity levels.

After preprocessing, a clean dataset of 1,204 complete samples was obtained for modeling.

## 3.3 | Feature Selection

Feature selection followed a structured three-phase process:

I. Initial reduction: variables with >80% missing data were removed, reducing features from 70 to 20.

II. Feature importance analysis: ensemble algorithms (RF, XGBoost) were used to evaluate feature relevance via Gini impurity and information gain.

III. Final selection: based on cross-validation and clinical significance, 8 features were retained:

- *Gender*
- *Education level*
- *Family history (diabetes/neuropathy)*
- *BMI*
- *Muscle cramp sensation*
- *Tuning fork test*
- *Left and Right Achilles reflexes*

## 3.4 | Feature Engineering

Feature engineering was conducted to extract higher-level and interaction-based indicators. All potential composite and interaction features were generated and statistically evaluated. The following engineered features were retained:

I. Metabolic Risk Score (MRS): combines BMI, gender, and family history.

II. Reflex Symmetry Index (RSI): difference between left and right Achilles reflexes.

III. Sensory Composite Index (SCI): mean of tuning fork and reflex scores.

IV. MNSI composite score: integrated reflex and symptom metrics.

These engineered variables improved both model interpretability and predictive consistency.

## 3.5 | Model Development

Six algorithms were trained for comparison: RF, XGBoost, LightGBM, SVM, MLP, and a Voting Classifier (ensemble). Each model was trained using 5-fold cross-validation, with hyperparameters optimized through Grid Search CV.

## 3.6 | Model Evaluation

Models were assessed using multiple performance metrics: accuracy, sensitivity (Recall), specificity, F1-Score, and AUC-ROC. This evaluation ensured robust performance measurement across all neuropathy severity levels without overemphasizing specific model outcomes.

## 3.7 | Model Interpretability Analysis (Feature Importance and SHapley Additive exPlanations Analysis)

To enhance the interpretability of the predictive framework, two complementary techniques were applied. First, a Feature Importance analysis was conducted to quantify the relative contribution of each input variable to the model's predictive accuracy. Second, the SHapley Additive exPlanations (SHAP) method was employed to determine both the magnitude and direction of each feature's influence on the model's output.

The SHAP analysis provided a more granular understanding of how individual clinical and lifestyle features affected neuropathy severity predictions.

# 4 | Results and Discussion

## 4.1 | Class Balancing Using Synthetic Minority Over-sampling Technique

Before applying SMOTE, the dataset exhibited a pronounced imbalance, with a higher prevalence of non-neuropathic cases. Following preprocessing (sorting, imputation, encoding, and outlier removal), SMOTE was implemented to generate synthetic samples for minority classes, achieving a balanced class distribution as shown in *Table 1*.

**Table 1. Number of samples before and after applying SMOTE.**

| Neuropathy Severity | Before SMOTE | After SMOTE |
|---|---|---|
| 0: No Neuropathy | 418 | 418 |
| 1: Mild | 339 | 418 |
| 2: Moderate | 256 | 418 |
| 3: Severe | 191 | 418 |
| Total | 1204 | 1672 |

Balancing the dataset significantly improved the model's ability to identify minority-class patterns and enhanced predictive fairness across severity levels. Post-balancing analysis highlighted that sensory and neurological variables—particularly the tuning fork test, reflexes, and BMI—were the most discriminative indicators of neuropathy progression.

## 4.2 | Feature Selection

The initial dataset contained over 70 attributes. Features with excessive missingness (>80%) were removed, reducing the set to 20. Feature importance was then evaluated using RF and XGBoost algorithms, which provided embedded measures of variable relevance based on impurity reduction and information gain. Consistent results across both methods confirmed the reliability of selected predictors.

## 4.3 | Random Forest Feature Importance

**Table 2. Feature importance (RF model).**

| Rank | Feature | Importance |
|---|---|---|
| 1 | BMI | 0.4346 |
| 2 | Tuning fork test | 0.1790 |
| 3 | Left Achilles Reflex | 0.1475 |
| 4 | Right Achilles Reflex | 0.0991 |
| 5 | Muscle cramp sensation | 0.0228 |
| 6 | Gender | 0.0223 |
| 7 | Muscle cramp (left) | 0.0158 |
| 8 | Education (low level) | 0.0138 |

BMI and neurological indices (reflexes, tuning fork) emerged as dominant predictors, aligning with clinical evidence linking metabolic dysfunction and neural impairment to DPN progression.

## 4.4 | eXtreme Gradient Boosting Feature Importance

**Table 3. Feature importance (XGBoost model).**

| Rank | Feature | Importance |
|---|---|---|
| 1 | Left Achilles Reflex | 0.3627 |
| 2 | Tuning fork test | 0.1870 |
| 3 | Right Achilles Reflex | 0.1130 |
| 4 | Muscle cramp sensation | 0.0514 |
| 5 | Education (associate) | 0.0321 |
| 6 | Education (low level) | 0.0317 |
| 7 | Family history (yes) | 0.0284 |
| 8 | BMI | 0.0236 |

XGBoost results were highly consistent with RF, confirming the stability of neurological and metabolic predictors in defining neuropathy severity.

## 4.5 | Final Selected Features

**Table 4. Final selected features.**

| No. | Feature | Type | Description |
|---|---|---|---|
| 1 | Left Achilles Reflex | Numeric | Motor–sensory nerve response |
| 2 | Tuning fork test | Numeric | Vibration sensitivity |
| 3 | Right Achilles Reflex | Numeric | Bilateral reflex assessment |
| 4 | Muscle cramp sensation | Categorical | Neuromuscular symptom |
| 5 | BMI | Numeric | Metabolic and obesity-related risk |
| 6 | Gender | Categorical | Demographic factor |
| 7 | Education | Categorical | Health literacy proxy |
| 8 | Family history of diabetes | Categorical | Hereditary predisposition |

These eight features collectively captured the clinical, neurological, behavioral, and genetic dimensions of DPN progression.

## 4.6 | Feature Engineering and Composite Indices

All composite (engineered) features were initially designed and tested. After evaluation through correlation (Pearson, Spearman), ANOVA, and feature importance analysis, four were selected for the final model.

**Table 5. Final composite features.**

| No. | Feature Name | Construction | Rationale |
|---|---|---|---|
| 1 | SCI | Mean of bilateral reflexes + tuning fork test | Captures sensory degradation |
| 2 | Symptom severity index | Reflex + muscle cramp score | Represents overall neurological impairment |
| 3 | Reflex symmetry index | Difference between left and right reflex | Indicates asymmetry in nerve function |
| 4 | MRS | BMI + gender + family history | Integrates metabolic and hereditary risk |

These engineered indicators enhanced model accuracy from ~65% to over 86%, demonstrating their value in representing multi-domain pathophysiological interactions.

## 4.7 | Model Execution and Results

The dataset was split into 80% for training and 20% for testing to ensure a reliable generalization capability. For final evaluation, 5-fold cross-validation was used — training the model on four folds and testing on the remaining one, repeated five times, with average results reported for robustness.

The study evaluated several ML algorithms, including RF, XGBoost, LightGBM, SVM, MLP, and Voting Classifier, using Accuracy, Sensitivity, Specificity, F1-Score, and ROC-AUC as metrics.

**Table 6. Final model comparison based on 5-fold cross-validation.**

| Model | Accuracy | Sensitivity | Specificity | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| RF | 0.861 | 0.859 | 0.864 | 0.861 | 0.855 |
| XGBoost | 0.841 | 0.839 | 0.844 | 0.841 | 0.845 |
| LightGBM | 0.833 | 0.829 | 0.838 | 0.833 | 0.846 |
| SVM | 0.829 | 0.825 | 0.833 | 0.829 | 0.841 |
| MLP | 0.827 | 0.823 | 0.830 | 0.826 | 0.843 |
| Voting classifier | 0.821 | 0.818 | 0.824 | 0.821 | 0.845 |

## 4.8 | Model Comparison and Selection of the Optimal Model

I. RF achieved the highest accuracy (0.861) and F1-score (0.861), along with excellent sensitivity (0.859) and specificity (0.864).

II. XGBoost performed comparably (ROC-AUC = 0.845), maintaining strong balance across metrics.

III. LightGBM achieved the highest ROC-AUC (0.846), showing great discrimination capability.

IV. SVM and MLP showed moderate but consistent results, with non-linear adaptability in MLP reflected in ROC-AUC (0.843).

V. Voting Classifier integrated multiple model strengths with ROC-AUC of 0.845, though slightly lower overall accuracy (0.821).

Based on the comparative results presented in *Table 6*, the RF model demonstrated the strongest overall predictive performance and was therefore selected as the final classification model. Feature importance analysis was subsequently conducted using the RF model to identify the most influential variables contributing to the classification task.

To further enhance model interpretability, SHAP analysis was additionally performed using the XGBoost model, as it provides a stable and well-established framework for SHAP-based explanations. The SHAP findings were consistent with the feature importance patterns obtained from the RF model and were used as a complementary interpretability approach, rather than to redefine the final predictive model.

## 4.9|Feature Importance and SHapley Additive exPlanations–Based Model Interpretation

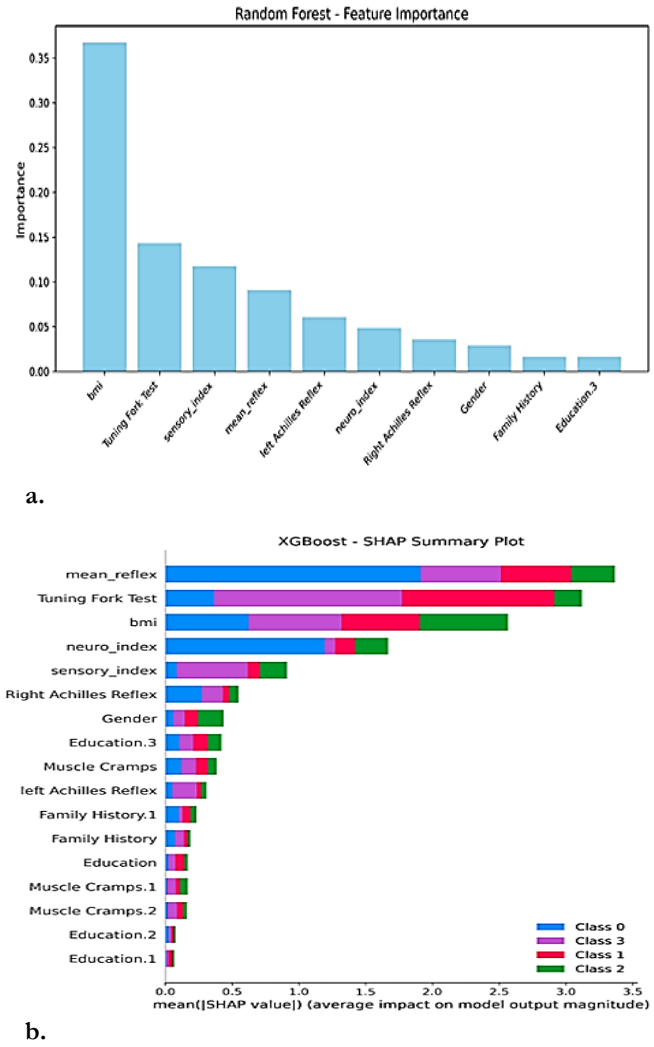


**Fig. 1. Feature importance and SHAP results showing BMI, tuning fork test, and mean reflex as the main predictors of DPN severity; a. feature importance – RF, and b. SHAP summary plot.**

RF and XGBoost models consistently identified BMI, tuning fork (vibration) test, and Mean Reflex as the most influential predictors of DPN severity. In the RF model, BMI exhibited the highest feature importance, followed by the tuning fork test and composite sensory indices, emphasizing the combined contribution of metabolic status and clinical–sensory measurements.

To investigate the contribution and direction of each feature, SHAP were applied; due to variability in RF SHAP values, the main interpretability analysis was conducted using XGBoost. SHAP results showed that lower Mean Reflex values were strongly associated with higher neuropathy severity, the tuning fork test effectively distinguished mild from severe cases, and higher BMI was directly associated with increased risk of severe DPN.

Additional composite indices (Neuro Index and Sensory Index) demonstrated clinically consistent effects, whereas demographic variables had minimal influence. The agreement between SHAP findings and RF feature importance confirms the robustness, interpretability, and clinical coherence of the proposed models.

## 4.10 | Comparison with Previous Studies and Study Strengths/Limitations

In recent years, several studies have applied ML techniques to predict DPN or related diabetic complications. For instance, a recent systematic review reported that most ML-based DPN prediction models rely primarily on clinical and biochemical parameters [7]. Another study employed ML models to predict Diabetic Sensorimotor Polyneuropathy (DSPN) and concluded that algorithms such as RF perform effectively, particularly when electrophysiological or biochemical features are incorporated [14]. Moreover, a recent risk-prediction model based on laboratory and clinical data demonstrated the feasibility of ML-based approaches for DPN prediction in real-world clinical settings [3].

Compared with these earlier works, the present study introduces two key methodological improvements. First, a wider and more diverse feature set was utilized. While many previous studies focused mainly on biochemical markers (e.g., HbA1c, serum glucose) or standard clinical and demographic risk factors, our model integrates neurological examination outcomes (reflex measurements and tuning-fork vibration test), anthropometric data (BMI), lifestyle and demographic variables (education and gender), and family history. Feature-importance analysis using RF and SHAP-based interpretability revealed that neurological and metabolic indicators—particularly Mean Reflex, vibration perception, and BMI—play a dominant role in predicting DPN severity, underscoring the clinical relevance of the selected features.

Second, a multi-level severity classification strategy was adopted. Instead of a binary classification (presence vs. absence of neuropathy), patients were categorized into multiple severity levels (none, mild, moderate, and severe). SHAP analysis further supported this approach by demonstrating how key features contribute differently across severity levels, providing more clinically actionable insights and better reflecting the progressive nature of DPN.

These methodological enhancements likely explain the improved predictive performance of the proposed model (Accuracy = 0.861, AUC = 0.855) and the consistent identification of neurological, metabolic, and demographic variables as jointly significant predictors.

## 5 | Conclusion

This study developed and validated a multi-level ML model for classifying the severity of DPN using a rich integration of clinical, neurological, metabolic, lifestyle, and familial indicators. Through comprehensive preprocessing, feature engineering, and ensemble-based selection, the RF model demonstrated the best overall predictive performance (Accuracy = 0.861, AUC = 0.855).

To enhance interpretability and clinical transparency, SHAP analysis was conducted using the XGBoost framework, which provides superior compatibility with gradient-based TreeExplainer methods.The SHAP findings confirmed that lower Mean Reflex values, reduced vibration sensitivity, and higher BMI were strongly associated with more severe neuropathy levels. Composite indices such as the Sensory Index and Neuro Index also exhibited consistent, clinically meaningful effects, while demographic factors like gender and education showed minimal influence. The close alignment between SHAP and RF feature importance validates the robustness of the selected predictors and underscores the potential of combining interpretability with predictive accuracy in AI-based clinical tools.

In summary, integrating interpretable ML with domain-informed feature engineering enables both accurate and explainable prediction of neuropathy severity. These insights pave the way for developing CDSS that can assist healthcare professionals in early detection, risk stratification, and personalized intervention for diabetic patients.

Future research should focus on validating this model across multi-center datasets, incorporating biochemical and genetic markers, and exploring deep learning architectures to further enhance model adaptability and real-world clinical utility.

## Acknowledgment

The authors sincerely thank the Diabetes Research and Treatment Center of Shahid Sadoughi University of Medical Sciences, Yazd, for providing the clinical data used in this study. Their support and collaboration made the completion of this research possible.

## References

[1] Liu, L., Bi, B., Gui, M., Zhang, L., Ju, F., Wang, X., & Cao, L. (2025). Development and internal validation of an interpretable risk prediction model for diabetic peripheral neuropathy in type 2 diabetes: A single-centre retrospective cohort study in China. *British medical journal open*, *15*(4), e092463. https://doi.org/10.1136/bmjopen-2024-092463

[2] Sun, M., Sun, X., Wang, F., & Liu, L. (2025). Machine learning-based prediction of diabetic peripheral neuropathy: Model development and clinical validation. *Frontiers in endocrinology*, *16*, 1614657. https://doi.org/10.3389/fendo.2025.1614657

[3] Lian, X., Qi, J., Yuan, M., Li, X., Wang, M., Li, G., ... & Zhong, J. (2023). Study on risk factors of diabetic peripheral neuropathy and establishment of a prediction model by machine learning. *BMC medical informatics and decision making*, *23*(1), 146. https://doi.org/10.1186/s12911-023-02232-1

[4] Basebaa, A. S., Musiaan, N. S., & Mahross, A. H. (2024). Prevalence and risk factors of diabetic peripheral neuropathy: A cross-sectional study from Yemen. *Ain shams medical journal*, *75*(1), 201-213. https://doi.org/10.21608/asmj.2024.237487.1171

[5] Wu, R. L., Chen, N., Chen, Y., Wu, X., Ko, C. Y., & Chen, X. Y. (2024). Visceral adiposity as an independent risk factor for diabetic peripheral neuropathy in type 2 diabetes mellitus: A retrospective study. *Journal of diabetes research*, *2024*(1), 9912907. https://doi.org/10.1155/2024/9912907

[6] Wei, Z., Wang, X., Lu, L., Li, S., Long, W., Zhang, L., & Shen, S. (2024). Construction of an early risk prediction model for type 2 diabetic peripheral neuropathy based on random forest. *CIN: Computers, informatics, nursing*, *42*(9), 665-674. https://doi.org/10.1097/cin.0000000000001157

[7] Ma, Y., Wang, Z., Yao, Z., Lu, B., & He, Y. (2025). Machine learning in the prediction of diabetic peripheral neuropathy: A systematic review. *BMC medical informatics and decision making*, *25*(1), 344. https://doi.org/10.1186/s12911-025-03201-6

[8] Jian, Y., Pasquier, M., Sagahyroon, A., & Aloul, F. (2021). A machine learning approach to predicting diabetes complications. *Healthcare*, *9*(12), 1712. https://doi.org/10.3390/healthcare9121712

[9] Yu, X., Wu, Z., & Zhang, N. (2024). Machine learning-driven discovery of novel therapeutic targets in diabetic foot ulcers. *Molecular medicine*, *30*(1), 215. https://doi.org/10.1186/s10020-024-00955-z

[10] Haque, F., Bin Ibne Reaz, M., Chowdhury, M. E. H., Srivastava, G., Hamid Md Ali, S., Bakar, A. A. A., & Bhuiyan, M. A. S. (2021). Performance analysis of conventional machine learning algorithms for diabetic sensorimotor polyneuropathy severity classification. *Diagnostics*, *11*(5), 801. https://doi.org/10.3390/diagnostics11050801

[11] Sheikh, M. M., Balachandra, M., VG, N., & Maiya, A. G. (2025). Predicting diabetic peripheral neuropathy through advanced plantar pressure analysis: A machine learning approach. *Scientific reports*, *15*(1), 20962. https://doi.org/10.1038/s41598-025-07774-0

[12] Gao, L., Liu, Z., Han, S., & Wang, J. (2025). A machine-learning-based clinical decision model for predicting amputation risk in patients with diabetic foot ulcers: Diagnostic performance and practical implications. *Diagnostics*, *15*(24), 3142. https://doi.org/10.3390/diagnostics15243142

[13] Almutairi, E., Abbod, M., & Hunaiti, Z. (2025). Prediction of diabetes using statistical and machine learning modelling techniques. *Algorithms*, *18*(3), 145. https://doi.org/10.3390/a18030145

[14] Shin, D. Y., Lee, B., Yoo, W. S., Park, J. W., & Hyun, J. K. (2021). Prediction of diabetic sensorimotor polyneuropathy using machine learning techniques. *Journal of clinical medicine*, *10*(19), 4576. https://doi.org/10.3390/jcm10194576