Prediction of the risk for diabetic neuropathy among diabetes mellitus patients at selected referral hospitals, in Amhara regional state Northwest Ethiopia, January, 2005- December 2021; Development and validation of risk prediction model

Negalg B yardgie Gelaw*1* Achenef Asmamaw Muche2 Adugnaw Zeleke Alem3 Nebiyu Bekele Gbei4 Yazachew Moges Chekol5 Tigabu Kidie Tesfie6 Tsion Mulat Tebeje7

1Department of public health, Mizan- Aman College of health sciences, Mizan- Aman, Ethiopia
2Department of Epidemiology and biostatics, College of medicine and health sciences, University of Gondar, Gondar, Ethiopia
3Department of Epidemiology and biostatics, College of medicine and health sciences, University of Gondar, Gondar, Ethiopia
4Department of internal medicine, School of Medicine, University of Gondar comprehensive specialized hospital, Gondar, Ethiopia
5Department of Health Information Technology, Mizan Aman College of health sciences, Mizan-Aman, Ethiopia
6Department of Public Health, College of Health Sciences, Debre Markos University, Debre Markos, Ethiopia
7Unit of Epidemiology and biostatistics, School of public health, college of medicine and health sciences, Dilla University, Dilla, Ethiopia

Corresponding Author *1* (negalgnbyadgie21@gmail.com)

Emails:
NBG; negalgnbyadgie21@gmail.com
AAM; ashua2014@gmail.com
AZA; aduzeleke2201@gmail.com
NBG; dawabekele@gmail.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background:- Diabetic neuropathy is the most common complication in both Type-1 and Type-2 DM patients with more than one half of all patients developing nerve dysfunction in their lifetime. Although, risk prediction model was developed for diabetic neuropathy in developed countries, it is not applicable in clinical practice, due to poor data, methodological problems, inappropriately analyzed and reported. To date, no risk prediction model developed for diabetic neuropathy among DM in Ethiopia. Therefore, this study aimed prediction the risk of diabetic neuropathy among DM patients, used for guiding in clinical decision making for clinicians.

Objective:- Prediction of the risk for diabetic neuropathy among diabetes mellitus patients, in selected referral hospitals, Amhara regional state, northwest Ethiopia, 2022.

Methods:- A retrospective follow up study was conducted with a total of 808 DM patients were enrolled from January 1, 2005 to December 30, 2021 at two selected referral hospitals in Amhara regional state. Multi-stage sampling techniques were used and the data was collected by checklist from medical records by Kobo collect and exported to STATA version-17 for analysis. Lasso method were used to select predictors and entered to multivariable logistic regression with P-value < 0.05 was used for nomogram development. Model performance was assessed by AUC and calibration plot. Internal validation was done through bootstrapping method and decision curve analysis was performed to evaluate net benefit of model.

Results:- The incidence proportion of diabetic neuropathy among DM patients was 21.29% (95% CI; 18.59, 24.25). In multivariable logistic regression glycemic control, other comorbidities, physical activity, hypertension, alcohol drinking, type of treatment, white blood cells and red blood cells count were statistically significant. Nomogram was developed, has discriminating power AUC; 73.2% (95% CI; 69.0%, 77.3%) and calibration test (P-value = 0.45). It was internally validated by bootstrapping method with discrimination performance 71.7(95% CI; 67.2%, 75.9%). It had less optimism coefficient (0.015). To make
nomogram accessible, mobile based tool were developed. In machine learning, classification and regression tree has discriminating performance of 70.2 % (95 % CI; 65.8%, 74.6%). The model had high net benefit at different threshold probabilities in both nomogram and classification and regression tree.

Conclusion:- The developed nomogram and decision tree, has good level of accuracy and well calibration, easily individualized prediction of diabetic neuropathy. Both models had added net benefit in clinical practice and to be clinically applicable mobile based tool were developed.

Key-words:-Prediction, Diabetes mellitus, Diabetic neuropathy, Ethiopia
Background

Diabetes mellitus is a metabolic disorder characterized by increasing blood glucose level in the body resulting from either defect insulin secretion, insulin action, or both (1) with around 90% having type 2 DM and 10% having type-1 DM (2). According to IDF, the global prevalence of diabetes mellitus in 20-79 year olds at 2021 was estimated to be 536.6 million people, rising to 783.2 million in 2045 (3).

Globally, around 3 million deaths annually are due to DM and 49.7% living with diabetes undiagnosed (4). It is one of the most common problems in the world that has faced both in developed and developing countries (5). In Africa, by 2019, around 19 million adult populations were estimated to have diabetes and it is expected that by 2045 it will be around 47 million and similarly in Ethiopia, an estimated 1,699,400 adults were living with diabetes mellitus (6).

The chronic hyperglycemia of diabetes mellitus had long-term damage, dysfunction, and failure of different organs such as nerves, eyes, kidneys (7). Diabetic neuropathy defined as presence of symptoms and signs of peripheral nerve dysfunction in people with diabetes mellitus after exclusion of other causes (8). The mechanism of nerve dysfunctions was through metabolic process by hyperglycemia due to oxidative stress, directly affecting nerve fibers (9). It is the most common complication of both type-1 and type-2 DM with more than one half of all patients developing nerve dysfunction in their lifetime (10).

The burden of diabetic neuropathy common in industrialized countries ranges from in China, 33.1% (11) to 40.3% among patients with type-1 diabetes mellitus and 42.2% with type-2 diabetes mellitus in Germany (12). Evidence from previous study documented shows that in Latin American and Caribbean countries, the incidence of diabetic neuropathy was 29.2% (13) and in Sri Lankan 28.8% (14).

In Africa, according to a systematic review and meta-analysis done, the overall pooled prevalence of diabetic neuropathy was 46% (15). In South Africa 30.3% (16), in Ghana 8.1% (17), in Northern Africa, incidence ranges from 21.9% to 60% (18).

In Ethiopia, the prevalence of diabetic neuropathy ranges from 16.63% at university of Gondar comprehensive specialized referral hospital to 52.2% at Bahir-Dar city (19), at Tikur-Anbesa and St. Paul's specialized university hospital was 48.2% with 53.6% in type 2 and 33.3% in type 1 diabetes mellitus patients (20).

As part of NCD prevention and treatment strategies, diabetic neuropathy and related complications are getting attention at the national and international levels (21). Guidelines are recommending that by multidisciplinary team approach can treating and preventing diabetic
neuropathy. WHO and other professional society guideline are also recommended that strengthening of the surveillance & management of diabetic neuropathy was important to reduce the incidence and its consequence(22).

Despite different strategies and initiatives implemented to reduce the Burdon of diabetic neuropathy, up to one third of patients with diabetes have neuropathic pain(23) which often leads to sleep disturbance, poor quality of life, depression, and unemployment. Both acute and chronic diabetic neuropathies had occur in DM patients (24) that results diabetic foot disease, including ulceration and non-traumatic amputations that might be due to lack of screening tool that can be used to identify diabetic patients who is at risk of diabetic neuropathy.

Although, risk prediction model for diabetic neuropathy among DM patients in developed countries was developed, it is not applicable in clinical practice due to poor data, methodological problems, inappropriately analyzed and poorly reported(25). Delay and lack of detection of the complication was mostly resulted from patients being asymptomatic during the early stage of the disease(26). So that a simple and accurate risk prediction tool to identify those at high risk of patents develop to diabetic neuropathy had great value.

Early identification of high risk DM patients using simple screening tool , avoids late diagnosis of diabetic neuropathy that reduces huge economic costs and serious complications (27). Thus, a method used to provide information about their level of risk was needed to take timely intervention measures to prevent occurrence of diabetic neuropathy.

Many epidemiological studies on diabetic neuropathy, including cross-sectional studies are carried out around the world to explore the risk factors associated with diabetic neuropathy (12, 28) but level of risk to develop diabetic neuropathy for DM patients was unknown. Hence studies suggest that patient specific model like individual patients risk prediction model is useful to alleviate the problem.

Risk prediction models have been used to predict the probability of risk of diseases used in medicine and public health to guide clinical decision-making(29). However, risk prediction model for diabetic neuropathy among DM patients have not been developed in Ethiopia. There was variability in the clinical diagnosis of diabetic neuropathy in different setting (30). Thus, this prediction model considers prognostic factors that have been generally monitored in clinical practice and precisely measured to ensure its feasibility and accuracy for clinical application. Currently an urgent change for improvement of the diagnosis and management of diabetic neuropathy were needed(31).
According to a previous risk prediction model conducted to identify DM patients at risk of developing diabetic neuropathy being old age (32), sex (33), educational level (34) were significant predictors with area under ROC curve of 0.957. In another study, age (35), residence (3), BMI (36), educational level (34), poor diabetic control (37), type of DM (38), duration of diabetes mellitus (39), number of hypoglycemic drugs used (40) with AUC 0.859 were also determinates of diabetic neuropathy. Hemoglobin level (41), presence of other micro and macro vascular complications (13), hypertension (39), type of DM (33) baseline comorbidities (42), complications such as diabetic nephropathy, diabetic-retinopathy (12), lipid profile (triglyceride, cholesterol) levels (43), fasting blood glucose (44, 45), HbA1c (46), high level of alcohol drinking (13) physical inactive (39), eating fat containing foods (47) were the most common predictors of diabetic neuropathy.

Although, different strategies and interventions have been made so far, to reduce the Burdon diabetic neuropathy, Overtime at least 50% of individuals with diabetes develop diabetic neuropathy (48). It is preventable through detected early the high risk patients, by providing timely intervention measures. However, identifying high risk patients for diabetic neuropathy was not performed as expected (26).

Until an advanced stage, diabetic neuropathy is asymptomatic but it can be prevented through identifying high risk patients and prompt treatment before their complication. Up to my knowledge there is no risk prediction model for diabetic neuropathy among DM patients in Ethiopia. As a result, this risk prediction models alleviate the problem by developing nomogram and decision tree, estimate disease risk, can guide healthcare providers for disease intervention and arrangement future health care needs by providing treatment.

Monitoring biomarkers and other predictors is also very essential to be aware of metabolic abnormality by providing meaningful prognostic information that can help to differentiate patients with regard complication and lead the way to change other intervention. This risk prediction model was developed by applying sound statistical methods and analysis to identify DM patients at which level of risk for the development of diabetic neuropathy (49). Besides, the finding of this study was highly useful as a simple clinical tool to guide clinicians for decision making, aid for specific screening of high-risk patients and for their informed choice of treatment to the patients. Therefore, these studies aimed development and validation of risk prediction model for diabetic neuropathy among diabetes mellitus patients in selected referral hospitals at Amhara regional state, Ethiopia.
Methods and materials

Study design and area

An institution-based retrospective follow up study was conducted among patients diagnosed with diabetes mellitus from January 1, 2005 to December 30, 2021. The study area was selected referral hospitals in northwest part of Amara regional state, which includes University of Gondar Comprehensive specialized and referral hospital, Debre markos comprehensive specialized hospital, Tibebe-Gihon comprehensive specialized referral hospital, Felge-Hiwot comprehensive specialized referral hospital, and Debre tabor comprehensive specialized referral hospital.

Using lottery Method, University of Gondar comprehensive specialized referral hospital and Felege-Hiwot comprehensive specialized referral hospital was selected. University of Gondar comprehensive specialized hospital found in Gondar town, far from Addis Abeba, the capital city of Ethiopia 750 km and 200 km from Bahir-dar, the capital city of Amhara regional state. It serves for more than 7 million people in northwest Ethiopia. The hospital serves around 24,862 numbers of people are having chronic follow-up per year, and among this, 4760 were DM patients.

Felege-Hiwot Referral Hospital is found in Bahir-Dar, which is the capital city of Amhara regional state located at 565 km from Addis Ababa, Northwest Ethiopia. It serves as over 7 million people from the surrounding area. Around 21,218 people had a chronic follow-up in this hospital and among these 4200 were DM patients. The study setting was displayed below

Population

The source population was all people diagnosed with diabetes mellitus having a follow up at selected referral hospitals in Amhara regional state, North West Ethiopia and all people having confirmed diabetes mellitus started treatment and follow up at selected referral hospitals in Amhara regional state with in follow up period were study population. A diabetes mellitus patient who was enrolled from January 1, 2005 to December 30, 2021 with minimum of two years follow up, at university of Gondar comprehensive specialized hospital and Felege-Hiwot comprehensive specialized referral hospital were included. Diabetes mellitus patients whose date of initiation was not recorded and incomplete charts, those DM patients develop diabetic neuropathy at beginning, transferred in and gestational diabetes mellitus was excluded.
Sample size calculation and sampling technique

In risk prediction model development sample size was calculated through different approaches/methods either considering Minimum Mean Absolute Prediction Error (MAPE) to be minimum and use shrinkage factor to be minimize the issue of over fitted model.

Therefore, by applying the expected average error of predicting the outcome when the developed prediction model is going to be applied to new individuals should be considered, that is called Mean Absolute Prediction Error (MAPE) should be minimum. Sample size also can be calculated using the following formula by taking the prevalence according to a systematic review and meta-analysis study done in Ethiopia, the prevalence of diabetic neuropathy among diabetes mellitus patients was 22% (50).

\[
n = \exp \left( \frac{-0.508 + 0.259 \ln(\phi) + 0.504 \ln(p) - \ln(MAPE)}{0.544} \right)
\]

........................................Equation 1

Here \(\phi\) -is the proportion of diabetic neuropathy \(P\)- is the number of predictors used to predict the diabetic neuropathy and MAPE is the mean absolute prediction error of maximum 5% considered in our case.

\[
n = \exp \left( \frac{-0.508 + 0.259 \ln(0.22) + 0.504 \ln(22) - \ln(0.05)}{0.544} \right) = 475
\]

The second method is taking the issue of over fitted model into consideration in prediction model. We have targeted less than 10% over fitted model for predicting diabetic neuropathy in DM patients. The amount of sample size required to minimize the problem of over fitting should be identified. We consider the following sample size formula.

\[
n = \frac{P}{(S - 1)\ln \left(1 - \frac{R^2_{CS}}{S}\right)}
\]

........................................Equation 2

\[
R^2_{CS} = R^2_{nagelkerke} \max (R^2_{CS})
\]

........................................Equation 2a

\[
\max (R^2_{CS}) = 1 - \exp \left( \frac{-2\ln(\text{null})}{n} \right)
\]

........................................Equation 2b

\[
\ln(\text{null}) = \ln \left( \frac{E}{n} \right) + (n - E)\ln \left(1 - \frac{E}{n}\right)
\]

........................................Equation 2c

\[
\ln(\text{null}) = 22\ln \left( \frac{22}{100} \right) + (100 - 22)\ln \left(1 - \frac{22}{100} \right) = -52.7
\]

\[
\max (R^2_{CS}) = 1 - \exp \left( \frac{-2 \times -52.7}{100} \right) = 0.65
\]
Here, $\ln_{\text{null}}$ is the log-likelihood of an intercept-only model; $E$ is the total numbers of DM patients develop diabetic neuropathy, $S$- shrinkage factor taken as 0.9 for optimum non-over fitted model $n$- is the arbitrary sample size from it E occurs. Max $R^2_{CS}$ is the maximum cox-Snell variability explained by the model $\delta$ - is the expected model optimism, $R^2_{nagelkerke}$ is also a fundamental overall measure of model fitness, it is usually considered as default of 0.5.

From the above different types of approaches for sample size calculation in risk prediction model the largest sample size is 492 and since multi stage sampling technique was used, design effect taken as equal to 1.5, Then 492*1.5=738 Finally by adding 15% contingency because of while during data extraction checklist pretesting, to check completeness and clarity of checklist, most of the patient medical record number was absent in the medical record room by taking in to consideration such missing charts, adding 10-15% contingency was recommended so, the total minimum sample size was 849.

Multi stage sampling techniques were used by first selecting the referral hospitals in North West part of Amhara regional state. Among these referral hospitals by using lottery method UOG comprehensive specialized referral hospital and Felege-Hiwot comprehensive specialized referral hospital were selected. Then, study participates were selected by first proportion allocation to each selected referral hospitals through the following proportional allocation formula.

$$ni = \frac{n}{N} \times Ni$$

Where $n$=Total sample size $N$-total population $ni$- sample size for each hospital

Finally the study participates was selected by using computer random generated number through preparing sampling frame by arranging his/her medical record number order.

A total of 8960 Type-1 and Type-2 diabetes mellitus patients were enrolled from January 1, 2005 up to December 30,2021 and had follow up at university of Gondar comprehensive specialized referral hospital and Felege Hiwot comprehensive specialized referral hospital. The flow chart of participates selection was displayed (Figure-1).

Figure 1:-Flowchart of participant selection for the development of nomogram for prediction of diabetic neuropathy among DM patients in selected referral hospitals, Amhara regional state, North West Ethiopia, 2005-2021.
Variables

Dependent variable
Diabetic neuropathy

Independent variable

Socio demographic variables: Baseline age (year), sex, residence, body mass index

Laboratory related variables: Mean fasting blood glucose level, Mean arterial blood pressure, baseline total cholesterol, baseline triglyceride level, baseline value of WBC, RBC, platelets count, baseline hemoglobin and creatinine level

Diagnosis and treatment related variables: Type of DM, duration of DM, Type of treatment for DM, hypertension, baseline other comorbidities and adherence

Behavioral related variables: Alcohol drinking, unhealthy diet, physical activity

Operational definition

Diabetic neuropathy: It can be either small fiber neuropathy or large fiber neuropathy. Small fiber neuropathy manifested by pain, tingling, paraesthesia and confirmed by pinprick and temperature examination. Large fiber neuropathy is manifested by numb feet and gait ataxia and confirmed by touch sensation by 10g monofilament, vibration sense by biothesiometer and ankle reflex. A patient is considered as having diabetic neuropathy if he/she is diagnosed as diabetic neuropathy on his/her medical record (51).

Other comorbidities: if other baseline disease (except hypertension) in DM patients were present such as ischemic heart diseases, stroke, e.t.c

Mean arterial blood pressure; calculated as MABP = (SBP + (2 * DBP))/3

- Normal----- 70-100 mmHg
- High-------- >100mm Hg

Glycemic control: Patients was classified as per the WHO criteria into:

- Good glycemic control = fasting blood glucose of 80–130 mg/dl
- Poor glycemic control = fasting blood glucose of > 130 mg/dl------ (52)

Total cholesterol: Total Triglyceride

- Normal = <200 mg/dl
- High -- >200 mg/dl -----(53)

White blood cells Red blood cells
Normal ---- 4,500-11,000 cells/micro liter    Normal = 4-6 million cells/ml
Low ------    <4,500 cells/ml                        Low =<4 million cells/ml
High ------    >11,000 cells/ml--(54)           High >6 million cells/ml----(54)

Platelets:

Normal ----150,000-400,000 cells/ml               Underweight = BMI < 18.5 kg/m²
Low ------   <150,000 cells/ml                            Normal weight = BMI 18.5 to 24.9 kg/m²
High -- >400,000 cells/ml------(54)                Overweight = BMI >=25 kg/m²-- (55)

Data collection tools and quality control

This risk prediction model was used secondary data source, collected from June 8 to July 10, 2022 using the developed data extraction check lists. The reviewed records were identified by their medical registration number. Patient intake form, follow-up card, and DM registration book were used as data sources. Socio-demographic characteristics, behavioral characteristics, baseline, and follow-up clinical and laboratory data were collected from patient cards. The date that patients start regular follow-up treatment until the end of the study to the confirmation of a final event in the study period was collected. Four trained BSc nursing health professional and two public health officer supervisors were recruited for data collection. Training on the objective of the study and how to retrieve records as per data extraction checklists were given to data collectors and supervisors one day before data collection. As well as, random sample from data extracted was crosschecked for its consistency. The information formats were crosschecked with the source card on the spot, and regular supervisions were done.

Data processing and analysis

Data was coded and entered to Kobo collect Version v2022.1.2 for clean-up and then exported to Stata version-17 for analysis. Descriptive statistics (frequency, percentage, proportion, and mean with standard deviation, median with IQR) were used to describe the clinical course of diabetic neuropathy diseases. The candidate prognostic factors were selected by using the least absolute shrinkage and selection operator (LASSO) method, by shrinking variables with very unstable estimates towards zero, it can effectively exclude some irrelevant variables and produce sparse estimations while variables with nonzero
regression coefficient was considered as most strongly associated with the response variable. Then selected variables by (LASSO) method were entered into multivariable logistic regression and model reduction was done using log likelihood ratio test >0.15 and finally p-value < 0.05 in model reduction were taken as statistically significant predictors and beta coefficients with 95% CI was reported. Multi-collinerity was checked by variance inflation factor where VIF > 10 indicate the presence of multi-collinearity.

**Missing data management**

Missing variable with value <50%, the missing data was handled by multiple imputation through checking the type of missing data. In this risk prediction model the missing variable was checked by fitting statistical test (logistic regression) found a result, the missing variable was associated with the observed variables, indicates that Missing at random (MAR). The percentage of missing value was done by “mdesc” package by STATA version-17 and if the missing value >= 50%, were not included in our model. From 22 predictors, six of them had complete data: baseline age, sex, residence, type of DM, type of treatment for DM and duration. The remaining missing variables were managed by multiple imputation technique, better that considers mean and variance and decreases uncertainty or bias. The missing variables and its percentage value was found in (Supplementary file S-1)

**Risk prediction model development**

In risk prediction model development both the prognostic markers and determinates that predict the diabetic neuropathy was selected from socio-demographic factors, diagnosis and treatment related factors, laboratory related factors and behavioral related factors. The determinants factors included in the multivariable logistic regression analysis were selected based on the results of lasso regression method. The theoretical design was the incidence of diabetic neuropathy at a future time “t” is a function of prognostic determinants ascertained at one time (baseline) points before the occurrence of the diabetic neuropathy (“t0”). The occurrence relation was incidence of DN = f (Comorbidities, glycemic control, alcohol drinking+ physical activity + hypertension+ type of treatment+ WBC count+ RBC count)

Domain: All type-1 and type-2 DM patients at risk for developing diabetic neuropathy

**Nomogram and classification and regression tree development**

Nomogram, which are reliable prediction tools, with the ability to generate an individual probability of a clinical event by incorporating a number of prognostic determinant variables
(56) was developed to predict individual risk of diabetic neuropathy among DM patients at diabetic clinic of university of Gondar comprehensive specialized hospital and Felege Hiwot referral hospital. The total nomogram score were classified as low risk, intermediate and high-risk scores to see size of each category. Classification and Regression Tree (CRT) analysis was performed to determine the complex interactions among the most potent candidate predictors in the final reduced model to build the classification trees. Youden-index was used to estimate cut off point for predictive probability of diabetic neuropathy at different threshold. By choosing the cutoff value get from Youden’s index the sensitivity, specificity, NPV, PPV and accuracy was calculated and cut off value was selected depend availability of resources and aim of implemented program.

**Risk prediction model performance**

The reduced multivariable model was evaluated by calibration and discrimination (57). Model discrimination is the model’s ability to discriminate between subjects with and without the diabetic neuropathy was estimated through the areas under the ROC curve. The area under the receiver operating characteristic curve (AUC) ranges between 0.5-1. Model calibration is examined the level of agreement between the diabetic neuropathy probabilities estimated by the model versus the observed diabetic neuropathy frequencies through graphical comparison, calibration plot. Good calibration means that the estimated probability of diabetic neuropathy is similar to the observed diabetic neuropathy frequency. Statistically we checked by the Homer and Lemeshow test with a significant P- value implies that the model is not well calibrated as it performs differently for different risk categories. Finally the results were presented using statement, table, figures and reported according to transparent reporting of multivariable prediction model for individual progress or diagnosis (TRPOID) statement(58)

**Internal validation**

In risk prediction model, validating the model expected from the researchers in order to apply in clinical practice. In this study by using bootstrapping method internal validation was done to avoid over fitting of the model (57). The first step in bootstrapping method is the model was developed in the full sample. Then multiple random minimum samples (1000) draw from the full sample. The calibration and discrimination of each bootstrap model were compared to the corresponding estimates of the bootstrap models when applied in the
original full sample. These differences were averaged, and provide an indication of the average optimism of the bootstrap models.

**Decision curve analysis**

Decision curve analysis was done to evaluate prediction models that is important for medical decision making to overcome the limitations discrimination and calibration\(^{(59)}\). It was done graphically for model, treatment for all, and treatment for none. The net benefit of the developed model for carrying out a certain intervention measure in DM patients at risk of developing diabetic neuropathy was done by decision curve plot.

**Mobile based application of Nomogram**

Application of the nomogram by mobile based is useful to facilitate limited resources in poor healthcare systems to get evidence based clinical decision making. To make the nomogram accessible to physicians and patients, mobile based application tool were developed for predicting risk of diabetic neuropathy in diabetes mellitus patients. The mobile based tool was developed by “sublime text” application with java script and used to calculate the risk of diabetic neuropathy for DM patients without requiring internet service “off line” in clinical setting. Where clinicians or diabetes mellitus patients enter his/her selected clinical and laboratory data then developed mobile based application toll giving alert as low risk or high risks of diabetic neuropathy to keep or improve their health condition.

**Results**

**Socio-demographic and behavioral characteristics of DM patients**

A total of 808 diabetes mellitus patients who had follow up visit were included. The mean baseline age of diabetes mellitus patients was 45.6±3.1 years. More than three-fourth 616(76.2%) diabetes mellitus patients were from urban residents \((\text{Table-1})\)

**Table 1:** Socio-demographic and behavioral predictors of diabetic neuropathy among diabetes mellitus patients in North West, Ethiopia, 2005-2021 \((n=808)\)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Diabetic Neuropathy</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Yes</td>
<td>73</td>
<td>400</td>
</tr>
</tbody>
</table>
## Diagnosis and treatment related characteristics of DM patients

More than two-third of patients 545(67.5%) were type-2 DM patients and half of them 401(49.6%) diabetes mellitus patients were used oral drugs medication. Nearly half of the diabetes mellitus patients 385(47.6%) had other comorbidities. Majority of diabetes mellitus patients 557(71.4%) had good adherence (Table-2).

### Table 2:-Clinical predictors of diabetic neuropathy among diabetes mellitus patients
**North West, Ethiopia, 2005-2021 (n=808)**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Diabetic Neuropathy</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Type of diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type-1 DM</td>
<td>35</td>
<td>228</td>
<td>263</td>
</tr>
<tr>
<td>Type-2 DM</td>
<td>137</td>
<td>408</td>
<td>545</td>
</tr>
<tr>
<td>Type of treatment for diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin only</td>
<td>55</td>
<td>293</td>
<td>348</td>
</tr>
<tr>
<td>Oral drugs only</td>
<td>96</td>
<td>305</td>
<td>401</td>
</tr>
<tr>
<td>Predictor</td>
<td>Mean ± SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean fasting blood glucose (mg/dl)</td>
<td>92 ± 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemoglobin (g/dl)</td>
<td>13.9 ± 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol level (mg/dl)</td>
<td>189.9 ± 58.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean value of fasting blood glucose level was 92 ± 15 and the mean total cholesterol and total triglyceride level had 189.9 ± 58.7 and 170.5 ± 73.7 mg/dl respectively. (Table-3)
Total triglyceride level(mg/dl) 170.5 ± 73.7
Creatinine(mg/dl) 0.9 ± 0.6
White blood cells(Cells/microliter) 6.9 ± 2.8 x10^3 cells/mcl
Red blood cells(Cells/microliter) 4.8 ± 0.9x10^6 cells/mcl
Platelets(Cells/microliter) 233 ± 74x10^3 cells/mcl

**Incidence of diabetic neuropathy**

The incidence proportion of diabetic neuropathy was 172 (21.29 %, 95% CI; 18.59, 24.25)

Type-2 diabetes mellitus 137(79.65%) had highest incidence of diabetic neuropathy whereas 35(20.35%) were found in type-1 diabetes mellitus patients. (Figure-2)

**Figure 2; Pie chart for the proportion of diabetic neuropathy among diabetes mellitus patients North West, Ethiopia, 2005-2021.**

**Prediction model development and validation**

**Selected predictors**

In this risk prediction model socio-demographic, diagnosis and treatment related, laboratory and behavioral related predictors were considered for development of prediction model of diabetic neuropathy.

The candidate predictors were selected by least absolute shrinkage and selection operator (LASSO) regression for the development and validation of model for diabetic neuropathy (Table-4). Multi- collinerity test was checked by VIF and unhealthy diet was not included in our model due to high VIF (17.8).

**Table 4:- Optimum shrinkage factor (lambda) and potential predictors identified by lasso regression by 10-fold cross validation selection method for diabetic neuropathy among diabetes mellitus patients North west, Ethiopia 2005-2021.(n=808)**

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Lambda</th>
<th>No. of non-zero coefficients</th>
<th>0ut of sample dev. Ratio</th>
<th>CV mean deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First lambda</td>
<td>0.09923</td>
<td>0</td>
<td>0.0023</td>
<td>1.037898</td>
</tr>
<tr>
<td>21</td>
<td>Lambda before</td>
<td>0.015437</td>
<td>12</td>
<td>0.0575</td>
<td>0.9759118</td>
</tr>
<tr>
<td><strong>22</strong></td>
<td>selected lambda</td>
<td><strong>0.0140656</strong></td>
<td><strong>13</strong></td>
<td><strong>0.0578</strong></td>
<td><strong>0.0756381</strong></td>
</tr>
<tr>
<td>23</td>
<td>Lambda after</td>
<td>0.012816</td>
<td>13</td>
<td>0.0576</td>
<td>0.9757846</td>
</tr>
<tr>
<td>27</td>
<td>Last lambda</td>
<td>0.0088336</td>
<td>19</td>
<td>0.0553</td>
<td>0.9782416</td>
</tr>
</tbody>
</table>
Thirteen variables were selected by Lasso regression and included in the multiple logistic regression models (Table-5). Finally, we construct the clinical risk prediction model with eight statistically significant predictors including hypertension, glycemic control (FBG), other comorbidities, Alcohol drinking, physical activity, type of treatment for DM, WBC and RBC count. Goodness of fit test was checked had insignificant p-value 0.3077, best fitted model.

Table 5: Multi variable logistic regression for variables retained in the final reduced model to predict diabetic neuropathy among DM patients Northwest Ethiopia, 2005-2021. (n=808)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Diabetic neuropathy</th>
<th>Multi variable analysis</th>
<th>Model reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Beta coefficients (95 % CI)</td>
</tr>
<tr>
<td>Baseline age in years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45 years</td>
<td>321</td>
<td>62</td>
<td>0</td>
</tr>
<tr>
<td>&gt;=45 years</td>
<td>315</td>
<td>110</td>
<td>0.15(-0.28,0.59)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>309</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td>Male</td>
<td>327</td>
<td>73</td>
<td>-0.26(-0.63,0.10)</td>
</tr>
<tr>
<td>Residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>166</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Urban</td>
<td>470</td>
<td>146</td>
<td>0.28(-0.23,0.79)</td>
</tr>
<tr>
<td>Type of DM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type-1 DM</td>
<td>228</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>Type-2 DM</td>
<td>408</td>
<td>137</td>
<td>0.024(-0.6, 0.7)</td>
</tr>
<tr>
<td>Mean arterial blood pressure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-100 mmHg</td>
<td>484</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>&gt;100 mmHg</td>
<td>152</td>
<td>62</td>
<td>0.15(-0.25,0.56)</td>
</tr>
<tr>
<td>Type of treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both insulin and oral drug</td>
<td>38</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Insulin only</td>
<td>305</td>
<td>96</td>
<td>-0.49(-1.32,0.10)</td>
</tr>
</tbody>
</table>
Oral drugs only | 293 | 55 | -0.61(-1.09,0.15) | 0.139

Baseline glycemic control

<130mg/dl | 148 | 25 | 0

>=130mg/dl | 488 | 147 | 0.58(-0.01,1.16) | 0.04 | 0.04*

Other comorbidities

No | 373 | 50 | 0

Yes | 263 | 122 | 1.29(0.85,1.68) | 0.000 | 0.00*

Alcohol drinking

No | 349 | 287 | 0

Yes | 98 | 74 | 1.13(0.91,2.25) | 0.04 | 0.003*

Physical Activity

No | 97 | 75 | 0

Yes | 325 | 311 | -1.79(-2.94,-0.7) | 0.002 | 0.001*

Hypertension

No | 392 | 73 | 0

Yes | 244 | 99 | 0.412(0.00,0.82) | 0.04 | 0.01*

White blood cells/microliter

4500-11000 | 266 | 82

<4500 | 33 | 20 | -1.14(-2.2,-0.01) | 0.04 | 0.03*

>11000 | 337 | 70 | 0.15(-0.58,0.89) | 0.684

Red blood cells/microliter

4-6 million | 270 | 80 | 0

<4 million | 34 | 20 | 0.90(0.56,1.74) | 0.03 | 0.03*

>6 million | 332 | 72 | -0.24(-0.99,0.52) | 0.539

_cons | -1.96(-2.97,0.94) | 0.000

**Prediction model development using original beta coefficients**

By using area under ROC curve the discriminating power of the model was evaluated and individual predictors in the final reduced model had poor performance discriminating the risk of diabetic neuropathy among DM patients starting from 51.1% to 62.4% but they had good discriminating ability in combined effect. The predictive performance of model using combination of other comorbidities, glycemic control, physical activity and WBC had 70.0%. The predictive performance powers of each predictor were displayed (supplementary file S-
The area under curve (AUC) of the final reduced model using original beta coefficients were 73.49 % (95% CI; 69.3%, 77.6%). (Figure-3, A). The developed model had calibration test value (p-value=0.451) well calibrated model, the model well represented the data that was agreement between observed and the predicted probability (Figure-3, B).

Figure 3: The performance of the model using original beta coefficients A) AUC B) calibration plot, for prediction of diabetic neuropathy among DM patients North West Ethiopia, 2005-2021.

The probability for risk of diabetic neuropathy using original beta coefficients was

\[ P(\text{Diabetic-neuropathy})=\frac{1}{1+\exp(-1.96+0.9\cdot\text{RBC(low)}-1.14\cdot\text{WBC(low)} +0.4\cdot\text{Hypertension (yes)}-1.79\cdot\text{Physical activity (yes)} +1.13\cdot\text{Alcohol drinking (yes)} +1.29\cdot\text{other comorbidities (yes)} +0.58\cdot\text{Glycemic control (poor)}-0.72\cdot\text{type of treatment (insulin only)}) \]

Prediction model development using nomogram

For the sake of simplicity and easy to use in clinical practice the model was developed by nomogram (Logistic regression). All significant regression coefficients in the final reduced model were used for model development. The model presented by monogram through eight significant variables that predicts diabetic neuropathy was displayed (Figure-4). From nomogram the possible minimum and maximum score was 0 to 47.2 with different threshold probabilities to develop diabetic neuropathy. The cut off value to nomogram score as the probability of 0.24 was 26. Hence, DM patients who had <26 score from nomogram total score was low risk whereas those who had >=26 score from total score was as high risk to develop diabetic neuropathy.

Figure 4:- Nomogram for develop and validate risk prediction model for diabetic neuropathy among diabetes mellitus patients, northwest Ethiopia, 2005-2021.

The performance of developed nomogram in predicting diabetic neuropathy was AUC= 73.2 % (95% CI; 69.0%, 77.3%), had good level of accuracy (Figure-5, A) and calibration test (p-value 0.502), the model was well calibrated model, indicates the observed probability was agreement with expected probability (Figure-5, B).

Figure 5: Performance of model using nomogram A) AUC B) Calibration plot, prediction for the risk of diabetic neuropathy among DM patients North West Ethiopia 2005-2021.
Prediction model development using machine learning by classification and regression tree

Besides the model developed by nomogram, it also presented by machine learning algorism with classification and regression tree analysis suitable when continuous and categorical variables used that predict diabetic neuropathy. The prediction rate (accuracy) for diabetic neuropathy in classification and regression tree was 78.7% (Table-6).

Table 6:- Prediction rate of diabetic neuropathy among DM patients using machine learning by CRT decision tree in North West Ethiopia, 2005-2021

<table>
<thead>
<tr>
<th>Observed</th>
<th>Classification</th>
<th>Predicted</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>636</td>
<td>100.0%</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

From eight significant variables used for nomogram development, the classification and regression tree (CRT) select five most potent predictors includes other comorbidities, glycemic control, physical activity, White blood cells and red blood cells count. Presence of comorbidities was the most important node that predicts diabetic neuropathy, followed by glycemic control and physical activity; low RBC and high WBC count. For instance, from 172 having diabetic neuropathy cases, 59(43.1%) had presence other comorbidities and no doing physical activity were develop diabetic neuropathy. In classification and regression tree there were four depths, six terminal nodes used for decisions (Figure-6).

Figure 6:-Classification and regression decision tree for prediction of diabetic neuropathy among diabetes mellitus patients North West Ethiopia 2005-2021

The performance of the classification and regression decision tree was good accuracy with AUC=70.2 %( 95% CI; 65.8%, 74.6%) (Figure:-7, A) and well calibrated model (p-value -0.412) (Figure:-7, B), which means the predicting and observed probability was the same.

Figure 7:-The Performance of the model using decision tree A) AUC B) calibration plot to prediction of the risk for diabetic neuropathy among DM patients North West Ethiopia 2022.
Internal validation

In order to checking over fitting and bias of the developed model using nomogram and decision tree, bootstrapping technique were used for internal validation. The area under ROC curve was assessed based on the bootstrap dataset coefficients. The developed nomogram was internally validated had AUC; 71.7(95% CI; 67.3 %, 76.0%) (Figure-8, A) and well calibration model (p-value- 0.945)(Figure-8, B)

After internal validation the discriminating performance of nomogram was comparable with less optimism coefficients found to be 0.015, which indicates less likely over fitting model and bias. The average calculated value of optimism to the performance model was 0.0075.

**Figure 8: The performance of the model using nomogram after bootstrapping A) Area under the curve (AUC) B) calibration plot, for the probability of the risk of diabetic neuropathy among DM patients North West Ethiopia, 2005-2021.**

The performance of the classification and regression decision tree after bootstrapping was the same as the developed CRT decision tree ,good accuracy with AUC=70.2 %( 95 % CI; 65.8%, 74.6%) (Figure:-9 A) and well calibrated model (p-value:-0.389) (Figure:-9 B). The optimism coefficient was 0.000.

**Figure 9:-The performance of the model using classification and regression decision tree after bootstrapping A) AUC B) Calibration plot to predict diabetic neuropathy among DM patients northwest Ethiopia 2022.**

In both nomogram and classification and regression tree (CRT) the discrimination and calibration performance power were comparable, good accuracy so that, they were used to predict the risk of diabetic neuropathy among DM patients (Figure-10)

**Figure 10: Comparison of nomogram and decision tree for prediction of diabetic neuropathy among DM patients North West Ethiopia 2005-2021.**

Cutoff point for probability of diabetic neuropathy by nomogram and CRT

Using the nomogram, optimal cutoff point for predicted probability for risk of diabetic neuropathy by using Youden’s index(J) was 0.2447 and the sensitivity, specificity, positive predictive value and negative predictive value was 65.2%(95% CI; 58;0%, 72.2%),
71.7% (95% CI; 68.0%, 75.2%), 38.4% (95% CI; 32.7%, 44.2%), 88.4% (95% CI; 85.3%, 91.0%) respectively. (Table 7)

Table 7: Sensitivity, Specificity, PPV, NPV, LR+, LR- and Accuracy of the nomogram at different cut-off points for prediction of diabetic neuropathy among DM patients, North West Ethiopia 2005-2021.

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;=0.1</td>
<td>72.1%</td>
<td>63.5%</td>
<td>65.4%</td>
<td>35.3%</td>
<td>90.4%</td>
<td>1.9</td>
<td>0.4</td>
</tr>
<tr>
<td>&gt;=0.2</td>
<td>70.4%</td>
<td>66.4%</td>
<td>67.2%</td>
<td>36.1%</td>
<td>89.2%</td>
<td>2.1</td>
<td>0.4</td>
</tr>
<tr>
<td>&gt;=0.24</td>
<td>65.2%</td>
<td>71.7%</td>
<td>70.3%</td>
<td>38.4%</td>
<td>88.4%</td>
<td>2.3</td>
<td>0.4</td>
</tr>
<tr>
<td>&gt;=0.3</td>
<td>54.7%</td>
<td>78.6%</td>
<td>73.5%</td>
<td>41.0%</td>
<td>86.5%</td>
<td>2.5</td>
<td>0.5</td>
</tr>
<tr>
<td>&gt;=0.4</td>
<td>34.3%</td>
<td>88.4%</td>
<td>77.0%</td>
<td>44.4%</td>
<td>83.3%</td>
<td>2.9</td>
<td>0.7</td>
</tr>
</tbody>
</table>

In classification and regression tree (CRT) using optimal cutoff point for predicting the probability risk of diabetic neuropathy by using Youden’s index was 0.1847 and the sensitivity, specificity, positive predictive value and negative predictive value was 72.09 (95% CI; 64.8%, 78.6%), 57.7% (95% CI; 53.8%, 61.6%), 31.5% (95% CI; 27.0%, 36.4%), 88.4% (95% CI; 84.9%, 91.3%) respectively.

Risk classification for diabetic neuropathy using nomogram

For easily practical utility we developed nomogram by using regression coefficients in the final reduced model. The proportion of diabetes mellitus patients to risk of diabetic neuropathy at low risk was 516 (63.8%), intermediate risk 282 (34.90%) and high risk group 10 (1.24%). When the risk classified as two categories the proportions of DM patients for risk of diabetic neuropathy at low risk was 516 (63.86%) and high risk 292 (36.14%). (Table 8)

Table 8: Risk classification for predicting diabetic neuropathy among DM patients using nomogram, Northwest Ethiopia, 2005-2021 (n=808)

<table>
<thead>
<tr>
<th>Risk categories’</th>
<th>Total number of DM patients</th>
<th>Incidence of diabetic neuropathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk (&lt;0.2)</td>
<td>516 (63.86%)</td>
<td>60 (7.4%)</td>
</tr>
<tr>
<td>Intermediate risk (0.2-0.6)</td>
<td>282 (34.90%)</td>
<td>107 (13.3%)</td>
</tr>
<tr>
<td>High risk (&gt;0.6)</td>
<td>10 (1.24%)</td>
<td>5 (0.6%)</td>
</tr>
<tr>
<td>Total</td>
<td>808 (100%)</td>
<td>172 (21.3%)</td>
</tr>
</tbody>
</table>
In classification and regression tree (CRT), the proportion diabetes mellitus patients to the risk of diabetic neuropathy at low risk was 150 (18.56%), intermediate risk 636 (78.72) and high risk 22 (2.72%).

Decision Curve analysis

Besides model performance assessed by AUC, calibration plot, sensitivity, specificity and accuracy, clinical and public health utility of the model was also assessed by decision curve analysis. The developed model using nomogram and classification and regression tree (thick red line) had highest net benefit from threshold probabilities greater than 0.1 (10%) compared treatments none (thick black line), had high cost benefit ratio. The model has no or similar net benefit to treatment all (thin black line) regardless of their risk across thresholds below 0.1 (10%). (Figure-11, A, B)

Mobile-based application of nomogram

To make the nomogram accessible to physicians and patients, mobile based application tool were developed for predicting the risk of diabetic neuropathy among diabetes mellitus patients by using “sublime text” application with java script. Where physicians’ or diabetes mellitus patients enter their selected clinical and laboratory data to the developed mobile based tool, then the tool was giving alert as cutoff value 26 get from nomogram, weather they are at low risk (<26) or high risk (>=26) of diabetic neuropathy by calculating cumulating score in order to keep or improve their health condition. In Figure:-15 when we enter data to the tool had score 21.2

Figure 12:-Mobile based application outputs for predicting risk of diabetic neuropathy among DM patients Northwest Ethiopia 2005-2021.

Discussion

This risk prediction model used a multicenter retrospective follow up study to develop a practical tool for prediction of diabetic neuropathy among both type-1 and type-2 DM patients. Our study established a diabetic neuropathy risk prediction model using combined
set predictors to be used in the primary care settings, by developing nomogram and decision
tree, aid in public health and clinical decision making for clinicians and to patients.

There were challenges to diagnose diabetic neuropathy, which needs electro physical and
autonomic tests by mono filaments, is not available in low level health care system (60, 61)
Hence, predicting the risk of diabetic neuropathy in diabetes mellitus patients using easily
measurable predictors were essential to take appropriate measures accordingly.

The study finding showed that the incidence proportion of diabetic neuropathy was 21.29
%(95%CI; 18.59, 24.25), which is consistent with SRMA study in North Africa, 21.9% (18),
Ethiopia, 22% (50), Qatar 23.0%(62) but lower than Tikur Anbesa and St. Paulous
hospital,48.2% (20), SRMA in Africa 46% (15), Srilanka 28.8% (14).

These disparities might be number of participates involved with sample size difference, in
case small sample size in this study(63), difference way of life style (64) and different level of
health care system used for diagnosis. For instance, monofilament examination was not
performed in our study area, results asymptomatic patients may not detect.

The model was developed by reducing twenty-two candidate predictors to thirteen potential
predictors, which were passed to multivariable analysis. The model comprised eight variables
identified as independent predictors of diabetic neuropath including glycemic control (FBG),
other comorbidities, alcohol drinking, Hypertension, type of treatment for DM, physical
activity, WBC and RBC count. This result is supported by previous studies conducted (65),
(66), (37), (42), (47), (67).

The prediction model developed from eight predictors and nomogram was developed found
to have discrimination accuracy of AUC; 73.2% This result showed that a good accuracy
discrimination power according to diagnostic accuracy classification criteria(68). The model
calibration assessed by calibration plot and Hosmer and Lemeshow test had p-value; 0.502,
which had good agreement between observed and expected probability.

The predictive role of each predictors was assessed for the identification of individuals at
greater risk of diabetic neuropathy and its capacity of score from nomogram for glycemic
control (poor), physical activity (no), other comorbidities (yes), alcohol drinking (yes), use
combined insulin and oral drugs, hypertension (yes), WBC (high) and RBC (low) was 2.6,
10, 7, 6.4, 4, 2.6, 7.8 and 6.8 respectively.

The model was internally validated using the bootstrapping method of 1000 repetitions with
replacement. The discriminating performance of the model after internal validation had also
good discrimination power of AUC ;71.7% with well calibrated model (p value = 0.945). The
The optimism coefficient was found to be 0.015 (1.5%), ensures that the model is less likely over fitting and less sample dependent, can easily and accurately individualized prediction of the risk of diabetic neuropathy.

The performance of our model was found to be consistent with other prediction models developed to predict diabetic neuropathy among DM patients, using hypertension, age, heart rate and BMI as predictors, AUC; 71% in China (69) using hypertension, comorbidities, gender, age, obesity, abnormal triglycerides as predictors, AUC; 75% in China (70), but better than a study using glomerular filtration rate, glibenclamide and creatinine as predictors AUC = 66.05% in Mexico (71), using age, FBG, PBG, HbA1c, LDL, HDL and BMI as predictors AUC; 55.6% in China (72), using FBG, BMI, age as predictors AUC; 63.50 in Korea (73). This might be due to difference in study participants involved through socio-demographic characteristics, difference number of predictors used in the model development (74). The model calibration based on Hosmer and Lemshow test P-value; 0.945, is consistent another study 0.52 in China (72), indicates well calibrated model.

However, the discriminative performance of the model was lower than a study done in Italy using type of DM, smoke, BMI and HbA1c as predictors AUC; 76.9% (75) and AUC; 85.9% in China (40). This might sample size variation, use variety of validation techniques, the lower number of predictors incorporated in case of our study compared to others, use machine learning algorism, to select the most potent predictors increases the power of the study (74).

In our nomogram prediction score, using 0.24 as cutoff point has an acceptable level of the sensitivity; specificity was 65.2%, 71.2%, respectively. Using 0.18 as cut off point in classification and regression tree (CRT) had also acceptable level of sensitivity; specificity was 72.09% and 57.7% respectively. Thus it was better to identify diabetic neuropathy cases and possible to shift the cutoff point to increase either of the accuracy measures depending on the aim of program and availability of resources.

The benefit that the developed nomogram and decision tree would add to clinical practice was also presented in the form of a decision curve analysis. In our study the decision curve analysis showed that there was high net benefit than using treat all or treat none strategies when the threshold probability of the patient greater than 0.1 (10%) in both nomogram and classification and regression tree (CRT). However, the model is not useful for threshold probabilities below 10%. Thus, threshold probabilities are the most important components of the decision curve analysis depending on which a clinician can decide on whether to use the
Although the autonomic tests by monofilament examination and nerve conduction studies gives better diagnosis of diabetic neuropathy (76). However, prediction of diabetic neuropathy among type-1 and type-2 DM patients using clinical and behavioral characteristics of alone enabled to identify low and high risk diabetes mellitus patients. This prediction model is not a replacement of electro physical diagnosis and monofilament examination of diabetes mellitus patients to diagnose diabetic neuropathy (77); however, it will be as screening tool in resource poor settings for further diagnostic workup and management options. Besides the nomogram and decision tree is easier to use in routine clinical and public health practice than regression models and has comparable discrimination and calibration.

This study has following strengths. Firstly, it was conducted multicenter study and mobile based application tool was developed for easily applicability. Secondly, the model was internally validated using bootstrapping technique, resulted small optimism coefficient, indicating our model is less likely over fitted model, can easily predict the risk of diabetic neuropathy. Thirdly, this prediction model is constructed from easily obtainable clinical and behavioral predictors that make it applicable in primary care settings. Finally the model was also developed by classification and regression tree analysis to show the interaction effect of each predictor to the risk of diabetic neuropathy.

However, the findings from this study should be interpreted with the perspective of the following limitations. Even if multicenter studies, the model did not validate in separate datasets so that it needs external validation before using it in another context. In addition most diabetic neuropathy patient was asymptomatic may missed cases, who had no sign and symptoms. The study had many missing predictors due to retrospective study design. Generally the model will provide its maximum benefit in clinical practice provided that all the required predictor information is collected and developing mobile based application makes it the nomogram easily applicable for clinicians and to patients.

**Conclusion and recommendation**

The model was developed from precisely measured, clinical and behavioral predictors found in primary health care setting includes glycemic control (FBG), hypertension, physical activity, other comorbidities, alcohol drinking, type of treatment for DM, WBC and RBC
The model was presented by nomogram and decision tree with having good discriminating performance power and well calibrated. It was internally validated by bootstrapping techniques with small optimum coefficients, less likely of over fitting of the model. The nomogram was better than decision tree, helps us to do a risk stratification of diabetes mellitus patients and to identify those at higher risk of having diabetic neuropathy. Subsequently, high-risk groups can be linked to a center, which is equipped with electro physical diagnosis and monofilament examination centers for further assessment and better management. The nomogram and decision tree had added net benefit in clinical practice as it was assurred by decision curve analysis across different threshold probabilities. To make easily accessible and applicable of nomogram for clinicians and to patients, mobile based application tool (off line) were developed, that works for both type-1 and Type-2 DM patients to calculate their risk in order to take appropriate intervention measures timely.

Based on study findings, we recommend that for health professionals, who are providing health care services at diabetes mellitus chronic follow up clinic, use the nomogram developed by mobile based application tool to identify high risk of diabetic neuropathy patients for early management and treatment. Additionally Patients use this developed mobile based application tool for their informed choice of treatment by calculating their individualized risk either to keep or improve their health condition. Finally for future researchers improve the model by adding additional predictors like HbA1c, family history of DM and smoking through prospective study design and using monofilament examination by detect enough number of events. Besides, validate externally by using independent sample datasets to be applicable in all health institution.

### Abbreviations

- **AUC**: Area under the Curve
- **BMI**: Body Mass Index
- **CI**: Confidence Interval
- **CRT**: Classification and Regression Tree
- **DBP**: Diastolic Blood Pressure
- **DCA**: Decision Curve Analysis
- **DM**: Diabetes Mellitus
- **DPN**: Diabetic Peripheral Neuropathy
- **FBG**: Fasting Blood Glucose
- **HDL**: High Density Lip-Protein
- **HbA1c**: Glycated Hemoglobin
- **ICT**: Information Communication Technology
- **IDF**: International Diabetes Federation
- **IQR**: Inter Quartile Range
- **LR+**: Log Likelihood Ratio Positive
- **LR-**: Log Likelihood Ratio Negative
- **MNSI**: Michigan Neuropathy Screening Instrument
- **NPV**: Negative Predictive Value
- **PBG**: Post Bradial Glucose
- **PPV**: Positive Predictive Value
- **ROC**: Receiver Operating Characteristics
- **SBP**: Systolic Blood Pressure
- **RBC**: Red Blood Cells
- **SBP**: Systolic Blood Pressure
- **SRMA**: Systematic
Review and Meta-Analysis T1DM; Type-1 Diabetes Mellitus T2DM; Type-2 Diabetes Mellitus TC; Total Cholesterol TG; Total Triglyceride UOG; University of Gondar WBC; White Blood Cells

Declarations

Ethics approval and consent to participate
The study use ethical clearance from university of Gondar for the collection of information from subjects (supplementary file-S3)

Availability of data and materials
Data is available on the manuscript and as supplementary files.

Competing Interests
The authors declare that they have no conflict of interest.

Funding
Funding was obtained from Mizan Aman college of health science, Mizan-Aman, Ethiopia (25,000 Ethiopian Birr)

Author’s Contributions
Conceptualization: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Data curation: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Formal analysis: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi
Investigation: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Methodology: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Software; Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Validation: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Visualization: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Writing original draft: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

Writing – review & editing: Negalgn Byadgie Gelaw, Achenef Asmamaw Muche, Adugnaw Zeleke Alem, Yazachew Moges Chekol, Tigabu Kidie Tesfie, Tsion Mulat Tebeje, Nebiyu Bekele Gebi

All authors read and approved the final manuscript.

Acknowledgments

We greatly acknowledge Mizan Aman College of health sciences for sponsorship and university of Gondar

Supplementary file


S3: - Ethical Clearance letter obtained from university of Gondar

References
22. (FMoH) FMoH. National Strategic Action Plan (Nsap) for Prevention & Control of Non-Communicable Diseases in Ethiopia. 2016.
31. Coppini DV. Diabetic neuropathy: are we still barking up the wrong tree and is change finally in sight? Diabetologia. 2020;63(9):1949-50.
32. Mendoza-Romo MA, Ramírez-Arriola MC, Velasco-Chávez JF, Nieva-de Jesús RN, Rodríguez-Pérez CV, Valdez-Jiménez LA. [Sensitivity and specificity of a utility model of


53. Carroll MD, Kit BK, Lacher DA. Total and high-density lipoprotein cholesterol in adults; national health and nutrition examination survey, 2009-2010. 2012.


Figure 1: Flowchart of participant selection for the development of nomogram for prediction of diabetic neuropathy among DM patients in selected referral hospitals, Amhara regional state, North West Ethiopia, 2005-2021.
Figure 2: Pie chart for the proportion of diabetic neuropathy among diabetes mellitus patients North West, Ethiopia, 2005-2021.
Figure 3: The performance of the model using original beta coefficients A) AUC B) calibration plot, for prediction of diabetic neuropathy among DM patients North West Ethiopia, 2005-2021.
Figure 4: Nomogram for developing and validating risk prediction model for diabetic neuropathy among diabetes mellitus patients, northwest Ethiopia, 2005-2021.
Figure 5: Performance of model using nomogram A) AUC B) Calibration plot, prediction for the risk of diabetic neuropathy among DM patients North West Ethiopia 2005-2021.
Figure 6: Classification and regression decision tree for prediction of diabetic neuropathy among diabetes mellitus patients North West Ethiopia 2005-2021
Figure 7.- The Performance of the model using decision tree A) AUC B) calibration plot to prediction of the risk for diabetic neuropathy among DM patients North West Ethiopia 2022.
Figure 8: The performance of the model using nomogram after bootstrapping A) Area under the curve (AUC) B) calibration plot, for the probability of the risk of diabetic neuropathy among DM patients North West Ethiopia, 2005-2021.
Figure 10: Comparison of nomogram and decision tree for prediction of diabetic neuropathy among DM patients North West Ethiopia 2005-2021.
Figure 11: Decision curve plot A) Nomogram B) Decision tree, showing the net benefit of the developed model for carrying out a certain intervention measure in DM patients at risk of diabetic neuropathy Northwest Ethiopia, 2005-2021
Diabetic Neuropathy Risk Prediction Calculator

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycemic Control</td>
<td>&lt;130 mg/dL □</td>
</tr>
<tr>
<td>Other comorbidities</td>
<td>Yes □</td>
</tr>
<tr>
<td>Alcohol Drinking</td>
<td>Yes □</td>
</tr>
<tr>
<td>Hypertension</td>
<td>□</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>Yes □</td>
</tr>
<tr>
<td>WBC</td>
<td>&gt;5,000 cells/mm³ □</td>
</tr>
<tr>
<td>RBC</td>
<td>&gt;4 million cells/mm³ □</td>
</tr>
</tbody>
</table>

Figure 12: Mobile based application outputs for predicting risk of diabetic neuropathy among DM patients Northwest Ethiopia 2005-2021.
Figure 9: The performance of the model using classification and regression decision tree after bootstrapping A) AUC B) Calibration plot to predict diabetic neuropathy among DM patients northwest Ethiopia 2022.