 doi:10.20944/preprints202111.0303.v1

Article

Machine Learning for Screening Microvascular Complications in Type-2 Diabetic Patients using Demographic, Clinical, and Laboratory Profiles

Mamunur Rashid[1,‡], Mohanad Alkhodari[2,‡,*], Abdul Mukit[1,3], Khawza Iftekhar Uddin Ahmed[1], Raqibul Mostafa[1], Sharmin Parveen[4], and Ahsan H Khandoker[2]

[1] Department of Electrical and Electronic Engineering, United International University, Dhaka, Bangladesh; mrashid152005@bseee.uiu.ac.bd; khawza@eee.uiu.ac.bd; rmostafa@eee.uiu.ac.bd
[2] Department of Biomedical Engineering, Khalifa University, Abu Dhabi, United Arab Emirates; mohanad.alkhodari@ku.ac.ae; ahsan.khandoker@ku.ac.ae
[3] Department of Electrical and Computer Engineering, University of Oklahoma, Tulsa, Oklahoma, USA; abdul.mukit64@gmail.com
[4] Department of Health Informatics, Bangladesh University of Health Sciences, Dhaka, Bangladesh; sharminparveen@yahoo.com
‡ These authors contributed equally to this work
* Correspondence: mohanad.alkhodari@ku.ac.ae

Abstract: Microvascular complications are one of the key causes of mortality among type-2 diabetic patients. This study was sought to investigate the use of a novel machine learning approach for predicting these complications from patient demographic, clinical, and laboratory profiles only. A total of 96 Bangladeshi participants having type-2 diabetes were recruited during their routine hospital visits. All patient profiles were assessed by using a Chi-squared (χ^2) test to statistically determine the most important markers in predicting four microvascular complications; namely cardiac autonomic neuropathy (CAN), diabetic peripheral neuropathy (DPN), diabetic nephropathy (NEP), and diabetic retinopathy (RET). A machine learning approach based on random forest (RF) and support vector machine (SVM) was then developed to ensure automated clinical testing for microvascular complication in diabetic patients. The highest prediction accuracies were obtained by RF using diastolic blood pressure, Albumin-Creatinine ratio, and gender for CAN testing (98.67%), Microalbuminuria, smoking history, and hemoglobin A1C for DPN testing (67.78%), Albumin-Creatinine ratio for NEP testing (100%), and hemoglobin A1C, Microalbuminuria, and smoking history for RET testing (84.38%). This study suggests machine learning as a promising automated tool for predicting microvascular complications in diabetic patients using their profiles, which could help prevent those patients from further microvascular complications leading to early death.

Keywords: Microvascular complications; Cardiac autonomic neuropathy; Diabetic peripheral neuropathy; Diabetic nephropathy; Diabetic retinopathy; patient profiles; machine learning

1. Introduction

Diabetes is called a ‘silent killer’ that is killing around 1.6 million people each year making it the 5th leading cause of death worldwide[1]. There are two types of diabetes, type-1 and type-2. Type-2 is a chronic metabolic disorder, an expanding global health problem from the past decades. It results in hyperglycemia which reduces body cells' ability to respond fully to insulin. This situation is called ‘insulin resistance’. In this state, insulin production of the body increases due to the inactiveness of the hormone. The global impact of type-2 diabetes prevalence in low and middle-income countries was estimated to be 415 million in 2015 and predicted to rise to 642 million in 2040[2]. Type-2 diabetes mellitus has been rapidly rising worldwide over the past three decades, particularly in developing countries including Bangladesh[3]. Type-2 diabetes prevalence in Bangladesh will be more than 50% by the next 15 years, placing Bangladesh as the 8th highest diabetic populous country in the world[4]. A study suggests that diabetic prevalence will be more than double between 2020 to 2030[5]. IDF (International Diabetes Federation) Diabetes Atlas has estimated that if nothing is done, the number of diabetes may rise to 629 million in 2045[6] which should be double of 151 million[7] from 2000 to 2025[8]. The prevalence of diabetes is higher in rural areas[9] but it was high for males in urban areas whereas lower in rural areas compared to females in Bangladesh[10], [11].

Neuropathies are a common persistent complication of both types of diabetes mellitus which conferred morbidity and mortality to diabetic patients. Cardiac autonomic neuropathy (CAN) is associated with an increased risk of mortality[12], [13]. A study including 1171 patients with type-1 and type-2 diabetes mellitus using predefined HRV and spectral analysis of R-R intervals reported abnormal findings for 34.3% of type-2 patients[14]. Neuropathy is the most common microvascular complication of both type-1 and type-2 diabetes mellitus[15]–[17]. A study found that 19.7% of total registered type-2 patients have diabetic peripheral neuropathy (DPN) conducted in the outpatient section of BIRDEM Hospital, Dhaka, Bangladesh[18]. The prevalence of DPN among type-2 diabetic patients is much higher in Europe. A study concludes that 32.1% of diabetic patients in the United Kingdom, 17.6% in Turkey, and 35.4% in Spain have DPN[19]. The prevalence of DPN increases with the age of the patient and also with the diabetic duration[20], [21]. A multi-country study conducted in Asia shows a 58.6% prevalence of micro or macroalbuminuria indicating an impending pandemic of diabetic renal (i.e., nephropathy) and cardiovascular diseases in Asia[22]. A cross-sectional study with 836 rural Bangladeshi patients shows a high prevalence of retinopathy in Bangladesh[23]. Results from 35 studies from 1980 to 2008 with 22,896 subjects with diabetes showed that the global prevalence for any RET was 34.6% (95% CI 34.5 – 34.8)[24]. Analyses of the exponential trend revealed an increase in diabetes prevalence among the urban and rural populations at a rate of 0.05% and 0.06% per year, respectively[25]. Increasing age, hypertension, and higher BMI were found to be significant risk factors in the urban and rural communities of Bangladesh[26]. But the patients with diabetic type-2 in Bangladesh have limited knowledge of its risk factors, cause, and also management[27], [28]. Depressive diabetic symptoms were found in 29% of males and 30.5% of female participants with diabetes and 6.0% of males and 14.6% of female subjects without diabetes[29].

Most recently, machine learning has emerged in many biomedical applications as a promising tool to aid in decision-making regarding many diseases including diabetes. In [30], authors have managed to implement a machine learning approach based on decision trees to identify diabetic patients with or without treatment procedures from their lipids profiles. In addition, Koren *et al.* [30] have developed a trained model capable of diagnosing diabetic patients with drugs that cause lowering blood glucose levels. Moreover, in [31], [32], authors have proposed a deep neural network to diagnose diabetic patients from clinical profiles. To recognize patterns among diabetic patients, Alloghani *et al.* [32] presented several machine learning models that were able of characterizing patients and explain the re-admission procedures. Several other studies [33]–[36] have utilized machine learning and deep neural network in many other applications in diabetes diagnostics. However, even though the implementation of machine learning models for diabetes diagnostics showed high levels of performance, there is still a lack of knowledge about its impact on discriminating between various microvascular complications. In addition, it is essential to be able of determining, both statistically as well as from a machine perspective, which features play a critical role in characterizing these complications in type-2 diabetic patients.

In this paper, a study is conducted to investigate the efficiency of applying a machine learning-based approach in discriminating between diabetic patients according to their microvascular complications status (Figure 1). The novelty of the presented approach lies in utilizing only demographic, laboratory, and clinical information of patients within the framework of machine learning for diabetes diagnostics. Therefore, time-consuming clinical testing using advanced medical equipment can be avoided, which is essential in communities with economic hardship or lack of clinical expertise. In addition, the proposed study allows for elaborating on the most important information within patient profiles when testing for each microvascular complication. To the best of the authors' knowledge, there have been very limited attempts towards identifying certain types of microvascular complications using machine learning. Therefore, there still exists a gap in the literature about how certain patient information impacts the discrimination between diabetes complications. The presented study herein provides a complete clinical testing approach for CAN, DPN, NEP, and RET positive cases by looking into patient information from a machine-based perspective. Further, with a focus on CAN cases, the study investigates the ability of trained models in deeply discriminating between CAN-only patients and patients with additional complications alongside CAN.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

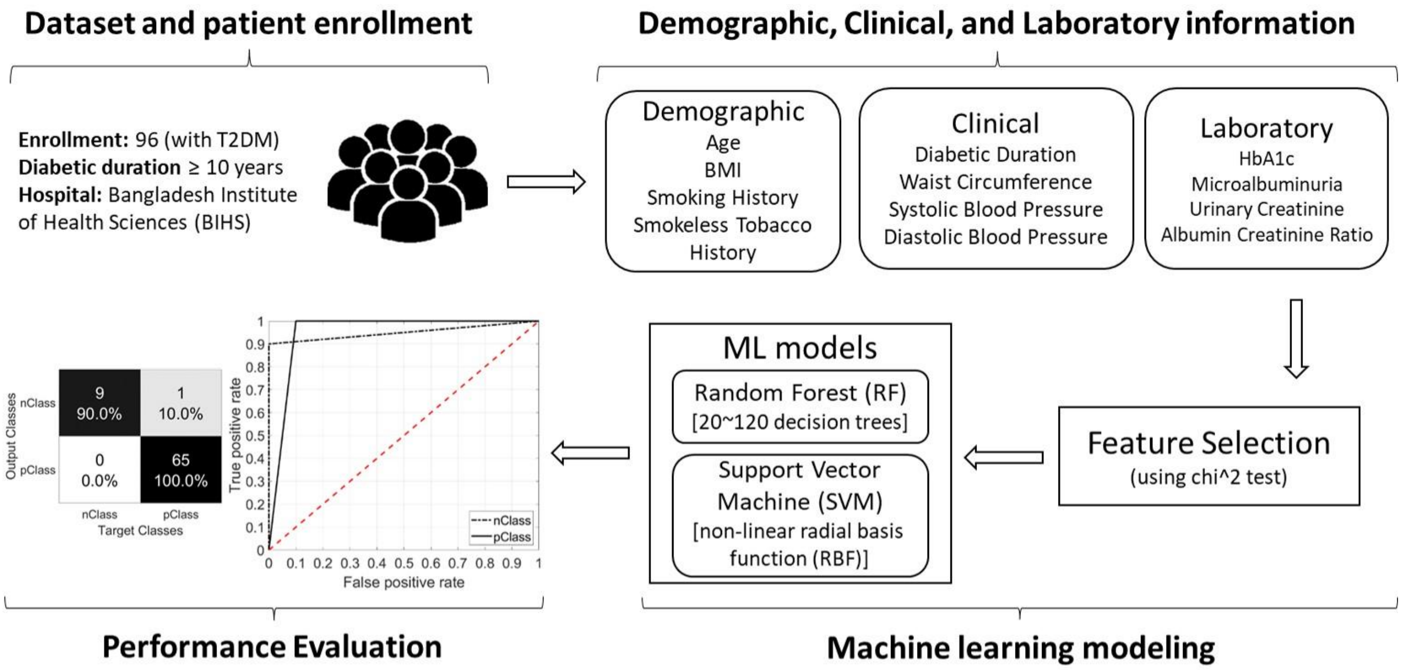

Figure 1: A graphical view of the complete research work in this study; patient enrollment, demographic clinical and laboratory information acquisition, machine learning modeling, and performance evaluation of the model

2. Materials and Methods

2.1. Study Type

This is a cross-sectional study of Bangladeshi patients from Dhaka with type-2 diabetes mellitus having more than 10 years. We followed the STROBE cross-sectional reporting guidelines[37]. The study was approved by the ethical review committee of the Bangladesh University of Health Sciences (BUHS/BIO/EA/17/01) and conforms to the ethical principles outlined in the declaration of Helsinki and the Ministry of Health and Family Welfare of Bangladesh.

2.2. Inclusion and Exclusion Criteria

The parameters that are included in the inclusion criteria: Bangladeshi national, diagnosis of type-2 diabetes mellitus, above 40 years of age, who were able to give written consent, and the diabetes duration was 10 years or more. The exclusion criteria include: stroke history, having any heart disease, not being able to give consent, diabetes duration less than 10 years, and the presence of any other pathophysiology that may lead to one or more similar complications such as cancer.

2.3. Participants and Complications

One hundred and three (47 males and 56 females), unrelated patients of age more than 40 years and having type-2 diabetes for 10 years or more were randomly selected and enrolled in the study during routine visits to the BIHS[38] Hospital between 18th December 2017 to 26th April 1018. This hospital is one of the most visited hospitals for diabetic patients in Bangladesh.

In this study, the recruited patients were diagnosed with complications such as CAN, DPN, NEP, and RET. The presence of these complications was confirmed by a qualified physician, based on the criteria outlined by the report of the WHO consultation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

group[39]. Diagnosis of cardiac autonomic neuropathy (CAN) was obtained from the Ewing test, which included five tests: deep breathing, lying to standing, Valsalva maneuver, lying to standing BP, and sustained handgrip BP[40]. Diagnosis of diabetic peripheral neuropathy (NCV) was performed by using a nerve conduction velocity (NCV) test. There were several tests for recognizing polyneuropathy, CTS (carpal tunnel syndrome), peroneal neuropathy, and other types of neuropathies. Diagnosis of nephropathy (NEP) was set by the ACR (albumin creatinine ratio) level >30 mg/mmol for microalbuminuria, and >300 mg/mmol for macroalbuminuria[41]. Diagnosis of retinopathy (RET) was obtained from the fundus image test and classified according to WHO criteria[42]. Fundus imaging is a process where 3-D retinal semi-transparent tissues projected onto the imaging plane are obtained using reflected light and represented in 2-D[43].

Among these subjects, 70 were able to complete diagnostic tests for all 4 complications (CAN, DPN, NEP, and RET). There were several combined complications found in some patients. The frequency of complications is shown in Table 2. To observe the importance of demographic, clinical, and laboratory profiles, a multiclass analysis is done using the classes marked in bold in the given table 2 (CAN vs CAN+DPN vs CAN+DPN+Others). CAN+DPN+Others are the combinations of CAN+DPN+NEP, CAN+DPN+RET, and CAN+DPN+NEP+RET.

Table 1: Types of complications of patients included in this study

Name of the Complication	Type	Number of Patients, N (%)
CAN	pCAN (with CAN)	65 (67.708)
	nCAN (without CAN)	10 (10.417)
	Test result unavailable	21 (21.875)
DPN	pDPN (with DPN)	44 (45.833)
	nDPN (without DPN)	46 (47.917)
	Test result unavailable	6 (6.250)
Nep	pNep (with Nep)	26 (27.083)
	nNep (without Nep)	70 (72.917)
Ret	pRet (with Ret)	7 (7.292)
	nRet (without Ret)	89 (92.708)

Table 2 Types and frequency of complications of diabetes patients

Types of Complications		Numerals, N(%)	Total, N
nComp (without any type of complication)		4 (5.714)	70
Single Complications	**CAN**	**21 (30.00)**	
	DPN	3 (4.286)	
	Nep	0 (0.00)	
	Ret	0 (0.00)	
	CAN, DPN	**16 (22.857)**	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

Combined Complications	CAN, Nep	6 (8.571)
	DPN, Nep	2 (2.857)
	CAN, DPN, Nep	**12 (17.143)**
	CAN, DPN, Ret	**2 (2.857)**
	CAN, DPN, Nep, Ret	**4 (5.714)**

2.4. Types of Variables

2.4.1. Demographic and clinical variable

Demographic data were collected at the time of enrollment from the patients. We measured the waist circumference, height, and weight at the time of enrollment and listed the value for the diabetic duration, age, gender, smoking history, and smokeless tobacco history. All of these data were verified from the necessary and relevant documents. Clinical data were measured at the time of enrollment. The blood pressure was measured on the first day before starting their Ewing test. If the systolic blood pressure is >130mm Hg and diastolic blood pressure is >80 mm Hg or taking antihypertensive medications, is called hypertension. Dyslipidemia was diagnosed from the medications of the patient or by checking the history of dyslipidemia of that patient. The data and its basic analysis are shown in Table 3.

Table 3: Demographic and clinical profile of patients.

Demographic variables						
Variables and their subdivisions	Male		Female		All	
	Mean ± SD	N (%of M)	Mean ± SD	N (% of F)	Mean ± SD	N(% of total)
Patients		47(45.63)		56(54.37)		103(100)
Age (years)	57.70±9.78	47(100)	54.60±7.93	56(100)	56.01±8.91	103(100)
≥40 and <50	44.8±3.22	*10(21.28)*	45.6±2.95	*15(26.79)*	45.28±3.02	*25(24.27)*
≥50 and <60	53.2±2.7	*15(31.91)*	52.86±3.17	*22(39.29)*	53±2.95	*37(35.92)*
≥60	66.63±4.78	*22(46.81)*	63.73±3.79	*19(33.93)*	65.29±4.54	*41(39.80)*
CAN	58.74±9.63	31(65.95)	53.32±7.40	37(66.07)	55.79±8.85	68(66.01)
DPN	58.95±10.33	21(44.68)	52.58±6.33	24(42.85)	55.55±8.93	45(43.68)
Nep	58.5±10.37	12(25.53)	54.37±8.75	16(28.57)	56.14±9.52	28(27.18)
Ret	56.8±11.64	5(10.63)	47.5±0.707	2(3.571)	54.14±10.54	7(6.796)
BMI (kg/m2)	25.53±3.47	47(100)	27.93±5.08	56(100)	26.84±4.56	103(100)
Underweight: <18.5	0	0(0)	0	0(0)	0	0(0)
Normal: ≥ 18.5, <25	23.54±1.45	**27(57.45)**	22.93±1.69	17(30.36)	23.31±1.56	44(42.72)
Overweight: ≥ 25.0, <30	26.54±1.03	15(31.91)	27.58±1.32	**24(42.86)**	27.18±1.31	39(37.86)
Obese: ≥ 30	33.23±4.09	5(10.638)	34.18±4.77	15(26.79)	33.94±4.52	20(19.42)
CAN	26.26±3.71	31(65.95)	27.94±5.82	37(66.07)	27.17±5.01	68(66.01)
DPN	25.52±3.56	21(44.68)	28.75±5.01	24(42.85)	27.24±4.64	45(43.68)
Nep	26.17±4.22	12(25.53)	29.18±5.60	16(28.57)	27.89±5.19	28(27.18)
Ret	26.79±5.53	5(10.63)	26.29±2.09	2(3.571)	26.65±4.60	7(6.796)
Smoking history		9(19.15)		0(0)		**9(8.74)**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

Name of the Variables and their subdivisions	Male		Female		All	
Smokeless tobacco history		10(21.28)		17(30.357)		**27(26.21)**
Clinical variables						
	Mean ± SD	N (%of M)	Mean ± SD	N (% of F)	Mean ± SD	N(% of total)
Diabetes duration (years)	16.17±6.07	47(100)	15.55±5.76	56(100)	15.83±5.88	103(100)
≥10 and <20	13.54±2.76	37(78.72)	12.60±2.64	41(73.21)	13.05±2.73	78(75.73)
≥20 and <30	24±3.116	8(17.02)	22.30±1.93	13(23.21)	22.95±2.52	21(20.39)
≥30	33.5±2.12	2(4.26)	32±2.828	2(3.57)	32.75±2.22	4(3.88)
CAN	16.54±6.20	31(65.95)	16.13±6.01	37(66.07)	16.32±6.05	68(66.01)
DPN	17.33±7.43	21(44.68)	**14.16±4.80**	24(42.85)	15.64±6.30	45(43.68)
Nep	18.91±8.11	12(25.53)	16.81±6.63	16(28.57)	17.71±7.24	28(27.18)
Ret	13±2.828	5(10.63)	17.5±3.535	2(3.571)	14.28±3.49	7(6.796)
Waist Circumference (cm)	90.84±8.61	47(100)	97.38±9.46	56(100)	94.39±9.61	103(100)
Men ≥90	97.40±6.7	23(48.94)				
Women ≥80			97.72±9.19	**55(98.21)**		
CAN	92.09±8.47	31(65.95)	96.58±9.30	37(66.07)	94.54±9.15	68(66.01)
DPN	92.64±8.13	21(44.68)	**98.63±9.07**	24(42.85)	95.84±9.06	45(43.68)
Nep	91.22±6.71	12(25.53)	97.31±9.80	16(28.57)	94.70±9.00	28(27.18)
Ret	89.91±5.26	5(10.63)	93.98±14.36	2(3.571)	91.07±7.53	7(6.796)
Systolic blood pressure(mmHg)	141.2±19.5	47(100)	136.0±20.14	56(100)	138.4±19.94	103(100)
≤119	108±5.29	4(8.51)	108.3±8.96	12(21.43)	108.2±8.03	16(15.53)
≥120 and <140	129.2±6.67	19(40.43)	130.1±4.98	19(33.93)	129.7±5.82	38(36.89)
≥140 and <160	148.2±7.52	15(31.91)	148.3±5.71	19(33.93)	148.2±6.47	34(33.01)
≥160	169.6±9.72	9(19.15)	171.3±6.40	6(10.714)	170.3±8.33	15(14.56)
CAN	145.0±20.16	31(65.95)	134.0±21.30	37(66.07)	139.0±21.35	68(66.01)
DPN	148.5±20.82	21(44.68)	134.8±15.96	24(42.85)	141.2±19.43	45(43.68)
Nep	153.0±15.16	12(25.53)	136.1±17.22	16(28.57)	143.4±18.19	28(27.18)
Ret	158.6±16.14	5(10.63)	137.5±17.67	2(3.571)	152.5±18.21	7(6.796)
Diastolic blood pressure(mmHg)	78.97±9.86	47(100)	76.42±11.96	56(100)	77.59±11.07	103(100)
≤79	71.36±7.45	22(46.81)	67.96±6.98	32(57.14)	69.35±7.30	54(52.43)
≥80-89	82.73±2.83	19(40.43)	83.81±3.08	16(28.57)	83.22±2.95	35(33.98)
≥90-99	94±3.39	5(10.64)	94.14±2.61	7(12.5)	94.08±2.81	12(11.65)
≥100	100±0	1(2.13)	105±0	1(1.79)	102.5±3.54	2(1.94)
CAN	78.45±11.40	31(65.95)	75.48±12.76	37(66.07)	76.83±12.16	68(66.01)
DPN	78.19±12.23	21(44.68)	76.87±10.63	24(42.85)	77.48±11.29	45(43.68)
Nep	74.91±13.48	12(25.53)	75.93±10.81	16(28.57)	75.5±11.79	28(27.18)
Ret	84.6±10.13	5(10.63)	72.5±3.54	2(3.571)	81.14±10.27	7(6.796)

2.4.2. Laboratory data

 doi:10.20944/preprints202111.0303.v1

The laboratory data were taken from the laboratory of the hospital after the enrollment. Laboratory test parameters are hemoglobin A1c (HbA1c), microalbuminuria, urinary creatinine, and albumin creatinine ratio. The data and its basic analysis are shown in Table 4.

Table 4: Laboratory variable of patients.

Types and their variables	**Male**		**Female**		**All**	
	Mean ± SD	N (%of	Mean ± SD	N (% of	Mean ± SD	N(% of
			HbA1c(%,mmol/mol)			
Not specified	9.066±1.944	47(45.63)	8.621±1.453	56(54.37)	8.824±1.701	103(100.0)
Optimal: <7		2(4.26)		8(14.29)		10(9.71)
Fair: 7-8		12(25.53)		11(19.64)		23(22.33)
High: >8		33(70.21)		37(66.07)		70(67.96)
CAN	9.213±1.790	31(45.59)	8.716±1.491	37(54.41)	8.943±1.640	68(66.02)
Optimal: <7		1(3.23)		4(10.81)		5(7.35)
Fair: 7-8		6(19.35)		8(21.62)		14(20.59)
High: >8		24(77.42)		25(67.57)		49(72.06)
DPN	9.291±1.988	21(46.67)	8.930±1.667	24(53.33)	9.098±1.810	45(43.69)
Optimal: <7		2(9.52)		3(12.50)		5(11.11)
Fair: 7-8		3(14.29)		4(16.67)		7(15.56)
High: >8		16(76.19)		17(70.83)		33(73.33)
Nephropathy	9.9750±2.221	12(42.86)	8.763±1.902	16(57.14)	9.282±2.094	28(27.18)
Optimal: <7		1(8.33)		3(18.75)		4(14.29)
Fair: 7-8		1(8.33)		4(25.00)		5(17.86)
High: >8		10(83.33)		9(56.25)		19(67.86)
Retinopathy	10.720±3.334	5(71.43)	11.100±1.980	2(28.57)	10.829±2.846	7(6.80)
Optimal: <7		0(0.00)		0(0.00)		0(0.00)
Fair: 7-8		2(40.00)		0(0.00)		2(28.57)
High: >8		3(60.00)		2(100.00)		5(71.43)
			Microalbuminuria (mg)			
Not specified	60.6164±99.49	47(46.08)	49.571±82.123	55(53.92)	54.661±90.247	102(99.03)
Optimal: <30		34(72.34)		38(69.09)		72(70.59)
Microalbuminuria: 30-300		10(21.28)		15(27.27)		25(24.51)
Macro albuminuria: >300		3(6.38)		2(3.64)		5(4.90)
CAN	88.439±113.17	31(45.59)	56.981±93.199	37(54.41)	71.322±103.204	68(66.02)
Optimal: <30		18(58.06)		25(67.57)		43(63.24)
Microalbuminuria: 30-300		10(32.26)		10(27.03)		20(29.41)
Macro albuminuria: >300		3(9.68)		2(5.41)		5(7.35)
DPN	121.925±124.4	21(47.73)	55.2565±87.479	23(52.27)	87.075±110.720	44(42.72)
Optimal: <30		10(47.62)		15(65.22)		25(56.82)
Microalbuminuria: 30-300		8(38.10)		7(30.43)		15(34.09)
Macro albuminuria: >300		3(14.29)		1(4.35)		4(9.09)
Nephropathy	210.308±91.41	12(42.86)	144.519±98.407	16(57.14)	172.7143±99.41	28(27.18)
Optimal: <30		0(0.00)		1(6.25)		1(3.57)
Microalbuminuria: 30-300		9(75.00)		13(81.25)		22(78.57)
Macro albuminuria: >300		3(25.00)		2(12.50)		5(17.86)
Retinopathy	158.62±140.29	5(71.43)	136.15±178.691	2(28.57)	152.20±136.247	7(6.80)
Optimal: <30		2(40.00)		1(50.00)		3(42.86)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

Microalbuminuria: 30-300		2(40.00)		1(50.00)		3(42.86)
Macro albuminuria: >300		1(20.00)		0(00.00)		1(14.28)
Urinary Creatinine (mg/dl)						
Not specified	194.46±139.83		130.87±117.85		160.17±131.70	102(99.03)
Target 20-320 mg/dl		41(87.23)		50(90.91)		91(89.22)
Non-Target >320mg/dl		6(12.77)		4(7.27)		10(9.80)
CAN	236.15±150.39	31(45.59)	123.28±107.24	37(54.41)	174.74±139.68	68(66.02)
Target 20-320 mg/dl		25(80.65)		34(91.89)		59(86.76)
Non-Target >320mg/dl		6(19.35)		2(5.41)		8(11.76)
DPN	236.84±160.20	21(47.73)	157.52±149.63	23(52.27)	195.34±158.11	44(42.72)
Target 20-320 mg/dl		17(80.95)		20(86.96)		37(84.09)
Non-Target >320mg/dl		4(19.05)		3(13.04)		7(15.91)
Nephropathy	256.43±205.44	12(42.86)	152.65±77.99	16(57.14)	197.13±152.68	28(27.18)
Target 20-320 mg/dl		9(75.00)		16(100.0)		25(89.29)
Non-Target >320mg/dl		3(25.00)		0(0.00)		3(10.71)
Retinopathy	211.36±55.58	5(71.43)	159.95±135.98	2(28.57)	196.67±75.96	7(6.80)
Target 20-320 mg/dl		5(100.0)		2(100.0)		7(100.0)
Non-Target >320mg/dl		0(0.00)		0(0.00)		0(0.00)
Albumin Creatinine Ratio (mg/mmol)						
Not Specified	32.09±52.45	47(46.08)	39.28±74.58	55(53.92)	35.97±65.11	102(99.03)
Optimal: <3		12(25.53)		10(18.18)		22(21.57)
Borderline high: 3-30		23(48.94)		29(52.73)		52(50.98)
High: >30		12(25.53)		16(29.09)		28(27.45)
CAN	44.35±60.99	31(45.59)	45.36±86.19	37(54.41)	44.90±75.22	68(66.02)
Optimal: <3		7(22.58)		6(16.22)		13(19.12)
Borderline high: 3-30		12(38.71)		20(54.05)		32(47.06)
High: >30		12(38.71)		11(29.73)		23(33.82)
DPN	60.73±68.22	21(47.73)	35.97±51.28	23(52.27)	47.79±60.55	44(42.72)
Optimal: <3		6(28.57)		5(21.74)		11(25.00)
Borderline high: 3-30		4(19.05)		11(47.83)		15(34.09)
High: >30		11(52.38)		7(30.43)		18(40.91)
Nephropathy	105.960±57.95	12(42.86)	111.404±109.67	16(57.14)	109.071±89.771	28(27.18)
Optimal: <3		0(0.00)		0(0.00)		0(0.00)
Borderline high: 3-30		0(0.00)		0(0.00)		0(0.00)
High: >30		12(100.0)		16(100.0)		28(100.0)
Retinopathy	86.567±87.999	5(71.43)	58.923±61.616	2(28.57)	78.671±77.312	7(6.80)
Optimal: <3		0(0.00)		0(0.00)		0(0.00)
Borderline high: 3-30		2(40.00)		1(50.00)		3(42.86)
High: >30		3(60.00)		1(50.00)		4(57.14)

2.5. Machine Learning Modeling

2.5.1. Clinical Testing Approach

To provide a complete diagnosis of a diabetes type-2 patient, 5 tests in two steps were applied sequentially (Figure 2) on patients' demographic, clinical, and laboratory information. This study supports diabetic type-2 patients with microvascular complications to have a better screening from their DCL information. The approach combines a single-class binary classification model with four different classifiers and a multi-class classification model. The single class classification model can run four tests parallelly to classify CAN, DPN, NEP, and RET separately. If all four tests result in a negative class means the patient with type-2 diabetes has no microvascular complication.

If the test shows positive results, the patient goes for that specific complication treatment. However, resulting in a positive class from the CAN test leads to a multiclass classification model. This model allows knowing whether the patient has other microvascular complications with him along with CAN. Thus, this model results: CAN (having only CAN), CANDPN (having DPN with CAN), CANDPN+ (having NEP or RET with CAN and DPN). The resulting class determines the treatment that should be provided to the patient.

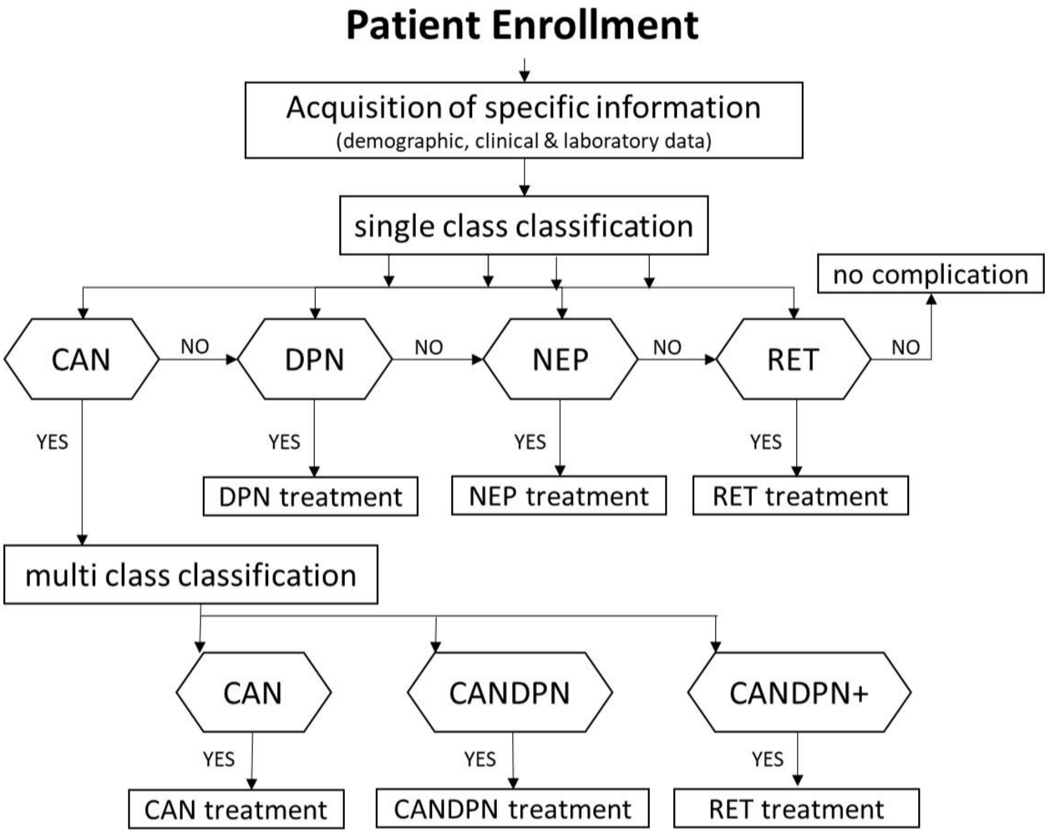

Figure 2: The proposed procedure for screening diabetic patients. Every patient initially goes through information acquisition of this clinical diagnosis flowchart. Five tests are then applied in two stages. The second stage (multiclass class is only for the patients who go through the CAN test and have positive CAN. Single class classification can predict the presence of microvascular complications (CAN, DPN, NEP, or RET) and can predict whether there is any presence of complications. Multiple complications with CAN could be classified using the multiclass classifier.

2.5.2. Analysis of Demographic Clinical and Laboratory profiles

Demographic variable (such as gender, height, age, weight, smoking history, tobacco history, and diabetes duration), clinical measurement (waist circumference, BMI, systolic blood pressure, and diastolic blood pressure), measured laboratory values (such as HbA1c, microalbuminuria, urinary creatinine, albumin creatinine ratio) were selected for further analysis as patient information.

A feature selection approach was then followed based on the univariate Chi-squared test to choose the foremost critical factors among all the demographic, clinical, and laboratory variables. In this test, a statistical hypothesis investigation is performed for each DCL feature to test whether the observed calculations coordinate with the anticipated ones, i.e. patient's complication type. Moreover, it gives a noteworthy distinction p-value measure (p-value <0.05) between categories based on the statistical calculations and desire[44]. Feature having a lower p-value signifies that this variable is most likely dependent on the complication label, hence, it is vital for anticipating the complication and has discriminatory characteristics. In this way, a score of significance is returned for each DCL

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

profile utilized within the test as score = -log(p). In this work, we call this score as importance. We calculated this importance using a function called fscchi2() in MATLAB 2021a.

2.5.3. Support vector machine (SVM)

SVM is an exceedingly popular machine learning algorithm in the use of classification and regression problems. It is one of the classic machine learning techniques that can help to solve big data classification problems. SVM allows classifying single-class as well as multi-class classification problems. It is commonly utilized as an exception finder, where the model is prepared to recognize training data from any other irrelevant information[45]. The model tends to distinguish which unused objects are closely representing the selected class in the training phase, which is generally called a positive class[46]. A set of probabilities has been returned by the model to show the degree of matching between the testing and training samples. In this paper, a single-class SVM was used for the training model in CAN DPN, NEP, and RET tests. Having the complication has been considered as the positive class in the single-class classification. However, a multi-class SVM was for training in the CANDPNOthers test. To guarantee the highest performance from the model, a non-linear RBF (radial basis function) kernel has been used with fine-tuned hyper-parameters.

2.5.4. Random forest (RF)

Random forest(RF) is also known as classification and regression tree (CART), is a form of decision trees, where a set of tree-like trait nodes is associated by a set of sub-trees of decision nodes[47][48]. This algorithm is considered a conglomeration strategy that employs the concepts of bagging. All the decision tree is calculated based on the corresponding resource cost, outcome chances, and utility to provide a prediction. The prediction preparation begins by doling out an occasion at each tree to its root node. At that point, for each of the taking after sub-nodes, the results are calculated successively. Once a leaf is experienced, the tree-like nodes halt and an occasion is relegated with a prediction. The whole occasions and predictions shape the ultimate choice made by the tree model[49]. In this work, 20 – 120 decision trees were utilized to construct the model. The choice of the number of trees for each single-class test as well as the multi-class test was fine-tuned to guarantee the greatest conceivable performance from the model.

2.5.5. Training and testing

A leave-one-out scheme was followed in the single-class models as well as in the multi-class model to ensure the incorporation of the highest possible number of samples within the prepared models. Besides, it was fundamental to supply a prediction for each and every patient. An iterative process is applied in this scheme by selecting one subject as testing data, whereas the remaining subjects are used for training. The method repeats on each cycle until a prediction is given for every subject.

2.5.6. Parameters optimization

In each test, several model parameters were fine-tuned to ensure the highest acquirable model performance. Performance was measured in the form of accuracy, sensitivity, specificity, precision, f1-score and are under the curve(AUC). To handle data imbalance (65 positive class vs 10 negative class in CAN test; 7 positive class vs 89 negative class in RET test), a model parameter called 'prior probability' was introduced in the algorithm during the training phase. The prior probabilities were found observationally, where the initial weight was set to each class that equals to its number of samples relative to the whole number of samples[50]. Prior probability was not used in DPN and NEP tests due to having a balanced class.

3. Results

3.1. Demographic, Clinical, and Laboratory Profiles

Demographic, and clinical data along with major comorbidities with type-2 diabetes are shown in Table 3, and laboratory profiles are shown in Table 4. There were 47 (45.63%) male patients and 56 (54.37%) female patients. The mean age of the patients was 56 years (±8.913), the mean age of the male and female patients was 57.1 years (±9.78) and 54.6 years (±7.93), respectively. It is consistent with the finding that the diabetic population in Bangladesh, as well as south Asia, are comparatively younger than the west[51], [52]. The sub-variables under ‘Age' show that 46.8% of the male subjects are greater than 60 years old but about 40% female subjects aged between 40 to 50 though overall patient shows the increasing prevalence for a higher age. A study in Spain also shows that an increase in patient age increases the prevalence of diabetic complications[19]. 27(57.45%) males, 35(62.50%) females, and a total of 62(60.19%) patients had a history of hypertension (mean systolic blood pressure was 138.4 mm Hg). A total of 35(33.98%) patients had dyslipidemia where 14(29.79%) were male and 21(37.5%) were female. A very low number of patients 9(8.74%) had a smoking history where everyone was male. Also, overweight condition (42.86%) is common for female diabetic patients. Having waist circumference higher than 80cm for more than 98% female subjects. 57.45% of the male subject has normal weight. Though obesity is relatively common for female patients (27%), a total of 20(19.42%) patients had obese (mean BMI (body mass index) = 33.94 kg/m^2, mean waist circumference = 90.84 cm for male and 97.38 cm for female. For the Retinopathy patients, waist circumference was 89.91 cm for males and 93.98 cm for females) where 15 (26.79%) were female and 5(10.638%) were males.

More than 67% of patients for any type of complication had a high HbA1c (mean HbA1c = 8.824, for male mean HbA1c = 9.066, and female mean HbA1c = 8.621 for the patients with CAN). The retinopathy patients had very high HbA1c (mean HbA1c = 10.829, for male mean HbA1c = 10.720, and for female mean HbA1c = 11.100). Microalbuminuria was found in 25(24.51%) patients where 10 were male and 15 were female. In the case of nephropathy total of 22(78.57%) patients had microalbuminuria. All the retinopathy patients had a creatinine level of 20 to 320 mg/dl. The mean ACR (albumin creatinine ratio) for the patients is 35.967 mg/mmol where 47(46.08%) males had the mean ACR of 32.092 mg/mmol and 55(53.92%) females had a mean ACR of 39.280 mg/mmol. Neuropathy is the most common complication in Bangladeshi diabetic type-2 patients of more than 40 years old who have diabetes duration of more than 10 years. Besides, there are very few retinopathy patients, so it implies that the rate of retinopathy in Bangladeshi diabetes type-2 patients is very low.

3.2. Complications of Type-2 Diabetes

Overall, more than one clinically diagnosed complications were present in 99 subjects out of 103 diabetes cohort included in this study. Most of the subjects had CAN (66.02%) followed by diabetic peripheral neuropathy (43.69%), nephropathy (27.18%), and retinopathy (6.8%). Those patients who had retinopathy also had CAN and DPN. The rate of retinopathy complication is very low. Only 7 retinopathy patients were found. 5 patients out of them had all type of complication and the rest two had CAN and DPN. This trend suggests that the RET should be the final stage of above four diabetes microvascular complications in Bangladesh. We did not find any subject with only NEP or only RET. If a patient had RET, we can say that he/she had CAN and DPN both or CAN, DPN, and NEP i.e., all the complications. The average diabetic duration of male patients with CAN and DPN is high (17.33 years for CAN and 18.91 years for DPN) and comparatively lower for RET (13 years). Female patients with retinopathy had a high diabetic duration of 17.5 years. They did not check DM until they became very ill. So, their reported DM duration is from the day "they" found out first not the form the actual moment of DM development.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

The overall result indicates a high prevalence of complications in Bangladeshi type-2 diabetes patients.

3.3. Classification of Cardiac-related Microvascular Complications

To assess the association between any complication (as an outcome) and significant demographic, clinical and laboratory variables of the patients, several machine learning models (RF, SVM) were trained by changing the model parameters in an iterative way and observe sensitivity, specificity, precision, f1-score and accuracy of the model. The Chi-squared (χ^2) test has been used to choose significant variables and we use only these significant variables to determine the classification accuracy. The threshold for significant importance level is different for each test.

Table 5: Comparison between two machine learning models for each test

Tests	SVM			RF		
	Accuracy,%	Sensitivity,%	Specificity,%	Accuracy,%	Sensitivity,%	Specificity,%
CAN(pCAN vs nCAN)	77.33	29.41	91.34	98.67	100	98.48
DPN(pDPN vs nDPN)	67.8	68.89	66.67	67.78	68.09	67.44
NEP(pNEP vs nNEP)	100	100	100	100	100	100
Ret (pRET vs nRET)	80.5	96.05	20	84.38	97.44	27.78

3.3.1. CAN

We found diastolic BP (importance 2.1), albumin creatinine ratio (importance 1.6), and gender (importance 1) as the significant predictors for screening CAN which is the most common complication among Bangladeshi patients having type-2 diabetes. We have 65 positive and 10 negative CAN patients in our study. To find the best suitable result and to maximize the model performance, we have to use prior probability in the classification model of CAN. We have found RF as the best model at the weight of [1.05 0.9]. The performance (shown in Table 5) of the model is obtained as, accuracy 98.68%, sensitivity 98.48%, and specifity100%. The performance is shown in Figure 3.

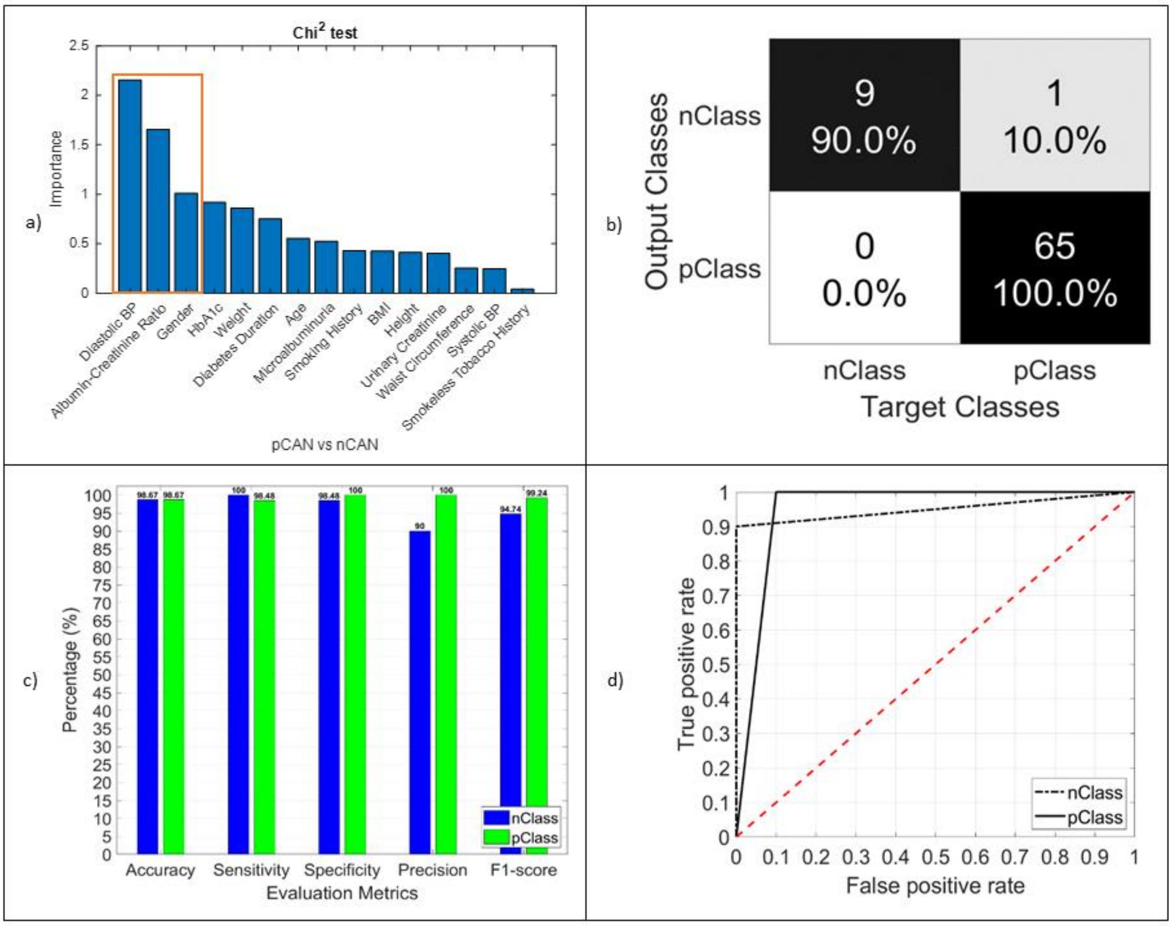

Figure 3: a) Chi-squared test result, the importance of different features- marked features were used in the model as an identifier; b) confusion matrix of CAN test (pClass vs nClass); c) performance evaluation matrices; d) TPR vs FPR, graphical view of CAN classifier model performance

3.3.2. DPN

Similarly, microalbuminuria (importance 5.1), smoking history (importance 2.9), smokeless tobacco history (importance 2.7), HbA1c (importance 2.4), albumin creatinine ratio (importance 1.9), systolic BP (importance 1.8), diastolic BP (importance 1.4), and urinary creatinine (importance 1.4) were found to be the most significant predictors for determining DPN from type-2 diabetes patients in Bangladesh. It is consistent with other finding that age and diabetic duration is insignificant [15], [53]–[58] here since all the patients are more than 40 years of age and the diabetic duration is minimum of 10 years. RF and SVM both model shows the highest accuracy for classifying DPN in the patients with type-2 diabetes mellitus from Bangladesh. Figure 4 illustrates the result of classifying DPN and the numeric values are stored in Table 5.

 doi:10.20944/preprints202111.0303.v1

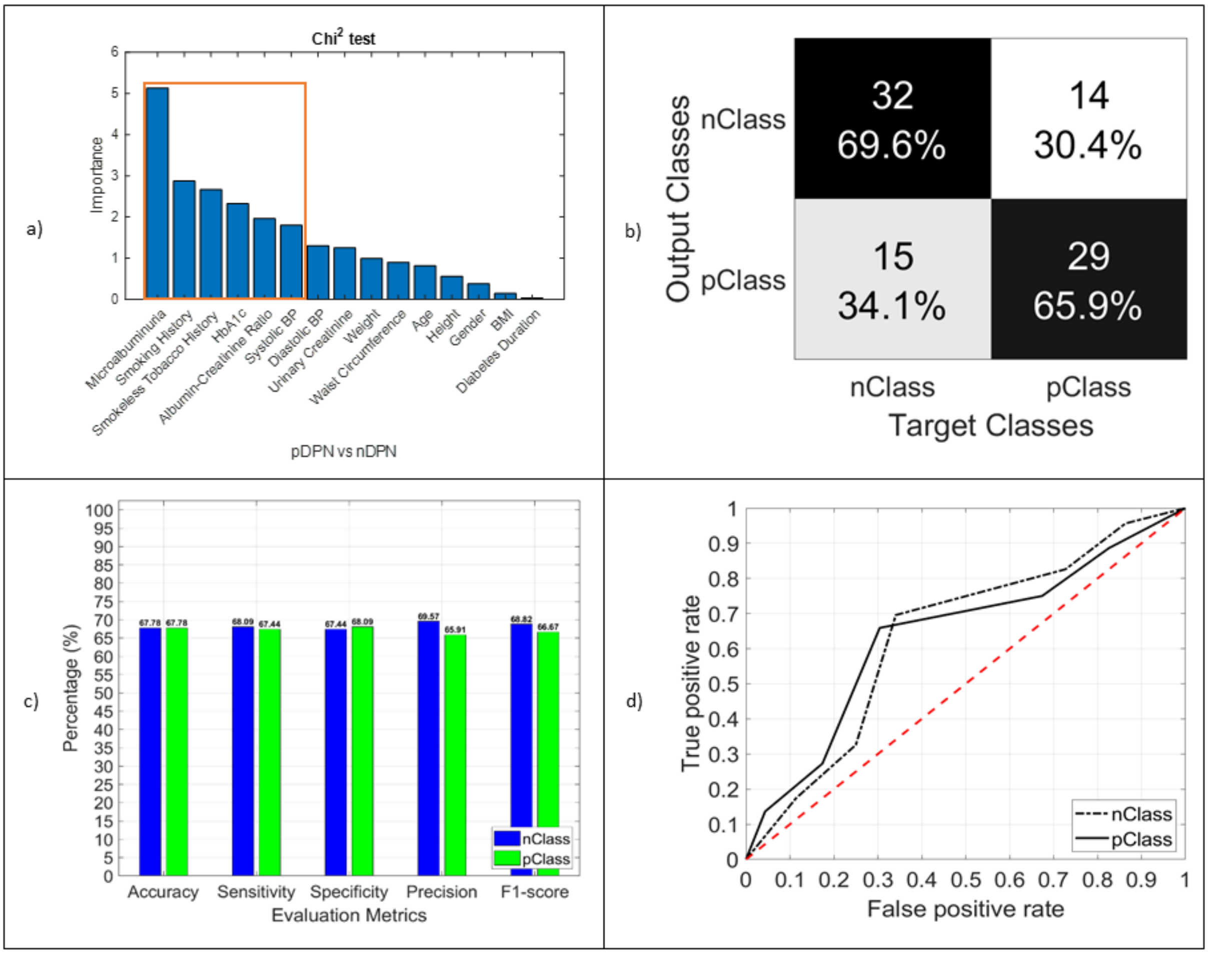

Figure 4: a) Chi-squared test result, the importance of different features- marked features were used in the model as an identifier; b) confusion matrix of DPN test (pClass vs nClass); c) performance evaluation matrices; d) TPR vs FPR, graphical view of DPN classifier model performance

3.3.3. NEP

Albumin-creatinine ratio (importance 33) was found to be the most significant predictor variable whether microalbuminuria (importance 30) was also a significant predictor for the classification of nephropathy. Albumin-creatinine ratio is a globally used test to determine kidney disease. It is obvious having 100% accuracy by using the albumin-creatinine ratio as a predictor (shown in Table 5). The confusion matrix shows the accuracy for classifying Nep is in Figure 5.

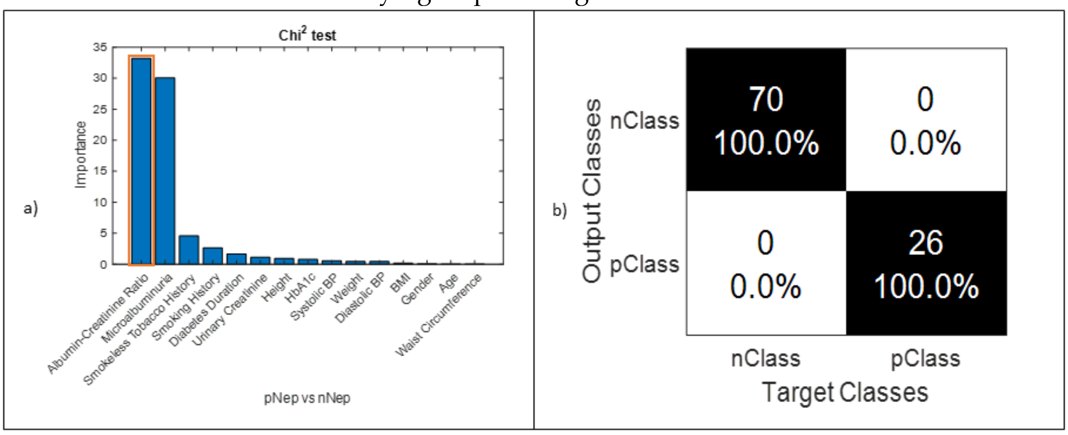

Figure 5: a) Chi-squared test result, the importance of different features- marked features were used in the model as an identifier; b) confusion matrix of DPN test (pClass vs nClass)

3.3.4. RET

In the case of diabetic retinopathy (RET), HbA1c (importance 6.1), microalbuminuria (importance 4.7), smokeless tobacco history (importance 2.8), weight (importance 1.9), gender (importance 1.8), urinary creatinine (importance 1.7), and albumin creatinine ratio (importance 1.7) were found to be significant predictors to classify whether a type-2 diabetes mellitus patients have retinopathy or not. A previous study in Bangladesh shows a 5.4% prevalence of retinopathy patients[23] and in our study, we have 6.8% of retinopathy patients with type-2 diabetes. The accuracy (shown in Table 5) of the RF model is 84.38%. The performance is shown in Figure 6 below.

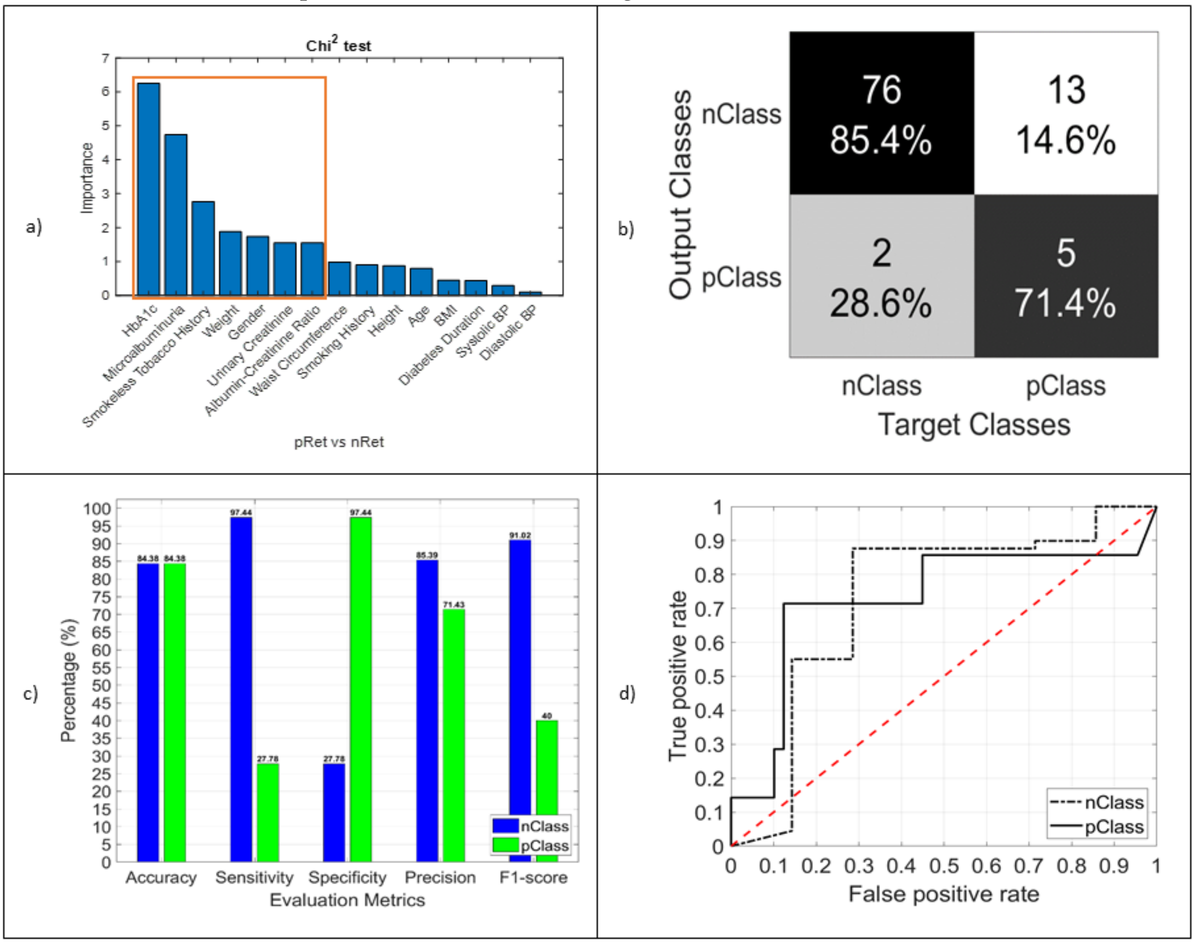

Figure 6: a) Chi-squared test result, the importance of different features, marked features were used in the model as an identifier; b) confusion matrix of RET test (pClass vs nClass); c) performance evaluation matrices; d) TPR vs FPR, graphical view of RET classifier model performance

3.3.5. CANDPNOthers

In the CANDPNOther test, three classes were assigned to the model, i.e. CAN vs CANDPN (the patients having CAN and DPN both complications) vs CANDPN+ (the patients having CAN, DPN and NEP or CAN, DPN and RET or CAN, DPN, NEP, and RET). We only include these classes due to the insufficient number of patients in the other classes. SVM performs better in this multiclass classification rather than RF. The confusion matrix (Figure 7) illustrates CAN and CANDPN+ classes could be classified effectively. However, identifying CANDPN patients using this model might be inefficient.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

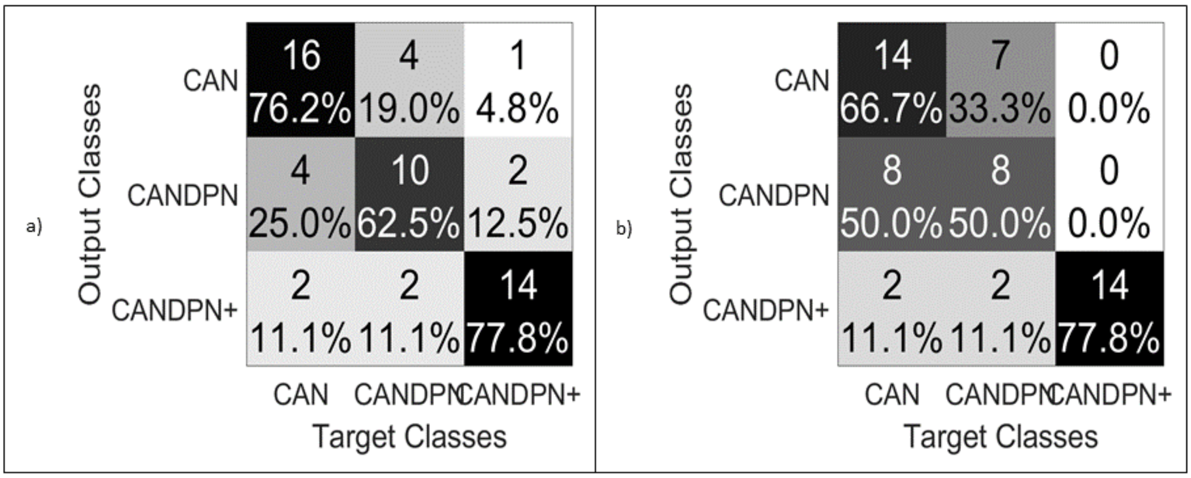

Figure 7: Performance comparison (confusion matrix) between multiclass SVM classifier and multiclass RF classifier a) confusion matrix of SVM classifier; b) confusion matrix of RF classifier; classes: 1. CAN – patients having CAN; 2. CANDPN – having DPN with CAN; 3. CANDPN+ - having NEP and/or RET with CAN and DPN

4. Discussion

This study demonstrated the importance of demographic, clinical, and laboratory profiles in the machine learning domain for the classification of diabetic microvascular complications. Moreover, this study illustrated a complete machine-learning-based clinical approach to screen diabetic patients suffering from diabetic microvascular complications. It also provides an association of other microvascular complications along with CAN. It provides a stepwise clinical approach to screen diabetes microvascular complications for the Bangladeshi type-2 diabetic cohort. The high performance achieved in each test strongly suggests to include DCL profiles as features in the machine learning approach to ensure a high classification accuracy. In our study, we also showed the DCL profiles that are highly associated with a kind of complication. Thus, the clinician can easily coordinate the physiological grounds.

4.1. Demographic, Clinical and Laboratory profiles

In this study, we demonstrated the significance of DCL profiles to screen a microvascular complication of the type-2 diabetic population in Bangladesh. Profiles that are used in this study are easily collectible to any hospital in Bangladesh. Moreover, gathering this information from a patient is not costly. Furthermore, all the DCL profiles used in this study are not required to be collected for screening using the proposed method. Only the significant features that are listed (Section 3.3) for each test will be needed to execute the test. However, to execute all the tests proposed in this study, a mathematical union of all the significant features that are used in each test would require.

Diastolic BP, Albumin Creatinine Ratio, and Gender were highly associated with CAN in our CAN test. Thus we have found the highest accuracy for CAN classifier by using these three predictors. In [59], and [60] authors showed the influence of hypertension on diabetic complications and our study found Diastolic BP as a good predictor variable to classify CAN. However, we have found HbA1c unrequired for testing CAN. On the other hand, microalbuminuria was significantly associated with peripheral neuropathy, nephropathy, and retinopathy. This finding supports both the studies in [61], [62], where authors showed the association of microalbuminuria with nephropathy and retinopathy. Moreover, in [63], *Bell et. al.* observed the significant association of microalbuminuria with diabetic neuropathy. HbA1c was significantly associated with retinopathy in our study. It was also associated with peripheral neuropathy. In [64], [65], the authors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 November 2021 doi:10.20944/preprints202111.0303.v1

established the relationship between HbA1c and microvascular complications. Many authors show the significance and association of different demographic, clinical, and laboratory parameters with diabetic microvascular complications. However, in our study, we found significant DCL profiles using a statistical model and used these significant profiles into a machine learning model to show the performance.

4.2. Machine Learning as a Screening Tool

This work describes the application of a modern machine learning model, combining the use of statistically significant features to exploit demographic, clinical, and laboratory data to extract a classifier that can classify type-2 diabetes microvascular complications. To address the class imbalance, a machine learning hyper-parameter 'prior probability' had been used. Picking up the benefits of the recent advances of machine learning in the area of diabetes diagnosis is considered to be fundamental. It makes a difference within the investigation of colossal healthcare records and changes over them into clinical experiences that can help healthcare experts in prompt and intelligent decision-making. Even though the involvement of a clinician within the diagnosis and treatment of diabetic patients may be a must, machine learning models might be able to provide an early-stage screening that can avoid numerous complications from further development. Besides, when there is a tremendous request for medical specialists or unbounded data available, it is quite hard to provide a complete diagnostic for each quite effectively. In this manner, pre-trained machine learning models can make the process faster and less rigorous for healthcare suppliers and practitioners. Different sorts of machine learning algorithms such as support vector machines (SVM), K-nearest neighbor (KNN), choice trees, etc. have been utilized broadly within the research associated with type-2 diabetes microvascular complications[66].

CAN play a major role in myocardial ischemia and infarction, heart arrhythmias, hypertension. heart disappointment as well as it increases the risk of sudden cardiac death. Jelinek and Cornforth[67] proposed a novel clustering technique using a graph-based machine learning system that enables the identification of severe diabetic neuropathies in 2016. This proposed model outperforms SVM, RF, and KNN. Cho et al.[68] showed an accuracy of 88.7% (AUC 0.969 and specificity 0.85) using SVM classifiers along with a feature selection method for the prediction of diabetic nephropathy from the data of 4321 patients. Reedy et. al.[69] proposed a multi-model ensemble-based machine learning algorithm to classify diabetic retinopathy. The authors included several machine learning classifiers in their research and concluded that ensemble model provides better accuracy with better sensitivity and specificity. Sambyal et. al.[66] provided a review of using machine learning models to classify diabetes microvascular complications in 2020. The authors showed that most of the work for classifying RET had been done using fundus image as the input and he compares the different achieved accuracy of different classifiers by different authors. By only using the demographic, clinical, and laboratory profiles, our model outperforms all the models reviewed by the author in terms of classifying diabetic retinopathy. The authors also have reviewed several machine learning models proposed by different authors for classifying cardiac autonomic neuropathy and nephropathy. However, we only use DCL profiles as independent features to classify microvascular complications.

4.3. Clinical Relevance

The test schemes followed in this work offer physicians an important clinical diagnostic method in the evaluation and diagnosis of type-2 diabetes microvascular complications. The single class classifiers can operate individually in parallel or sequentially. However, for finding combined complications with CAN, the model work sequentially with the CAN testing classifier. Since this model is sequential, this digs deeper by analyzing diabetics with single and combined complications. Single class classifiers identify any microvascular complication is present in a patient whether the multiclass classifier predicts the presence of other complications with CAN. Such a clinical test will ensure a better diagnose of type-2 microvascular complications, by distinguishing the cause of single CAN and other related complications. Furthermore, the silent nature of these complications makes it difficult to diagnosis correctly, especially when combined with other microvascular complications. The performance achieved through machine learning using only DCL profiles in this study provides a path to prevent many undiagnosed CAN-only cases. Since a CAN-only medical procedure may not provide effective treatment if additional complications are not properly identified, it is vital to know about combined complications. Multiclass classification test helps to identify multiple complications with autonomic neuropathy.

4.4. Key Message to the Health Community of Bangladesh

Globally, healthcare stakeholders are entering a new era of data-driven clinical detection and prognostication. The application of modern machine learning-based approaches offers great promises for early diagnosis or prognosis of various health complications. Early identification of patients at risk of microvascular complications due to type-2 diabetes can mitigate the burden of the healthcare system especially in the context of resource-limited setup. As the present study shows that screening is feasible from the demographic, clinical, and laboratory (DCL) variables using a proper machine learning classification model, the health community can utilize this benefit for screening that can avoid numerous complications from further development. It also can help healthcare experts in prompt and intelligent decision-making and save the patients from incurring greater healthcare costs.

5. Conclusions

This study explored the present status of microvascular complications in a cohort of type-2 diabetes patients in Bangladesh. Higher comorbidities and microvascular complications were found as compared with neighboring countries, most likely, due to the increased levels of hypertension in this cohort. This study also suggests high diastolic BP, and albumin-creatinine ratio are related to CAN; high Microalbuminuria, HbA1c, and blood pressure are related to DPN; albumin creatinine ratio is related to NEP; high HbA1c and Microalbuminuria are related to RET. These findings may be useful in finding risk factors for the development of diabetic complications. Using these risk factors as the independent features, a machine learning model could be designed to screen microvascular complications. This study shows a machine learning model could be utilized to identify diabetes complications in Bangladesh, where the majority of its population is poor. We believe this study could contribute to more effective and affordable screening techniques for diabetes-related microvascular complications.

Author Contributions: Conceptualization, Mamunur Rashid, Abdul Mukit, Khawza Iftekhar Uddin Ahmed and Ahsan H. Khandoker; Data curation, Mamunur Rashid and Abdul Mukit; Formal analysis, Mamunur Rashid and Mohanad Alkhodari; Investigation, Mamunur Rashid and Mohanad Alkhodari; Methodology, Mamunur Rashid, Mohanad Alkhodari, Abdul Mukit, Khawza Iftekhar Uddin Ahmed and Ahsan H. Khandoker; Project administration, Khawza Iftekhar Uddin Ahmed, Raqibul Mostafa and Ahsan H. Khandoker; Software, Mamunur Rashid and Mohanad Alkhodari; Supervision, Khawza Iftekhar Uddin Ahmed, Raqibul Mostafa and Ahsan H. Khandoker; Validation, Mamunur Rashid, Mohanad Alkhodari, Abdul Mukit, Khawza Iftekhar Uddin Ahmed, Raqibul Mostafa, Sharmin Parveen and Ahsan H. Khandoker; Visualization, Mamunur Rashid and Mohanad Alkhodari; Writing – original draft, Mamunur Rashid; Writing – review & editing, Mohanad Alkhodari, Abdul Mukit, Khawza Iftekhar Uddin Ahmed, Raqibul Mostafa, Sharmin Parveen and Ahsan H. Khandoker.

Funding: This work was funded by United International University (www.uiu.ac.bd) [UIU-RG-162013] and was partially supported by another grant (award number 8474000132) from Healthcare Engineering Innovation Center (HEIC) at Khalifa University, Abu Dhabi, UAE.

Institutional Review Board Statement: The study was approved by the ethical review committee of Bangladesh university of health sciences (BUHS/BIO/EA/17/01) and conforms to the ethical principles outlined in the declaration of Helsinki and the Ministry of Health and Family Welfare of Bangladesh.

Informed Consent Statement: A consent form was taken from every participant to be eligible for enrollment in the study.

Data Availability Statement: Data and models could be shared with any other researchers working in non-profit organizations under research agreement.

Acknowledgments: This work would not have been possible without the research funding from Institute of Advance Research (IAR), United International University. Their support for this study is greatly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

List of Abbreviations:

CAN – cardiac autonomic neuropathy

DPN – diabetic peripheral neuropathy

Nep – nephropathy

Ret – retinopathy

NCV – nerve conduction velocity

CTS – carpal tunnel syndrome

ACR – albumin creatinine ratio

References

[1] "Diabetes." [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1. [Accessed: 23-Aug-2021].

[2] K. Ogurtsova *et al.*, "IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040.," *Diabetes Res. Clin. Pract.*, vol. 128, pp. 40–50, Jun. 2017.

[3] L. Chen, D. J. Magliano, and P. Z. Zimmet, "The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives.," *Nat. Rev. Endocrinol.*, vol. 8, no. 4, pp. 228–36, Nov. 2011.

[4] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030.," *Diabetes Res. Clin. Pract.*, vol. 94, no. 3, pp. 311–21, Dec. 2011.

[5] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030.," *Diabetes Care*, vol. 27, no. 5, pp. 1047–53, May 2004.

[6] International Diabetes Federation, *Eighth edition 2017*. 2017.

[7] I. D. Federation, *Diabetes Atlas 2000 Diabetes Atlas 2000*. 2000.

[8] International Diabetes Federation, *Seventh Edition 2015*. 2015.

[9] R. Hira, M. A. W. Miah, and D. H. Akash, "Prevalence of Type 2 Diabetes Mellitus in Rural Adults (>_31years) in Bangladesh," *Faridpur Med. Coll. J.*, vol. 13, no. 1, pp. 20–23, 2018.

[10] N. Saquib, J. Saquib, T. Ahmed, M. A. Khanam, and M. R. Cullen, "Cardiovascular diseases and type 2 diabetes in Bangladesh: a systematic review and meta-analysis of studies between 1995 and 2010.," *BMC Public Health*, vol. 12, p. 434, Jun. 2012.

[11] P. Katulanda, P. Ranasinghe, R. Jayawardena, G. R. Constantine, M. H. R. Sheriff, and D. R. Matthews, "The prevalence, patterns and predictors of diabetic peripheral neuropathy in a developing country.," *Diabetol. Metab. Syndr.*, vol. 4, no. 1, p. 21, May 2012.

[12] R. E. Maser, B. D. Mitchell, A. I. Vinik, and R. Freeman, "The association between cardiovascular autonomic neuropathy and mortality in individuals with diabetes a meta-analysis," *Diabetes Care*, vol. 26, no. 6, pp. 1895–1901, Jun. 2003.

[13] G. A. Suarez *et al.*, "Sudden cardiac death in diabetes mellitus: risk factors in the Rochester diabetic neuropathy study.," *J. Neurol. Neurosurg. Psychiatry*, vol. 76, no. 2, pp. 240–5, Feb. 2005.

[14] D. Ziegler, K. Dannehl, H. Mühlen, M. Spüler, and F. A. Gries, “Prevalence of Cardiovascular Autonomic Dysfunction Assessed by Spectral Analysis, Vector Analysis, and Standard Tests of Heart Rate Variation and Blood Pressure Responses at Various Stages of Diabetic Neuropathy,” *Diabet. Med.*, vol. 9, no. 9, pp. 806–814, 1992.
[15] C. A. Abbott *et al.*, “The North-West Diabetes Foot Care Study: incidence of, and risk factors for, new diabetic foot ulceration in a community-based patient cohort.,” *Diabet. Med.*, vol. 19, no. 5, pp. 377–84, May 2002.
[16] C. Daousi, I. A. MacFarlane, A. Woodward, T. J. Nurmikko, P. E. Bundred, and S. J. Benbow, “Chronic painful peripheral neuropathy in an urban community: a controlled comparison of people with and without diabetes.,” *Diabet. Med.*, vol. 21, no. 9, pp. 976–82, Sep. 2004.
[17] A. A. F. Sima, “Diabetic Neuropathy, 2nd Edition. P.J. Dyck and P.K. Thomas. Philadelphia: W.B. Saunders, 1999. No. of pages: 560. Price: £85.00. ISBN: 0721661823,” *Diabetes. Metab. Res. Rev.*, vol. 15, no. 5, pp. 379–379, Sep. 1999.
[18] K. Mørkrid, L. Ali, and A. Hussain, “Risk factors and prevalence of diabetic peripheral neuropathy: A study of type 2 diabetic outpatients in Bangladesh,” *Int. J. Diabetes Dev. Ctries.*, vol. 30, no. 1, p. 11, Jan. 2010.
[19] J. Cabezas-Cerrato, “The prevalence of clinical diabetic polyneuropathy in Spain: a study in primary care and hospital clinic groups. Neuropathy Spanish Study Group of the Spanish Diabetes Society (SDS).,” *Diabetologia*, vol. 41, no. 11, pp. 1263–1269, 1998.
[20] K. Mørkrid, L. Ali, and A. Hussain, “Risk factors and prevalence of diabetic peripheral neuropathy: A study of type 2 diabetic outpatients in Bangladesh.,” *Int. J. Diabetes Dev. Ctries.*, vol. 30, no. 1, pp. 11–7, Jan. 2010.
[21] A. Hussain, S. Vaaler, M. A. Sayeed, H. Mahtab, S. M. K. Ali, and A. K. A. Khan, “Type 2 diabetes and impaired fasting blood glucose in rural Bangladesh: A population-based study,” *Eur. J. Public Health*, vol. 17, no. 3, pp. 291–296, Sep. 2007.
[22] A. Y. T. Wu *et al.*, “An alarmingly high prevalence of diabetic nephropathy in Asian type 2 diabetic patients: the MicroAlbuminuria Prevalence (MAP) Study.,” *Diabetologia*, vol. 48, no. 1, pp. 17–26, Jan. 2005.
[23] A. Akhter, K. Fatema, S. F. Ahmed, A. Afroz, L. Ali, and A. Hussain, “Prevalence and Associated Risk Indicators of Retinopathy in a Rural Bangladeshi Population with and without Diabetes,” *Ophthalmic Epidemiol.*, vol. 20, no. 4, pp. 220–227, Aug. 2013.
[24] J. W. Y. Yau *et al.*, “Global prevalence and major risk factors of diabetic retinopathy.,” *Diabetes Care*, vol. 35, no. 3, pp. 556–64, Mar. 2012.
[25] T. Biswas, A. Islam, L. B. Rawal, and S. M. S. Islam, “Increasing prevalence of diabetes in Bangladesh: a scoping review,” *Public Health*, vol. 138. Elsevier B.V., pp. 4–11, 01-Sep-2016.
[26] M. M. Rahman, M. A. Rahim, and Q. Nahar, “Prevalence and risk factors of Type 2 diabetes in an urbanizing rural community of Bangladesh,” *Bangladesh Med. Res. Counc. Bull.*, vol. 33, no. 2, pp. 48–54, Jan. 1970.
[27] S. M. S. Islam *et al.*, “Diabetes knowledge and glycemic control among patients with type 2 diabetes in Bangladesh,” *Springerplus*, vol. 4, no. 1, Dec. 2015.
[28] F. Saleh, S. J. Mumu, F. Ara, H. A. Begum, and L. Ali, “Knowledge and self-care practices regarding diabetes among newly diagnosed type 2 diabetics in Bangladesh: A cross-sectional study,” *BMC Public Health*, vol. 12, no. 1, 2012.
[29] S. Asghar, A. Hussain, S. M. K. Ali, A. K. A. Khan, and A. Magnusson, “Prevalence of depression and diabetes: a population-based study from rural Bangladesh.,” *Diabet. Med.*, vol. 24, no. 8, pp. 872–7, Aug. 2007.
[30] V. Alcalá-Rmz *et al.*, “Identification of people with diabetes treatment through lipids profile using machine learning algorithms,” *Healthc.*, vol. 9, no. 4, 2021.
[31] V. Alcalá-Rmz *et al.*, “Identification of Diabetic Patients through Clinical and Para-Clinical Features in Mexico: An Approach Using Deep Neural Networks,” *Int. J. Environ. Res. Public Heal. 2019, Vol. 16, Page 381*, vol. 16, no. 3, p. 381, Jan. 2019.
[32] H. Zhou, R. Myrzashova, and R. Zheng, “Diabetes prediction model based on an enhanced deep neural network,” *EURASIP J. Wirel. Commun. Netw. 2020 20201*, vol. 2020, no. 1, pp. 1–13, Jul. 2020.
[33] S. Bae and T. Park, “Risk prediction of type 2 diabetes using common and rare variants,” *Int. J. Data Min. Bioinform.*, vol. 20, no. 1, pp. 77–90, 2018.

[34] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clin. Epidemiol. Glob. Heal.*, vol. 7, no. 4, pp. 530–535, Dec. 2019.

[35] A. Alharbi and M. Alghahtani, "Using Genetic Algorithm and ELM Neural Networks for Feature Extraction and Classification of Type 2-Diabetes Mellitus," *https://doi.org/10.1080/08839514.2018.1560545*, vol. 33, no. 4, pp. 311–328, 2018.

[36] M. Alkhodari *et al.*, "Screening Cardiovascular Autonomic Neuropathy in Diabetic Patients With Microvascular Complications Using Machine Learning: A 24-Hour Heart Rate Variability Study," *IEEE Access*, vol. 9, pp. 119171–119187, Aug. 2021.

[37] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenbroucke, "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies," *J. Clin. Epidemiol.*, vol. 61, no. 4, pp. 344–349, Apr. 2008.

[38] "Bangladesh Institute of Health Sciences Hospital." [Online]. Available: http://www.bihsh.org.bd/. [Accessed: 12-Sep-2019].

[39] P. A. James *et al.*, "Guía basada en la evidencia de 2014 para el manejo de la presión arterial alta en adultos: informe de los miembros del panel designados para el Octavo Comité Nacional Conjunto (JNC 8).," *JAMA*, vol. 311, no. 5, pp. 507–20, 2014.

[40] K. Lin, L. Wei, Z. Huang, and Q. Zeng, "Combination of Ewing test, heart rate variability, and heart rate turbulence analysis for early diagnosis of diabetic cardiac autonomic neuropathy," *Med. (United States)*, vol. 96, no. 45, Nov. 2017.

[41] P. D. Study, A. Americans, N. Americans, and A. D. Asso-, "Kaplow Shavell Fairness v Welfare Chapter 3.pdf," 2005.

[42] C. A. McCarty, K. I. Taylor, R. McKay, J. E. Keeffe, and Working Group on Evaluation of NHMRC Diabetic Retinopathy Guidelines, "Diabetic retinopathy: effects of national guidelines on the referral, examination and treatment practices of ophthalmologists and optometrists.," *Clin. Experiment. Ophthalmol.*, vol. 29, no. 2, pp. 52–8, Apr. 2001.

[43] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3. pp. 169–208, 2010.

[44] J. Antoch, "A Guide to Chi-Squared Testing," *Comput. Stat. Data Anal.*, vol. 23, no. 4, pp. 565–566, Feb. 1997.

[45] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2. pp. 181–201, Mar-2001.

[46] M. M. Moya, M. W. Koch, L. D. Hostetler, M. M. Moya, M. W. Koch, and L. D. Hostetler, "One-class classifier networks for target recognition applications," *STIN*, vol. 93, p. 24043, 1993.

[47] T. K. Ho, "Ho, T.K. (1995) Random Decision Forest. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16 August 1995, 278-282. - References - Scientific Research Publishing," p. undefined-undefined, 1995.

[48] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Dec. 2008.

[49] J. R. Quinlan, "{C4}.5 - Programs for Machine Learning," p. undefined-undefined, 1993.

[50] R. Nisbet, G. Miner, and K. Yale, "Data Understanding and Preparation," in *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier, 2018, pp. 55–82.

[51] A. Ramachandran, "Specific problems of the diabetic foot in developing countries.," *Diabetes. Metab. Res. Rev.*, vol. 20 Suppl 1, pp. S19-22.

[52] A. Hussain, M. A. Rahim, A. K. Azad Khan, S. M. K. Ali, and S. Vaaler, "Type 2 diabetes in rural and urban population: diverse prevalence and associated risk factors in Bangladesh.," *Diabet. Med.*, vol. 22, no. 7, pp. 931–6, Jul. 2005.

[53] A. J. M. (Andrew J. M. Boulton, P. R. Cavanagh, and G. Rayman, *The foot in diabetes*. Wiley, 2006.

[54] U. T. Börü *et al.*, "Prevalence of peripheral neuropathy in type 2 diabetic patients attending a diabetes center in Turkey.," *Endocr. J.*, vol. 51, no. 6, pp. 563–7, Dec. 2004.

[55] O. Mimi, C. L. Teng, and Y. C. Chia, "The prevalence of diabetic peripheral neuropathy in an outpatient setting.," *Med. J. Malaysia*, vol. 58, no. 4, pp. 533–8, Oct. 2003.

[56] C. P. Yang *et al.*, "Cardiovascular risk factors increase the risks of diabetic peripheral neuropathy in patients with type 2 diabetes mellitus," *Med. (United States)*, vol. 94, no. 42, p. e1783, Oct. 2015.

[57] M. J. Young, A. J. Boulton, A. F. MacLeod, D. R. Williams, and P. H. Sonksen, "A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population.," *Diabetologia*, vol. 36, no. 2, pp. 150–4, Feb. 1993.

[58] S. Ashok, M. Ramu, R. Deepa, and V. Mohan, "Prevalence of neuropathy in type 2 diabetic patients attending a diabetes centre in South India.," *J. Assoc. Physicians India*, vol. 50, pp. 546–50, Apr. 2002.

[59] I. B. Hirsch and M. Brownlee, "Beyond hemoglobin A1c - Need for additional markers of risk for diabetic microvascular complications," *JAMA - J. Am. Med. Assoc.*, vol. 303, no. 22, pp. 2291–2292, 2010.

[60] F. Ayad, M. Belhadj, J. Pariés, J. R. Attali, and P. Valensi, "Association between cardiac autonomic neuropathy and hypertension and its potential influence on diabetic complications," *Diabet. Med.*, vol. 27, no. 7, pp. 804–811, Jul. 2010.

[61] C. M. Florkowski, P. E. Jennings, B. Rowe, N. Lawson, S. Nightingale, and A. H. Barnett, "Microalbuminuria in diabetic subjects with chronic peripheral neuropathy," *Diabetes Res. Clin. Pract.*, vol. 5, no. 1, pp. 45–48, May 1988.

[62] H.-H. Parving *et al.*, "Prevalence of microalbuminuria, arterial hypertension, retinopathy, and neuropathy in patients with insulin dependent diabetes," *Br Med J (Clin Res Ed)*, vol. 296, no. 6616, pp. 156–160, Jan. 1988.

[63] D. S. Bell, C. H. Ketchum, C. A. Robinson, L. E. Wagenknecht, and B. T. Williams, "Microalbuminuria Associated With Diabetic Neuropathy," *Diabetes Care*, vol. 15, no. 4, pp. 528–531, Apr. 1992.

[64] J. Škrha, J. Šoupal, J. Škrha, and M. Prázný, "Glucose variability, HbA1c and microvascular complications," *Rev. Endocr. Metab. Disord. 2016 171*, vol. 17, no. 1, pp. 103–110, Mar. 2016.

[65] P. M. M. S. J. L. DM Nathan, "Relationship of glycated albumin to blood glucose and HbA(1c) values and to retinopathy, nephropathy and cardiovascular outcomes in the DCCT/EDIC study," *Diabetes*, vol. 63, no. 1, pp. 282–90, Jan. 2014.

[66] N. Sambyal, P. Saini, and R. Syal, "Microvascular Complications in Type-2 Diabetes: A Review of Statistical Techniques and Machine Learning Models," *Wirel. Pers. Commun.*, vol. 115, no. 1, 2020.

[67] H. F. Jelinek, D. J. Cornforth, and A. V Kelarev, "Machine Learning Methods for Automated Detection of Severe Diabetic Neuropathy," *J. Diabet. Complicat. Med.*, vol. 01, no. 02, 2016.

[68] B. H. Cho, H. Yu, K. W. Kim, T. H. Kim, I. Y. Kim, and S. I. Kim, "Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods," *Artif. Intell. Med.*, vol. 42, no. 1, pp. 37–53, Jan. 2008.

[69] G. T. Reddy *et al.*, "An Ensemble based Machine Learning model for Diabetic Retinopathy Classification," in *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, 2020.