Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(1): 61-76.

Original Article

An application of CART algorithms for detection of an association between VDR polymorphisms and reduced bone density in individuals with type 2 diabetes: a population-based cross-sectional study

Maryam Ghodsi[1], Bagher Larijani[2], Shahin Roshani[3], Mahsa Mohammad Amoli[4], Farideh Razi[1], Abbas Ali Keshtkar[5], Patricia Khashayar[6,7], Fariba Zarrabi[8,9], Mohamad Reza Mohajeri-Tehrani[2*]

[1]Diabetes Research Center (DRC), Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran.
[2]Endocrinology and Metabolism Research Center (EMRC), Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences (TUMS), Tehran, Iran.
[3]Non-Communicable Diseases Research Center (NCDRC), Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences (TUMS), Tehran, Iran.
[4]Metabolic Disorders Research Center (MDRC), Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences (TUMS), Tehran, Iran.
[5]Department of Health Sciences Education Development, School of Public Health (SPH), Tehran University of Medical Sciences (TUMS), Tehran, Iran.
[6]Center for microsystem technology, Imec and Ghent University, Zwijnaarde, Ghent, Belgium.
[7]Osteoporosis Research Center (ORC), Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran.
[8]Thrombosis Hemostasis Research Center, Tehran University of Medical Sciences, Tehran, Iran.
[9]Department of Biology, College of Basic Science, Tehran Science and Research Branch, Islamic Azad University, Tehran, Iran.

ARTICLE INFO

Received 16.11.2021
Revised 28.11.2021
Accepted 22.12.2021
Published 15.03.2022

Key words:
Data mining;
Osteoporosis;
Bone density;
Type 2;
Diabetes;
Vitamin D receptor

ABSTRACT

Introduction: An important part of preventing major common diseases is identifying genetic factors that contribute to their occurrence. For the first time in our knowledge, we investigated the association between five polymorphisms of vitamin D receptor (VDR) gene (ApaI, BsmI, FokI, EcoRV, and TaqI) and low bone density/osteopenia/osteoporosis in individuals with type 2 diabetes using classification and regression tree (CART) algorithms.

Methods: Data from 158 participants with T2D were used to develop the CART analysis. The binary output variable was "bone state" with low or normal values. Age and BMI (continuous variables), vitamin D deficiency (yes/no), and gender (binary variables), as well as the studied polymorphism of the VDR gene (categorical variables) all played a role in the explanatory model. A 5-fold cross-validation process was used for model validation.

Results: Participants were divided into three groups: men, women, and both sexes. In all groups, age was the major factor predicting the low state in the final obtained tree model. The second most significant predictor in each model was BMI in both sexes (accuracy:75.30% ± 2.80%, AUC: 0.740 ± 0.064), EcoRV polymorphism in women (accuracy: 80.79% ± 6.58%, AUC:0.785 ± 0.063), and TaqI polymorphism in men (accuracy: 76.36% ± 3.05%, AUC:0.706 ± 0.125).

Conclusion: Model validation of the final tree models demonstrated that the use of CART algorithms could be an acceptable technique for risk factors of osteoporosis among individuals with T2D. Our recommendation is to conduct more population-based studies. We hope this study will serve as a basis for future research.

*.Corresponding Author: mrmohajeri@tums.ac.ir

Copyright © 2022 Tehran University of Medical Sciences. Published by Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/).
Noncommercial uses of the work are permitted, provided the original work is properly cited.

Introduction

Osteoporosis is a multifactorial disease greatly affected by a variety of factors, including genetics.[1]To develop methods that reduce the burden of this disease, it is necessary to better understand the factors causing osteoporosis. Hence, there is a focus on deciphering new productive genetic components associated with bone density. Data mining has benefits in the field of genetic research, especially where clinicians try to deal with huge data and translate knowledge from population-based to personalized medicine.[2]

Up to 60% of the bone character is attributed to genetic factors.[1] In this context, the genes of vitamin D receptor (*VDR*) genes are the most investigated genes for potential links to low bone density and osteoporosis.[3] The active form of vitamin D exerts the majority of its effect through its receptor.[3] Previously published works have highlighted the association between polymorphism of the *VDR* gene and low bone density (LBD) or osteoporosis.[4] Results indicate that polymorphism of the *VDR* gene may be linked to chronic inflammatory diseases such as diabetes.[5] Amongst almost 200 discovered polymorphisms of the *VDR* gene, ApaI (rs7975232), BsmI (rs1544410), EcoRV (rs4516035), FokI (rs2228570), and TaqI (rs731236) are those that can influence the role of *VDR* protein and modulate susceptibility to type 2 diabetes (T2D).[6] The connection between particular polymorphisms of *VDR* gene (BsmI[7, 8], FokI[9, 10] and, TaqI[10]) and the occurrence of T2D has been suggested by some published data; however, some studies have offered different conclusions.[11, 12]

We use descriptive analysis to help us understand the nature of data and visualize potential relationships. We usually conduct hypothesis testing and regression analysis to validate the root causes. However, there are instances where it is more acceptable to use other nonparametric techniques. These include the violation of the normality assumptions, multiple categorical explanatory variables, considerable multicollinearity and outliers, and low sample size. The utility of the traditional methods is reduced and becomes problematic in such cases. These problems were addressed by Breiman's method of classification and regression tree (CART) analysis.[13] The CART algorithm for fitting a classification tree is a useful nonparametric technique suitable for medical research where many potential causes of variation and defects are categorical in nature.[14]

Despite the increasing interest in utilizing the CART method in medical research, there have been few published studies practising this method in diabetes and osteoporosis fields. Our aim was to investigate the possible association between five polymorphisms of the *VDR* gene (ApaI rs7975232, BsmI rs1544410, EcoRV rs4516035, FokI rs2228570, and TaqI rs731236) and the occurrence of LBD/osteopenia/osteoporosis using the decision tree fitted by the CART algorithms in T2D subjects who participated in the third phase of the Iranian Multicenter Osteoporosis Study (IMOS) in Sanandaj, Iran.[15]

Methods

The Study population

In 2015, the Endocrinology and Metabolism Research Institute (EMRI) Ethics Committee approved the protocol of the current population-based cross-sectional study. We used data of samples from Sanandaj city from the

IMOS study (phase III) conducted by EMRI researchers in 2010 in Arak and Sanandaj cities. The *VDR* gene was examined only in the samples from Sanandaj city.[15] The primary recruited participants were healthy individuals above 20 years of age, selected by cluster random sampling method.[15] Our study sample included the primary participants who lived in Sanandaj city, were at least 26 years old, had been diagnosed with T2D, and had bone mineral densitometry (BMD) reports.

To obtain a reliable estimate for individuals with T2D, our approach was based on related studies such as the article conducted in the NHANES population.[16, 17] According to the exclusion criteria, participants who self-reported having diabetes in the primary questionnaire but did not meet the criteria for identification as type 1 diabetic were screened for T2D even if they did not report taking diabetes medications. Participants without a positive history of diabetes who had fast blood sugar (FBS) ≥126 mg/dl or glycated haemoglobin (A1C) ≥ 6.5% were considered to have undiagnosed T2D unless they met the criteria for type 1 diabetes. Any primary participant record was excluded from our study if diabetes was not mentioned in the questionnaire, and laboratory data for FBS and A1C also excluded diabetes. The next exclusion criteria were mention of type 1 diabetes in the questionnaire and use of insulin without concomitant use of other oral hypoglycaemic agents.

The studied SNPs of the *VDR* gene included ApaI (rs7975232), BsmI (rs1544410), EcoRV (rs4516035), FokI (rs2228570), and TaqI (rs731236). The genetic study of the *VDR* gene polymorphism was performed on whole blood samples of the Sanandaj participants which were previously stored in Ethylenediaminetetraacetic acid (EDTA) at -70°C. DNA extraction and *VDR* genotyping method are described in detail in the published protocol of the study IMOS III.[15] Genotypes Nomenclature was as follows; the restriction endonucleases enzyme of ApaI (allele A/a), BsmI (allele B/b), EcoRV (allele E/e), FokI (allele F/f), and TaqI (allele T/t) are recognized as allelic variants of the *VDR* polymorphism.

Each eligible participant had a report of BMD analysis at three sites (lumbar (l2-L4) spine, Hip, and femoral neck), performed by Dual-Energy X-ray Absorptiometry (Norland XR46) in 2011.[15] The DEXA variables were expressed as T-scores and Z-scores.

Measurements

Response variable and risk factors

In two categories of response variable and risk factors, we defined the study measurements for current research as follows:

Response variable

Post-menopausal women and men over 50 years of age were categorized into three groups based on their T-scores at any of the three BMD sites: "normal", "osteopenia", and "osteoporosis". Then, premenopausal women and men younger than 50 years of age were divided into two groups according to their Z-scores: "normal" and" low bone mass (density)".[18] Lastly, the binary outcome variable called “bone state” was formed with either "normal" or "low" values. The "normal" group consisted of participants with normal results of BMD at all sites; the "low" group consisted of participants with low bone density (LBD), osteopenia, or osteoporosis.

Risk factor

There were three kinds of potential risk factors as follows: continuous (age and BMI), binary (gender and vitamin D deficiency), and categorical (the five studied polymorphisms of the *VDR* gene).
Based on the Endocrine Society Clinical Practice Guideline,[19] we defined vitamin D deficiency as 25-hydroxyvitamin D below 20 ng/ml (50 nmol/litre).

Analyzing method

The missing values of the risk factors were imputed by median values for the numerical variables and the most frequent class for the categorical variables.[20] According to the SAMPLE guideline,[21] we summarized the risk factors in the two groups of the main binary outcome variable as follow: in case of continuous variables with normal distribution, we summarized the data as the mean (standard deviations); otherwise, we reported it as the median (interquartile range). We expressed the distribution of categorical variables as numbers (percentages).
We checked the normality assumption by applying both statistical tests (test Kolmogorov-Smirnov/Shapiro-Wilk) and graphical assessments (histograms, Q–Q plots, and box plots). Comparing the two groups, the t-test was used when the distribution of a continuous variable was normal in both groups and Mann-Whitney's U test when the distribution was skewed in any of the groups. Chi-square tests (Pearson/Fisher test) were used to study the unadjusted effects of categorical variables.
A 3*2 Chi-square test was used to determine the genotype association from the overall genotype frequencies. Allele and genotype frequencies were tested for Hardy-Weinberg Equilibrium (HWE).
To perform decision tree analysis, evaluate crucial variables, and find the cut-off point for the continuous variables, the CART algorithm was applied using the Gini index as the main criterion for recursive partition.[22] The pruning rules were set as follows. To build a tree with the best size and lowest misclassification rate, the maximum depth of 1-20, minimum gain index of 0.01-0.21, and minimum leaf size of 1-30 were utilized. The 5-fold cross-validation was used as the model selection method based on the criteria of overall accuracy estimation of models. Other evaluation parameters such as AUC (area under the curve), sensitivity, specificity, etc. of each selected best-fitted model were calculated through this method as well.
All statistical tests were two-tailed and, the P-value (P) < 0.05 was considered significant. The primary statistical analyses were conducted using STATA)ver.12). The CART Analysing method was performed by applying the RapidMiner (ver.9) software.

Results

We studied data of 158 T2D subjects (99 women; age 26–83 years) to determine the impact of age, sex, BMI, and the polymorphisms of the *VDR* gene (ApaI, BsmI, EcoRV, FokI, and TaqI) on the “bone state” by the CART analysis. As defined previously, 50.63% (84/158) of the participants had "low" bone states. Compared to the "normal" classification (Table 1), the "low" group was significantly older (P < 0.001), a lower percentage had vitamin D deficiency (P = 0.029), and a higher percentage were women (P = 0.044).
Allele and genotype frequencies conform to

Table 1. Differences between the two “bone state” groups in the distribution of age, BMI, and vitamin D deficiency according to sex in participants with type 2 diabetes

variables			Bone state			P
			low	normal	total	
In both genders	Age (year); Mdn(Q1,Q3)		55.00 (51.00,61.00)	45.00 (39.00,51.00)	51.00 (43.00,58.00)	<0.001[a]
	BMI (kg/m^2); Mdn(Q1,Q3)		28.63 (25.40,31.24)	29.42 (27.48,32.83)	29.18 (26.70,31.60)	0.340[a]
	Serum Vitamin D (nmol/l); N(%)	<50	60(75.00)	69(88.46)	129(81.65)	0.029[b]
		≥50	20(25.00)	9(11.54)	29(18.35)	
		Total	80(100)	78(100)	158(100)	
	Gender; N(%)	Female	44(55.00)	55(70.51)	99(62.66)	0.044[b]
		Male	36(45.00)	23(29.49)	59(37.34)	
		Total	80(100)	78(100)	158(100)	
In women	Age (year); mean(SD)		56.20(8.47)	45.71(7.69)	50.37(9.57)	<0.001[c]
	BMI (kg/m^2); Mdn(Q_1,Q_3))		30.08 (27.96,31.64)	29.78 (27.97,34.15)	29.78 (27.97,32.83)	0.928[a]
	Serum Vitamin D (nmol/l) ; N(%)	<50	29(65.91)	49(89.09)	78(78.79)	0.005[b]
		≥50	15(34.09)	6(10.91)	21(21.21)	
		tTotal	44(100)	55(100)	99(100)	
In men	Age (year); Mdn(Q_1,Q_3)		55.50 (50.50,55.50)	42.00 (36.00,42.00)	52.00 (40.00,52.00)	0.005[a]
	BMI (kg/m^2); mean(SD)		27.07 (4.17)	28.50 (3.19)	27.63 (3.86)	0.167[c]
	Serum Vitamin D (nmol/l); N(%)	<50	31(86.11)	20(86.96)	51(86.44)	0.624[d]
		≥50	5(13.89)	3(13.04)	8(13.56)	
		Total	36(100)	23(100	59(100)	

P-value(P)<0.05 is bolded and assumed as significant. Mdn, Median; Q1, 25th percentile; Q3, 75th percentile; SD, Standard deviation; [a], Based on Mann-Whitney test; [b], Based on Pearson chi2; [c], Based on t-test; [d], Based on Fisher exact t-test

HWE in controls, P>0.05, except for the BsmI variant (X2:11.96, P<0.001). Table 2 displays the frequencies of alleles and genotypes of polymorphisms in the *VDR* gene between the two sexes and bone density groups. Women in the "low" group had a frequency of the *EE* genotype of the EcoRV that was more than twice that of the "normal" group (61.36% vs 34.55; P=0.028). In the comparison of the dominant model (*EE* vs *Ee+ee*), a significant difference was detected in categories of both sexes (55% vs 45%, P=0.037) and women (61.36% vs 38.64%, P=0.008). In both categories, the frequency of the *EE* variant was significantly higher than *Ee+ee* in the “low” group (Table2). Comparison of the allele-frequency genetic model (*E* vs. e allele) in the women category showed the frequency of the *E* allele in the “low” group was significantly higher than the *e* (75% vs 25%, P=0.026).

Across all gender categories, the genotype *ff* of the FokI polymorphism was noticeably lower in the “low” class than in the “normal” class and, the difference was significant in the category of both sexes (2.50% versus 12.82%; P=0.048). In the comparison of the recessive model (*ff* vs *FF+Ff*), a significant difference was detected in the categories of both sexes (2.5% vs 97.5%, P=0.014) and women (2.27% vs 97.73% P=0.041). In both categories, the

Table 2. Differences in allele and genotype frequencies of the studied polymorphisms (ApaI, BsmI, EcoRV, FokI and, TaqI) between the two groups of the bone state according to sex in participants with type 2 diabetes

Variables			in both sexes				in women				in men			
			Bone state			P	Bone state			P	Bone state			P
			Low	Normal	Total		Low	Normal	Total		Low	Normal	Total	
ApaI	Genotypes	AA	26(32.50)	27(34.62)	53(33.54)	0.794[a]	14(31.82)	18(32.73)	32(32.32)	0.837[a]	12(33.33)	9(39.13)	21(35.59)	0.936[b]
		Aa	41(51.25)	36(46.15)	77(48.73)	-	23(52.27)	26(47.27)	49(49.49)	-	18(50.00)	10(43.48)	28(47.46)	-
		aa	13(16.25)	15(19.23)	28(17.72)	-	7(15.91)	11(20.00)	18(18.18)	-	6(16.67)	4(17.39)	10(16.95)	-
		Total	80(100)	78(100)	158(100)	-	44(100)	55(100)	99(100)	-	36(100)	23(100)	59(100)	-
	Alleles	A	93(58.13)	90(57.69)	183(57.91)	0.938[a]	51(57.95)	62(56.36)	113(57.07)	0.822[a]	42(58.13)	28(60.87)	70(59.32)	0.784[a]
		a	67(41.88)	66(42.31)	133(42.09)	-	37(42.05)	48(43.64)	85(42.93)	-	30(41.67)	18(39.13)	48(40.68)	-
		Total	160(100)	156(100)	316(100)	-	88(100)	110(100)	198(100)	-	72.00(100)	46(100)	118(100)	-
		X^2 (P)[c]	0.223(0.636)	0.232 (0.630)	0.000 (0.997)	-	0.232 (0.630)	0.084 (0.772)	0.010 (0.920)	-	0.029 (0.863)	0.175 (0.675)	0.016 (0.898)	-
	R.M.	AA+ Aa	67(83.75)	63(80.77)	130(82.28)	-	37(84.09)	44(80)	81(81.82)	0.600[a]	30(83.13)	19(82.61)	49(83.05)	0.942[a]
		aa	13(16.25)	15(19.23)	28(17.72)	0.624[a]	7(15.91)	11(20.00)	18(18.18)	-	6(16.67)	4(17.39)	10(16.95)	-
	D.M.	AA	26(32.50)	27(34.62)	53(33.54)	-	14(31.82)	18(32.73)	32(32.32)	0.923[a]	12(33.33)	9(39.13)	21(35.59)	0.650[a]
		Aa+ aa	54(67.50)	51(65.38)	105(66.46)	0.788[a]	30(68.18)	37(67.27)	67(67.68)	-	24(66.67)	14(60.87)	38(64.41)	-
BsmI	Genotypes	BB	24(30)	21(26.92)	45(28.48)	0.922[b]	16(36.36)	15(27.27)	31(31.31)	0.418[b]	8(22.22)	6(26.09)	14(23.73)	0.576[b]
		Bb	52(65.00)	53(67.95)	105(66.46)	-	25(56.82)	38(69.09)	63(63.64)	-	27(75.00)	15(65.22)	42(71.19)	-
		bb	4(5.00)	4(5.13)	8(5.06)	-	3(6.82)	2(3.64)	5(5.05)	-	1(2.78)	2(8.70)	3(5.08)	-
		Total	80(100)	78(100)	158(100)	-	44(100)	55(100)	99(100)	-	36(100)	23(100)	59(100)	-
	Alleles	B	100(62.50)	95(60.90)	195(61.71)	0.770[a]	57(64.77)	68(61.82)	125(63.13)	0.669[a]	43(59.72)	27(58.70)	70(59.32)	0.912[a]
		b	60(37.50)	61(39.10)	121(38.29)	-	31(35.23)	42(38.18)	73(36.87)	-	29(40.28)	19(41.30)	48(40.68)	-
		Total	160(100)	156(100)	316(100)	-	88(100)	110(100)	198(100)	-	29(40.28)	19(41.30)	48(40.68)	-
		X^2 (P)	11.96(0.000)	14.20 (0.000)	26.07 (0.000)	-	2.642 (0.104)	11.82 (0.000)	13.33 (0.000)	-	11.24 (0.000)	2.73 (0.098)	13.31 (0.000)	-
	R.M.	BB+ Bb	76(95.00)	74(94.87)	150(94.94)	0.626[b]	41(93.18)	53(96.36)	94(94.95)	0.653[b]	35(97.22)	21(91.30)	56(94.92)	0.554[b]
		bb	4(5.00)	4(5.13)	8(5.06)	-	3(6.82)	2(3.64)	5(5.05)	-	1(2.78)	2(8.70)	3(5.08)	-
	D.M.	BB	24(30)	21(26.92)	45(28.48)	0.668[a]	16(36.36)	15(27.27)	31(31.31)	0.332[a]	8(22.22)	6(26.09)	14(23.73)	0.734[a]
		Bb+ bb	56(70.00)	57(73.08)	113(71.52)	-	28(63.64)	40(72.73)	68(68.69)	-	28(77.78)	17(73.91)	45(76.27)	-
EcoRV	Genotypes	EE	44(55.00)	30(38.46)	74(46.84)	0.101[a]	27(61.36)	19(34.55)	46(46.46)	0.028[a]	17(47.22)	11(47.83)	28(47.46)	0.869[b]
		Ee	26(32.50)	37(47.44)	63(39.87)	-	12(27.27)	27(49.09)	39(39.39)	-	14(38.89)	10(43.48)	24(40.68)	-
		ee	10(12.50)	11(14.10)	21(13.29)	-	5(11.36)	9(16.36)	14(14.14)	-	5(13.89)	2(8.70)	7(11.86)	-
		Total	80(10.00)	78(100)	158(100)	-	44(100)	55(100)	99(100)	-	36(100)	23(100)	59(100)	-
	Alleles	E	114(71.25)	97(62.18)	211(66.77)	0.087[a]	66(75.00)	65(59.09)	131(66.16)	0.019[a]	48(66.67)	32(69.57)	80(67.80)	0.742[a]
		e	46(28.75)	59(37.82)	105(33.23)	-	22(25.00)	45(40.91)	67(33.84)	-	24(33.33)	14(30.43)	38(32.20)	-
		Total	160(100)	156(100)	316(100)	-	88(100)	110(100)	198(100)	-	72(100)	46(100)	118(100)	-
		X^2 (P)	3.418 (0.064)	0.006 (0.939)	1.625 (0.202)	-	3.273 (0.070)	0.013 (0.909)	1.430 (0.231)	-	0.563 (0.453)	0.017 (0.897)	0.276 (0.599)	-

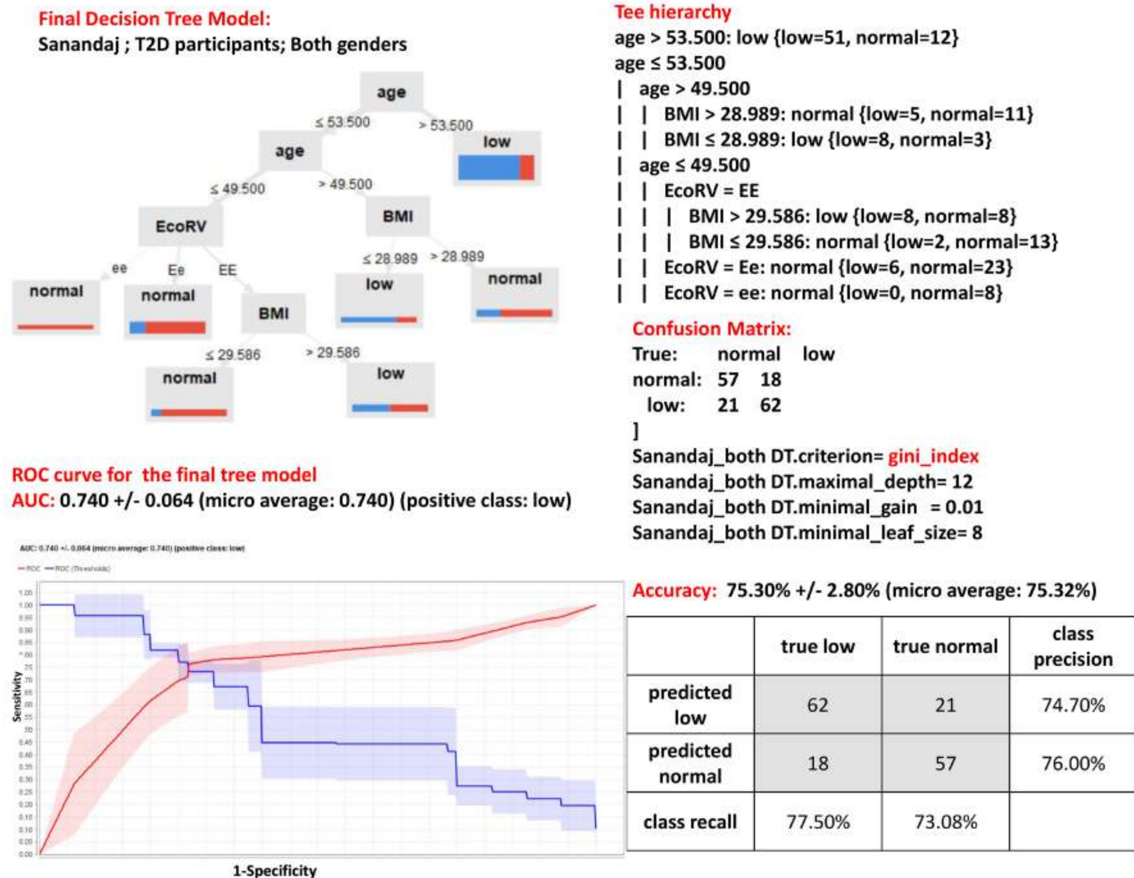
EcoRV	R.M.	EE+ Ee	70(87.50)	67(85.90)	137(86.71)	0.767^{a}	39(88.64)	46(83.64)	85(85.86)	0.478^{a}	31(86.11)	21(91.30)	52(88.14)	0.694^{b}
		ee	10(12.50)	11(14.10)	21(13.29)	-	5(11.36)	9(16.36)	14(14.14)	-	5(13.89)	2(8.70)	7(11.86)	-
	D.M.	EE	44(55.00)	30(38.46)	74(46.84)	0.037^{a}	27(61.36)	19(34.55)	46(46.46)	0.008^{a}	17(47.22)	11(47.83)	28(47.46)	0.964^{a}
		Ee+ ee	36(45.00)	48(61.54)	84(53.16)	-	17(38.64)	36(65.45)	53(53.54)	-	19(52.78)	12(52.17)	31(52.54)	-
FokI	Genotypes	FF	45(56.25)	41(52.56)	86(54.43)	0.048^{b}	26(59.09)	29(52.73)	55(55.56)	0.115^{b}	19(52.78)	12(52.17)	31(52.54)	0.738^{b}
		Ff	33(41.25)	27(34.62)	60(37.97)	-	17(38.64)	18(32.73)	35(35.35)	-	16(44.44)	9(39.13)	25(42.37)	-
		ff	2(2.50)	10(12.82)	12(7.59)	-	1(2.27)	8(14.55)	9(9.09)	-	1(2.78)	2(8.70)	3(5.08)	-
		Total	80(100)	78(100)	158(100)	-	44(100)	55(100)	99(100)	-	36(100)	23(100)	59(100)	-
	Alleles	F	123(76.88)	109 (69.87)	232 (73.42)	0.159^{a}	69(78.41)	76(67.86)	145(72.50)	0.097	54(75.00)	33(71.74)	87(73.73)	0.695
		f	37(23.13)	47(30.13)	84(26.58)		19(21.59)	36(32.14)	55(27.50)	-	18(25.00)	13(28.26)	31(26.27)	-
		Total	160(100)	156(100)	316(100)	-	88(100)	112(100)	200(100)	-	72(100)	46(100)	118(100)	-
	X^2 (P)		2.053 (0.151)	2.467 (0.116)	0.116 (0.733)	-	0.876 (0.349)	3.878 (0.048)	0.956 (0.328)	-	1.235 (0.266)	0.028 (0.866)	0.519 (0.471)	-
	R.M.	FF+ Ff	78(97.50)	68(87.18)	146(92.41)	0.014^{b}	43(97.73)	47(85.45)	90(90.91)	0.041^{b}	35(97.22)	21(91.30)	56(94.92)	0.313^{a}
		ff	2(2.50)	10(12.82)	12(7.59)	-	1(2.27)	8(14.55)	9(9.09)	-	1(2.78)	2(8.70)	3(5.08)	-
	D.M.	FF	45(56.25)	41(52.56)	86(54.43)	0.642^{a}	26 (59.09)	29(52.73)	55(55.56)	0.527^{a}	19(52.78)	12(52.17)	31(52.54)	0.964^{a}
		Ff+ ff	35(43.75)	37(47.44)	72(45.57)	-	18 (40.91)	26(47.27)	44(44.44)	-	17(47.22)	11(47.83)	28(47.46)	-
TaqI	Genotypes	TT	32 (40.00)	24 (30.77)	56 (35.44)	0.451^{a}	20 (45.45)	17 (30.91)	37 (37.37)	0.112^{b}	12 (33.33)	7 (30.43)	19 (32.20)	0.602^{b}
		Tt	37 (46.25)	40 (51.28)	77 (48.73)		21 (47.73)	27 (49.09)	48 (48.48)		16 (44.44)	13 (56.52)	29 (49.15)	
		tt	11 (13.75)	14 (17.95)	25 (15.82)		3 (6.82)	11 (20.00)	14 (14.14)		8 (22.22)	3 (13.04)	11 (18.64)	
		Total	80 (100)	78 (100)	158 (100)		44 (100)	55 (100)	99 (100)		36 (100)	23 (100)	59 (100)	
	Alleles	T	101 (63.12)	88 (56.41)	189 (59.81)	0.224^{a}	61(69.32)	61 (55.45)	122 (61.62)	0.046^{a}	40 (55.56)	27 (58.70)	67 (56.78)	0.737^{a}
		t	59 (36.88)	68 (43.59)	127 (40.19)		27(30.68)	49(44.55)	76(38.38)		32 (44.44)	19 (41.30)	51 (43.22)	
		Total	160 (100)	156 (100)	316 (100)		88(100)	110(100)	198(100)		72(100)	46(100)	118(100)	
	X^2 (P)		0.003 (0.935)	0.143 (0.705)	0.30 (0.863)		0.655 (0.418)	0.002 (0.962)	0.062 (0.803)		0.360 (0.548)	0.631 (0.426)	0.000 (0.991)	
	R.M.	TT+ Tt	69 (86.25)	64 (82.05)	133 (84.18)	0.470^{a}	41 (93.18)	44 (80.00)	85 (85.86)	0.045^{b}	28 (77.78)	20 (86.96)	48 (81.36)	0.502^{b}
		tt	11 (13.75)	14 (17.95)	25 (15.82)		3 (6.82)	11 (20.00)	14 (14.14)		8 (22.22)	3 (13.04)	11 (18.64)	
	D.M.	TT	32 (40.00)	24 (30.77)	56 (35.44)	0.225^{a}	20 (45.45)	17 (30.91)	37 (37.37)	0.137^{a}	12 (33.33)	7 (30.43)	19 (32.20)	0.816^{a}
		Tt+ Tt	48 (60.00)	54 (69.23)	102 (64.56)		24 (54.55)	38 (69.09)	62 (62.63)		24 (66.67)	16 (69.57)	40 (67.80)	

P-Value(P)<0.05 is assumed as significant and bolded.

Each VDR gene and its genotypes are as follows; ApaI (*AA, Aa*, and *aa*), BsmI (*BB, Bb*, and *bb*), EcoRV (*EE, Ee*, and *ee*), FokI (*FF, Ff,* and *ff*), and TaqI (*TT, Tt*, and *tt*).

According to sex and “bone state”, the frequency of each allele/genotype is presented as a number (%) in the frequency columns.

a,Based on Pearson Chi2 test; b, Based on Fisher's exact test; c, X2(P-value) for Hardy-Weinberg Equilibrium (HWE); D. M, Dominant model; R.M, Recessive model



	true low	true normal	class precision
predicted low	62	21	74.70%
predicted normal	18	57	76.00%
class recall	77.50%	73.08%	

Figure 1. Final decision tree model for predicting being in the "low" group in both sexes through classification and regression tree algorithms with a predictive accuracy of 75.30%±2.80% and AUC of 0.740±0.064

frequency of *ff* was significantly lower than *FF+Ff* in the “low” group (Table 2).

Comparison of the allele-frequency (*t* vs. *T* allele) in the women category showed the frequency of the *t* allele in the “low” group was significantly higher than the *T* (30.68% vs 69.32%, P=0.046) In the comparison of the recessive model (*tt* vs *TT+Tt*) significant difference was detected in the women category (6.82% vs 93.18%, P=0.045); the frequency of tt combination was significantly lower than the *TT+Tt* combination in the “low” group (Table2).

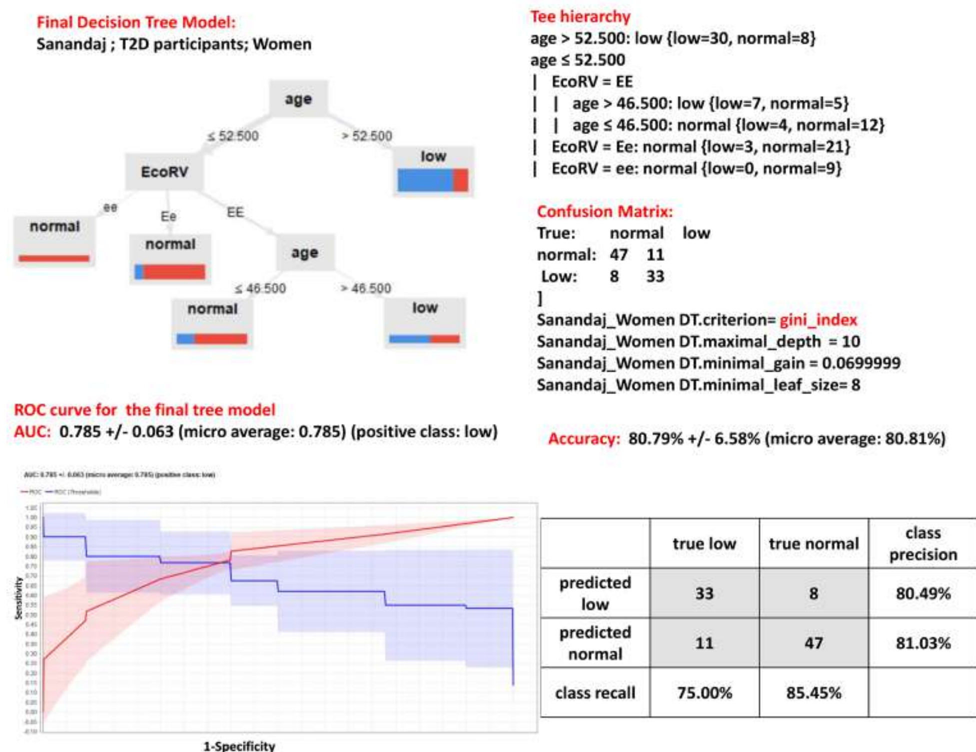
In both sexes, the final fitted tree model shows that age, with the cut-off point of 53.5, was the strongest predictor to be in the “low” group (Figure 1). When 49.5<age≤53.5, the CART algorithm identified the BMI as the next major predictor, and BMI≤28.9 predicted the group of “low”. When age<49.5 the EcoRV became the third strong predictor and, the EE variant predicted the “low” class. The final tree model had a predictive accuracy of 75.30%±2.80% and an area under the curve (AUC) of 0. 0.740±0.064 (Table 3).

For women, age>52.5 was the best predictor for the class of “low” (Figure 2). When age≤52.5, the polymorphism of the EcoRV became the second important predictor to be in the “low” group and, the *EE* variant strongly predicted

Table 3. Overall accuracy and other parameters of model evaluation that calculated by 5-fold cross-validation method for the final tree models

	Both sexes	Women	Men
Accuracy	75.30%±2.80% (micro average: 75.32%)	80.79%±6.58% (micro average: 80.81%)	76.36%±3.05% (micro average: 76.27%)
Sensitivity	77.50%±5.59% (micro average: 77.50%)	74.72%±15.23% (micro average: 75.00%)	97.14%±6.39% (micro average: 97.22%)
Specificity	73.08%±2.56% (micro average: 73.08%)	85.18%±17.39% (micro average: 85.45%)	43.00%±13.04% (micro average: 43.48%)
PPV	74.70%±1.90% (micro average: 74.70%)	83.85%±17.35% (micro average: 80.49%)	73.10%±3.35% (micro average: 72.92%)
NPV	76.18%±4.54% (micro average: 76.00%)	83.10%±9.90% (micro average: 81.03%)	95.00%±11.18% (micro average: 90.91%)
AUC	0.740±0.064 (micro average: 0.740)	0.785±0.063 (micro average: 0.785)	0.706±0.125 (micro average: 0.706)
F Measure	76.00%±3.24% (micro average: 76.07%)	77.33%±8.36% (micro average: 77.65%)	83.28%±2.79% (micro average: 83.33%)
Classification error	24.70%±2.80% (micro average: 24.68%)	19.21%±6.58% (micro average: 19.19%)	23.64%±3.05% (micro average: 23.73%)
Decision Tree Parameters	Criterion= Gini Index Maximal Depth= 12 Minimal Gain= 0.01 Minimal Leaf Size= 8	Criterion= Gini Index Maximal Depth= 10 Minimal Gain= 0.069 Minimal Leaf Size= 8	Criterion= Gini Index Maximal Depth= 4 Minimal Gain= 0.03 Minimal Leaf Size= 4

PPV, Positive Predictive Value; NPV, Negative Predictive Value; AUC, Area Under Curve



	true low	true normal	class precision
predicted low	33	8	80.49%
predicted normal	11	47	81.03%
class recall	75.00%	85.45%	

Figure 2. Final decision tree model for predicting being in the "low" group in women through the classification and regression tree algorithms with a predictive accuracy of 80.79%±6.58% and AUC of 0.785±0.063

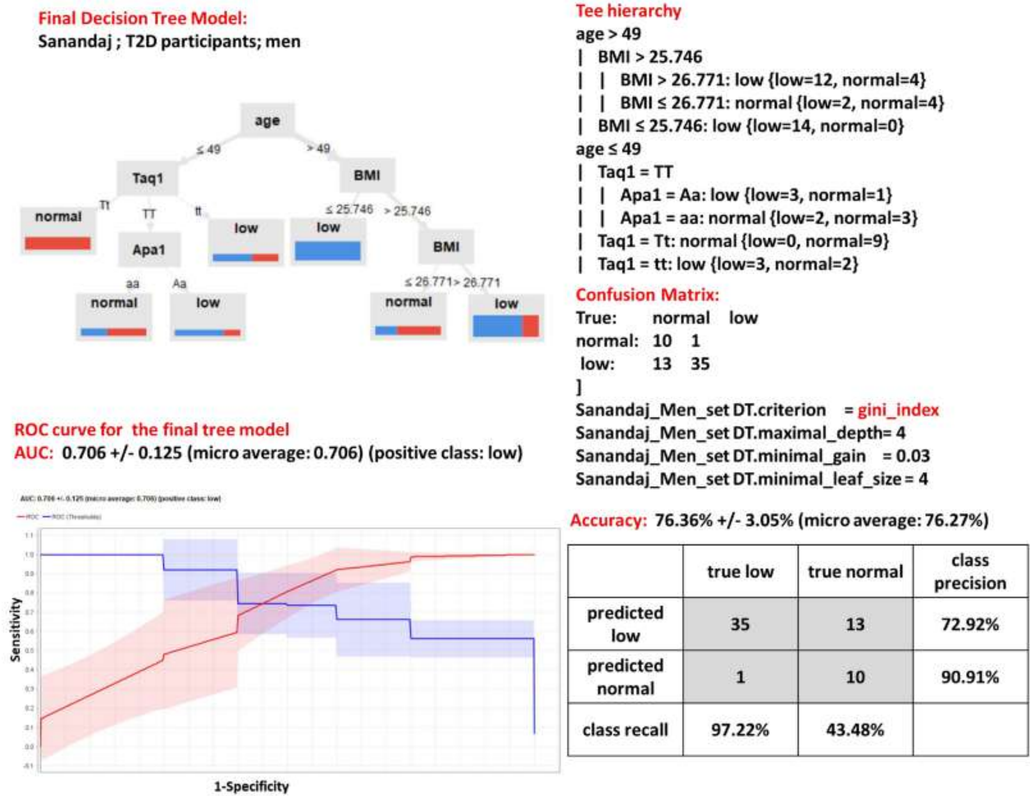


Figure 3. Final decision tree model for predicting being in the "low" group in men through the classification and regression tree algorithms with a predictive accuracy of 76.36%±3.05% and AUC of 0.706±0.125

the "low" class in women aged 46.5-52.5. The final tree model had a predictive accuracy of 80.79%±6.58% and an AUC of 0.785±0.063 (Table 3).

In men, the age variable with the cut-off point of 49 was identified as the most important predictor for being in the "low" group (Figure 3). In men under the age of 49, TaqI polymorphism was known as the next major predictor; the variant of tt strongly predicted being in the "low" class. The final tree model had a predictive accuracy of 76.36%±3.05% and an AUC of 0.706±0.125 (Table 3).

Final models were obtained using the 5-fold cross-validation method. A high value of accuracy performance estimation is reflected in all ultimate tree models, but suitability varied in terms of other parameters of model evaluation (Table 3).

Discussion

Using the CART analysis, we investigated whether there was an association between the five polymorphisms of the *VDR* gene (ApaI, BsmI, EcoRV, FokI, and TaqI) and the occurrence of LBD/osteopenia/osteoporosis in 158 T2D individuals who participated in the third phase of the IMOS study in the city of Sanandaj.[15] Participants from 26 to 83 years of age were mostly women (Table 1). It has been well established that ageing is positively correlated with the initiation and development of osteoporosis in both genders.[23] Therefore, it

is reasonable that the mean age of participants in the “low” classification was significantly higher than that of the “normal” group.

The lower percentage of participants in the “low” group had vitamin D deficiency compared to “normal” (Table 1). As noted, this group was also at a higher average age. This observation may be explained by more dietary supplements for this group, since older people usually receive more supplements, especially if they have diabetes.[24]

Our results demonstrate that the ultimate models of the CART method have acceptable accuracy performance estimation (Table 3). Overall, the CART analysis showed age was the most important predictor in all final models; however, the cut-off point for each model was different (Fig 1, 2, and 3). The BMI was the second major variable in the groups of both sexes (Figure 1) and men (Figure 3). Age is an unmodifiable risk factor for osteopenia/osteoporosis and the risk of bone density reduction is significantly increased by ageing for all ethnic groups. However, there is some debate on BMI in patients with T2D who have an increased double risk of fracture despite a higher BMI.[25] Aleti et al. suggested that normal BMI is an indicator of osteopenia/osteoporosis in patients with T2D.[26] Lee et al have declared an ideal range of BMIs that could avoid both T2D and osteoporotic disease in Korean men and postmenopausal women. The BMI range was from 23.0 to 24.9, part of the WHO overweight classification in the Asian ethnic group.[27] Our results have shown that BMI≤28.98 (the cut-off point of 28.98 classified as overweight[28]) was a strong predictor for decreasing BMD in individuals with T2D over 49.5 years of age (Figure 1).

Table 2 shows the results of the univariate analysis that revealed the unadjusted effect of polymorphisms of the *VDR* gene on bone density. The distribution of variants of the EcoRV in women was different in both groups; the frequency of the *EE* genotype in the “Low” group was almost double that in the normal group; this result was also complete by comparisons of the allele-frequency genetic model (Table 2). Following this result, the final tree model obtained in women detected the *VDR* EcoRV polymorphism as the second main predictor for the “low” category and the *EE* genotype increased the probability of being classified as low in T2D women ≤52 years of age (Figure 2). The *VDR* EcoRV polymorphism affects the activity of the vitamin D receptor and, can modulate susceptibility to T2D.[6] Interestingly, recently published results indicate that the frequency and severity of multiple sclerosis in women are influenced by *VDR* EcoRV polymorphism.[29] However, to our knowledge, the *VDR* EcoRV polymorphism has never been studied in connection with osteoporosis in patients with T2D, and this is the first study focused on this topic. To evaluate the impact of the *VDR* EcoRV polymorphism on the increase/decrease in the probability of low bone density further studies with larger sample sizes are needed.

A significant difference was found in the frequency of the *VDR* FokI polymorphism between "low" and "normal" groups among the category of both sexes; the distribution of the *ff* genotype was significantly lower in the “low” group than the “normal” group (Table 2). Previously, the *f* allele (*Ff*+*ff*) has been identified as a possible risk factor for T2D with increased effects on BMI and obesity as an explanatory mechanism.[5] According to Table 2, the frequency of the *ff* combination of alleles was significantly lower than the *FF*+*Ff* combination in "low" class participants, both

for men (P=0.014) and women (P=0.041). In support of this finding, a study of all participants from the IMOS III study from the Sanandaj city demonstrated a protective role for osteoporosis in postmenopausal women (*ff* vs. *FF*; adjusted OR:0.136, 95%CI: 0.023-0.810).[30] Thus, we hypothesize that women with T2D may be protected against osteoporosis by combining the *ff* alleles through an increase in BMI. Despite this, none of the final models in this study considered *VDR* FokI polymorphism as an effective predictor (figures 1 to 3). Further studies are needed to discover the rationale for the new finding and other inconsistencies in the results of the different investigations and determine the actual effective predictors.

We found that the frequency of the tt combination was significantly lower than the Tt+tt combination in women with T2D who were in the "low" group (Table 2). But, we did not observe similar results in the men group. Many discrepancies exist regarding the possible effects of the *VDR* TaqI polymorphism on BMD. For example, there is a study that demonstrated an association between the tt variants of TaqI and increased BMD; however, the study population was patients with ulcerative colitis.[31] Another study indicated the *VDR* TaqI polymorphism relates to both early-onset T2D and obesity.[10] In our study, the final tree model in men (figure 3) considered the VDR TaqI polymorphism as the second major predictor of the "low" bone density classification. It demonstrated that the Tt variant could have a protective role for the early-onset reduction of bone density in men with T2D aged ≤49. A similar role for the Tt variant has already been stated in another study; however, the study participants were Indian postmenopausal women.[32] Such inconsistencies could be explained by the difference in the ethnics/condition/diseases of their studied population which makes it impossible to confirm or rule out these findings; further studies with larger sample sizes are highly recommended.

All models were relatively accurate, indicating the validity of the analysis method (Table 3). Furthermore, the CART method succeeded in introducing *VDR* EcoRV and TaqI polymorphisms as potential risk factors for osteoporosis in individuals with T2D (Figures1-3). In women, the tree model achieved the highest accuracy (80.79%±6.58%), while the tree model in both sexes group showed the lowest accuracy (75.30%±2.80%) (Table 3). Other performance metrics were also acceptable: the Positive Predictive Value (PPV) ranged from 73.10% to 83.85% and the Negative Predictive Value (NPV) varied from 76.18% to 95.00%; all AUCs exceeded 0.70.

The CART method does not provide a p-value to test significance. However, this method remains appropriate for medical research, especially in conditions where traditional methods cannot be relied upon. This can occur, for example, in situations where there are many potential causes of variation, a large number of categorical predictors, and a small sample size.[13, 14, 33] Additionally, CART offers some advantages over logistic regression (LR) analysis: unlike LR, CART makes no assumptions about the distribution of the variables and is less affected by missing data, multicollinearity, and outliers. The LR analysis, however, has limitations relating to complicated interactions between the predictors, whereas the CART analysis can solve these issues by choosing an optimal splitter for each node. It represents the interactions in the final tree model and allows clinicians to decide whether a patient falls into one of the available subgroups. Consequently, the CART analysis method can

be used to develop medical guidelines and decisions.[34] We do not claim that CART is the best or most accurate method among widely used classification techniques such as Linear Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). This issue would require the evaluation of each of the above methods against a common evaluation criterion after they have been run and validated on the same dataset.

The findings of the current study must be viewed within the constraints of a cross-section design which cannot explain causality based on any association discovered through the analysis. To detect T2D participants, we used an approximate approach and considered the use of oral hypoglycaemic agents/insulin as a criterion for diagnosing diabetes. This may have led to an overestimation of diabetes. It should be emphasized that the IMOS study had a national, population-based cross-sectional design. Hence, the investigation of its data pool could provide valuable details in the field of osteoporosis. The other strength of our study was that considering the limited sample size, we were able to perform a high-precision analysis, which offers a new perspective in this area.

Conclusion

Since osteoporosis is a multifactorial disease influenced greatly by genetic factors, determining these factors can help health policymakers reduce the burden of osteoporosis by improving their ability to prevent/treat it. As the present study and its statistical population had conditions in which conventional statistical methods lacked effectiveness, we used the CART method to examine potential associations. Our results reflect acceptable accuracy in identifying osteoporosis risk factors among individuals with T2D who may be susceptible to early-onset osteoporosis by examining the polymorphism of the *VDR* gene. Further population-based studies with large data sets are strongly recommended.

Acknowledgements

The authors of this article consider it their duty to thank the staff of the Sanandaj health network, the participants, and all the EMRI staff who contributed to conducting the IMOS study.

Funding

The current study was funded by the Endocrinology and Metabolism Research Institute (EMRI).

Conflict of interest statement

There are no conflicts of interest and the authors have nothing to disclose.

References

1. Man SC, Chiriac M, Militaru MS, Trifa AP, Goia-Socol M, Georgescu CE. Association of Col1a1 Sp1 and Fok-I Vdr Genetic Polymorphisms in Young Male Idiopathic Osteoporosis. Acta endocrinologica. 2017;13(2):224-7.

2. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2011;1(5):416-30.

3. Horst-Sikorska W, Dytfeld J,

Wawrzyniak A, Marcinkowska M, Michalak M, Franek E, et al. Vitamin D receptor gene polymorphisms, bone mineral density and fractures in postmenopausal women with osteoporosis. Mol Biol Rep. 2013;40(1):383-90.

4. Pouresmaeili F, Kamalidehghan B, Kamarehei M, Goh YM. A comprehensive overview on osteoporosis and its risk factors. Ther Clin Risk Manag. 2018;14:2029-49.

5. Angel B, Lera L, Marquez C, Albala C. The association of VDR polymorphisms and type 2 diabetes in older people living in community in Santiago de Chile. Nutr Diabetes. 2018;8(1):31.

6. Al-Daghri NM, Al-Attas O, Alokail MS, Alkharfy KM, Draz HM, Agliardi C, et al. Vitamin D receptor gene polymorphisms and HLA DRB1*04 cosegregation in Saudi type 2 diabetes patients. J Immunol. 2012;188(3):1325-32.

7. Ortlepp JR, Lauscher J, Hoffmann R, Hanrath P, Joost HG. The vitamin D receptor gene variant is associated with the prevalence of type 2 diabetes mellitus and coronary artery disease. Diabetic medicine : a journal of the British Diabetic Association. 2001;18(10):842-5.

8. Malik R, Farooq R, Mehta P, Ishaq S, Din I, Shah P, et al. Association of Vitamin D Receptor Gene Polymorphism in Adults With Type 2 Diabetes in the Kashmir Valley2017.

9. Neyestani TR, Djazayery A, Shab-Bidar S, Eshraghian MR, Kalayi A, Shariatzadeh N, et al. Vitamin D Receptor Fok-I polymorphism modulates diabetic host response to vitamin D intake: need for a nutrigenetic approach. Diabetes Care. 2013;36(3):550-6.

10. Gnanaprakash V, Bodhini D, Kanthimathi S, Ginivenisha K, Shanthirani CS, Anjana RM, et al. Association of Vitamin D receptor (TaqI, BsmI, and FokI) polymorphisms with prediabetes and Type 2 diabetes in Asian Indians. Journal of Diabetology. 2019;10(1):29.

11. Shab-Bidar S, Neyestani TR, Djazayery A. Vitamin D Receptor Gene Polymorphisms, Metabolic Syndrome, and Type 2 Diabetes in Iranian Subjects: No Association with Observed SNPs. International journal for vitamin and nutrition research Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung Journal international de vitaminologie et de nutrition. 2016;86(1-2):71-80.

12. Malecki MT, Frey J, Moczulski D, Klupa T, Kozek E, Sieradzki J. Vitamin D receptor gene polymorphisms and association with type 2 diabetes mellitus in a Polish population. Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association. 2003;111(8):505-9.

13. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees (Wadsworth and Brooks/Cole, Pacific Grove, CA). CA Mathematical Reviews (MathSciNet): MR86b. 1984;62101.

14. Speybroeck N, Berkvens D, Mfoukou-Ntsakala A, Aerts M, Hens N, Van Huylenbroeck G, et al. Classification trees versus multinomial models in the analysis of urban farming systems in Central Africa. Agr Syst. 2004;80(2):133-49.

15. Keshtkar A, Khashayar P, Mohammadi Z, Etemad K, Dini M, Aghaei Meybodi H, et al. A Suggested Prototype for Assessing Bone Health. Arch Iran Med. 2015;18(7):411-5.

16. Demmer RT, Zuk AM, Rosenbaum M, Desvarieux M. Prevalence of diagnosed and undiagnosed type 2 diabetes mellitus among US adolescents: results from the continuous NHANES, 1999-2010. American journal of epidemiology. 2013;178(7):1106-13.

17. DeShields SC, Cunningham TD. Comparison of osteoporosis in US adults with type 1 and type 2 diabetes mellitus. J Endocrinol Invest. 2018;41(9):1051-60.

18. Dimai HP. Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases. Bone. 2017;104:39-43.

19. Holick MF, Binkley NC, Bischoff-Ferrari HA, Gordon CM, Hanley DA, Heaney RP, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. The Journal of clinical endocrinology and metabolism. 2011;96(7):1911-30.

20. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. Journal of clinical epidemiology. 2006;59(10):1087-91.

21. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. Int J Nurs Stud. 2015;52(1):5-9.

22. Prasad DVV, Venkataramana L, Balasubramanian P, Priyankha B, Rajagopal S, Dattuluri R, editors. An efficient pre-processing method for improved classification of diabetics using decision tree and artificial neural network. AIP Conference Proceedings; 2019: AIP Publishing LLC.

23. Khosla S, Riggs BL. Pathophysiology of age-related bone loss and osteoporosis. Endocrinology and metabolism clinics of North America. 2005;34(4):1015-30, xi.

24. Odegard PS, Janci MM, Foeppel MP, Beach JR, Trence DL. Prevalence and correlates of dietary supplement use in individuals with diabetes mellitus at an academic diabetes care clinic. The Diabetes Educator. 2011;37(3):419-25.

25. Vestergaard P. Discrepancies in bone mineral density and fracture risk in patients with type 1 and type 2 diabetes—a meta-analysis. Osteoporosis Int. 2007;18(4):427-44.

26. Aleti S, Pal R, Dutta P, Dhibar DP, Prakash M, Khandelwal N, et al. Prevalence and predictors of osteopenia and osteoporosis in patients with type 2 diabetes mellitus: a cross-sectional study from a tertiary care institute in North India. International Journal of Diabetes in Developing Countries. 2020:1-7.

27. Lee JH, Kim JH, Hong AR, Kim SW, Shin CS. Optimal body mass index for minimizing the risk for osteoporosis and type 2 diabetes. Korean J Intern Med. 2020;35(6):1432-42.

28. Purnell JQ. Definitions, Classification, and Epidemiology of Obesity. In: Feingold KR,

Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, et al., editors. Endotext. South Dartmouth (MA): MDText. com, Inc.; 2000.

29. Křenek P, Benešová Y, Bienertová-Vašků J, Vašků A. The impact of five VDR polymorphisms on multiple sclerosis risk and progression: a case-control and genotype-phenotype study. Journal of Molecular Neuroscience. 2018;64(4):559-66.

30. Mohammadi Z, Fayyazbakhsh F, Ebrahimi M, Amoli MM, Khashayar P, Dini M, et al. Association between vitamin D receptor gene polymorphisms (Fok1 and Bsm1) and osteoporosis: a systematic review. J Diabetes Metab Disord. 2014;13(1):98.

31. Szymczak-Tomczak A, Krela-Kazmierczak I, Kaczmarek-Rys M, Hryhorowicz S, Stawczyk-Eder K, Szalata M, et al. Vitamin D receptor (VDR) TaqI polymorphism, vitamin D and bone mineral density in patients with inflammatory bowel diseases. Adv Clin Exp Med. 2019;28(7):955-60.

32. Ahmad I, Jafar T, Mahdi F, Ameta K, Arshad M, Das SK, et al. Association of vitamin D receptor gene polymorphism (TaqI and Apa1) with bone mineral density in North Indian postmenopausal women. Gene. 2018;659:123-7.

33. Zimmerman RK, Balasubramani GK, Nowalk MP, Eng H, Urbanski L, Jackson ML, et al. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infectious Diseases. 2016;16(1):503.

34. Henrard S, Speybroeck N, Hermans C. Classification and regression tree analysis vs. multivariable linear and logistic regression methods as statistical tools for studying haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2015;21(6):715-22.