



Bone metabolic biomarker-based diagnosis of type 2 diabetes osteoporosis by support vector machine

Chuan Wang^{1#}, Taomin Zhang^{2#}, Peng Wang^{3^}, Xuan Liu⁴, Liming Zheng³, Lei Miao⁵, Deyu Zhou⁵, Yibo Zhang³, Yezi Hu⁶, Han Yin⁶, Qing Jiang³, Hui Jin⁶, Jianfei Sun²

¹Naval Medical Center of PLA, Shanghai, China; ²State Key Laboratory of Bioelectronics, Jiangsu Laboratory for Biomaterials and Devices, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China; ³Department of Sports Medicine and Adult Reconstructive Surgery, Drum Tower Hospital Affiliated to Medical School of Nanjing University, Nanjing, China; ⁴School of Medicine, Southeast University, Nanjing, China; ⁵School of Computer Science and Engineering, Southeast University, Nanjing, China; ⁶Department of Endocrine Secretion, Zhongda Hospital Affiliated to Southeast University, Nanjing, China

Contributions: (I) Conception and design: H Jin, J Sun; (II) Administrative support: C Wang, P Wang, X Liu; (III) Provision of study materials or patients: X Liu, Y Zhang, Q Jiang, H Yin, Y Hu, H Jin; (IV) Collection and assembly of data: T Zhang, L Zheng, Y Zhang, P Wang; (V) Data analysis and interpretation: T Zhang, J Sun; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jianfei Sun, PhD. State Key Laboratory of Bioelectronics Jiangsu Laboratory for Biomaterials and Devices, Department of Biological Science and Medical Engineering, Southeast University, Nanjing, China. Email: sunzaghi@seu.edu.cn; Hui Jin, MD. Department of Endocrine Secretion, Zhongda Hospital Affiliated to Southeast University, Nanjing, China. Email: jinhuison@126.com.

Background: Diabetes has significant effects on bone metabolism. Both type 1 and type 2 diabetes can cause osteoporotic fracture. However, it remains challenging to diagnose osteoporosis in type 2 diabetes by bone mineral density which lacks regular changes. Seen another way, osteoporosis can be ascribed to the imbalance of bone metabolism, which is closely related to diabetes as well.

Methods: Here, to assist clinicians in diagnosing osteoporosis in type 2 diabetes, an efficient and simple SVM (support vector machine) model was established based on different combinations of biochemical indexes, which were collected from patients who did the test of bone turn-over markers (BTMs) from January 2016 to March 2018 in the department of endocrine, Zhongda Hospital affiliated to Southeast University. The classification was done based on a software package of machine learning in Python. The classification performance was measured by SKLearn program incorporated in the Python software package and compared with the clinical diagnostic results.

Results: The predicting accuracy rate of final model was above 88%, with feature combination of sex, age, BMI (body mass index), TP1NP (total procollagen I N-terminal propeptide) and OSTEOC (osteocalcin).

Conclusions: Experimental results show that the model showed an anticipant result for early detection and daily monitoring on type 2 diabetic osteoporosis.

Keywords: Bone turnover markers; support vector machine; type 2 diabetes; osteoporosis

Submitted Apr 17, 2020. Accepted for publication Oct 16, 2020.

doi: 10.21037/atm-20-3388

View this article at: <http://dx.doi.org/10.21037/atm-20-3388>

[^] ORCID: 0000-0001-8597-9753.

Introduction

Osteoporosis and diabetes mellitus (DM) pose a major public health threat to modern society. An increase in the incidence of osteoporosis in DM patients has led clinicians to now view osteoporosis and the associated risk of bone fragility and fracture it imposes, as the most serious complication of type 2 diabetes mellitus (T2DM) (1). Considered to differ from common osteoporosis, the mechanism by which diabetic osteoporosis (DO) develops is unclear, rendering it a hotspot for both fundamental and clinical research (2). Some researchers believe the pathogenesis of DO to be related to a complex interaction between hyperglycemia, insulin, insulin-like growth factors and microvascular lesions (3,4). Others implicate antidiabetic drugs in causing bone loss (5). This uncertainty renders DO a diagnostic challenge. Dual-energy x-ray absorptiometry (DEXA) is the gold standard method of evaluating and diagnosing osteoporosis in the clinical setting. However, this and other bone mineral density (BMD)-based tests may not be adequate to diagnose some subtypes of DO. In type 1 DO, the formation of new bone is inhibited, causing the BMD to greatly decrease; a finding also seen in common osteoporosis patients (6). However, in patients with type 2 DO, the BMD often varies irregularly (7). In the early stage of type 2 DM, the BMD can actually increase as a result of insulin resistance and hyperinsulinemia (8,9). In this case, BMD based tests are unable to definitively diagnose type 2 DO. In the late stage of type 2 DO, both osteogenesis and bone quality are decreased, causing bone to become fragile and at a significantly greater risk of fracture (10). During the occurrence and development of DO, complex interactions between multiple factors, such as obesity, cytokines, and advanced glycation end products (AGEs) render BMD inutile as a diagnostic index (11,12). Due to the high risk of osteoporotic fracture in type 2 DM patients and differing treatment plans used in the management of common osteoporosis (13,14), it is imperative to establish an accurate and convenient method to evaluate bone health in these type 2 DM patients.

Serological testing is a common feature of routine physical examination. In comparison to DEXA testing, which is both radioactive and requires expensive clinical equipment, serological testing is cheap, safe, and extendible for emerging indexes. Serological tests also provide an indication of body metabolism at the systemic level in comparison to tests like DEXA. Studies have revealed, for

example, that glucose levels are strongly associated with bone health (15-17). Moreover, vitamin-D deficiency and hyperglycemia resulting from DM can suppress osteoblastic formation (18,19). However, because of the poor specificity (20,21), testing of only a small number of bone turn-over markers (BTMs) does not provide an accurate clinical diagnosis. On this basis, we analyzed the testing results of several common indexes with an artificial intelligence (AI) algorithm to yield a diagnostic indicator with high accuracy. Dramatic developments in machine learning have provided promising advances in AI-assisted diagnosis across diverse areas of healthcare (22-24).

We aimed to diagnose DO in type 2 DM patients by analyzing the BTMs of common serological tests with a support vector machine (SVM). The SVM is large margin classifier and a well-known algorithm in AI. The major functionality of the algorithm is in minimizing the distance between a classification hyperplane and the support vectors. The robust classification ability and excellent generalization performance of the SVM, mean that the setting of only a few parameters are required to tune the model based on hundreds of samples (25). We found that by using a certain combination of testing indexes, AI can effectively predict osteoporosis in type 2 DM patients. These results demonstrate that AI can greatly increase the efficiency of traditional diagnostic technology. We believe this strategy will play an increasingly important role in the fundamental and clinical research of osteoporosis. We present the following article in accordance with the TRIPOD reporting checklist (available at <http://dx.doi.org/10.21037/atm-20-3388>).

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the ethical committee of Zhongda Hospital affiliated to Southeast University (No.2018ZDKYSB092) and individual consent for this retrospective analysis was waived.

Datasets (patient sample)

Between January 2016 and March 2018, 289 patients receiving medical care for type 2 DM at the Endocrine Department of Zhongda Hospital affiliated to Southeast University, were initially recruited to the trial. A total of 82 patients with other chronic diseases such as hypertension, viral hepatitis, immune diseases and osteo-developmental disorder were excluded. A further five patients were

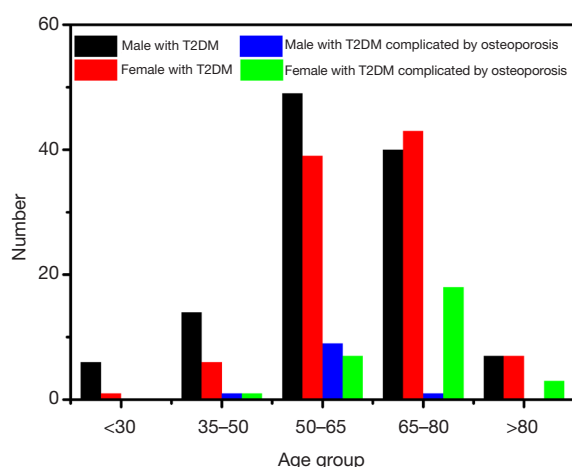


Figure 1 Sample distribution in the dataset. T2DM, type 2 diabetes mellitus.

excluded due to a lack of complete demographic data. In order to form the modelling dataset, the following information was then collected from the remaining 202 participants: gender; age; BMI; and levels of Ca, P, alkaline phosphatase (ALP), total procollagen I N-terminal propeptide (TP1NP), propeptide of type I procollagen (PICP), osteocalcin (OSTEOC), and vitamin D (Vit-D). The dataset distribution is shown in *Figure 1*.

Implementation design (measurement)

After the modeling dataset was established, an SVM algorithm was used for the classification task. The flowchart of data processing is shown in *Figure 2*. The classification was based on Scikit learning which is a software package of machine learning in Python. The steps in this process are described below.

Data preprocessing

Samples from patients diagnosed as having type 2 DM osteoporosis were labeled as positive 1 (1), and fell into the positive class. Samples in which the diagnosis was type 2 DM only, were labeled as negative 1 (-1), falling into the negative class. If the gender was female, the sample was labeled as 1; otherwise, it was labelled as 0. Age has a significant influence on the risk of osteoporosis both for men and women; therefore, age was grouped and the weight of each group was set, as shown in *Table 1*. The setting of weight was dependent upon the sample number of each group. For different ranges of attributes, the data

were normalized to avoid the influence of large numeric attributes on the calculation results (Equation 1)

$$y' = lower + (upper - lower) \times \frac{y - \min}{\max - \min} \quad [1]$$

This formula converted the eigenvalue to a specific interval, where y is the data before scaling, and y' is the scaled data. *Lower* and *upper* are the lower bound and upper bound of the given interval, respectively. In this study, the importance of all attributes was considered as the same at first. The data were scaled into [0, 1].

Modification of imbalanced data

Practical data are always imbalanced, especially when collected directly from the clinical setting. In addition, there is always a tendency for the classifiers to become biased to achieve higher prediction accuracy. There were 40 samples or 19.8% of positive class and 162 samples or 80.2% of negative class in our dataset. The Synthetic Minority Oversampling Technique (SMOTE) was adopted because of the limited sample size in this experiment (26). After SMOTE, the sample size of the minority class was increased from 40 to 162. The final dataset contained 324 samples.

Selecting of important features

The impact of each feature on the classification result differs. Therefore, based on the original data, the importance of features was judged using tree-based estimators. The features were ranked in the order of importance as shown in *Figure 3*. The weight was larger if the feature was more important. To ensure classification accuracy and reduce the cost of computing, data dimensions were reduced by ignoring less important features. Six combinations of the attributes were tested and named Test 1–6, as shown in *Table 2*. Here, 323 samples were used as a training set and 1 sample was used for testing. After tests were repeated 202 times, classification performance was finally obtained.

Parameter optimization

To map the original low-dimensional space into the high-dimensional feature space, the training set was modeled by various kernel functions including radial basis, polynomial, and sigmoid functions. In order to improve the generalization ability, a soft margin was introduced by adjusting the penalty coefficient C . Parameter C represents the relative importance of classification risk and error rate,

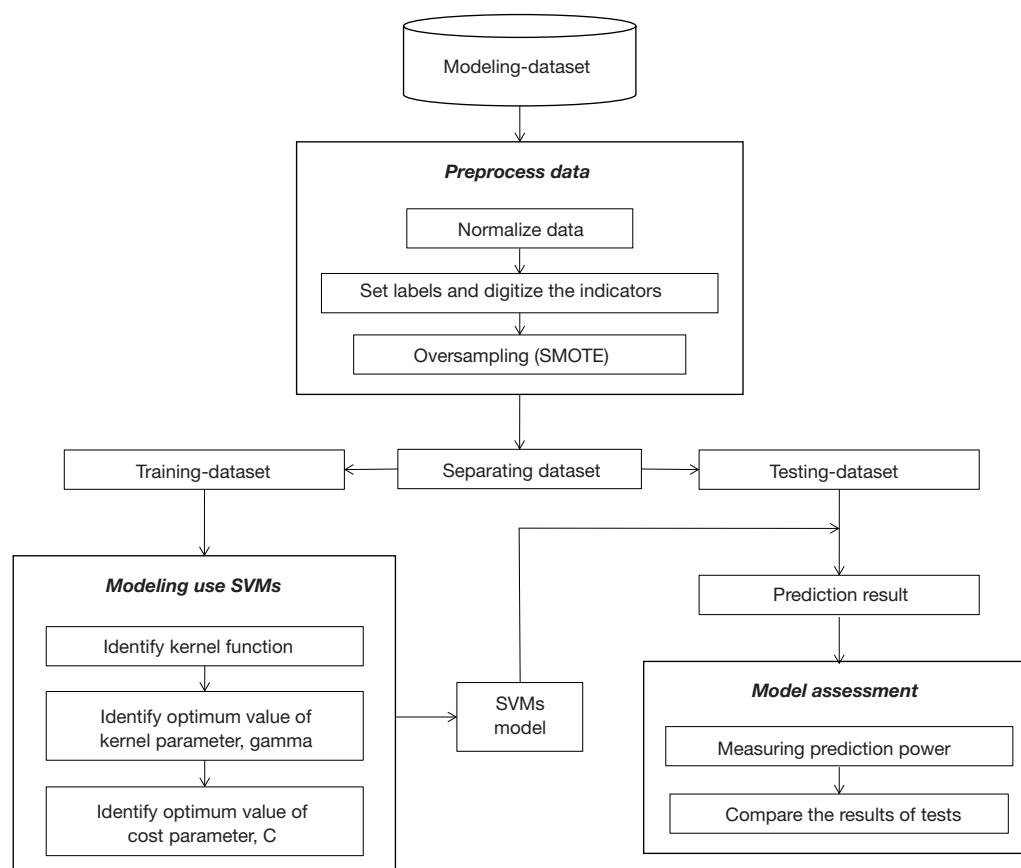


Figure 2 Flowchart of data processing for osteoporosis classification with SVM. SVM, support vector machine; SMOTE, Synthetic Minority Oversampling Technique.

Table 1 Weights setting for age groups

Age groups	The percentage of positive class (%)	The percentage of sample (%)	Weight
<35	0	3.47	0
35–50	10.00	9.90	0.5
50–65	18.18	43.56	0.75
65–80	26.03	36.14	1
≥80	21.43	6.93	0.75

which means C is the trade-off between the maximum margin and the noise tolerance. The larger C means the classification is more rigorous, yielding fewer mistakes. Secondly, when the Gaussian kernel function was selected, the complexity of the model could be adjusted by changing the parameter gamma. The larger value of gamma means the original data are mapped into the higher dimensions and the boundary of classification is more complex. To

obtain the best parameters of each model, cross validation was used. As mentioned above, 323 samples in the training set were divided into 5 sub-sample sets. One sub-sample set was selected randomly as the verification set while the others were used for training. After multiple training and verification, an average training score was obtained. The model with the highest training score was considered to be the best. By completing the above operations, the SVM

model was established.

Evaluation of classification performance

Classification performance was evaluated by using four metrics: accuracy (Acc), positive prediction value (PPV), sensitivity (Sen), and area under the receiver operating characteristic curve (ROC-AUC value) based on the confusion matrix. The Acc, PPV, and Sen were calculated by Equations 2–4:

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \quad [2]$$

$$PPV = \frac{TP}{TP + FP} \quad [3]$$

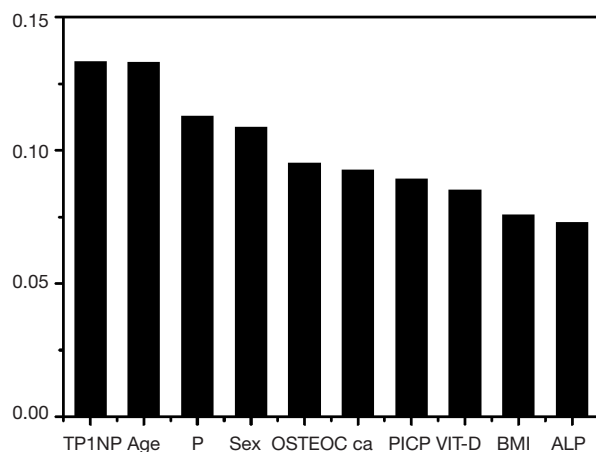


Figure 3 Ordering of importance for the testing items. TP1NP, total procollagen I N-terminal propeptide; OSTEOC, osteocalcin; PICP, propeptide of type I procollagen; VIT-D, vitamin D; BMI, body mass index; ALP, alkaline phosphatase.

$$Sen = \frac{TP}{TP + FN} \quad [4]$$

where *TP* refers to True Positive (number of positive-labeled samples that were predicted correctly), *TN* refers to True Negative (number of negative-labeled samples that were predicted correctly), *FP* refers to False Positive (number of negative-labeled samples that were predicted wrongly), and *FN* refers to False Negative (number of positive-labeled samples that were predicted wrongly). The Acc score is the percentage of correct classifications. One drawback of the Acc is a lack of the potential distribution of testing values. To obtain a comprehensive understanding of classification performance, the PPV and Sen score were used to ascertain whether the classifier identified a positive or negative case correctly. The percentage of real positive samples in samples marked as true positive is referred to as PPV. The percentage of true positive samples in the real positive data set is referred to as Sen. The ROC-AUC value is a score obtained from the receiver operating characteristic (ROC) analysis. ROC plots the function of true positive rate (TPR) and false positive rate (FPR) when verifying threshold. The increase of TPR comes at the cost of increasing FPR. The ROC-AUC was used as another evaluation of model accuracy. Values of ROC-AUC always lie between 0 and 1 among which the values above 0.9 indicate excellent prediction, those between 0.7 and 0.9 indicate good prediction, those between 0.5 and 0.7 indicated poor prediction, and any values below 0.5 are considered no better than a random guess (27).

Statistical analysis

All statistical analyses were performed using the SKLearn program incorporated in the Python software package.

Table 2 Different combinations of BMTs

	Sex	Age	BMI	TP1NP	PICP	OSTEOC	VIT-D	ALP	Ca	Phos.
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓		✓			✓	✓
3	✓	✓	✓	✓					✓	✓
4	✓	✓	✓		✓				✓	✓
5	✓	✓	✓					✓	✓	✓
6	✓	✓	✓	✓						

TP1NP, total procollagen I N-terminal propeptide; OSTEOC, osteocalcin; PICP, propeptide of type I procollagen; VIT-D, vitamin D; BMI, body mass index; ALP, alkaline phosphatase.

The SKLearn program is a powerful tool based on Python language to process data with a machine learning method, which includes six modules: classification, regression, clustering, dimensionality reduction, model selection and preprocessing. Thus, the data can be easily and automatically analyzed by computer. In our work, the performance of classification was evaluated using three critical outcome measures: accuracy, sensitivity, and positive prediction value. Additionally, a ROC curve was used to evaluate the efficiency of the model accuracy. The area AUC of the ROC with 90% confidence intervals (CI) was also applied to assess the classification.

Results

Importance of features

Different combinations were produced as input sets according to the importance of testing indexes (introduction to the common BTMs is shown in Supporting Information, [Supplementary File](#)). Based on *Figure 3*, the top five attributes were TP1NP, age, P content, gender, and OSTEOC, which are of great guiding significance in the diagnosis of type 2 DM osteoporosis. Interestingly, VIT-D, BMI, and especially ALP were found to be non-significantly important. Moreover, both TP1NP and PICP have been reported to indicate bone formation (28). However, the AI results showed that TP1NP was more sensitive than PICP in BMT-based diagnosis of type 2 DM osteoporosis. Surprisingly, ALP, as the commonly preferred biomarker of osteogenesis, was the least important index. In addition, BMI showed a more loose association with osteoporosis than is usually seen (29). Finally, Ca, PICP, and VIT-D were also found to have less importance than expected.

Classification results

Confusing metrics are often used to evaluate SVM-based classification algorithms, as shown in *Figure 4*. For the evaluation and comprehensive analysis of each classifier, the classification performances of six tests are listed in *Table 3* and plotted in *Figure 5*. As seen from Test 1, 2, and 3, the Acc was over 85% and the PPV was over 50%. There was a positive correlation between the number of attributes and classification accuracy. The Acc score of Test 1, which included 10 attributes, was improved remarkably in comparison to Test 4, 5, and 6, which included four or five attributes. At the same time, it should be noted that the

PPV in all tests was relatively low because of imbalanced data in the verification. As the number of features decreased, the true-positive samples were more difficult to be distinguished from the positively marked samples.

One important conclusion from this is that not all of the testing items are needed. Compared with Test 1, Test 2 with seven attributes showed nearly the same classification Acc and ROC-AUC value. The Sen of Test 2 was even higher than that of Test 1, indicating that it is feasible to use a small number of the most influential testing items for diagnosis. With the same number of dimensions, Test 3 obtained a higher score on classification Acc and ROC-AUC value than did Test 4 and 5, which demonstrated that the TP1NP from Test 3 is a better attribute than the PICP and ALP from Test 4 and 5. This may suggest that TP1NP is more specific as an evaluating indicator for bone metabolism in BTM testing.

In addition, for Tests 1–7, attributes were inadequate for yielding correct results with the SVM algorithm. Test 6, which included the important attributes, showed bad classification performance. Under the premise of not decreasing testing performance, the reduction of items was good. As a result of complex interactions among the organs and systems, the biochemical information from clinical tests may be redundant. The use of AI technology may allow some disease to be diagnosed by relatively simple testing items at a significantly reduced cost. Furthermore, AI can establish connections between the phenotype of serological testing and the development of disease. This is important for the diagnosis of degenerative diseases, such as osteoporosis, for which there remains no highly specific biomarker.

Discussion

The pathogenesis of osteoporosis has often been regarded as an imbalance of bone metabolic networks; however, in patients with type 2 DM, this appears to be more complicated. As some factors may simultaneously have both positive and negative effects, the testing value of common bone metabolic markers may be inaccurate, leading to a false-negative diagnosis. Using a combination of AI and serological testing may help reduce this uncertainty. Our results indicate that the diagnosis of osteoporosis in patients with type 2 DM can be achieved through the calculation and integration of serological test results using machine learning models based on BTMs. Among them, the SVM model, which combines features such as sex, age, BMI, TP1NP,

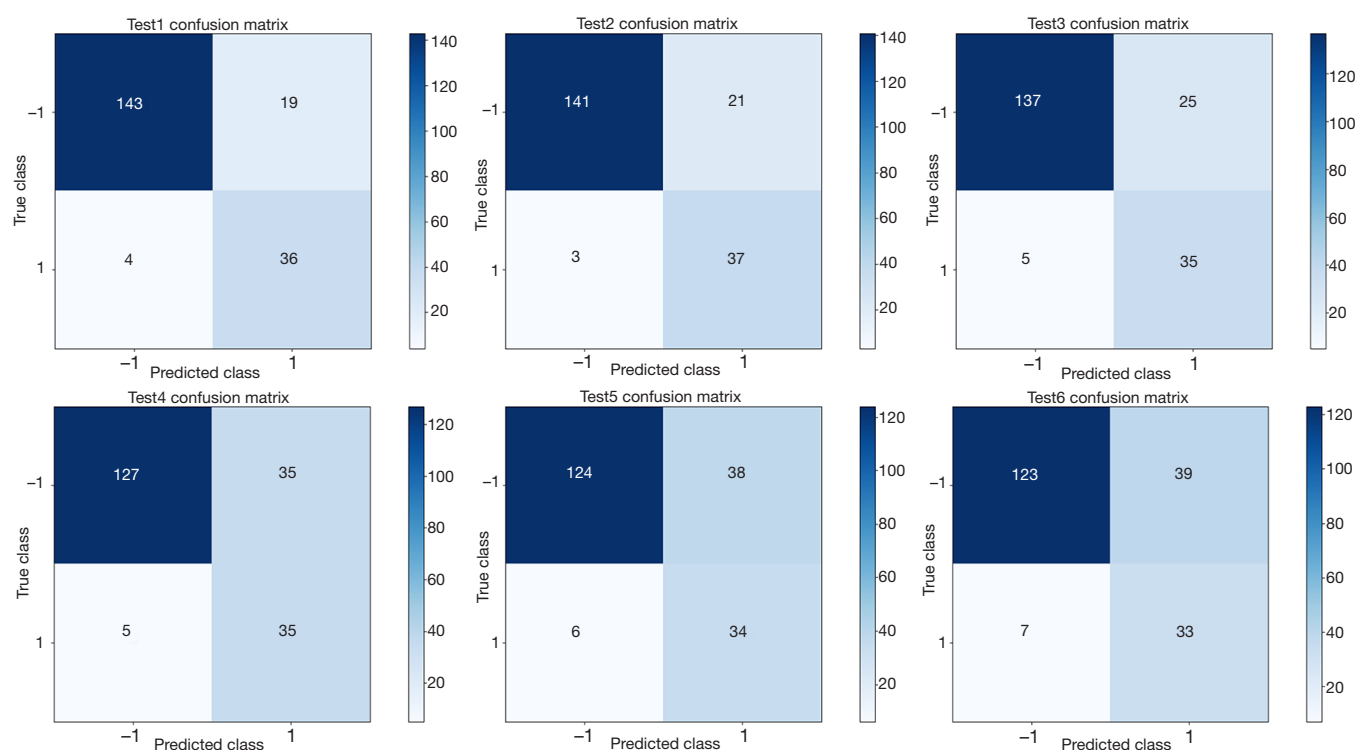


Figure 4 Confusion matrices of the tests.

Table 3 Classification performances

Test number	Acc	PPV	Sen	ROC-AUC values
1	0.8861	0.6545	0.9	0.8914
2	0.8812	0.6379	0.925	0.8977
3	0.8515	0.5833	0.875	0.8603
4	0.8020	0.5	0.875	0.8295
5	0.7821	0.4722	0.85	0.8077
6	0.7723	0.4583	0.8250	0.7921

Acc, accuracy; PPV, positive prediction value; Sen, sensitivity; ROC-AUC, area under the receiver operating characteristic curve.

and OSTEOC, has the best predictive performance with a primary accuracy reaching 88%. This result will guide more fundamental research to establish the connection between these factors. Furthermore, the SVM model found that Vit-D plays a greater role in the diagnosis of type 2 DO than commonly expected. The use of AI technology may, therefore cue researchers to investigate its role in the pathogenesis of type 2 DM. In summary, our view is that with the utilization of more advanced algorithms, AI-assisted diagnosis holds great promise.

Conclusions

The SVM algorithm was used to classify osteoporosis in patients with type 2 DM by relying on several serological items and personal information. This resulted in a diagnostic accuracy of 88%, suggesting that this method may hold promising potential for the diagnosis of type 2 DM complicated with osteoporosis in the clinical setting. This method is cheap, safe, and extensible. Interestingly, some results revealed unexpected findings, such as ALP playing an insignificant role in the AI-based diagnosis.

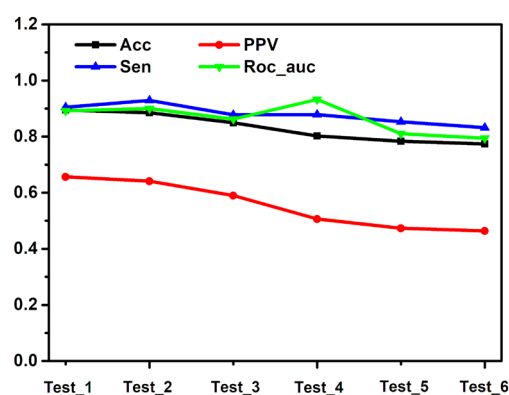


Figure 5 Classification performances of the six combinations with different attributes. Acc, accuracy; PPV, positive prediction value; Sen, Sensitivity; Roc_auc, area under the receiver operating characteristic curve.

These results will be helpful for the clinical and point-of-care diagnosis of osteoporosis, deepening the investigation of its pathological mechanism.

Acknowledgments

Funding: This work was supported by National Natural Science Foundation of China (81802135), the Key Technology RD Program of Jiangsu Province under Grant BE2018010-3.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <http://dx.doi.org/10.21037/atm-20-3388>

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/atm-20-3388>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-20-3388>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki

(as revised in 2013) and approved by the ethical committee of Zhongda Hospital affiliated to Southeast University (No.2018ZDKYSB092) and individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Chen Z, Zhao GH, Zhang YK, et al. Research on the correlation of diabetes mellitus complicated with osteoporosis with lipid metabolism, adipokines and inflammatory factors and its regression analysis. *Eur Rev Med Pharmacol Sci* 2017;21:3900-5.
- Wang N, Xu PC, Wang XP, et al. Timosaponin aiii attenuates inflammatory injury in ages-induced osteoblast and alloxan-induced diabetic osteoporosis zebrafish by modulating the rage/mapk signaling pathways. *Phytomedicine* 2020;75:153247.
- Poiana C, Capatina C. Osteoporosis and fracture risk in patients with type 2 diabetes mellitus. *Acta Endocrinol Buch* 2019;2:231-6.
- Ma R, Zhu R, Wang L, et al. Diabetic Osteoporosis: A review of its traditional Chinese medicinal use and clinical and preclinical research. *Evid Based Complement Alternat Med* 2016;2016:3218313.
- Napoli N, Chandran M, Pierroz D, et al. Mechanisms of diabetes mellitus-induced bone fragility. *Nat Rev Endocrinol* 2017;13:208-19.
- Miazgowski T, Pynka S, Noworyta Z. Bone mineral density and hip structural analysis in type 1 diabetic men. *Eur J Endocrinol* 2007;156:123-7.
- Rakic V, Davis WA, Chubb SAP. Bone mineral density and its determinants in diabetes: the Fremantle Diabetes Study. *Diabetologia* 2006;49:863-71.
- Hamann C, Kirschner S, Gunther KP, et al. Bone, sweet bone—osteoporotic fractures in diabetes mellitus. *Nat Rev Endocrinol* 2012;8:297-305.
- Si Y, Wang C, Guo Y, et al. Prevalence of osteoporosis in patients with type 2 diabetes mellitus in the Chinese

- mainland: A systematic review and meta-analysis. *Iran J Public Health* 2019;48:1203-14.
10. Shu A, Yin MT, Stein E, et al. Bone structure and turnover in type 2 diabetes mellitus. *Osteoporos Int* 2012;23:635-41.
 11. Yamamoto M. Insights into bone fragility in diabetes: The crucial role of bone quality on skeletal strength. *Endocr J* 2015;62:299-308.
 12. Iguacel I, Miguelberges ML, Gomezbruton A, et al. Veganism, vegetarianism, bone mineral density, and fracture risk: a systematic review and meta-analysis. *Nutr Rev* 2019;77:1-18.
 13. Schwartz AV, Sellmeyer DE, Ensrud KE. Older Women with Diabetes Have an Increased Risk of Fracture: A Prospective Study. *J Clin Endocrinol Metab* 2001;86:32-8.
 14. Yokomoto-Umakoshi M, Kanazawa I, Kondo S. Association between the risk of falls and osteoporotic fractures in patients with type 2 diabetes mellitus. *Endocr J* 2017;64:727-34.
 15. Kruse C. The New Possibilities from “Big Data” to Overlooked Associations Between Diabetes, Biochemical Parameters, Glucose Control, and Osteoporosis. *Curr Osteoporos Rep* 2018;16:320-4.
 16. Liu M, Lu Y, Cheng X, et al. Relationship between abnormal glucose metabolism and osteoporosis in Han Chinese men over the age of 50 years. *Clin Interv Aging* 2019;14:445-51.
 17. van Bommel EJM, de Jongh RT, Brands M, et al. The osteoblast: Linking glucocorticoid-induced osteoporosis and hyperglycaemia? A post-hoc analysis of a randomised clinical trial. *Bone* 2018;112:173-6.
 18. Lips P, Eekhoff M, van Schoor N, et al. Vitamin D and type 2 diabetes. *J Steroid Biochem Mol Biol* 2017;173:280-5.
 19. Jackuliak P, Payer J. Osteoporosis, Fractures, and Diabetes. *Int J Endocrinol* 2014;2014:820615.
 20. Kanis JA, Burlet N, Cooper C. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int* 2008;19:399-428.
 21. Kuo TR, Chen CH. Bone biomarker for the clinical assessment of osteoporosis: recent developments and future perspectives. *Biomark Res* 2017;5:18-9.
 22. Lewandowski J, Arochena HE, Naguib RNG. Logic-Centered Architecture for Ubiquitous Health Monitoring. *IEEE J Biomed Health Inform* 2014;18:1525-32.
 23. Li Q, Zhao K, Bustamante C, et al. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genet Med* 2019;21:2126-34.
 24. Safdar S, Zafar S, Zafar N, et al. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artif Intell Rev* 2018;50:597-623.
 25. Noble WS. What is a support vector machine? *Nat Biotechnol* 2006;24:1565-7.
 26. Verbiest N, Ramentol E, Cornelis C, et al. Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection. *Appl Soft Comput* 2014;22:511-7.
 27. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 2010;38:e132-e144.
 28. Price CP, Thompson PW. The role of biochemical tests in the screening and monitoring of osteoporosis. *Ann Clin Biochem* 1995;32:244-60.
 29. Frederiksen AG, Lysgaard AL, Iversen E, et al. BMI might be overestimated in elderly and old osteoporosis patients suffering vertebral fracture(s). *Eur Geriatr Med* 2013;4:S127-8.

Cite this article as: Wang C, Zhang T, Wang P, Liu X, Zheng L, Miao L, Zhou D, Zhang Y, Hu Y, Yin H, Jiang Q, Jin H, Sun J. Bone metabolic biomarker-based diagnosis of type 2 diabetes osteoporosis by support vector machine. *Ann Transl Med* 2021;9(4):316. doi: 10.21037/atm-20-3388

Alkaline phosphatase (ALP), widely distributed in bone, intestine, kidney and other tissues, is one of the diagnostic indexes for bone disease. The total alkaline phosphatase (TALP) level of patients with osteopenia or osteoporosis is higher than that of normal subjects.

Osteocalcin (OSTEOC), the most abundant proteins in bone and produced exclusively by osteoblasts, was initially believed to be an inhibitor of bone mineralization and now have proved to be a useful indicator of bone formation. The faster the bone regeneration rate, the higher the osteocalcin

value, and vice versa.

Propeptide of type I procollagen (PICP) and procollagen type 1 N-propeptide (P1NP) have been proven to be markers of bone formation. They reflect osteoblasts activity and bone formation rate.

Vitamin D (VIT-D) plays a key role in nutritional risk mediated osteoporosis. Vitamin D deficiency, associated with increased bone turnover, can be an important risk factor for osteoporosis.