Observational Study

Medicine®

OPEN

An investigation of machine learning algorithms for prediction of temporomandibular disorders by using clinical parameters

Nazım Tolgahan Yıldız, PhD[a,*], Hikmet Kocaman, PhD[a], Hasan Yıldırım, PhD[b], Mehmet Canlı, MSc[c]

Abstract

This study aimed to predict temporomandibular disorder (TMD) using machine learning (ML) approaches based on measurement parameters that are practically acquired in clinical settings. 125 patients with TMD and 103 individuals without TMD were included in the study. Pain intensity (with visual analog scale), maximum mouth opening (MMO) and lateral excursion movements (with millimeter ruler), cervical range of motion (with goniometer), pressure pain threshold (PPT; with algometer), oral parafunctional behaviors (with Oral Behaviors Checklist), psychological status (with Hospital Anxiety and Depression Scale), and quality of life (with Oral Health Impact Profile) were evaluated. The measurements were analyzed via over 20 ML algorithms, taking into account an extensive parameter tuning and cross-validation process. Results of variable importance were also provided. Bagging algorithm using Multivariate Adaptive Regression Spline (MARS) algorithm (accuracy = 0.8966, area under receiver operating characteristic curve = 0.9387, F1-score = 0.9032) was the best performing model regarding the performance criteria. According to this model, the 5 most important variables for predicting TMD were pain intensity, MMO, lateral excursion and PPT values of masseter and temporalis anterior muscles, respectively. The Bagging algorithm using the MARS algorithm is a robust model that, in combination with clinical parameters, assists in the detection of patients with TMD in settings with limited capabilities. The clinical parameters and ML algorithm proposed in this study may assist clinicians inexperienced in TMD to make a preliminary detection of TMD in clinics where diagnostic imaging tools are limited.

Abbreviations: AI = artificial intelligence, AUC = area under receiver operating characteristic curve, CBCT = cone beam computed tomography, CROM = cervical range of motion, CT = computed tomography, DC-TMD = diagnostic criteria for temporomandibular disorder, HADS = Hospital Anxiety and Depression Scale, MARS = Multivariate Adaptive Regression Spline, ML = machine learning, MMO = maximum mouth opening, MRI = magnetic resonance imaging, OBC = Oral Behaviors Checklist, OHIP-14 = Oral Health Impact Profile, OPBs = oral parafunctional behaviors, PPT = pressure pain threshold, TA = temporalis anterior, TMD = temporomandibular disorder, TMJ = temporomandibular joint.

Keywords: clinical measurement, machine learning, prediction, temporomandibular disorders, variables

1. Introduction

Temporomandibular disorders (TMDs), the second most prevalent musculoskeletal conditions, affect the temporomandibular joints (TMJs), masticatory muscles and related structures and are characterized by dysfunction, pain or both.[1,2] Patients typically suffer from 1 or more of the signs and symptoms of TMD, such as pain and tenderness in the TMJs or masticatory muscles, restrictions in the vertical and horizontal range of motion of the jaw, joint locking, joint sounds, headaches or ear pain.[1–3] TMDs can be broadly categorized as disorders involving arthralgia, myalgia, degenerative joint disease, and internal derangements (disc displacements).[2] The etiology of TMDs is multifactorial, and behavioral, biological and psychosocial factors (such as occlusal factors, deep pain input, emotional stress, parafunction, physical trauma such as whiplash injury, and psychological distress) may contribute to the onset or progression of TMD.[2,4] Among physical traumas, possible relationships between whiplash-related disorders and TMD have been the topic of research. Although there are studies suggesting that TMD may be more prevalent in patients with whiplash injuries than in those without whiplash injuries, the mechanism and pathophysiology

---

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

This cross-sectional and observational study was approved by Karamanoğlu Mehmetbey University Faculty of Medicine Clinical Research Ethics Committee (Number: 06-2022/13). Written and verbal informed consent was obtained from all participants. The study adhered to the principles outlined in the Declaration of Helsinki.

[a] Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Karamanoglu Mehmetbey University, Karaman, Turkey, [b] Department of Mathematics, Kamil Özdağ Faculty of Science, Karamanoğlu Mehmetbey University, Karaman, Turkey, [c] School of Physical Therapy and Rehabilitation, Kırşehir Ahi Evran University, Kırşehir, Turkey.

* Correspondence: Nazım Tolgahan Yıldız, Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Karamanoglu Mehmetbey University, Karaman, Turkey (e-mail: tolgafty@gmail.com).

How to cite this article: Yıldız NT, Kocaman H, Yıldırım H, Canlı M. An investigation of machine learning algorithms for prediction of temporomandibular disorders by using clinical parameters. Medicine 2024;103:41(e39912).

Received: 3 June 2024 / Received in final form: 27 August 2024 / Accepted: 13 September 2024

http://dx.doi.org/10.1097/MD.0000000000039912

of these associations are still not completely clarified.[5] In a recent systematic review and meta-analysis,[6] the prevalence of TMD was reported at 31% among the general population. In the management of TMD, conservative or surgical treatment methods are employed, depending on the patient's clinical situation. Conservative treatment consists of patient education, self-care techniques, intraoral appliances such as splints, physiotherapy interventions, psychological treatment, and pharmacologic treatment. Surgical treatment includes interventions such as arthrocentesis, arthroscopy, joint replacement, or arthroplasty.[1]

The complex structure of TMDs presents challenges in diagnosis. The diagnosis of TMD needs a detailed assessment of the patient's signs and symptoms (obtained by clinical examination and medical screening) as well as psychosocial and behavioral factors by a specialist physician experienced in TMD.[2] At present, the most commonly accepted diagnostic criteria is the diagnostic criteria for temporomandibular disorder (DC-TMD).[7] Although popular in the diagnosis of TMDs, the use of DC-TMD is limited by the fact that it is time-consuming and requires a specialized clinician trained in DC-TMD.[2] Diagnostic imaging is one of the methods utilized both to confirm DC-TMD and to assist the physician in the diagnosis of TMDs. Computed tomography (CT) scans, X-rays, cone beam computed tomography (CBCT), and magnetic resonance imaging (MRI) can be used in various conditions.[2,8] X-rays allow the detection of joint degeneration such as osteoarthritis, while CT scans provide a detailed evaluation of the cartilage and bone structures of the TMJ. CBCT offers high-resolution 3D images with a lower radiation exposure compared to CT scans. MRI allows detailed visualization of soft tissue and is useful for detecting inflammation of the joint and disc displacement.[8] Although diagnostic imaging methods are useful for the visualization of the TMJ and surrounding structures, they are costly, require special equipment and specialized physician training, and some of them cause radiation exposure, which limits their use.[2,9]

The TMD reduces the quality of life of patients and causes high treatment costs in communities. Early diagnosis and treatment of TMD can significantly improve the prognosis and quality of life of patients by preventing chronic pain and disability.[1,10,11] There are studies confirming that the disorder is largely undiagnosed by inexperienced dentists due to the complexity of TMD and the need for a multifactorial assessment.[12] Since DC-TMD[7] is still complex for general and inexperienced dentists in their daily clinical practices, novice and non-specialized dentists who have not been trained in DC-TMD are at high risk of making incorrect and inadequate decisions in the diagnosis of TMD.[11] Furthermore, in epidemiologic studies to determine the prevalence of TMD, it is important that more individuals can be screened for TMD in a shorter period of time and diagnosed in a more practical manner.[13] Considering all these, innovative technologies such as artificial intelligence (AI) with a high accuracy rate gain importance in terms of both diagnosing TMD in a more practical way and detecting TMD at an earlier stage by inexperienced dentists and referring patients to specialists in this field for diagnosis and treatment.[11,12] Currently, with the development of technology, the employment of AI and especially machine learning (ML) approaches in dentistry enables the processing of large data sets and assists in detecting and diagnosing TMDs in a more practical and accurate manner.[8,12] There is evidence in the literature of diagnosing various TMDs via AI utilizing CT, X-rays, CBCT, and MRI.[2,8]

It was reported that in clinics where diagnostic imaging methods are not available and there are no specialists in the field of TMD, inexperienced dentists can detect TMDs via AI methods utilizing a number of clinical measurements.[11,12] Predicting TMD via ML based on relatively practical and accessible clinical measurements may enable inexperienced dentists in general practice to detect TMD at an early stage. On the other hand, studies on predicting TMD via ML based on clinical measurements are quite limited.[11,12]

In light of the aforementioned, this study aimed to predict TMD using ML approaches, based on clinical and psychosocial variables to assist clinicians with limited experience in TMD in making a preliminary detection of TMD.

2. Materials and methods

2.1. Study design and ethical aspects

This cross-sectional and observational study was approved by Karamanoğlu Mehmetbey University Faculty of Medicine Clinical Research Ethics Committee (Number: 06-2022/13). Written and verbal informed consent was obtained from all participants. The study adhered to the principles outlined in the Declaration of Helsinki.

2.2. Participants

The study was conducted with 125 patients with TMD and 103 individuals without TMD who met the inclusion criteria. The patient group comprised patients who admitted to private dentistry clinics and Karamanoğlu Mehmetbey University Faculty of Dentistry with TMD complaints and were diagnosed with TMD. Specialist dentists who were experienced in TMD and calibrated according to DC-TMD Axis I[7] performed clinical examinations of the patients and diagnosed them with TMD by using imaging methods when necessary. The inclusion criteria for the patients with TMD were as follows: being aged between 18 and 55 years, having TMD complaints for at least 6 months, not receiving treatment for TMD for the last 3 months, and not using analgesic drugs, braces, and splints. A total of 147 individuals between the ages of 18 and 55 who had similar demographic characteristics (age, gender) to patients with TMD and who did not have TMD symptoms were considered for the group without TMD. These individuals were clinically examined by the same specialist dentists to ensure that they did not have TMD. Finally, 103 individuals who did not show any signs of TMD and met the inclusion criteria constituted the individuals without TMD. For both groups, those with a history of trauma or surgery including the neck, head and TMJ, cervical disc herniation, history of psychiatric or systemic illness, and those receiving treatment for the orofacial region were excluded from the study. After the clinical and demographic characteristics of the participants were noted, the following assessments were carried out.

2.3. Assessments

A visual analog scale was employed to measure the intensity of pain during jaw functions. Participants indicated their perceived pain intensity by marking a 10-cm line, ranging from "no pain" to "most severe pain," and this measurement, recorded in centimeters, represented the pain intensity.[14]

Maximum mouth opening (MMO; active pain-free) and TMJ lateral excursion movements were assessed using a millimeter ruler. MMO was determined by measuring the distance between the lower and upper central incisors. For TMJ lateral movement, the lateral distance between the midpoints of the lower and upper central incisors was measured separately for the left and right sides. The average of these values from both sides was utilized for analysis.[14]

Active neck extension, flexion, and right and left lateral flexion, representing cervical spine movements, were quantified using a cervical range of motion (CROM) goniometer (CROM Basic, Performance Attainment Associates, St. Paul), measured in degrees. While neck extension and flexion values were documented individually, for lateral flexion value right and left lateral flexion values were averaged. The CROM device has demonstrated validity and reliability in assessing cervical spine movements.[15]

The assessment of tenderness in the masticatory muscles and TMJ was performed by pressure pain threshold (PPT) measurement, which provides reliable and quantitative data.[16] The PPT was assessed using a digital pressure algometer (JTech Commander Algometer, JTech Medical, Salt Lake City) equipped with a 1 cm² probe applied to the anterior portion of the temporalis anterior (TA) muscle, the masseter muscle belly, and the lateral pole of the TMJ. After positioning the probe of the device at the designated measurement location, pressure was incrementally applied, and participants were instructed to indicate the onset of pain by saying "yes." The pressure level eliciting pain was recorded in Newton per square centimeter (N/cm²). Three assessments were taken from each of the 3 sites, with 30-second rest intervals between measurements. The average of these assessments was recorded individually for subsequent analysis.[16]

Participants were assessed for oral parafunctional behaviors (OPBs) utilizing the Oral Behaviors Checklist (OBC). The instrument consists of 21 items designed to evaluate the frequency of OPBs, with respondents providing ratings on a 5-point Likert scale ranging from 0 to 4 for each item. The total score achievable on the scale ranges from 0 to 84, with higher scores indicating a greater frequency of OPBs.[7]

The Hospital Anxiety and Depression Scale (HADS) was employed to assess the psychological status of the participants. The HADS comprises 14 items: 7 are dedicated to assessing the level of anxiety symptoms, and the remaining 7 are focused on evaluating the level of depression symptoms. The scores achievable from the anxiety (HADS-anxiety) and depression (HADS-depression) subscales range from 0 to 21, with higher scores reflecting a greater severity of symptoms.[17]

The Oral Health Impact Profile (OHIP-14) was employed for assessing oral health-related quality of life. The OHIP-14 is an instrument utilized for assessing oral health-related quality of life, providing insight into an individual's general health status through the analysis of functional limitations stemming from oral symptoms as well as social and emotional well-being. In the questionnaire consisting of 14 items, responses are provided on a 5-point Likert scale, yielding a total score ranging from 0 to 56. A higher total score signifies a lower level of oral health-related quality of life.[18]

2.4. Methodology

The methodological part of the study consists of 2 main parts: the statistical analysis and the development of ML models.

In the first part, statistical significance tests were performed between patients with TMD and individuals without TMD. The statistical assumption checks were performed with Shapiro–Wilk test for normality analysis and Levene's test for homogeneity of variance. The comparisons between patients with TMD and individuals without TMD were performed by independent samples *t* test. Differences between categorical variables were analyzed via Chi-square test.

In the second part of methodology, a comprehensive list of basic and advanced algorithms, which are the most widely used in the field, were used in the study and compared in details. While some algorithms are included as standalone, others are included as base learner in ensemble methods. K-nearest neighbors as an instance based algorithm, Naive Bayes, Logistic Regression, Partial Least Squares[19] as statistical algorithms, Classification and Regression Tree,[20] C5.0,[21] C5.0-rules,[21] RuleFit,[22] Bayesian Additive Regression Trees,[23] as tree and rule-based algorithms, Multilayer Perceptron[24] as a neural network based algorithm, Multivariate Adaptive Regression Spline (MARS)[25] as spline-based algorithm, Support Vector Machines[26] as Kernel-based algorithm, Bagging,[27] Random Forests,[28] and Boosting[29] as ensemble algorithms and NULL model as a base reference for benchmarking. As base learner, C5.0, Classification and Regression Tree, MARS and Multilayer Perceptron algorithms are utilized in Bagging ensemble algorithm and C5.0, Light Gradient Boosting Machine and Extreme Gradient Boosting algorithms are utilized in Boosting ensemble algorithm. Additionally, the Support Vector Machines algorithm assessed with 3 different kernels (linear, polynomial and radial). Thus, twenty-three different algorithms were extensively investigated. The statistical analysis comparisons of the study and the development of ML models were carried out with the R software and related packages including tidymodels[30] and tidyverse.

2.5. Parameter tuning and settings

The pre-processing and parameter tuning process carried out to systematically compare ML algorithms and eliminate possible unit differences. Firstly, the numerical data were standardized to have a mean of 0 and unit variance. Then the data set was split into training and test data at a ratio of 75% and 25% respectively. ML models were developed on the training data and validated on the test data. During the process of determination the optimal parameters, the candidate space of thirty elements for each model-specific parameters was examined by using a 10-fold cross-validation. Based on the optimal parameters, the performances were obtained according to the accuracy, F1-score and area under receiver operating characteristic curve (AUC) values on the test data. Finally, confusion matrix, receiver operating characteristic curve and variable importance plots were obtained with the overall best model.

2.6. Performance criteria

Numerous criteria such as accuracy, AUC value, F1-score, recall, precision, Brier score have been used to compare ML models. In this study, the most widely used and accepted accuracy, F1-score, and AUC values were computed and presented according to the confusion matrix given in Table 1.

- Accuracy is defined as the proportion of correctly identified cases (including true positives and true negatives) compared with all cases. As the accuracy value approaches 1, the predictive power of the model increases. Accuracy was calculated by using the components of the confusion matrix as follows:

$$\text{Accuracy} = \frac{(\text{true positive} + \text{true negative})}{(\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative})}$$

- The F1-score, usually referred to as the F-score or F-measure, is a robust statistic for evaluating a binary classification model's accuracy, particularly when there are class imbalances or when the impact of false positives and false negatives vary. The F1-score approaches its least accurate value at 0 and gets its optimal value at

Table 1

A representation of a classic confusion matrix.

		Predicted	
		Positive	**Negative**
Actual (truth)	Positive	True positive	False negative
	Negative	False positive	True negative

Table 2

The results of statistical comparisons based on demographic and clinical variables.

		Patients with TMD (n = 125)	**Individuals without TMD (n = 103)**	
		Mean ± SD	**Mean ± SD**	***P***
Age (yr)		39.05 ± 3.77	38.01 ± 3.29	.084*
Pain intensity, VAS (cm)		6.69 ± 1.12	0.4 ± 0.08	<.001*
MMO (mm)		18.70 ± 2.313	44.86 ± 4.59	<.001*
TMJ lateral movement (mm)		3.76 ± 0.56	9.23 ± 0.88	<.001*
PPT-TMJ (N/cm^2)		20.80 ± 2.34	40.55 ± 3.95	<.001*
PPT-masseter (N/cm^2)		16.70 ± 1.78	38.82 ± 3.29	<.001*
PPT-TA (N/cm^2)		18.70 ± 2.15	37.28 ± 3.34	<.001*
Neck flexion (°)		24.02 ± 2.89	52.74 ± 4.17	<.001*
Neck extension (°)		28.54 ± 2.31	55.86 ± 4.32	<.001*
Neck lateral flexion (°)		22.10 ± 2.11	44.89 ± 3.15	<.001*
OBC score (point)		44.37 ± 4.21	8.78 ± 2.04	<.001*
HADS-anxiety score (point)		14.97 ± 1.94	0.91 ± 0.17	<.001*
HADS-depression score (point)		12.86 ± 1.69	0.69 ± 0.13	<.001*
OHIP-14 score (point)		33.38 ± 3.27	5.14 ± 1.29	<.001*
		n (%)	n (%)	
Gender	Female	85 (68)	68 (66.02)	.895†
	Male	40 (32)	35 (33.98)	
TMD diagnoses	Muscle disorders	54 (43.2)		
	Joint disorders	43 (34.4)		
	Combined muscle-joint disorders	28 (22.4)		

HADS = Hospital Anxiety Depression Scale, MMO = maximum mouth opening, OBC = Oral Behaviors Checklist, OHIP-14 = Oral Health Impact Profile, PPT = pressure pain threshold, SD = standard deviation, TA = temporalis anterior, TMD = temporomandibular disorder, TMJ = temporomandibular joint, VAS = visual analog scale.

* Independent sample *t* test.

† Chi-square test, $P < .05$.

1, which represents excellent recall (also known as sensitivity) and precision and calculated via the following equation:

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

The formulas for the calculation of precision and recall used in the calculation of the F1-score were provided below:

$$\text{Precision} = \frac{\text{true positive}}{(\text{true positive} + \text{false positive})}, \quad \text{Recall} = \frac{\text{true positive}}{(\text{true positive} + \text{false negative})}$$

For the computation of specificity, which is another metric utilized, especially in the derivation of the receiver operating characteristic graph, the following formula was used:

$$\text{Specificity} = \frac{\text{true negative}}{(\text{true negative} + \text{false positive})}$$

- The AUC, a statistical metric, is used to evaluate a model's intrinsic ability to distinguish between positive and negative classes over a variety of classification thresholds. The receiver operating characteristic curve compares the actual positive rate (sensitivity) against the false positive rate (1-specificity) for various threshold settings. When the AUC value is 1.0, the classifier is considered perfect; when it is 0.5, the model's performance in classifying true positives and true negatives is no better than random chance.
- By providing a reference to improve the agreement performance between the model predictions and clinical outcomes for similarity and variability, 4 different metrics, including Jaccard, Cosine, Hamming, and Bernoulli variance scores, were calculated. For the Jaccard and Cosine similarity metrics, values above 0.80 and also for the Bernoulli variance score, values between 0 and 0.25 are indicative of a very good fit. The Hamming score ranges between 0 and the dimensionality of the data set, and generally, values <20% are considered as good fit.[31]

3. Results

The study was completed with 228 participants, including 125 patients with TMD and 103 individuals without TMD who met the inclusion criteria. In the post hoc power analysis process, the pain intensity measurements were evaluated, and the power value of 0.99 was achieved with an effect size = 7.92 from the data obtained from a total of 228 individuals at $\alpha = 0.05$ significance level. Therefore, regarding the level of power analysis after the study, the sample size was quite adequate. The analysis results, which cover the first part of the study and include group-specific statistical significance tests, are displayed in Table 2. According to the results in Table 2, there were statistically significant differences between patients with TMD and individuals without TMD in terms of clinical measurements ($P < .05$), except for age and gender distribution ($P > .05$).

The performance comparisons of the ML models, which constitute the second part of the study, according to 3 different criteria (accuracy, F1-score and AUC value) on the test data are presented in Table 3. The results show that although the models generally yield reasonable results, the Bagging algorithm, which uses the MARS algorithm as the base learner, is found to be the best model. The Bagging (MARS) algorithm outperformed all other models in all 3 criteria (accuracy = 0.8966, AUC = 0.9387, F1-score = 0.9032) and presented the best generalizability capability for unseen data. In addition, it achieved considerably higher scores than the NULL model given as the reference model in the study. Table 3 also presents the similarity and variance metrics for the Bagging (MARS) algorithm, which was the best-performing algorithm. Accordingly, the Jaccard, Cosine, and Hamming similarity and variance metrics obtained for the Bagging (MARS) algorithm revealed a good fit in terms of performance. On the other hand, the performance could be regarded as acceptable, as the Bernoulli variance value was borderline.

The confusion matrix and receiver operating characteristic curve of the Bagging (MARS) algorithm, which was identified as the best model, are illustrated in Figures 1 and 2, respectively, to further clarify the results of the study. It can be concluded from these figures that the performance of the model is fairly strong and the model is able to successfully identify the patients with TMD.

The majority of the ML models included in the study are intrinsically capable of variable importance assessment. The determination of important features in the identification of groups with and without TMD has the potential to make a number of valuable contributions, such as efficiency, rapid decision-making, and narrowing the focus of measurements. Therefore, the relative importance scores of the variables were obtained using the MARS algorithm and are given in Figure 3. These scores represent the proportion of their contribution to group identification rather than statistical significance. Based on the results of Figure 3, it is observed that the 5 most important features identified as pain intensity, MMO, TMJ lateral movement, PPT-masseter, and PPT-TA respectively, while the relatively least important features were neck extension, OHIP-14 score, and neck flexion.

4. Discussion

In this study, ML algorithms were employed to predict TMD using a set of clinical measurements that can be easily applied in clinical settings. All ML algorithms utilized demonstrated high accuracy. On the other hand, the Bagging algorithm, utilizing the MARS algorithm as the base learner, was found to be the best performing model in all 3 criteria (accuracy, AUC, F1-score). Moreover, considering the relative importance of the variables obtained through this model, it was revealed that the 5 most important variables for predicting TMD were pain intensity, MMO, TMJ lateral movement, PPT-masseter, and PPT-TA, respectively.

TMDs are the second most prevalent musculoskeletal disorders following chronic low back pain, affecting 5% to 12% of the population, and costing an estimated $4 billion annually.[32] Early diagnosis and management of TMD are important factors in achieving better patient outcomes. However, due to the complex nature of TMD and some factors that complicate the diagnosis (poor command of DC-TMD, lack of diagnostic imaging equipment and the ability to interpret it), especially inexperienced clinicians may experience varying difficulties in diagnosing TMD, and the majority of TMDs frequently stay undetected.[1,10] Today, along with the rapid progress in technology, AI applications are attracting great interest, especially in the healthcare industry.[8] It has been documented that AI may assist in the diagnosis of maxillofacial and oral diseases such as TMD by analyzing patients' imaging results and clinical parameters.[2,11] Currently, several studies have used a variety of AI

Table 3

The comparative test performance results of machine learning methods.

Model	Accuracy	AUC	F1-score
Bagging (C5.0)	0.8379	0.8373	0.8438
Bagging (Classification and Regression Tree)	0.8434	0.8935	0.8541
Bagging (MARS)	**0.8966**	**0.9387**	**0.9032**
Bagging (Multilayer Perceptron)	0.8671	0.8719	0.8692
Bayesian Additive Regression Trees	0.8692	0.8968	0.871
Boosting (C5.0)	0.8845	0.8989	0.8857
Boosting (Light Gradient Boosting Machine)	0.8348	0.8751	0.8417
Boosting (Extreme Gradient Boosting)	0.8845	0.9000	0.8862
C5.0	0.8512	0.8968	0.8561
C5.0 (Rules)	0.8589	0.8868	0.8681
Classification and Regression Tree	0.7914	0.8546	0.8000
K-nearest Neighbors	0.8225	0.8946	0.8227
Logistic Regression	0.8298	0.8859	0.8391
MARS	0.8712	0.9000	0.8724
Multilayer Perceptron	0.8203	0.8968	0.8215
Naive Bayes	0.7759	0.8308	0.7800
Null Model	0.5517	0.5000	0.7111
Partial Least Squares	0.8272	0.8886	0.8371
Random Forests	0.8655	0.8946	0.8672
RuleFit	0.8494	0.8978	0.8555
Support Vector Machines (linear)	0.8304	0.8448	0.8403
Support Vector Machines (polynomial)	0.8069	0.8448	0.8129
Support Vector Machines (radial)	0.8534	0.8978	0.8585

Similarity and variance metrics for Bagging (MARS) algorithm

Jaccard	Cosine	Hamming	Bernoulli variance
0.862	0.925	5	0.242

AUC = area under receiver operating characteristic curve, MARS = Multivariate Adaptive Regression Spline.

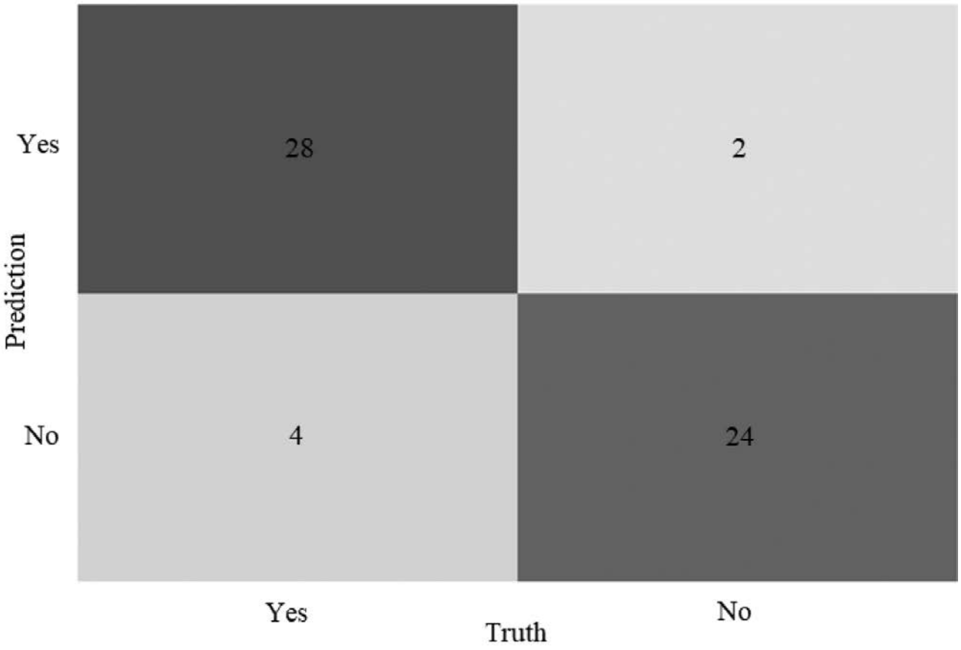

Figure 1. The confusion matrix of the MARS algorithm in Bagging ensemble. MARS = Multivariate Adaptive Regression Spline.

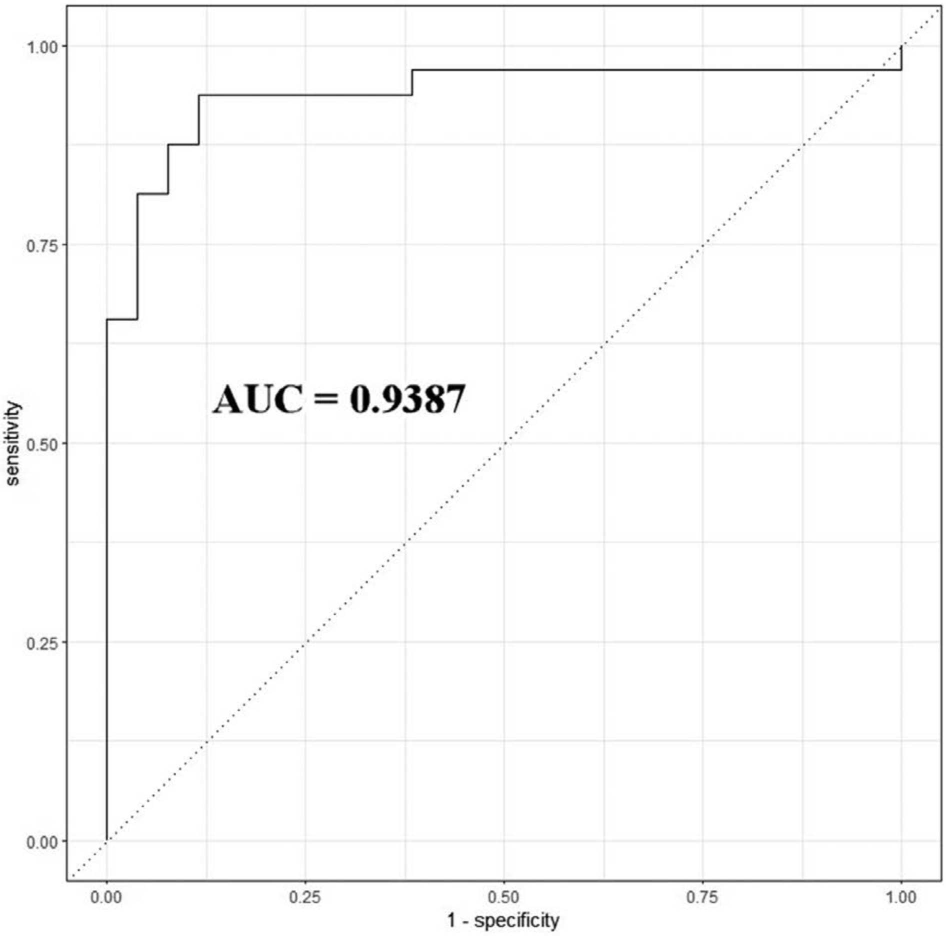

Figure 2. The receiver operating characteristic curve of the MARS algorithm in Bagging ensemble. AUC = area under receiver operating characteristic curve, MARS = Multivariate Adaptive Regression Spline.

algorithms such as ML and deep learning based on image and non-image data to facilitate the diagnosis of TMDs and support clinical decisions.[2]

Reda et al[12] used an AI-based system that can provide possible diagnoses based on a list of symptoms to support non-specialist dentists in the early identification of TMD. The authors reported that AI was a useful tool in facilitating primary diagnosis and improving TMD detection. In the study investigating the clinical utility of an AI deep learning diagnostic tool developed for the diagnosis of TMJ osteoarthritis from CBCT-verified orthopantomogram images, it was noted that AI can play an important role in the diagnosis of TMJ osteoarthritis from orthopantomograms in clinics where radiology specialists or CT are not available.[9] By using MRI images, TMJ disc displacement was estimated with the deep learning method in the study by Kao et al.[10] Conducted on 52 patients with TMJ disc displacement and 32 asymptomatic controls, this research showed that the best AI models were InceptionV3 and DenseNet169 and that these methods can support clinicians in diagnosing TMJ disc displacement from MRI images. In another study,[11] in which an artificial neural network model was developed based on clinical data, it was found that the artificial neural network model could predict TMD risks with a high accuracy rate of 90.91%. It was also suggested that the patient's psychological factors and clinical medical conditions might be parameters that can help determine the risk of TMD via AI.[11] Lee et al,[33] who investigated the usefulness of a deep learning algorithm based on a convolutional neural network in detecting TMJ disc displacement using MRI, determined that deep learning has high accuracy in detecting TMJ disc displacement. In the study performed by Bianchi et al[32] to predict TMJ osteoarthritis with 4 ML models using clinical, radiological features, and biomarkers, it was documented that the Extreme Gradient Boosting + Light Gradient Boosting Machine model had an accuracy rate of 0.823 in estimating TMJ osteoarthritis. In the current study, in which different ML algorithms were used based on clinical measurement data, it was determined that the Bagging algorithm using the MARS algorithm, as the best performing model, was able to predict TMD with a very high accuracy rate. Considering the evidence in the literature and the findings of this study, it seems that AI is a useful method for predicting and diagnosing TMD through various imaging and non-imaging variables. Previous studies have aimed to predict TMD with AI using methods that require more equipped tools.[2,9–12,32,33] Nevertheless, the lack of these equipments, especially imaging methods, and trained specialists in command of these tools limit the usability of these diagnostic approaches. In contrast, in this study, TMD was predicted with ML algorithms using data that can be practically obtained in general clinics. The results of the current study may be helpful for the preliminary detection of TMD in clinics where imaging methods and experts to interpret them are limited, especially for clinicians inexperienced in TMD.

The main symptom of TMD and the primary reason for patients to attend for treatment is pain affecting the masticatory muscles and TMJ. Also, other major complaints of patients are tenderness in the TMJ, or masticatory muscles, and restriction of jaw movements.[6] It has been suggested that a significant reduction in mandibular mobility may be an indicator of the

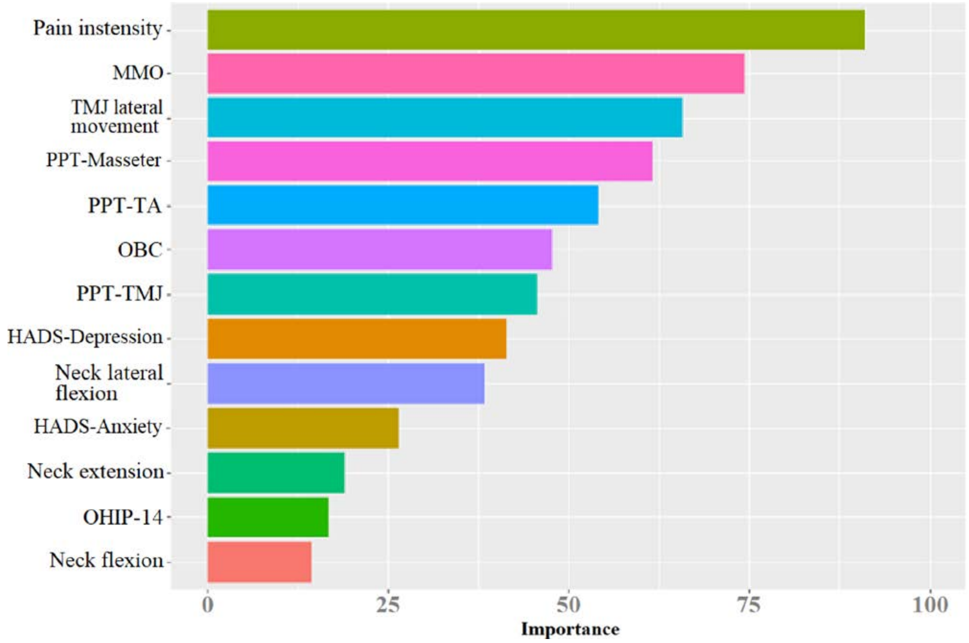

Figure 3. Variable importance plot of the MARS algorithm in Bagging ensemble. HADS = Hospital Anxiety and Depression Scale, MARS = Multivariate Adaptive Regression Spline, MMO = maximum mouth opening, OBC = Oral Behaviors Checklist, OHIP-14 = Oral Health Impact Profile, PPT-TA = pressure pain threshold temporalis anterior, PPT-TMJ = pressure pain threshold temporomandibular joint, TMJ = temporomandibular joint.

presence of TMD.[13] Studies comparing patients with TMD to individuals without TMD in terms of jaw movements and PPT measurements (assessed on TMJ and masticatory muscles) have documented that patients with TMD exhibit lower ranges of jaw movement[13,34] and lower PPT values.[34–36] According to the ML algorithm, which showed the best performance in the present study, the 5 most important variables in predicting TMD were pain intensity, MMO, TMJ lateral movement, and PPT values of the masseter and TA muscles, respectively. These findings, which are consistent with the literature, show that pain intensity, vertical and lateral jaw ranges of motion, and PPT measurements, which provide information about the tenderness in the masticatory muscles and TMJ, can be important parameters that allow differentiation between individuals with and without TMD and preliminary detection of TMD.

OPBs are considered among the important risk factors that may contribute to the onset or progression of TMD.[2,4] In the OPPERA study carried out by Ohrbach et al,[37] it was revealed that patients with TMD had significantly more OPBs compared to individuals without TMD, and OPBs were stated to be important risk factors for the onset of TMD. Research has documented that individuals with TMD have significantly higher levels of stress, anxiety, and depression than those without TMD and that there are consistent associations between psychological problems and TMD.[3] Slade et al[38] asserted that stress and depression lead to a 2- to 3-fold increase in the risk of TMD, suggesting that depression may be one of the risk factors for the initial onset of TMD. According to the Bagging algorithm using the MARS algorithm, which showed the best performance in this study, the OBC score and depression and anxiety scores obtained from HADS were found to be other important parameters in predicting TMD. In the light of these findings, which are consistent with previous research, it can be proposed that OPBs and psychological factors using AI algorithms can help clinicians in the preliminary detection of TMD by enabling them to discriminate between individuals with and without TMD.

Due to the direct anatomical associations between the muscles of the cervical spine and the muscles of the temporomandibular region, as well as the close neurological and biomechanical associations between the cervical spine and the TMJ, impairment in 1 region may often affects the other.[39] In line with this, previous studies have suggested that there may be significant associations between TMDs and cervical spine disorders, demonstrating significant limitations in cervical range of motion in individuals with TMD compared to those without TMD.[39,40] Based on the ML algorithm, which showed the best performance in the current study, it was determined that cervical range of motion values are important variables for predicting TMD, although they are relatively less important. Considering these findings, which support the results in the literature, it is thought that limitations in cervical range of motion may be valuable factors for discriminating between individuals with and without TMD.

Oro-facial function has a considerable influence on individuals' general health and quality of life. Impairment of essential oro-facial functions such as chewing, eating, and speaking and decreased social participation due to TMD may negatively influence the quality of life and general health of individuals.[41] Consistent with this, a systematic review documenting a possible association between TMD and a poorer quality of life suggested that physical and psychological impairments caused by TMD reduce the quality of life of individuals.[42] Almoznino et al[43] found that individuals with TMD had a significantly lower quality of life compared to those without TMD. Another finding of the present study was that quality of life could be used to discriminate between individuals with and without TMD depending on the Bagging algorithm using the MARS algorithm, which showed the best performance. As quality of life is significantly reduced in patients with TMD, and considering these findings of the present study, it can be suggested that quality of life measurements can be used in the estimation of TMD with AI algorithms.

This study had some limitations. Firstly, patients with TMD were not grouped according to variables such as gender and age during the analyses. Consequently, these variables could not be considered when interpreting the findings. A subgroup analysis based on gender and age may provide additional insights into the main determinants of TMDs. Additionally, considering that sleep disturbances may be associated with TMD, the lack of evaluation of participants' sleep disturbances may be counted as another limitation of the current study. Future research would benefit from taking sleep state into account.

5. Conclusion

In this study, which was designed to predict TMD using ML algorithms with a set of clinical measurement parameters that can be practically acquired in clinical settings, all ML algorithms showed high accuracy, but the Bagging algorithm using the MARS algorithm was found to be the best-performing model. In addition, considering the relative importance of the variables obtained through this model, the 5 most important variables for predicting TMD were pain intensity, MMO, TMJ lateral motion, and PPT values of the masseter and TA muscles, respectively. The high accuracy of the ML model employed in this study for predicting TMD indicated the potential of AI to assist in the preliminary detection of TMD using practically acquired clinical measurement parameters. Given the limited availability of well-equipped diagnostic imaging tools, experts to interpret them, and specialized clinicians experienced in TMD in general clinics, the clinical parameters and ML algorithm proposed in this study may help clinicians with limited experience in the field of TMD in making a preliminary detection of TMD.

Author contributions

Conceptualization: Nazım Tolgahan Yıldız, Hikmet Kocaman.
Data curation: Nazım Tolgahan Yıldız, Hikmet Kocaman.
Formal analysis: Hikmet Kocaman, Hasan Yıldırım.
Methodology: Nazım Tolgahan Yıldız, Hikmet Kocaman, Hasan Yıldırım, Mehmet Canlı.
Supervision: Nazım Tolgahan Yıldız, Hikmet Kocaman, Hasan Yıldırım, Mehmet Canlı.
Validation: Nazım Tolgahan Yıldız, Hikmet Kocaman, Hasan Yıldırım, Mehmet Canlı.
Visualization: Nazım Tolgahan Yıldız, Hikmet Kocaman, Hasan Yıldırım, Mehmet Canlı.
Writing – original draft: Nazım Tolgahan Yıldız, Hikmet Kocaman, Mehmet Canlı.
Writing – review & editing: Nazım Tolgahan Yıldız, Hikmet Kocaman, Hasan Yıldırım, Mehmet Canlı.

References

[1] Palmer J, Durham J. Temporomandibular disorders. BJA Educ. 2021;21:44–50.
[2] Jha N, Lee K-S, Kim Y-J. Diagnosis of temporomandibular disorders using artificial intelligence technologies: a systematic review and meta-analysis. PLoS One. 2022;17:e0272715.
[3] Rollman GB, Gillespie JM. The role of psychosocial factors in temporomandibular disorders. Curr Rev Pain. 2000;4:71–81.
[4] Chung J, Lobbezoo F, van Selms MK, Chattrattrai T, Aarab G, Mitrirattanakul S. Physical, psychological and socio-demographic predictors related to patients' self-belief of their temporomandibular disorders' aetiology. J Oral Rehabil. 2021;48:109–23.
[5] Corsalini M, Capodiferro S, dell'Olio F, et al. Cranio-mandibular disorders after whiplash injury: a mono-institutional clinical study on 31 patients. Int J Environ Res Public Health. 2022;19:901.
[6] Valesan LF, Da-Cas CD, Réus JC, et al. Prevalence of temporomandibular joint disorders: a systematic review and meta-analysis. Clin Oral Investig. 2021;25:441–53.
[7] Schiffman E, Ohrbach R, Truelove E, et al.; International RDC/TMD Consortium Network, International association for Dental Research. Diagnostic criteria for temporomandibular disorders (DC/TMD) for clinical and research applications: recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. J Oral Facial Pain Headache. 2014;28:6–27.
[8] Ozsari S, Güzel MS, Yılmaz D, Kamburoğlu K. A comprehensive review of artificial intelligence based algorithms regarding temporomandibular joint related diseases. Diagnostics (Basel). 2023;13:2700.
[9] Choi E, Kim D, Lee J-Y, Park H-K. Artificial intelligence in detecting temporomandibular joint osteoarthritis on orthopantomogram. Sci Rep. 2021;11:10246.
[10] Kao Z-K, Chiu N-T, Wu H-TH, et al. Classifying temporomandibular disorder with artificial intelligent architecture using magnetic resonance imaging. Ann Biomed Eng. 2023;51:517–26.
[11] Zou W, Mao B, Fadlullah ZM, Qi K. Assisting in diagnosis of temporomandibular disorders: a deep learning approach. IEEE Access. 2022;10:124076–82.
[12] Reda B, Contardo L, Prenassi M, Guerra E, Derchi G, Marceglia S. Artificial intelligence to support early diagnosis of temporomandibular disorders: a preliminary case study. J Oral Rehabil. 2023;50:31–8.
[13] Masumi S, Kim Y, Clark G. The value of maximum jaw motion measurements for distinguishing between common temporomandibular disorder subgroups. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2002;93:552–9.
[14] La Touche R, Fernández-de-Las-Peñas C, Fernández-Carnero J, et al. The effects of manual therapy and exercise directed at the cervical spine on pain and pressure pain sensitivity in patients with myofascial temporomandibular disorders. J Oral Rehabil. 2009;36:644–52.
[15] Micarelli A, Viziano A, Granito I, Micarelli RX, Augimeri I, Alessandrini M. Temporomandibular disorders and cervicogenic dizziness: relations between cervical range of motion and clinical parameters. Cranio. 2022;40:348–57.
[16] Yıldız NT, Kocaman H, Yıldırım H. Predictors of the masticatory muscle activity during chewing in patients with myogenous temporomandibular disorder. Clin Oral Investig. 2023;27:6547–58.
[17] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361–70.
[18] Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol. 1997;25:284–90.
[19] Wold HOA. Soft modelling: the basic design and some extensions. In: Joreskog KG, Wold HOA, eds. Systems Under Indirect Observation, Part II. Amsterdam: North-Holland; 1982:36–7.
[20] Razi MA, Athappilly K. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. Expert Syst Appl. 2005;29:65–74.
[21] Quinlan JR. Combining instance-based and model-based learning. Proceedings of the Tenth International Conference on Machine Learning. 1993.
[22] Friedman JH, Popescu BE. Predictive learning via rule ensembles. Ann Appl Stat. 2008;2:916–54.
[23] Chipman HA, George EI, McCulloch RE. BART: bayesian additive regression trees. Ann Appl Stat. 2010;4:266–98.
[24] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989;2:359–66.
[25] Friedman JH. Multivariate adaptive regression splines. Ann Stat. 1991;19:1–67.
[26] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing. Adv Neural Inf Process Syst. 1996;9:281–87.
[27] Breiman L. Bagging predictors. Mach Learn. 1996;24:123–40.
[28] Breiman L. Random forests. Mach Learn. 2001;45:5–32.
[29] Freund Y, Schapire RE. Experiments with a new boosting algorithm. ICML. 1996;96:148–56.
[30] Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Boston, MA, USA. 2020. Available at: https://www.tidymodels.org. Accessed December 8, 2023.
[31] Choi S-S, Cha S-H, Tappert CC. A survey of binary similarity and distance measures. J Syst Cybern Informatics. 2010;8:43–8.
[32] Bianchi J, de Oliveira Ruellas AC, Goncalves JR, et al. Osteoarthritis of the temporomandibular joint can be diagnosed earlier using biomarkers and machine learning. Sci Rep. 2020;10:8012.
[33] Lee Y-H, Won JH, Kim S, Auh Q-S, Noh Y-K. Advantages of deep learning with convolutional neural network in detecting disc displacement of the temporomandibular joint in magnetic resonance imaging. Sci Rep. 2022;12:11352.
[34] Hansdottir R, Bakke M. Joint tenderness, jaw opening, chewing velocity, and bite force in patients with temporomandibular joint pain and matched healthy control subjects. J Orofac Pain. 2004;18:108–13.
[35] Więckiewicz W, Woźniak K, Piątkowska D, Szyszka-Sommerfeld L, Lipski M. The diagnostic value of pressure algometry for temporomandibular disorders. Biomed Res Int. 2015;2015:575038.

[36] Santos Silva RS, Conti PCR, Lauris JRP, da Silva ROF, Pegoraro LF. Pressure pain threshold in the detection of masticatory myofascial pain: an algometer-based study. J Orofac Pain. 2005;19:318–24.

[37] Ohrbach R, Bair E, Fillingim RB, et al. Clinical orofacial characteristics associated with risk of first-onset TMD: the OPPERA prospective cohort study. J Pain. 2013;14:T33–50.

[38] Slade G, Diatchenko L, Bhalang K, et al. Influence of psychological factors on risk of temporomandibular disorders. J Dent Res. 2007;86:1120–5.

[39] Grondin F, Hall T, Laurentjoye M, Ella B. Upper cervical range of motion is impaired in patients with temporomandibular disorders. Cranio. 2015;33:91–9.

[40] Stiesch-Scholz M, Fink M, Tschernitschek H. Comorbidity of internal derangement of the temporomandibular joint and silent dysfunction of the cervical spine. J Oral Rehabil. 2003;30:386–91.

[41] Ohrbach R, Larsson P, List T. The jaw functional limitation scale: development, reliability, and validity of 8-item and 20-item versions. J Orofac Pain. 2008;22:219–30.

[42] Bitiniene D, Zamaliauskiene R, Kubilius R, Leketas M, Gailius T, Smirnovaite K. Quality of life in patients with temporomandibular disorders. A systematic review. Stomatologija. 2018;20:3–9.

[43] Almoznino G, Zini A, Zakuto A, et al. Oral health-related quality of life in patients with temporomandibular disorders. J Oral Facial Pain Headache. 2015;29:231–41.