Submitted: 20 January, 2026 Accepted: 20 April, 2026 Published: 03 September, 2026 DOI:10.22514/jocpd.2026.120

Journal of Clinical Pediatric Dentistry

ORIGINAL RESEARCH

Predictive performance of supervised machine learning models for temporomandibular joint disc displacement in adolescents diagnosed by magnetic resonance imaging

Seyit Erol[1,*], Halil Özer[1], İsmail Dilek[1], Abdi Gürhan[1], Ahmet Baytok[1], Mehmet Öztürk[1]

[1]Department of Radiology, Faculty of Medicine, Selcuk University, 42075 Konya, Turkey

***Correspondence**
seyit.erol@selcuk.edu.tr
(Seyit Erol)

Abstract

Background: Temporomandibular joint (TMJ) disc displacement is a common condition among adolescents and often requires accurate imaging-based diagnosis. This study aimed to assess the predictive performance of supervised machine learning (ML) algorithms in classifying TMJ disc displacement in adolescents based on morphological and signal intensity features derived from Magnetic Resonance Imaging (MRI). **Methods**: MRI data from 260 TMJs of 130 adolescent patients (aged 12–18 years) were retrospectively analyzed. Morphological variables included condylar shape, anteroposterior and mediolateral diameters, and a condylar ratio derived from these measurements. Signal intensity ratios of the superior and inferior heads of the lateral pterygoid muscle were also recorded. Disc status was initially categorized as normal, anterior displacement with reduction, or without reduction; for modeling purposes, dislocated groups were combined. ML analyses were performed using a patient-level data split to prevent information leakage, ensuring that both joints from the same patient were assigned to the same dataset. Fifteen supervised classification models were trained using group-aware cross-validation within the training set, followed by evaluation on a held-out test set. Hyperparameters were optimized exclusively within the training data. Model performance was assessed using standard classification metrics, with 95% confidence intervals estimated by patient-level bootstrap resampling. **Results**: Anteroposterior and mediolateral diameters differed significantly between disc status groups ($p < 0.001$). Among the evaluated models, MLPClassifier achieved the highest area under the receiver operating characteristic curve (ROC-AUC) (0.797), followed by Random Forest (0.777) and HistGradientBoosting (0.773). Although CatBoost ranked fourth in overall discrimination (ROC-AUC = 0.767), it demonstrated the highest recall (0.839), indicating superior sensitivity for detecting disc displacement. Ensemble and neural network-based models generally outperformed linear classifiers. **Conclusions**: Supervised ML algorithms showed balanced potential in diagnosing TMJ disc displacement in adolescents. Further validation in prospective clinical and multicenter studies is warranted to support clinical application.

Keywords

Temporomandibular joint disorders; Magnetic resonance imaging; Machine learning; Pediatric radiology

1. Introduction

Temporomandibular joint disorders (TMDs), among the most frequent causes of orofacial pain, are increasingly prevalent across all age groups. TMDs encompass a spectrum of conditions affecting the temporomandibular joint (TMJ), the masticatory muscles, or both [1]. The TMJ is a complex structure composed of a fibrocartilaginous disc, joint capsule, and supporting ligaments connecting the mandibular condyle to the temporal bone. Clinically, TMDs are characterized by joint and muscle pain, joint sounds (*e.g.*, clicking or popping), and limitations in mandibular movement, including joint locking [1, 2]. Prevalence estimates range from 10% to 70% in the general population and from 16% to 68% among pediatric and adolescent populations [3, 4]. Recent consensus statements emphasize that TMD diagnosis is primarily based on clinical assessment, rather than imaging, according to standardized criteria such as the Diagnostic Criteria for Temporomandibular

Disorders (DC/TMD) framework [5, 6]. Imaging is considered complementary and is indicated when structural characterization of the joint may influence treatment planning or when clinical findings are inconclusive [7].

Accurate diagnosis of TMDs is essential, as they significantly impact quality of life. Clinical examination complemented by radiological imaging is critical for precise assessment. Computed tomography (CT) is mainly employed for evaluating fractures, postoperative changes, and osseous pathologies. Bone scintigraphy aids in detecting inflammatory and osteoarthritic changes, whereas ultrasonography has shown some potential for joint evaluation, though its diagnostic capacity remains limited [8]. The gold standard for assessing the soft tissue structures of the TMJ, especially for identifying disc displacements, is still Magnetic Resonance Imaging (MRI) [9]. MRI's non-invasive nature and absence of ionizing radiation make it especially suitable for use in younger populations. Some studies have suggested that combining MRI with CT may improve diagnostic accuracy and inter-examiner reliability in TMJ evaluations [9]. Moreover, MRI is capable of identifying early indicators of joint dysfunction, such as osseous abnormalities, joint effusion, disc morphology changes, and thickening of the anterior or posterior disc bands [10]. MRI does not establish the clinical diagnosis of TMD, but provides valuable morphological information regarding internal derangement and structural changes when indicated [11]. Importantly, the presence of disc displacement alone is not sufficient to diagnose or manage TMD, as disc position and pain symptoms show only weak correlation in many cases [12]. Nevertheless, MRI has particular value in adolescents, where disc displacement may influence mandibular growth and remodeling due to incomplete skeletal maturation [13].

Recently, machine learning (ML) techniques have been increasingly explored to enhance diagnostic accuracy and automate the detection of TMJ disc displacement. Although MRI supports clinical evaluation, the diagnosis of TMD remains clinical, and artificial intelligence (AI) models should be interpreted as tools for standardized image assessment, rather than independent diagnostic systems [6]. Previous AI studies have primarily focused on adult populations, and limited evidence exists regarding the performance of supervised ML models in adolescents, whose ongoing skeletal growth may affect both morphology and imaging characteristics [14, 15]. AI-assisted diagnostic tools have shown promise, especially in cases where clinical findings are inconclusive or subjective [16]. Several studies have reported that AI models can detect TMJ-related abnormalities with accuracies as high as 0.93 [14]. Recent research has focused on segmenting the joint space and localizing disc position through deep learning-based neural networks applied to MRI data [16]. Additionally, ultrasonography-based approaches incorporating AI have demonstrated high diagnostic accuracy in assessing TMJ morphology by segmenting anatomical structures such as the mandibular condyle and joint space [17].

This study aimed to evaluate the predictive performance of supervised ML models in classifying TMJ disc displacement based on MRI findings in adolescents aged 12 to 18 years. Morphological and signal intensity-based features—including mandibular condyle shape, condylar anteroposterior and mediolateral dimensions, and signal intensities from the superior and inferior heads of the lateral pterygoid muscle—were extracted as input variables. Fifteen ML algorithms were developed by performing class balancing and hyperparameter optimization. Their classification performances were compared to assess the potential of ML models to support clinical decision-making in the diagnosis of TMJ disc displacement in the adolescent population. Rather than diagnosing TMD, our aim was to classify disc displacement patterns based on quantitative MRI features and explore whether supervised ML models could provide standardized image-based characterization to support clinical decision-making. This approach aligns with recent recommendations that emphasize imaging-based classification of internal derangement, rather than MRI-based "diagnosis" of TMD [18].

2. Materials and methods

Study approval was obtained from the local clinical research ethics committee (file number: 2024/316). We have read the Helsinki Declaration and have followed the guidelines in this investigation.

2.1 Patient selection

This retrospective study included adolescent patients aged 12 to 18 years who underwent TMJ MRI for various clinical indications—including persistent orofacial pain, limited mouth opening, joint sounds (clicking), locking, or suspected internal derangement—at Selcuk University Faculty of Medicine Hospital between January 2018 and February 2024, provided that the MRI examinations were of sufficient image quality and that mouth opening was adequate for reliable assessment of disc position. A total of 130 patients (260 TMJs) were initially evaluated for disc displacement, condylar morphology, and lateral pterygoid muscle (LPM) status. Exclusion criteria included non-diagnostic or poor-quality MRI scans, a history of TMJ trauma or surgery, presence of rheumatologic or systemic diseases, prior orthodontic treatment, and the presence of lesions within the TMJ region. Based on these criteria, 10 patients were excluded: 8 due to inadequate mouth opening on MRI preventing disc displacement evaluation, 1 due to a traumatic mandibular fracture, and 1 due to confirmed mandibular osteomyelitis.

2.2 MRI protocol

MRI of the TMJ was performed using a 1.5 T (Aera, Siemens Healthcare, Erlangen, BY, Germany) MRI scanner with a 20-channel head coil. All scans were acquired in the supine position, without sedation or intravenous contrast administration. In accordance with routine clinical practice, mouth images were taken with the mouth closed and at the maximum voluntary mouth opening that the patient could tolerate. The routine TMJ MRI protocol in our institution includes the following sequences:

- Coronal T2-weighted images (T2-W) with the mouth closed: field of view (FOV) = 150 mm; matrix/number of excitations (NEX) = 320 × 180/2; slice thickness/inter-slice gap (ST/IP) = 4.0/0.4 mm; repetition time (TR) = 4000 ms;

echo time (TE) = 90 ms; echo train length (ETL) = 15; bandwidth = 195 Hz/pixel.

- Sagittal oblique T1-weighted images with the mouth closed: FOV = 100 mm; matrix/NEX = 190 × 130/2; ST/IP = 3.0/0.3 mm; TR = 1100 ms; TE = 12 ms; ETL = 3; bandwidth = 135 Hz/pixel.

- Sagittal oblique proton density-weighted (PD-W) images with the mouth open and closed: FOV = 100 mm; matrix/NEX = 190 × 130/2; ST/IP = 3.0/0.3 mm; TR = 2100 ms; TE = 35 ms; ETL = 8; bandwidth = 180 Hz/pixel.

MRI provides visualization of the TMJ disc, periarticular bone marrow, joint space, masticatory muscles, and surrounding soft tissues.

2.3 MRI evaluation

All MRI images were jointly evaluated by two radiologists during the same session, and consensus-based decisions were recorded. Morphometric measurements were performed using dedicated imaging software (Syngo.via, version VB30, Siemens Healthineers, Erlangen, BY, Germany). TMJs were categorized into three groups based on disc position, following standardized nonosseous component assessment criteria [15].

- Normal: Normal articular disc position.

- ADwR: Anterior disc displacement with reduction.

- ADwoR: Anterior disc displacement without reduction (Fig. 1).

On sagittal PD-W images, the shape of the mandibular condyle was classified as round (convex), flat (blunt), or beak (spur-like) (Fig. 2). The anteroposterior and mediolateral diameters of the condyle were measured from cortex to cortex on sagittal and coronal sections, respectively (Fig. 3). Measurements were performed in a consensus setting by two radiologists to minimize variability. A derived variable, CondyleRatio, was calculated by dividing the mediolateral diameter by the anteroposterior diameter.

Additionally, signal intensities of the superior and inferior heads of the LPM were measured on coronal T2-W images. Slice selection was standardized by identifying the coronal section demonstrating the maximum cross-sectional area of the LPM belly, avoiding partial volume effects at the muscle origin or insertion. Circular regions of interest (ROIs) with a fixed area of 4.97 mm^2 were manually placed within the central portion of each LPM head, taking care to avoid adjacent fat planes, visible blood vessels, and susceptibility artifacts. For signal normalization, reference ROIs of identical size were placed in visually homogeneous regions of cerebral gray matter (GM) and white matter (WM), avoiding cortical boundaries and cerebrospinal fluid contamination (Fig. 4). Signal intensity ratios (SIRs) were calculated separately for the superior and inferior heads of the LPM using the following normalisation formula [19]: SIR: (SI(measured) − SI(GM))/(SI(WM) − SI(GM)). All measurements were performed by two radiologists in a consensus reading setting.

2.4 Machine learning for TMJ dislocation description, model assessment, and performance measures

This study aimed to assess the predictive performance of supervised ML algorithms in classifying TMJ disc displacement in adolescents using MRI-derived morphological and signal-based features. Predictor variables included anteroposterior and mediolateral condylar diameters, the derived CondyleRatio (mediolateral-to-anteroposterior ratio), condylar and disc morphology, and normalized SIRs of the superior and inferior heads of the lateral pterygoid muscle (LPMsupSIR and LPMinfSIR).

2.4.1 Outcome definition

Initially, TMJ disc position was categorized into three classes: Normal, anterior disc displacement with reduction (ADwR), and anterior disc displacement without reduction (ADwoR). Due to class imbalance and overlapping clinical characteristics between the ADwR and ADwoR groups, these two categories were merged into a single Dislocated disc group. This binary classification (Normal *vs.* Dislocated) provided a more stable predictive setting and improved model generalizability in the supervised learning context. A three-class classification (Normal, ADwR, ADwoR) was additionally explored as a secondary analysis.

2.4.2 Patient-level data splitting

To prevent data leakage arising from correlated bilateral TMJs, all data splitting was performed at the patient level. The dataset was divided into training (80%) and testing (20%) sets using GroupShuffleSplit, with the patient identifier used as the grouping variable. This ensured that TMJs from the same patient were never distributed across both the training and test sets.

2.4.3 Cross-validation and preprocessing

Model development and hyperparameter optimization were conducted exclusively within the training set using group-aware cross-validation. All preprocessing steps, including feature standardization with StandardScaler, were implemented within the cross-validation pipeline and fitted only on the corresponding training folds to avoid information leakage.

2.4.4 Model training and hyperparameter optimization

A total of 15 supervised ML classifiers were implemented: Random Forest, Logistic Regression, Ridge Classifier, Linear SVC, K-Nearest Neighbor, Gaussian Naive Bayes, Multilayer Perceptron, Stochastic Gradient Descent, AdaBoost, Gradient Boosting, HistGradientBoosting, Extra Trees, CatBoost, LightGBM, and XGBoost.

Hyperparameter optimization was performed using the Optuna framework with five-fold cross-validation, adopting a nested search strategy to minimize model selection bias. The test set was not accessed during model selection or optimization. After optimization, each model was refit once on the full training set using the selected hyperparameters. Final model performance was then evaluated a single time on the

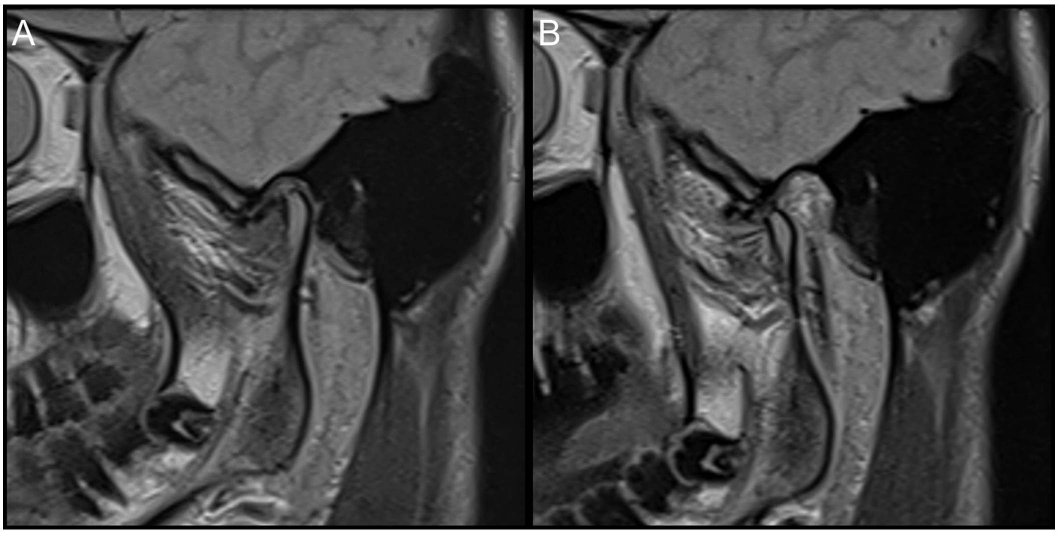

FIGURE 1. Oblique sagittal proton density-weighted magnetic resonance images of the temporomandibular joint. (A) Closed-mouth image showing an anteriorly displaced disc. (B) Open-mouth image showing a non-reduced disc position, consistent with anterior disc displacement without reduction.

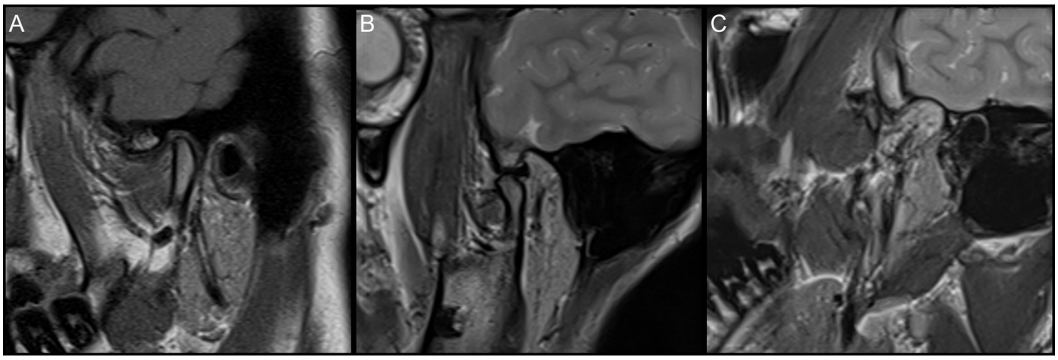

FIGURE 2. Oblique sagittal proton density-weighted magnetic resonance images demonstrating examples of mandibular condyle morphologies. (A) Closed-mouth image showing a round condyle. (B) Open-mouth image showing a flat condyle. (C) Open-mouth image showing a beak-shaped condyle.

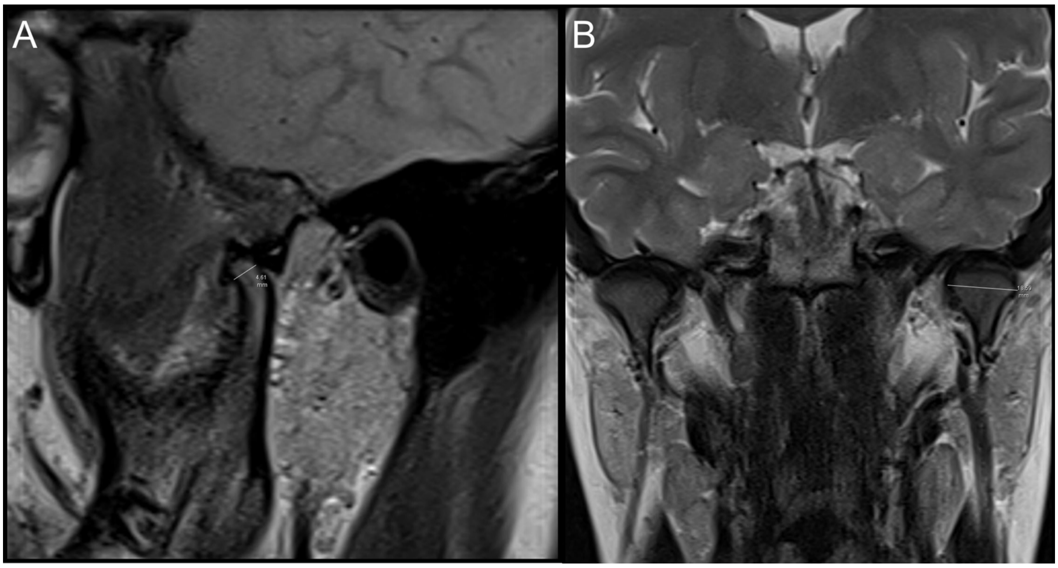

FIGURE 3. Condylar dimension measurements on magnetic resonance images. (A) Anteroposterior diameter measured on oblique sagittal proton density-weighted image. (B) Mediolateral diameter measured on coronal T2-weighted image.

FIGURE 4. Coronal T2-weighted axial magnetic resonance image showing signal intensity measurements obtained using regions of interest. Measurements were taken from the superior and inferior heads of the lateral pterygoid muscle, as well as from the cerebral white and grey matter.

held-out test set, without any further tuning. Model performance was evaluated using Accuracy, Precision, Recall, F1-score, and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC), all calculated on the withheld test set. To quantify model uncertainty, 95% confidence intervals (CIs) for Accuracy and ROC-AUC were estimated using 1000 bootstrap resamples. The six best-performing classifiers based on ROC-AUC were selected for extended evaluation.

To further interpret error patterns, confusion matrices for these four classifiers were generated from test-set predictions (Fig. 5), providing a detailed visualization of true positive, true negative, false positive, and false negative distributions. Model explainability was assessed using SHAP (SHapley Additive exPlanations). Global feature importance was computed using mean absolute SHAP values derived from the positive (Dislocated) class, and feature ranking profiles were visualized for each model (Fig. 6).

All preprocessing, model development, and statistical analyses were performed in Python (3.11.5) using the following libraries: pandas (1.5.3), numpy (1.24.0), scikit-learn (1.1.3), imbalanced-learn, Optuna (3.1.1), matplotlib (3.6.3), seaborn (0.11.2), and statsmodels (0.13.5).

2.5 Statistical analysis

Categorical variables were summarized as frequencies and percentages, and comparisons between groups were conducted using the chi-square test. For continuous variables, means, standard deviations, and 95% CIs were calculated. Effect sizes were reported using eta-squared (η^2) and Cohen's f statistics.

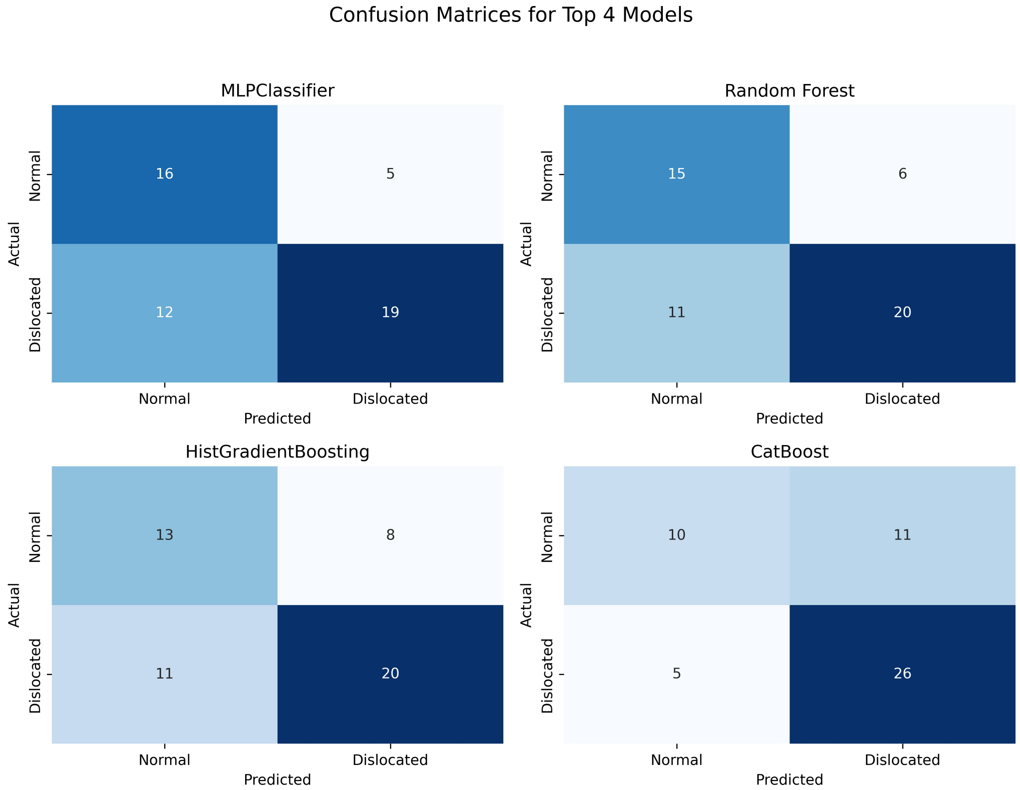

FIGURE 5. Confusion matrices of the four best-performing classifiers (MLPClassifier, Random Forest, HistGradientBoosting, and CatBoost) evaluated on the held-out patient-level test set. The split was performed at the patient level to prevent information leakage between contralateral temporomandibular joints. CatBoost demonstrated the lowest number of false-negative predictions for disc displacement, indicating high sensitivity for pathology detection, while MLPClassifier showed the highest overall discriminative balance. Random Forest and HistGradientBoosting exhibited comparable error patterns.

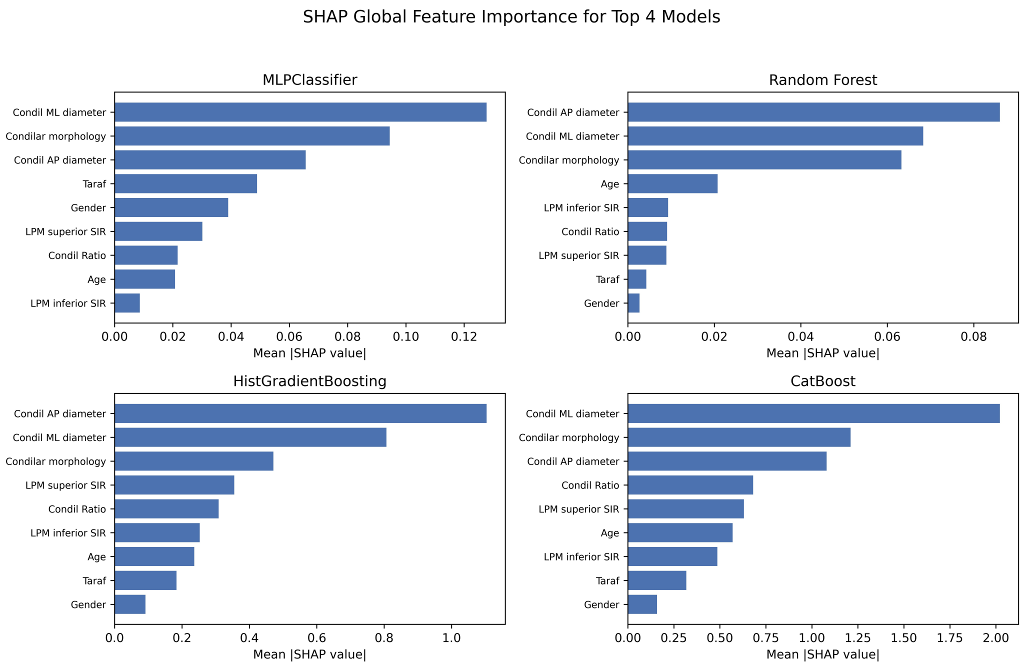

FIGURE 6. SHAP-based global feature importance plots for the top four classifiers. Across all models, mediolateral (CondilML) and anteroposterior (CondilAP) condylar diameters consistently emerged as the most influential predictors of disc displacement, followed by condylar morphology and lateral pterygoid muscle signal intensity ratios. Demographic variables (age and sex) showed minimal contribution. SHAP values are presented on model-specific scales and are not intended for direct quantitative comparison across models. SHAP: SHapley Additive exPlanations; ML: Machine learning; AP: anteroposterior; SIR: Signal intensity ratio; LPM: lateral pterygoid muscle.

The Kolmogorov-Smirnov test was applied to assess the normality of continuous variables. For normally distributed variables, one-way analysis of variance (ANOVA) was performed to compare means across the three disc displacement groups. *Post-hoc* pairwise comparisons were conducted using Tukey's Honestly Significant Difference (HSD) test. For variables not normally distributed, the Kruskal-Wallis test was employed, followed by Bonferroni-corrected *post-hoc* comparisons where appropriate. A two-tailed p-value < 0.05 was considered statistically significant. All statistical analyses and visualizations were performed using Python (version 3.11.5), with the assistance of the scipy, statsmodels, pandas, seaborn, and matplotlib libraries.

3. Results

3.1 Demographic and morphological characteristics of the study cohort

A total of 260 TMJs from 130 patients were analyzed, including 66 males (25.4%) and 194 females (74.6%), with a mean age of 15.66 ± 1.31 years. The unit of analysis was the individual temporomandibular joint, while data splitting for ML purposes was performed at the patient level, ensuring that both joints from the same patient were assigned to the same dataset (training or test) to prevent information leakage. The distribution of TMJ disc displacement was as follows: 103 (39.6%) normal, 88 (33.8%) ADwR, and 69 (26.5%) ADwoR.

Regarding condylar morphology, 131 (50.4%) were classified as round, 91 (35%) as flat, and 38 (14.6%) as beak-shaped. Beak-shaped condyles and folded discs were more commonly found in cases with ADwoR ($p < 0.001$) (Table 1).

Both the anteroposterior and mediolateral diameters of the mandibular condyle significantly differed among the disc displacement groups ($p < 0.001$). *Post-hoc* analysis revealed that the AP diameter was significantly different across all three groups (Normal > ADwR > ADwoR; $p < 0.05$), while the mediolateral diameter was significantly smaller in both dislocated groups (ADwR and ADwoR) compared with the Normal group ($p < 0.001$), with no difference between ADwR and ADwoR ($p > 0.05$). Effect size analysis indicated a moderate effect for AP diameter ($\eta^2 = 0.093$, Cohen's f = 0.32) and a large effect for mediolateral diameter ($\eta^2 = 0.116$, Cohen's $f = 0.36$).

3.2 Machine learning model performance

The classification performance of fifteen supervised ML models in detecting TMJ disc displacement was assessed using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. All performance estimates were derived from a patient-level held-out test set, and 95% CIs were computed using patient-level bootstrap resampling. A comprehensive summary of the leakage-free model performances is presented in Table 2.

3.2.1 Binary classification (normal *vs.* dislocated)

Among all classifiers, MLPClassifier achieved the highest ROC-AUC (0.797; 95% CI: 0.647–0.917), followed by Random Forest (0.773; 95% CI: 0.626–0.901), HistGradientBoosting (0.770; 95% CI: 0.621–0.900), and CatBoost (0.765; 95% CI: 0.598–0.907).

CatBoost demonstrated the highest recall (0.842; 95% CI: 0.667–1.000), indicating strong sensitivity for detecting dislocated discs. k-nearest neighbors (KNN) showed the highest recall overall (0.907), although with comparatively lower ROC-AUC (0.722).

Accuracy values ranged from 0.631 to 0.694 across models, with CatBoost (0.694), KNN (0.694), and AdaBoost (0.689) achieving the highest test accuracies.

Linear models (Logistic Regression, Ridge, LinearSVC, Stochastic Gradient Descent (SGD)) showed relatively lower discriminative performance (ROC-AUC range: 0.681–0.702), whereas ensemble-based tree models demonstrated more stable performance across metrics. ROC curves of the four best-performing models are shown in Fig. 7, and comparative metric visualizations are provided in Figs. 8,9.

3.2.2 Multiclass classification (normal *vs.* ADwR *vs.* ADwoR)

For the three-class classification task, macro-averaged one-vs-rest ROC-AUC was used as the primary discrimination metric.

CatBoost achieved the highest ROC-AUC (0.677; 95% CI: 0.539–0.801), followed closely by Gradient Boosting (0.671; 95% CI: 0.544–0.789), AdaBoost (0.668; 95% CI: 0.540–0.794), and Random Forest (0.662; 95% CI: 0.533–0.796).

Overall accuracy values ranged from 0.369 to 0.540, with KNN (0.540) and AdaBoost (0.535) demonstrating the highest accuracies. However, these models did not correspondingly achieve the highest ROC-AUC values, reflecting differences in class-wise discrimination balance.

Compared with binary classification, multiclass performance was consistently lower across all algorithms, reflecting the increased complexity of distinguishing between ADwR and ADwoR subtypes.

3.3 SHAP-based feature importance

SHAP analysis was performed for the three best-performing binary classifiers based on ROC-AUC (MLPClassifier, Random Forest, and CatBoost).

Across models, mediolateral (CondilML) and anteroposterior (CondilAP) condylar diameters consistently emerged as the most influential predictors. LPMsupSIR demonstrated moderate importance, followed by LPMinfSIR and CondyleRatio (mediolateral-to-anteroposterior ratio).

The consistency of feature ranking across heterogeneous model architectures supports the robustness of these MRI-derived morphometric biomarkers. Demographic variables contributed minimally to predictive performance.

3.4 Confusion matrix analysis

Confusion matrices for the three top-performing binary models (MLPClassifier, Random Forest, and CatBoost) revealed model-specific error patterns.

CatBoost demonstrated high sensitivity for the dislocated class, consistent with its elevated recall. MLPClassifier exhibited balanced precision and recall, resulting in the highest overall ROC-AUC. Random Forest displayed stable classifica-

TABLE 1. Comparison of disc locations with condyle morphology (n = 260).

	Normal (n = 103)	AdwR (n = 88)	AdwoR (n = 69)	*p*-value*
Round	70 (68.0%)	42 (47.7%)	19 (27.5%)	
Flat	29 (28.2%)	33 (37.5%)	29 (42.0%)	<0.001
Beak	4 (3.9%)	13 (14.8%)	21 (30.4%)	

**: Chi-Square Test p-value. ADwR: Anterior dislocation with reduction; ADwoR: Anterior dislocation without reduction.*

TABLE 2. Performance metrics of fifteen supervised machine learning models in the binary classification of TMJ disc displacement diagnosed on MRI.

Model	Accuracy (95% CI)	Recall (95% CI)	Precision (95% CI)	F1-Score (95% CI)	ROC-AUC (95% CI)
MLPClassifier	0.675 (0.538–0.808)	0.619 (0.412–0.813)	0.788 (0.600–0.952)	0.689 (0.511–0.842)	0.797 (0.647–0.917)
RandomForest	0.671 (0.538–0.788)	0.649 (0.458–0.839)	0.763 (0.600–0.917)	0.697 (0.540–0.828)	0.773 (0.626–0.901)
HistGradientBoosting	0.631 (0.500–0.769)	0.648 (0.469–0.818)	0.710 (0.524–0.889)	0.673 (0.519–0.806)	0.770 (0.621–0.900)
CatBoost	0.694 (0.558–0.827)	0.842 (0.667–1.000)	0.705 (0.553–0.842)	0.764 (0.640–0.870)	0.765 (0.598–0.907)
LightGBM	0.653 (0.519–0.788)	0.682 (0.500–0.848)	0.721 (0.548–0.885)	0.697 (0.545–0.827)	0.748 (0.589–0.889)
XGBoost	0.652 (0.500–0.788)	0.714 (0.500–0.889)	0.705 (0.529–0.875)	0.706 (0.537–0.841)	0.738 (0.588–0.885)
DecisionTree	0.655 (0.519–0.788)	0.620 (0.429–0.808)	0.757 (0.583–0.920)	0.677 (0.520–0.815)	0.729 (0.569–0.864)
GradientBoosting	0.618 (0.481–0.769)	0.717 (0.516–0.906)	0.669 (0.500–0.839)	0.687 (0.542–0.815)	0.735 (0.572–0.892)
KNN	0.694 (0.577–0.827)	0.907 (0.750–1.000)	0.684 (0.537–0.826)	0.777 (0.656–0.884)	0.722 (0.552–0.868)
AdaBoost	0.689 (0.538–0.827)	0.776 (0.577–0.938)	0.722 (0.545–0.882)	0.744 (0.581–0.867)	0.723 (0.562–0.867)
LinearSVC	0.654 (0.500–0.808)	0.650 (0.423–0.871)	0.738 (0.545–0.905)	0.686 (0.490–0.842)	0.702 (0.524–0.861)
LogisticRegression	0.635 (0.481–0.788)	0.682 (0.471–0.882)	0.699 (0.517–0.880)	0.686 (0.518–0.828)	0.696 (0.517–0.860)
Ridge	0.634 (0.481–0.788)	0.713 (0.518–0.886)	0.685 (0.500–0.865)	0.695 (0.528–0.833)	0.690 (0.504–0.858)
SGD	0.637 (0.481–0.788)	0.713 (0.518–0.886)	0.688 (0.500–0.861)	0.697 (0.533–0.839)	0.681 (0.493–0.853)
ExtraTrees	0.638 (0.481–0.788)	0.619 (0.400–0.833)	0.731 (0.550–0.889)	0.666 (0.480–0.818)	0.675 (0.492–0.837)

Performance metrics are presented as means with 95% confidence intervals (CI).
ROC-AUC: Receiver Operating Characteristic-Area Under the Curve; MLP: Multilayer Perceptron; SGD: Stochastic Gradient Descent; SVC: Support Vector Classifier; KNN: k-nearest neighbors.

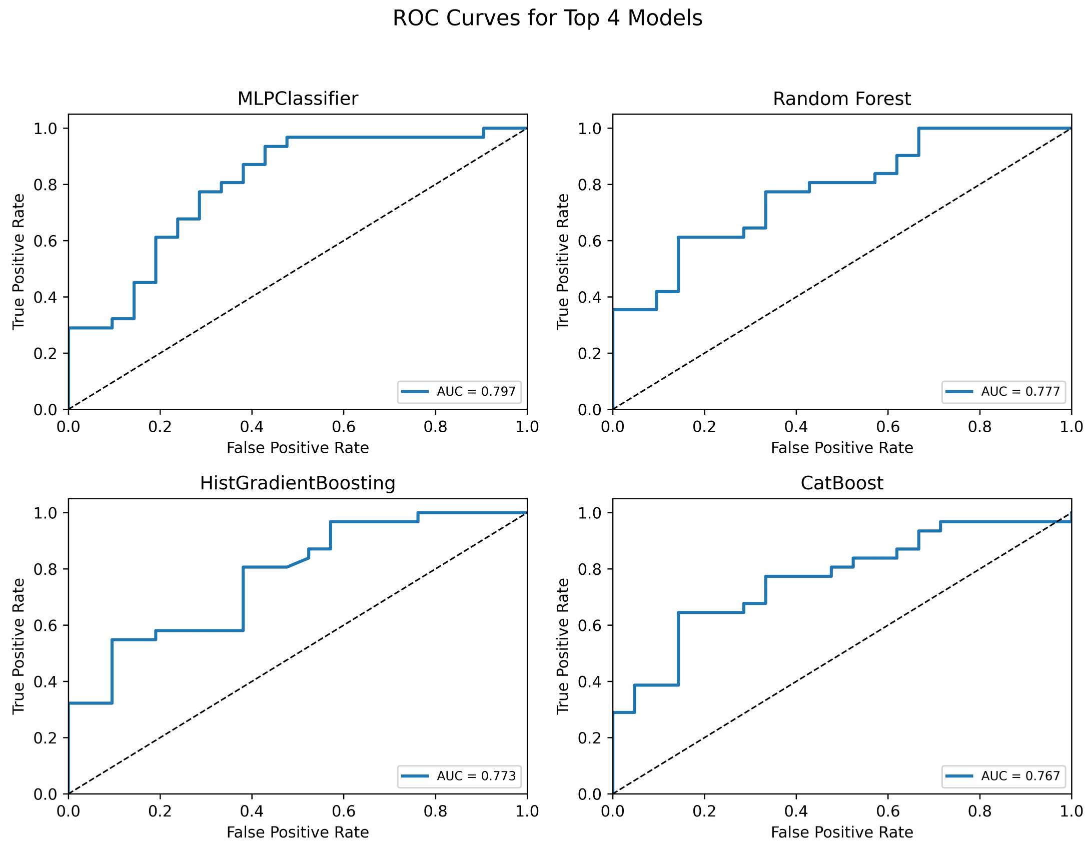

FIGURE 7. Receiver operating characteristic (ROC) curves of the four best-performing machine learning models for binary classification of temporomandibular joint disc status (normal *vs.* dislocated). Curves were generated using predictions from the held-out test set, which was not involved in model training, hyperparameter optimization, or threshold selection. ROC: Receiver Operating Characteristic; AUC: Area Under the Curve; MLP: Multilayer Perceptron.

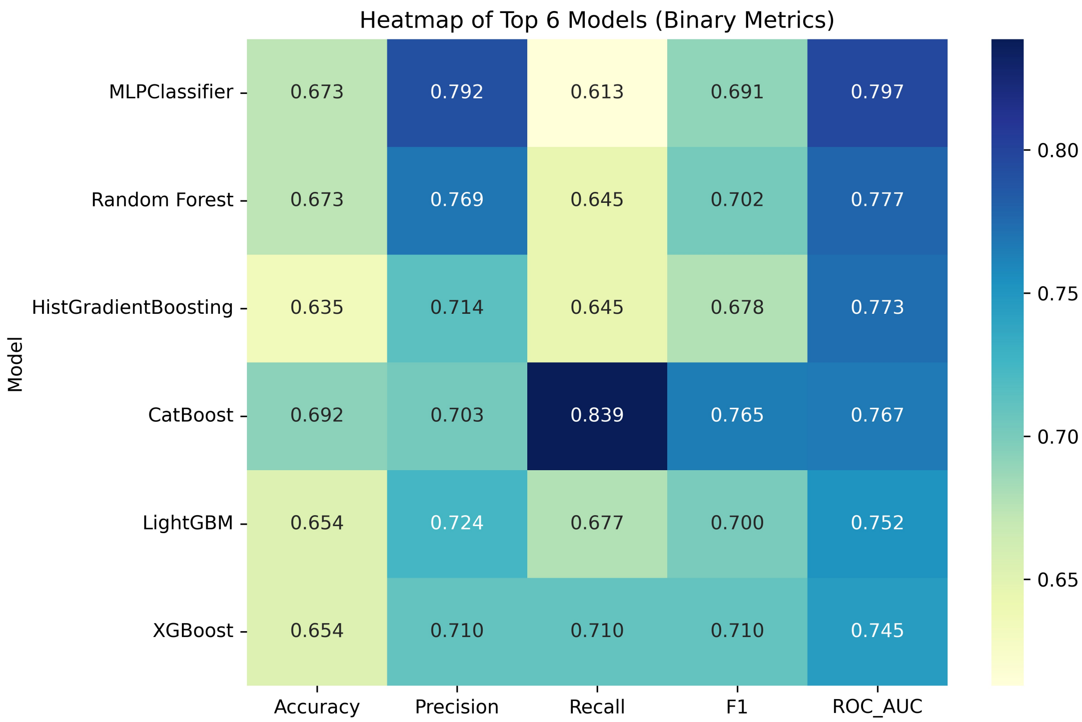

FIGURE 8. Heatmap summarizing the classification performance of the top six models across five evaluation metrics: accuracy, precision, recall, F1-score, and area under the ROC curve (ROC-AUC). Metric values are derived from the patient-level test set. The heatmap provides a comparative overview of model-specific performance trade-offs, rather than a ranking based on a single metric. ROC: Receiver Operating Characteristic; AUC: Area Under the Curve; MLP: Multilayer Perceptron.

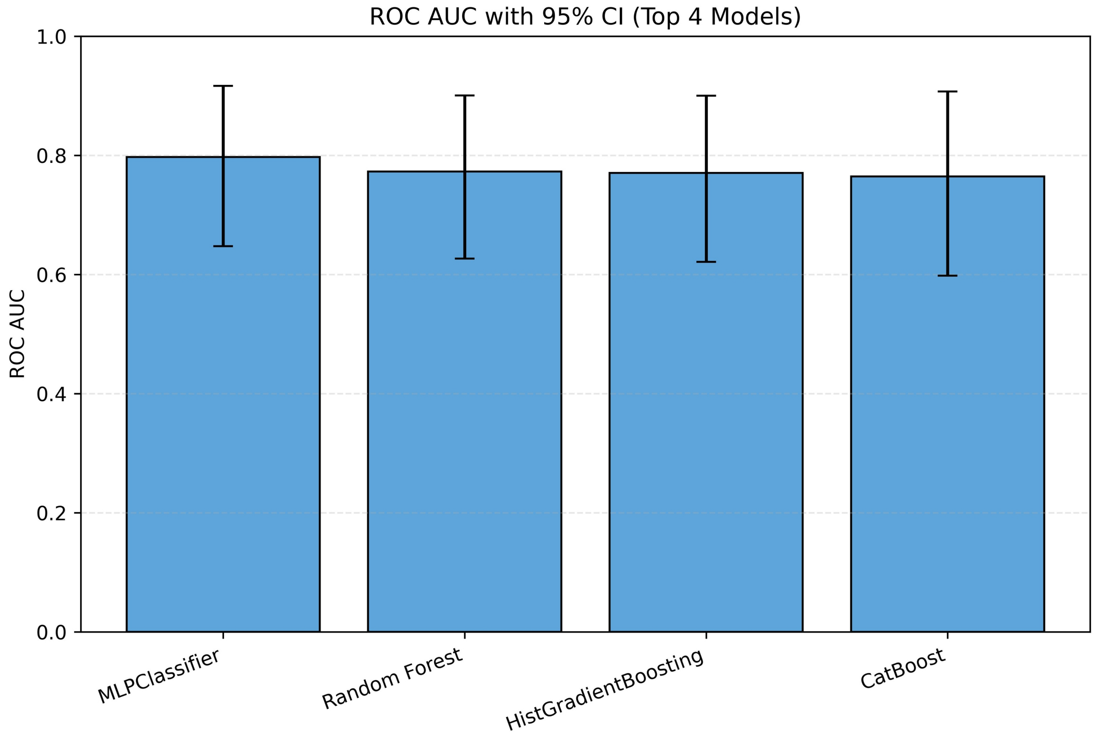

FIGURE 9. Bar plots showing ROC-AUC values of the four best-performing classifiers, along with 95% confidence intervals estimated using patient-level bootstrap resampling of the held-out test set (1000 iterations). Overall, performance differences between the top five models were moderate, and confidence intervals partially overlapped, suggesting comparable discriminative ability under the leakage-free evaluation framework. ROC: Receiver Operating Characteristic; AUC: Area Under the Curve; MLP: Multilayer Perceptron; CI: confidence intervals.

tion behavior with moderate false-positive and false-negative rates.

Although no single model clearly outperformed all others across every metric, ensemble-based and neural network-based classifiers consistently provided more balanced diagnostic performance compared with linear methods.

4. Discussion

Temporomandibular disorders (TMDs) constitute a heterogeneous group of musculoskeletal conditions affecting the masticatory system, frequently emerging during adolescence and early adulthood. TMDs are common, multifactorial conditions that primarily affect individuals during adolescence and early adulthood, often presenting with joint clicking, pain, and limited mandibular mobility [2]. Although TMJ disc dislocation can occur in various directions, anterior displacement is the most frequently observed subtype [20]. Epidemiological studies consistently demonstrate a higher prevalence of TMD among females. A systematic review and meta-analysis reported that women have approximately twice the risk of TMDs compared with men [21]. In clinical series, TMD signs and symptoms have also been found to be significantly more frequent in females [22]. Pediatric and adolescent cohorts similarly show increased TMD prevalence in females, suggesting that sex-related factors may influence both symptom burden and imaging-based presentations. Early recognition of structural alterations may be relevant in adolescents, given the ongoing craniofacial growth. Pathological alterations in the mandibular condyle prior to skeletal maturity can lead to long-term asymmetries and growth disturbances in the mandible [23]. MRI is the reference imaging modality for visualizing TMJ disc morphology and joint structures; however, contemporary guidelines indicate that imaging findings should complement, rather than define, the clinical diagnosis of TMD. The International Association for Dental Research (IADR) recently articulated updated standards of care, highlighting that imaging findings must be interpreted within a clinical context, as many structural variations detected on MRI occur in asymptomatic individuals and do not, by themselves, mandate treatment [6]. In the present study, we investigated the predictive performance of fifteen supervised ML algorithms in classifying TMJ disc displacement in adolescents using MRI-derived morphometric and signal-based features.

While MRI is valuable for visualizing TMJ structures, contemporary orofacial pain guidelines emphasize that disc displacement is not a diagnostic or therapeutic target, and its association with pain is weak [11, 12]. Recent recommendations from the IADR also highlight that TMD management should not be based solely on structural imaging findings [6]. Therefore, MRI should be prescribed only when clinical findings warrant its use, in accordance with established imaging appropriateness frameworks [7]. Because disc displacement and other MRI findings may also be observed in asymptomatic individuals, disc position should be interpreted cautiously in the clinical context [24].

Several studies have reported that the mandibular condyle is smaller in joints exhibiting anterior disc displacement compared with those with a normal disc position, both in mediolateral and anteroposterior dimensions [25]. Experimental evidence further suggests that TMJ disc displacement can negatively affect mandibular development, potentially resulting in reduced mandibular height and length due to impaired condylar growth [13]. Our findings align with this body of evidence, as both anteroposterior and mediolateral diameters of the mandibular condyle were significantly smaller in joints diagnosed with ADwR and ADwoR compared with those with normal disc position. These results support the hypothesis that disc displacement, particularly in adolescents, may interfere with normal growth and remodeling processes of the TMJ.

It has been suggested that, due to the morphological and functional characteristics of the TMJ, adjacent anatomical structures in joints with disc displacement may be subject to changes in response to altered functional loading patterns [26]. Prior research has identified a relationship between osteoarthropathic changes and specific condylar morphologies, particularly flattened or beak-like condyle [27]. This association is believed to be of particular importance during adolescence, a critical period for craniofacial growth and joint development. In our study, approximately 50% of the TMJs exhibited a rounded condylar morphology. However, the prevalence of beak-shaped and flat condyles was markedly higher in joints diagnosed with ADwR and ADwoR. No statistically significant association was observed between condyle morphology and patient sex. These findings are consistent with the literature and further support the hypothesis that disc displacement may influence not only the size, but also the shape of the mandibular condyle, potentially contributing to early degenerative or adaptive joint changes.

In recent years, ML models have attracted increasing attention for their potential to enhance sensitivity and overall diagnostic performance in detecting TMJ disc displacement. As demonstrated by multiple studies, the integration of ML techniques into TMJ imaging analysis may contribute to more accurate diagnoses and better treatment planning. In a study by Orhan *et al.* [28], the diagnostic validity of MRI-based ML models was evaluated using support vector machines, random forest, and k-nearest neighbor algorithms to classify TMJ disorders associated with condylar changes and disc displacement. The study reported promising results, with an AUC of 0.89 in the training set and 0.77 in the test set, indicating strong generalizability of the model [28]. Similarly, Yüksel *et al.* [15] employed a radiomics-based ML approach to extract quantitative imaging features from MRI scans and classify TMJ disc displacement. Among the algorithms tested, the k-nearest neighbor classifier demonstrated the highest predictive performance and was identified as the most suitable model for this task [15]. In our study, non-linear and ensemble-based models demonstrated slightly superior overall discrimination compared with traditional linear classifiers. The MLPClassifier achieved the highest ROC-AUC (0.797), followed by Random Forest and HistGradientBoosting, while CatBoost showed the highest recall, indicating enhanced sensitivity for detecting disc displacement. These findings suggest that advanced non-linear approaches may provide a more adaptable and clinically balanced solution for classifying adolescent TMJ disc status based on MRI-derived morphometric and signal intensity features.

ML applications in TMDs are not limited to imaging-based evaluations such as MRI, but have also been applied to the analysis of clinical and functional parameters. In the present study, the proposed model is intended to function as a decision-support tool in cases where TMJ MRI is already clinically indicated, assisting radiologists in identifying morphometric patterns associated with disc displacement, rather than replacing expert interpretation. Taborri *et al.* [29] investigated the diagnostic performance of nine ML algorithms using mobility-related and baropodometric measurements and found that the k-nearest neighbors algorithm yielded the best performance, with an accuracy of 0.94 and an F1-score of 0.91 [29]. Complementing this, Yildiz *et al.* [30] evaluated over 20 ML models to predict TMDs using easily obtainable clinical parameters such as pain intensity and psychological factors. Their results indicated that a Bagging classifier using the Multivariate Adaptive Regression Spline (MARS) algorithm achieved the highest performance, with an accuracy of 0.89 and ROC-AUC of 0.90 [30]. In our study, non-linear ML models demonstrated modest, but consistent, advantages over linear classifiers in classifying TMJ disc displacement. The MLPClassifier achieved the highest overall discriminative performance based on ROC-AUC, while Random Forest and HistGradientBoosting showed comparable results with balanced accuracy and F1-scores. Notably, CatBoost exhibited the highest recall among the evaluated models, indicating enhanced sensitivity for detecting displaced discs, albeit with slightly lower specificity. HistGradientBoosting and LightGBM also performed competitively across multiple metrics, whereas linear classifiers, such as Logistic Regression and Ridge Classifier, yielded comparatively lower performance. Confusion matrix analysis provided complementary insight into model behavior, revealing that CatBoost minimized false-negative predictions, while MLPClassifier and Random Forest demonstrated more balanced error distributions. Collectively, these findings suggest that advanced non-linear and ensemble-based approaches may offer a more flexible and clinically relevant framework for TMJ disc displacement classification using MRI-derived morphometric and signal intensity features, particularly when sensitivity to pathology is prioritized.

Zhang *et al.* [31] conducted a systematic review and meta-analysis evaluating the diagnostic accuracy of ML models in distinguishing TMDs using imaging data. Although the review encompassed both traditional ML and deep learning algorithms, the meta-analysis did not provide a definitive conclusion as to which type of model demonstrated superior diagnostic performance for TMDs [31]. Similarly, Farook and Dudley reviewed twenty studies focused on automation and deep learning applications in TMJ radiomics. Their findings indicated that deep learning models performed similarly to clinicians in diagnosing TMJ osteoarthritis from both 2D and 3D radiographs. However, these models exhibited reduced performance when applied to MRI data for the detection of disc displacement or internal derangement [32].

This study has several limitations that should be acknowledged. First, the retrospective nature of the study limited the inclusion of important clinical variables, such as pain severity, functional limitations, and psychological status, which may influence the diagnosis and progression of TMDs. Second, although all TMJ MRI examinations were evaluated by two radiologists in consensus, inter-observer agreement was not formally assessed, and the evaluations were not performed independently at separate time-point factors that may introduce interpretative bias. Accordingly, model performance reflects agreement with expert consensus, rather than robustness to inter-reader variability, and clinical applicability should be interpreted cautiously. Additionally, all MRI data were obtained from a single institution, which may affect the generalizability of the findings to wider populations and imaging protocols. Future research should aim to incorporate multicenter datasets, integrate relevant clinical variables, and include inter-observer variability assessment to validate and strengthen the utility of ML models in TMJ diagnostics.

5. Conclusions

This study demonstrates the potential of supervised ML algorithms in classifying TMJ disc displacement in adolescents using MRI-derived morphometric and muscle signal features. Among the fifteen evaluated models, non-linear approaches showed modest, but consistent, advantages over traditional linear classifiers. The MLPClassifier achieved the highest overall discriminative performance, while tree-based ensemble models demonstrated comparable and stable results across evaluation metrics. Notably, CatBoost exhibited the highest sensitivity for detecting disc displacement, suggesting potential utility in scenarios where minimizing false-negative findings is clinically important.

Morphometric parameters—particularly condylar diameters—were strongly associated with disc displacement, underscoring their diagnostic relevance. To our knowledge, this is one of the largest ML-based TMJ studies conducted in an adolescent MRI cohort incorporating both structural and muscle signal characteristics. Future research should focus on external validation in multi-institutional datasets and on integrating automated segmentation and radiomics-based features to further enhance diagnostic robustness and generalizability.

ABBREVIATIONS

TMJ, Temporomandibular joint; TMDs, Temporomandibular joint disorders; ML, Machine learning; MRI, Magnetic Resonance Imaging; CT, Computed Tomography; AI, artificial intelligence; LPM, lateral pterygoid muscle; ADwR, Anterior disc displacement with reduction; ADwoR, Anterior disc displacement without reduction; GM, gray matter; WM, white matter; ROIs, regions of interest; SIRs, Signal intensity ratios; ROC-AUC, Receiver Operating Characteristic-Area Under the Curve; CIs, confidence intervals; FOV, field of view; T2-W, T2-weighted; ST/IP, slice thickness/inter-slice gap; TR, repetition time; TE, echo time; ETL, echo train length; PD-W, proton density-weighted; SHAP, SHapley Additive exPlanations; ANOVA, one-way analysis of variance; HSD, Honestly Significant Difference; SGD, Stochastic Gradient Descent; IADR, International Association for Dental Research; MARS, Multivariate Adaptive Regression Spline; DC/TMD, Diagnostic Criteria for Temporomandibular Disorders; NEX, number of excitations; KNN, k-nearest neighbors.

AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

SE—Conceptualization, Data Curation, Methodology, Writing–Original Draft, Software, Formal analysis, Visualization, Resources, Supervision. HÖ—Methodology, Formal Analysis, Visualization, Supervision. İD—Investigation, Supervision, Project administration. AG—Investigation, Validation, Data curation. AB—Investigation, Supervision, Validation, Resources. MÖ—Conceptualization, Supervision, Resources. All authors contributed to writing–review & editing and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Approval was obtained from the clinical research ethics committee of Selçuk University Faculty of Medicine (file number: 2024/316). Due to the retrospective design of the study and the use of anonymized MRI data, individual written informed consent was not obtained.

ACKNOWLEDGMENT

The authors would like to acknowledge the institutional support provided by Selcuk University Faculty of Medicine during this study.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Manfredini D, Guarda-Nardini L, Winocur E, Piccotti F, Ahlberg J, Lobbezoo F. Research diagnostic criteria for temporomandibular disorders: a systematic review of axis I epidemiologic findings. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 2011; 112: 453–462.

[2] Cong N, Wang N, Huang S, Cheng T, Yan X. Diagnostic significance of magnetic resonance imaging in distinguishing temporomandibular disorders: a retrospective chart review. BMC Oral Health. 2021; 21: 481.

[3] Kundu H, Basavaraj P, Kote S, Singla A, Singh S. Assessment of TMJ disorders using ultrasonography as a diagnostic tool: a review. Journal of Clinical and Diagnostic Research. 2013; 7: 3116–3120.

[4] Sena MF, Mesquita KS, Santos FR, Silva FW, Serrano KV. Prevalence of temporomandibular dysfunction in children and adolescents. Revista Paulista de Pediatria. 2013; 31: 538–545.

[5] Ahmad M, Hollender L, Anderson Q, Kartha K, Ohrbach R, Truelove EL, *et al*. Research diagnostic criteria for temporomandibular disorders (RDC/TMD): development of image analysis criteria and examiner reliability for image analysis. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 2009; 107: 844–860.

[6] Manfredini D, Häggman-Henrikson B, Al Jaghsi A, Baad-Hansen L, Beecroft E, Bijelic T, *et al*. Temporomandibular disorders: INfORM/IADR key points for good clinical practice based on standard of care. CRANIO®. 2025; 43: 1–5.

[7] Petersson A. What you can and cannot see in TMJ imaging—an overview related to the RDC/TMD diagnostic system. Journal of Oral Rehabilitation. 2010; 37: 771–778.

[8] Klatkiewicz T, Gawriołek K, Pobudek Radzikowska M, Czajka-Jakubowska A. Ultrasonography in the diagnosis of temporomandibular disorders: a meta-analysis. Medical Science Monitor. 2018; 24: 812–817.

[9] ElShennawy EM, Hamed WM, Samir SM. Diagnostic accuracy of MRI-CBCT fused images in assessment of clinically diagnosed internal derangement of the temporomandibular joint. Oral Radiology. 2024; 40: 226–233.

[10] Tomas X, Pomes J, Berenguer J, Quinto L, Nicolau C, Mercader JM, *et al*. MR imaging of temporomandibular joint dysfunction: a pictorial review. Radiographics. 2006; 26: 765–781.

[11] Mercuri LG, Greene CS, Manfredini D. The temporomandibular joint disc: a complex fable about an elusive butterfly. CRANIO®. 2026; 44: 228–235.

[12] Guarda Nardini L, Meneghini M, Guido M, Baciorri F, Manfredini D. Histopathology of the temporomandibular joint disc: findings in 30 samples from joints with degenerative disease. Journal of Oral Rehabilitation. 2021; 48: 1025–1034.

[13] Legrell PE, Reibel J, Nylander K, Hörstedt P, Isberg A. Temporomandibular joint condyle changes after surgically induced non-reducing disk displacement in rabbits: a macroscopic and microscopic study. Acta Odontologica Scandinavica. 1999; 57: 290–300.

[14] Jha N, Lee KS, Kim YJ. Diagnosis of temporomandibular disorders using artificial intelligence technologies: a systematic review and meta-analysis. PLOS ONE. 2022; 17: e0272715.

[15] Duyan Yüksel H, Orhan K, Evlice B, Kaya Ö. Evaluation of temporomandibular joint disc displacement with MRI-based radiomics analysis. Dentomaxillofacial Radiology. 2025; 54: 19–27.

[16] Kao ZK, Chiu NT, Wu HH, Chang WC, Wang DH, Kung YY, *et al*. Classifying temporomandibular disorder with artificial intelligent architecture using magnetic resonance imaging. Annals of Biomedical Engineering. 2023; 51: 517–526.

[17] Lasek J, Nurzynska K, Piórkowski A, Strzelecki M, Obuchowicz R. Deep learning for ultrasonographic assessment of temporomandibular joint morphology. Tomography. 2025; 11: 27.

[18] Nagamatsu-Sakaguchi C, Maekawa K, Ono T, Yanagi Y, Minakuchi H, Miyawaki S, *et al.* Test-retest reliability of MRI-based disk position diagnosis of the temporomandibular joint. Clinical Oral Investigations. 2012; 16: 101–108.

[19] Soydan Çabuk D, Etöz M, Akgün İE, Doğan S, Öztürk E, Coşgunarslan A. The evaluation of lateral pterygoid signal intensity changes related to temporomandibular joint anterior disc displacement. Oral Radiology. 2021; 37: 74–79.

[20] Manfredini D. Etiopathogenesis of disk displacement of the temporomandibular joint: a review of the mechanisms. Indian Journal of Dental Research. 2009; 20: 212–221.

[21] Bueno CH, Pereira DD, Pattussi MP, Grossi PK, Grossi ML. Gender differences in temporomandibular disorders in adult populational studies: a systematic review and meta-analysis. Journal of Oral Rehabilitation. 2018; 45: 720–729.

[22] Minervini G, Franco R, Marrapodi MM, Fiorillo L, Cervino G, Cicciù M. Prevalence of temporomandibular disorders in children and adolescents evaluated with Diagnostic Criteria for Temporomandibular Disorders: a systematic review with meta-analysis. Journal of Oral Rehabilitation. 2023; 50: 522–530.

[23] Olate S, Ravelo V, Alister JP, Netto HD, Haidar ZS, Sacco R. Early treatment of unilateral condylar hyperplasia in adolescents: preliminary results. Journal of Clinical Medicine. 2023; 12: 3408.

[24] Tallents RH, Katzberg RW, Murphy W, Proskin H. Magnetic resonance imaging findings in asymptomatic volunteers and symptomatic patients with temporomandibular disorders. Journal of Prosthetic Dentistry. 1996; 75: 529–533.

[25] Kurita H, Ohtsuka A, Kobayashi H, Kurashina K. Alteration of the horizontal mandibular condyle size associated with temporomandibular joint internal derangement in adult females. Dentomaxillofacial Radiology. 2002; 31: 373–378.

[26] Katsavrias EG, Halazonetis DJ. Condyle and fossa shape in Class II and Class III skeletal patterns: a morphometric tomographic study. American Journal of Orthodontics and Dentofacial Orthopedics. 2005; 128: 337–346.

[27] Alexiou K, Stamatakis H, Tsiklakis K. Evaluation of the severity of temporomandibular joint osteoarthritic changes related to age using cone beam computed tomography. Dentomaxillofacial Radiology. 2009; 38: 141–147.

[28] Orhan K, Driesen L, Shujaat S, Jacobs R, Chai X. Development and validation of a magnetic resonance imaging-based machine learning model for TMJ pathologies. BioMed Research International. 2021; 2021: 6656773.

[29] Taborri J, Molinaro L, Russo L, Palmerini V, Larion A, Rossi S. Comparison of machine learning algorithms fed with mobility-related and baropodometric measurements to identify temporomandibular disorders. Sensors. 2024; 24: 3646.

[30] 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: e39912.

[31] Zhang Y, Zhu T, Zheng Y, Xiong Y, Liu W, Zeng W, *et al.* Machine learning-based medical imaging diagnosis in patients with temporomandibular disorders: a diagnostic test accuracy systematic review and meta-analysis. Clinical Oral Investigations. 2024; 28: 186.

[32] Farook TH, Dudley J. Automation and deep (machine) learning in temporomandibular joint disorder radiomics: a systematic review. Journal of Oral Rehabilitation. 2023; 50: 501–521.

How to cite this article: Seyit Erol, Halil Özer, İsmail Dilek, Abdi Gürhan, Ahmet Baytok, Mehmet Öztürk. Predictive performance of supervised machine learning models for temporomandibular joint disc displacement in adolescents diagnosed by magnetic resonance imaging. Journal of Clinical Pediatric Dentistry. 2026; 50(5): 105-117. doi: 10.22514/jocpd.2026.120.