Journal of Dental Sciences 19 (2024) 186–195

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

A family-centered orthodontic screening approach using a machine learning-based mobile application

Banu Kılıç [a*], Ahmed Hassan İbrahim [b], Selahattin Aksoy [b], Mehmet Cihan Sakman [b,c], Gül Sude Demircan [d], Tuğba Önal-Süzek [b]

[a] *Bezmialem Vakif University, Istanbul, Turkey*
[b] *Muğla Sıtkı Koçman University, Muğla, Turkey*
[c] *Zurich University of Applied Sciences, Zurich, Switzerland*
[d] *Denmark Technical University, Copenhagen Denmark, Denmark*

Received 9 April 2023; Final revision received 1 May 2023
Available online 17 May 2023

KEYWORDS
Artificial intelligence;
Early diagnosis;
Telemedicine;
Machine learning;
Malocclusion;
Angle Class III

Abstract *Background/purpose:* Skeletal orthodontic deformities can have functional and aesthetic consequences, making early detection critical. This study aimed to address the issue of parents bringing their children for routine orthodontic checkups after the ideal treatment age has passed. To address this, we developed a mobile application that uses machine-learning to make a preliminary diagnosis of skeletal malocclusion using just one photograph.
Materials and methods: A retrospective study was conducted on 524 pre-pubertal children, aged between 5 and 12 years, to evaluate the accuracy of the machine learning based mobile application. The application detects multiple points in photographs taken from the mobile camera and generates a signal indicating the diagnosis of skeletal malocclusion.
Results: The final accuracy of the Class III vs not Class III model deployed to the mobile application was above 81%, indicating its ability to accurately identify skeletal malocclusion. On a separate validation dataset of 145 patients diagnosed by 5 different clinicians, the accuracy of Class II vs Class I model was 69%; And pg 4, ln 61: as Class II vs Class I with 69% accuracy.
Conclusion: The application provides parents with important information about the orthodontic problem, age of treatment, and various treatment options. This enables parents to seek further advice from an orthodontist at an earlier stage and make informed decisions. However,

* Corresponding author. Department of Orthodontics, Dentistry Faculty, Bezmialem Vakif University, 113, Adnan Menderes Vatan Street, Istanbul, 34093, Turkey.
E-mail addresses: bkilic@bezmialem.edu.tr (B. Kılıç), tugbasuzek@mu.edu.tr (T. Önal-Süzek).

https://doi.org/10.1016/j.jds.2023.05.001

the diagnosis should still be confirmed by an orthodontist. This approach has the potential to improve access to orthodontic care, especially in underserved communities.

Introduction

An increase at parent's awareness of their children's growth pattern at a young age has a significant impact on preventing the malocclusion in its early stages. On the other hand, delaying care until a later age may make malocclusion treatment more challenging and, in some instances, necessitate surgical intervention.[1] Although there is no clear consensus on the need for early treatment,[2] the American Association of Orthodontists recommends that all children be screened by an orthodontist at age 7.[3] According to existing literature, the most favorable time to address skeletal issues in Class III malocclusions is during the deciduous dentition period or between 6 and 8 years of age.[4—6] Adolescence is a critical period for Class III malocclusion, as the condition can worsen during this growth phase. Delayed treatment can lead to greater orthodontic effects and reduced orthopedic effects, which can have negative consequences for the patient's esthetics and self-esteem.[7] However, due to limitations in financial resources and the current capacity of the orthodontic workforce, it is impractical to screen all children worldwide who are over the age of 7 for orthodontic assessment. To address this issue, an artificial intelligence (AI)-based pre-diagnostic tool is needed, which requires a dataset consisting of prepubertal-aged participants.

AI is the branch of computer science that deals with the development of intelligent systems that can perform tasks that would normally require human intelligence, such as learning, problem-solving, and decision-making. Although AI has promising capabilities to transform various fields, including orthodontics,[8—20] current machine learning-based orthodontic applications primarily rely on cephalometric or intraoral photographic images that require specialized equipment.[21] This limits their accessibility, making it difficult for families to obtain remote pre-diagnosis screenings via mobile devices.

While the term AI encompasses a wide range of technologies, from a methodological perspective, there are two main categories of AI: symbolic AI and machine learning.[22] Machine learning enables computers to learn and improve their performance without being explicitly programmed. It involves the use of algorithms and statistical models to analyze data, identify patterns and make predictions or decisions.

The aim of this study was to provide easy access to early screening for skeletal malocclusion in remote areas, by utilizing profile photos through a machine learning-based mobile application. The application automatically detects multiple points in the photograph taken from the mobile camera, generates a signal that expresses the diagnosis of skeletal malocclusion according to predetermined criteria, and shares information about the correct timing of treatment.

Materials and methods

Ethical approval

This study was a retrospective study that was approved by the Non-interventional Research Ethics Committee of the Technology Transfer Office at Bezmialem Vakif University (approval number: E—54022451-050.05.04—87993) and was conducted in compliance with the principles of the Declaration of Helsinki.

Participants

The study included 524 pre-pubertal children aged between 5 and 12 years, with a mean age of 8.3, who were diagnosed as Class I, Class II Division 1, Class II Division 2, and Class III by the same orthodontist through clinical examination, extraoral evaluation of both parents and siblings, intraoral examination, Temporomandibular Joint (TMJ) examination, and evaluation of cephalometric films with Dolphin Imaging 11.8 (Imaging and Management Solutions, Chatsworth, CA, USA).

A separate dataset of 43 Class III, 76 Class II, and 26 Class I patients diagnosed by a committee of 5 orthodontists during the weekly committee meetings at the same institute is collected for independent validation.

Inclusion and exclusion criteria

Patients were included in the study if they met the criteria of being under 13 years of age and applied to the clinic for diagnosis and treatment, while patients were excluded if they had difficulty following instructions, congenital craniofacial deformity, facial swelling due to inflammation, or had undergone previous orthodontic treatment. A project flow chart of this study is presented (Fig. 1).

Sample and data collection

In this study, we paid particular attention to ensure that our sample was representative of the population, in terms of the distribution of gender, and we made sure to have a balanced number of male and female patients (Table 1). The data was collected through patient's non-professional cameras in the orthodontics clinic (Fig. 2). For all patient photos in the dataset, the profile photos were taken facing right. There was no other restriction in patients' own mobile camera type, focal length, angle, background, or

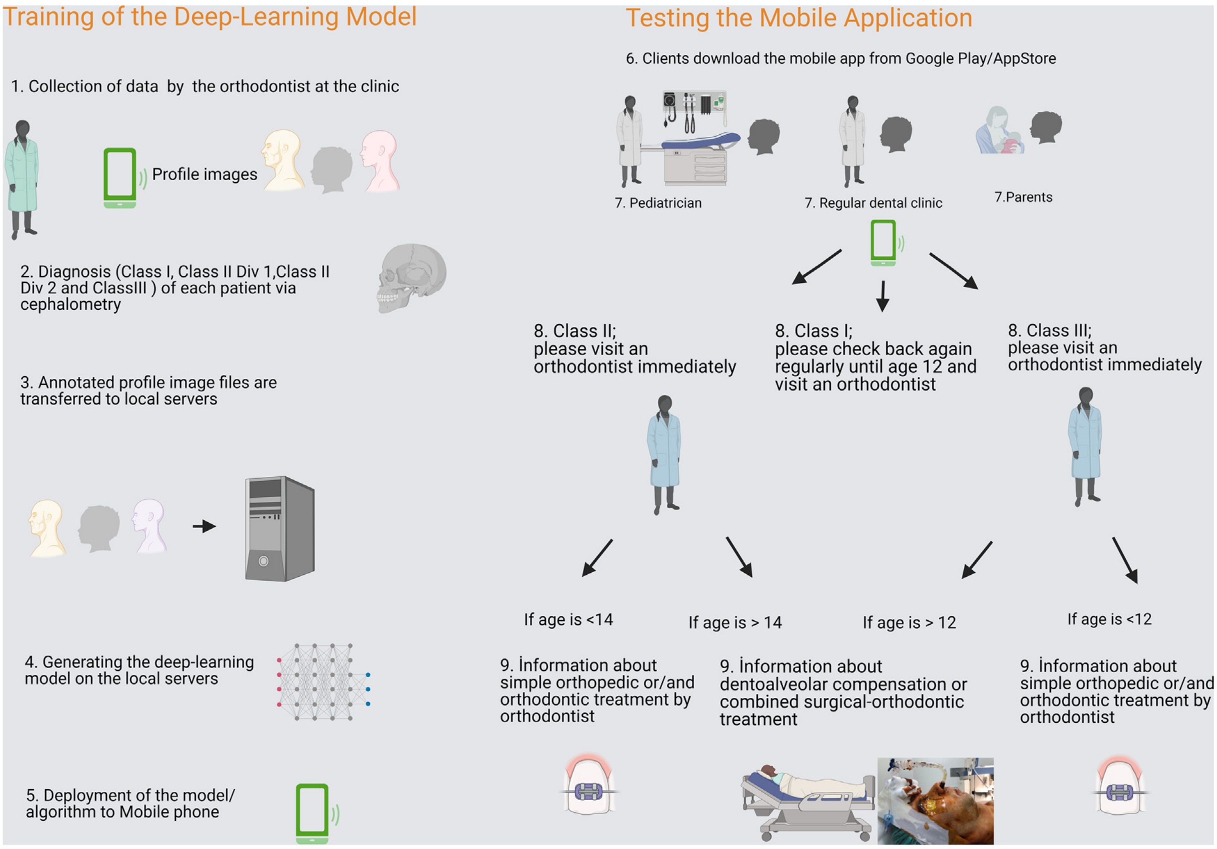

Figure 1 Project flow chart (created with Biorender.com).

Table 1 Age and sex distribution of patients by Angle classification.

		Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Total
Female	Class I	15	23	15	23	21	13	18	6	134
	Class II Division 1	4	8	16	12	11	8	8	6	73
	Class II Division 2		3	3	1	5	5	4	2	23
	Class III	4	6	5	8	4	5	7	4	43
Male	Class I	14	19	18	22	17	8	8	6	112
	Class II Division 1		3	9	5	13	13	7	3	53
	Class II Division 2	3	2	4	5	5	6	3	3	31
	Class III	5	6	8	12	10	3	9	2	55
	Total	45	70	78	88	86	61	64	32	524

lighting as the image feature selection algorithm is based on normalized ratios of facial areas or angles which are all selected due to their independence of the patient's distance or environmental settings. An additional software module is implemented asking the parent to adjust the head position if the patient has over-rotated to the left. The number of features in the dataset was augmented by adding 10 new "normalized" features, dividing each feature's value by the 10 reference distances that the orthodontist provided. To assess whether the Gonial angle feature can help distinguish Class II Division 2 from non-Class II Division 2 cases, we employed an automated method to place points on the photos and calculate the intersection of lines (Fig. 3).

Machine learning implementation

The study trained two machine learning models to distinguish Class III vs not Class III and Class II vs Class I. The data set was divided into 80% training and 20% testing, and model selection evaluation, parameter optimization, and other operations were performed on 80% of the training data. The best-performing model for diagnosis was selected, and hyper-parameters of that model were optimized (Fig. 4). To classify different classes of malocclusion, the study utilized the PyCaret library, which automates machine-learning workflows. The library trained 14 different classification models with 5-fold cross-validation and ranked these models according to model accuracy.

Figure 2 Collection of data in the clinic.

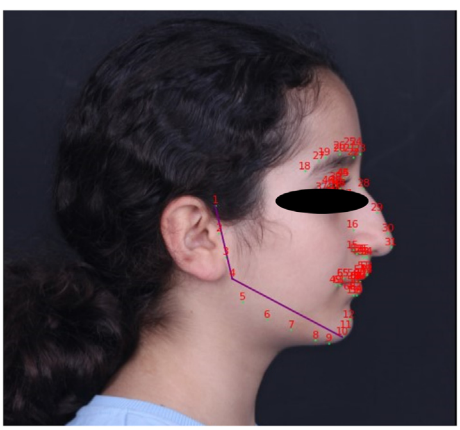

Figure 3 Artificial Intelligence-based detection of the Gonial angle on a photograph.

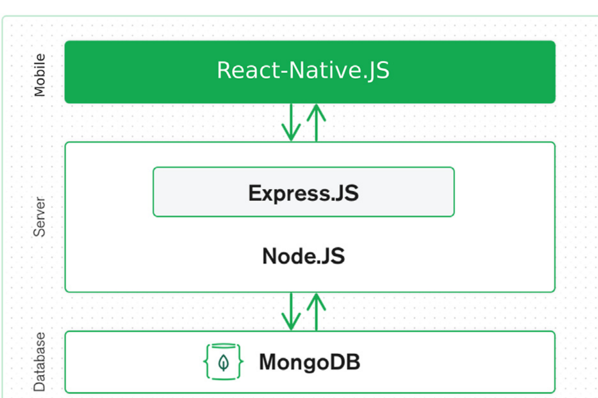

Figure 4 Summary of the MERN (MongoDB, Express, React, Node) stack technology used in the project.

The best model was selected and hyperparameter tuning was applied to improve the model accuracy using PyCaret's tune model function. After both models were deployed to production, they are tested on an independent validation dataset of 145 patients.

The mobile application was developed using the React-Native (open-source iOS application framework), MongoDB (open-source database) for data storage, and Node. JS Express (open-source web application) to build the REST (REpresentational State Transfer) APIs (API: Application Programming Interface) (Fig. 5).

The contact physical addresses, emails, and phone numbers of orthodontists registered to Turkish orthodontic society were scraped from the publicly available national orthodontic society website. Physical street and city addresses were converted to geographical locations using the Google Geocoder (service that provides geocoding and reverse geocoding of addresses) to be displayed in Google Map inside the mobile application, enabling easy-reach of parents to an orthodontist near-by. The application, Orthodontist, is freely available from https://apps.apple.com/tr/app/orthodontist/id6444021818.

Results

Classification of class iii vs not class iii

The accuracies of all machine learning models are listed for Class III vs not Class III along with Boosting and Bagging options (Table 2). The Logistic Regression model, which produced the best results in cross-validation, achieved an accuracy of 80% on the test results (Fig. 6). The highest accuracy of 84% was achieved by applying Bagging to Logistic Regression. The final accuracy of the model deployed to the mobile application is above 81%.

Classification of class i & class ii division 1 vs class ii division 2

We also attempted to classify Class I vs Class II and Class I & Class II Division 1 vs Class II Division 2 using the Gonial angle. The mean accuracy of Class I vs Class II classification using the Gonial angle was 63% (Table 3), while without the Gonial Ridge Classifier, accuracy was 65% (Table 4). Classifying Class I & Class II Division1 vs Class II Division 2 using the Gonial angle had 66% accuracy (Table 5), while not using the Gonial angle had 67% accuracy (Table 6). Due to the low accuracy, the model for Class I vs Class II without the Gonial angle was chosen for the mobile application. We utilized Logistic Regression as a model for diagnosing Class I versus Class II and found it to be the best-performing model for diagnosing Class III with optimized hyperparameters.

Validation on an independent dataset

Our first machine learning model has classified the independent dataset as either ClassIII vs not Class III with 76.4% accuracy. Our second machine learning model has classified

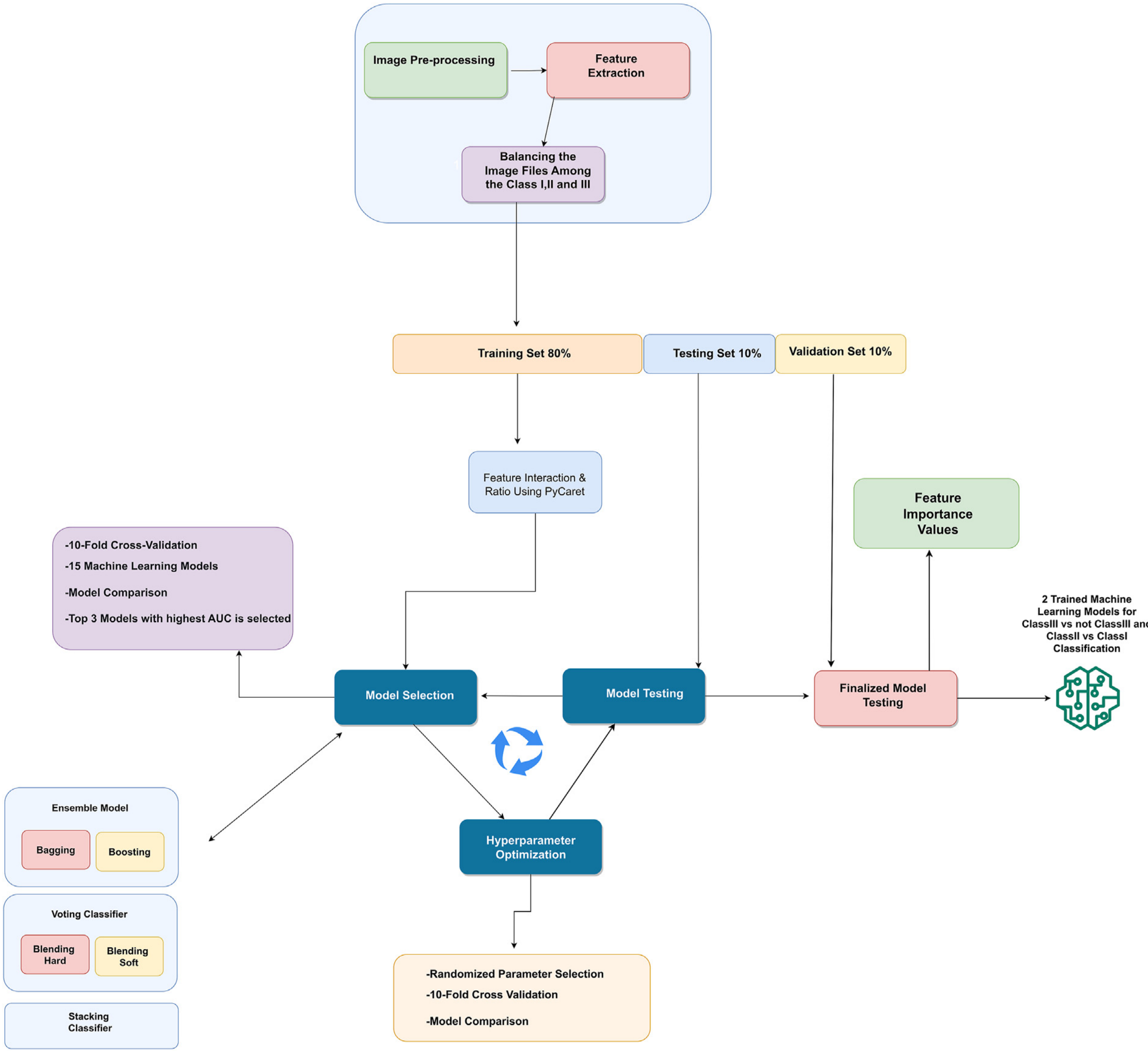

Figure 5 Pipeline of the machine learning algorithm.

the independent dataset as Class II vs Class I with 79.4% accuracy.

Discussion

It's difficult to estimate the ratio of parents willing to use internet platforms such as Instagram (Meta Platforms, Menlo Park, CA, USA) or mobile applications to make the decision to take their children to an orthodontics clinic. The ratio would likely vary on the location and demographics of the population surveyed. While the exact number of parents who would rely on online platforms for their children's orthodontic treatment is difficult to determine, there is a growing trend towards using technology in healthcare. Many parents seek information about their child's health online, and studies have shown that they have a positive attitude towards using AI for their children.[23,24] The current perspectives and future directions of AI in orthodontics and dentistry has been well discussed before.[16,25,26] Although the application of artificial intelligence and machine learning in orthodontics has mainly focused on cephalometric analysis and diagnosis, there have been recent studies that have explored the potential of these techniques in more specific areas. For example, Kamrani and colleagues have proposed a machine learning-based approach to classify Class III malocclusion. While their method has demonstrated promising results, it still requires the use of lateral cephalometric records. In contrast, our research team is striving to develop a pre-diagnostic tool that utilizes profile photographs alone, which would not only reduce costs and logistical barriers but also enable more widespread use among parents.[27] Rao et al. utilized a machine learning approach like ours to identify and analyze facial landmarks. However, their study solely utilized frontal facial photographic images and did not focus on any orthodontic diagnosis. Unlike our study, they employed the "You Only Look Once" (YOLO) deep learning methodology to detect the human face, followed using the "Active Shape Model" to recognize the facial

Table 2 All model accuracies for Class III vs not Class III.

Model	Method	Accuracy	AUC	Recall	Precision	F1-score	Kappa	MCC
LRC	-	0.8023	0.8108	0.7556	0.85	0.8	0.6059	0.61
LRC	Bagging	0.814	0.8103	0.7556	0.8718	0.8095	0.6295	0.6357
LRC	Boosting	0.7209	0.8022	0.7111	0.7442	0.7273	0.4419	0.4423
RFC	-	0.8023	0.8528	0.7333	0.8684	0.7952	0.6068	0.6149
RFC	Bagging	0.7907	0.8388	0.7333	0.8462	0.7857	0.5832	0.5889
RFC	Boosting	0.8023	0.8477	0.7333	0.8684	0.7952	0.6068	0.6149
ETC	-	0.8023	0.8152	0.7556	0.85	0.8	0.6059	0.61
ETC	Bagging	0.8023	0.8179	0.7556	0.85	0.8	0.6059	0.61
ETC	Boosting	0.7907	0.819	0.7556	0.8293	0.7907	0.5823	0.5848
GBC	-	0.7907	0.8309	0.7556	0.8293	0.7907	0.5823	0.5848
GBC	Bagging	0.7326	0.8352	0.6	0.8438	0.7013	0.4714	0.494
GBC	Boosting	0.8023	0.8369	0.7556	0.85	0.8	0.6059	0.61
ABC	-	0.8023	0.8065	0.7556	0.85	0.8	0.6059	0.61
ABC	Bagging	0.7791	0.8228	0.7333	0.825	0.7765	0.5596	0.5634
ABC	Boosting	0.7093	0.7442	0.6222	0.7778	0.6914	0.423	0.4324
Stacking Classifier	-	0.7907	0.8179	0.7556	0.8293	0.7907	0.5823	0.5848
Voting classifier	Blending soft	0.8023	0.8206	0.7556	0.85	0.8	0.6059	0.61
Voting Classifier	Blending hard	0.814	0.8168	0.7556	0.8718	0.8095	0.6295	0.6357

LRC: Linear Regression Classifier.
RFC: Random Forest Classifier.
ETC: Extra Trees Classifier.
GBC: Gradient Boosting Classifier.
ABC: AdaBoostClassifier is an ensemble boosting classifier combining multiple classifiers to increase the accuracy of classifiers iteratively.
Stacking Classifier: A technique that combines multiple sub-models to improve classification performance.
Voting Classifier: Each individual model is trained on the same dataset, but with different algorithms or hyperparameters.
Cross-validation: A resampling method that uses different portions of the data to test and train a model on different iterations.
AUC: The area under the receiver operating characteristic curve (ROC) is also called Area Under Curve illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
Recall: A performance metric, calculated as the number of true positive (tp) results divided by the number of all positive objects (tp + fp).
Precision: A performance metric, calculated as the number of true positive (tp) results divided by the number of all predicted positive results (tp + fp).
F1-score:Harmonic Mean of the Precision and Recall.
Kappa: The Kappa statistic (or value) is a metric that compares an Observed Accuracy with an Expected Accuracy (random chance).
MCC: Matthew's correlation coefficient is a statistical tool to gauge or measure the difference between the predicted values and actual values and is equivalent to chi-square statistics for a 2 x 2 contingency table.
Boosting: Boosting is an ensemble meta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms that convert weak learners to strong ones.
Bagging: A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregates their predictions (either by voting or by averaging) to form a final prediction.
Hard Voting(Blending Hard): Hard Voting picks up the prediction with the highest number of votes as the final prediction.
Soft Voting(Blending Soft): Soft Voting combines the probability of each class in each model and picks the class with the highest probability as the final prediction.

landmarks.[13] Focusing on camouflage treatments, Jung et al. proposed a machine learning methodology for diagnosing teeth extractions. Their study utilized the V-ceph program and neural network models to diagnose extractions, but with a smaller sample size compared to our study. Additionally, their study also required cephalometric facial images of patients.[28] Our study stands out as the first in the literature to predict the skeletal class of the patients using an explainable machine learning algorithm assessing the mobile app-generated profile pictures of the patients without the requirement of any medical apparatus. This approach was used for patients with skeletal Class I, Class II Division 1, Class II Division 2, and Class III malocclusions. By utilizing a significantly larger pre-pubertal patient population, we aimed to increase computational accuracy by providing a more diverse set of input for the supervised learning algorithm.

Early diagnosis of skeletal malocclusions is crucial in orthodontics. Treating skeletal Class II malocclusions during growth modification is highly preferred due to the wider range of therapeutic options available at this stage.[29,30] Our machine learning-based application aimed to detect Class II patients early and distinguish between Class II Division 1 and Class II Division 2 at a pre-pubertal age. The most distinctive feature of Angle's Class II Division 2 malocclusion is the decrease in the lower facial thirds.[31]

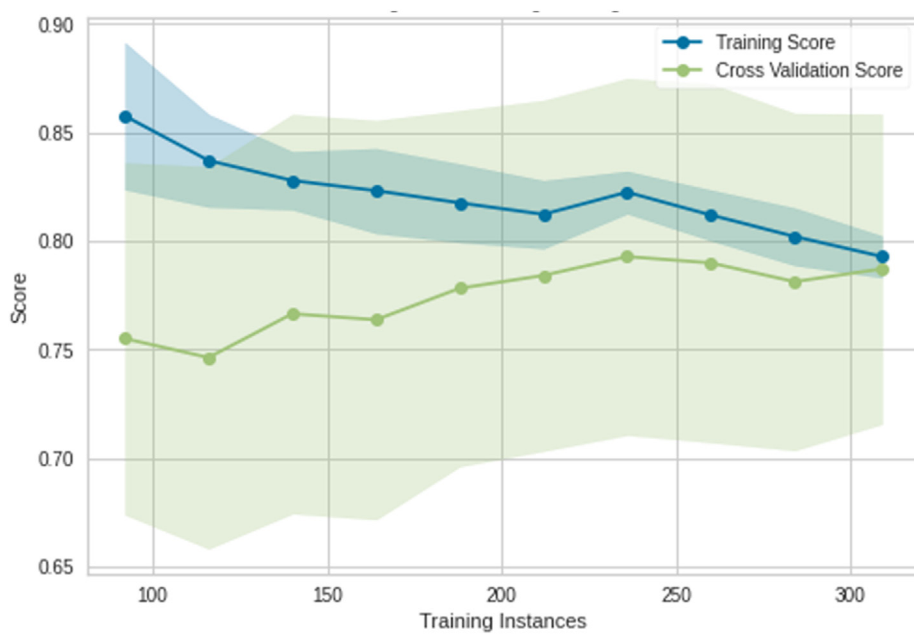

Figure 6 Learning curve for logistic regression.

Table 3 Results obtained from a 5-fold cross-validation experiment of Class I vs Class II using Logistic Regression with Gonial Angle feature.

Fold	Accuracy	Recall	F1-score	Precision	AUC
Fold-1	0.6842	0.5600	0.6087	0.6667	0.7500
Fold-2	0.7544	0.5769	0.6818	0.8333	0.8288
Fold-3	0.5439	0.4615	0.4800	0.5000	0.6129
Fold-4	0.5614	0.6923	0.5902	0.5143	0.6253
Fold-5	0.6429	0.6400	0.6154	0.5926	0.7071
Mean	**0.6373**	**0.5862**	**0.5952**	**0.6214**	**0.7048**
SD	**0.0780**	**0.0781**	**0.0654**	**0.1217**	**0.0802**

Fold: A division of a sample used in cross-validation. The use of multiple folds helps to mitigate the effects of random variations in the data and to obtain a more robust estimate of the model's performance.
Recall: A performance metric, calculated as the number of true positive (tp) results divided by the number of all positive objects (tp + fp).
F1-score: Harmonic Mean of the Precision and Recall.
Precision: A performance metric, calculated as the number of true positive (tp) results divided by the number of all predicted positive results (tp + fp).
AUC: The area under the receiver operating characteristic curve (ROC) is also called Area Under Curve illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
SD: Standard Deviation.

Table 4 5-fold-cv results of Class I vs Class II using RidgeClassifier without Gonial Angle feature.

Fold	Accuracy	Recall	F1-score	Precision	AUC
Fold-1	0.6316	0.3600	0.4615	0.6429	0.6637
Fold-2	0.6491	0.4400	0.5238	0.6471	0.5762
Fold-3	0.6140	0.5417	0.5417	0.5417	0.6566
Fold-4	0.6842	0.6667	0.6400	0.6154	0.6843
Fold-5	0.6964	0.4583	0.5641	0.7333	0.8203
Mean	**0.6551**	**0.4933**	**0.5462**	**0.6361**	**0.6802**
SD	**0.0311**	**0.1041**	**0.0580**	**0.0616**	**0.0791**

Fold: A division of a sample used in cross-validation. The use of multiple folds helps to mitigate the effects of random variations in the data and to obtain a more robust estimate of the model's performance.
Recall: A performance metric, calculated as the number of true positive (tp) results divided by the number of all positive objects (tp + fp).
F1-score: Harmonic Mean of the Precision and Recall.
Precision: A performance metric, calculated as the number of true positive (tp) results divided by the number of all predicted positive results (tp + fp).
AUC: The area under the receiver operating characteristic curve (ROC) is also called Area Under Curve illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
SD: Standard Deviation.

Although we experimented with the gonial angle feature as a skeletal vertical parameter due to its easy detectability from extraoral photographs this feature is not included in the final Class II vs not Class II model due to decrease of accuracy of the model on the test set. Additionally, we incorporated a feature that identified Class II Division 2 malocclusion by detecting a decrease in lower face height relative to the total face height. Although studies have shown that distinct differences in hypodivergent and hyperdivergent face shapes begin to emerge in both sexes around the age of 6,[32] predicting the direction of mandibular growth and the shape of the face at this early age in a pre-pubertal patient was not predictable. As a result, our application had a lower success rate in distinguishing between Class II Division 1 and Class II Division 2. As a future work, we plan to collect a pubertal patient dataset to train the software with the Gonial angle and lower face height

Table 5 Average metrics of five runs with 5-fold-cv of Class I &Class II Division 1 vs Class II Division 2 with Gonial Angle feature.

Run	Model	Accuracy	Recall	F1-score	Precision	AUC
Run-1	GradientBoostingClassifier	0.7	0.71	0.72	0.74	0.71
Run-2	AdaBoostClassifier	0.6	0.59	0.62	0.66	0.54
Run-3	RandomForest	0.64	0.6	0.63	0.69	0.59
Run-4	LGBMClassifier	0.66	0.85	0.74	0.66	0.66
Run-5	KNeighborsClassifier	0.69	0.79	0.74	0.71	0.67
Mean		**0.66**	**0.71**	**0.69**	**0.69**	**0.63**
SD		**0.04**	**0.1**	**0.05**	**0.03**	**0.06**

Run - The process of training on a specific data set, involving adjusting hyperparameters and testing different model architectures.
Recall: A performance metric, calculated as the number of true positive (tp) results divided by the number of all positive objects (tp + fp).
F1-score: Harmonic Mean of the Precision and Recall.
Precision: A performance metric, calculated as the number of true positive (tp) results divided by the number of all predicted positive results (tp + fp).
AUC: The area under the receiver operating characteristic curve (ROC) is also called Area Under Curve illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
SD: Standard Deviation.
AdaBoostClassifier: An ensemble boosting classifier combining multiple classifiers to increase the accuracy of classifiers iteratively.
KNeighborsClassfier: A non-parametric supervised learning method that requires the input consists of the k closest training examples in a data set.
LGBMClassifier: Light Gradient Boosting Machine.
GradientBoostingClassifier: A machine learning technique used in regression and classification tasks in the form of an ensemble of weak prediction models, which are typically decision trees.
RandomForest: An ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

Table 6 Average metrics of five runs with 5-fold-cv of Class I &Class II Division 1 vs Class II Division 2 without Gonial Angle feature.

Run	Model	Accuracy	Recall	F1-score	Precision	AUC
Run-1	KNeighborsClassifier	0.68	0.74	0.72	0.69	0.66
Run-2	AdaBoostClassifier	0.64	0.7	0.68	0.68	0.62
Run-3	KNeighborsClassifier	0.72	0.74	0.74	0.77	0.71
Run-4	KNeighborsClassifier	0.65	0.94	0.76	0.64	0.59
Run-5	LGBMClassifier	0.64	0.76	0.71	0.66	0.67
Mean		**0.67**	**0.78**	**0.72**	**0.69**	**0.65**
SD		**0.03**	**0.08**	**0.03**	**0.04**	**0.04**

Run: The process of training on a specific data set, involving adjusting hyperparameters and testing different model architectures.
Recall: A performance metric, calculated as the number of true positive (tp) results divided by the number of all positive objects (tp + fp).
F1-score: Harmonic Mean of the Precision and Recall.
Precision: A performance metric, calculated as the number of true positive (tp) results divided by the number of all predicted positive results (tp + fp).
AUC: The area under the receiver operating characteristic curve (ROC) is also called Area Under Curve illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
SD: Standard Deviation.
AdaBoostClassifier: An ensemble boosting classifier combining multiple classifiers to increase the accuracy of classifiers iteratively.
KNeighborsClassfier: A non-parametric supervised learning method that requires the input consists of the k closest training examples in a data set.
LGBMClassifier: Light Gradient Boosting Machine.

features for our model to better distinguish Class II subdivisions.

Patients and laypeople are only accurate in identifying their own skeletal class around 48% and 43% of the time, respectively. Even among orthodontists, there is only 60% agreement in determining a profile's skeletal class.[33] Despite the complexity of soft tissue differences at a young age, our application's accuracy rate for distinguishing between Class II Division 1 and Division 2 malocclusions is higher than that of laypeople. Our application recommends orthodontic evaluation before puberty for all patients, providing them with general information about any potential problems detected. By doing so, we aim to increase the number of patients seeking orthodontic care by raising awareness among families who may not have previously considered orthodontic evaluation before puberty. Our goal is to facilitate timely access to orthodontic specialists and improve the overall oral health of a greater number of individuals. When a patient visits an orthodontist with a recommendation from our application, the orthodontist can evaluate additional resources such as intraoral examination, TMJ evaluation, parents' faces, and radiographs if needed.

The results of Takada et al.'s study suggest that orthopedic changes in dentofacial morphology can be achieved with the use of a maxillary protraction headgear in skeletal Class III patients, particularly when the treatment is initiated before the age of 12.[34] Therefore, we set the age limit as 12 years in the inclusion criteria of our study. Growth estimation for Class III patients is challenging but critical due to the possibility of orthognathic surgery.[35] During growth, the mandible tends to move forward and become less retruded, which is generally true for all facial types. This forward growth is stronger in the mandible than the maxilla, resulting in a less convex facial profile. The late detection of Class III malocclusion by families is often due to this differential growth pattern. Class III malocclusion can originate from the maxilla, mandible, or both, and can also be caused by a pseudo-Class III.[36,37] Early diagnosis of pseudo-Class III malocclusion is crucial to prevent it from progressing to a more severe form. The condition can be diagnosed through the De Nevreze maneuver, which involves assessing the ability of the patient to bring their upper and lower incisors into contact without difficulty during closure of the mouth.[38] The mobile application uses only profile photos of the patients for diagnosis, which makes it impossible to accurately diagnose pseudo-Class III. However, referring patients with a preliminary diagnosis of skeletal Class III to an orthodontist through the application may increase the chance of early treatment, even if the case is not skeletal. If the pre-diagnosis is Class III in the application, it gives general information about skeletal Class III problems and emphasizes the importance of early treatment, while explaining the need to see an orthodontist before the age of 7. To compensate for the intrinsic instability of patients taking their own pictures, we computed extra facial features to make our model independent of any distance or position requirements by the patients. The final accuracy of the model for distinguishing between Class III and non-Class III malocclusions in the mobile application was above 81%, which is higher than the accuracy reported in comparable studies in the literature.[18]

Our study has several limitations. The sample size was relatively small for deep-learning models, and further studies with larger sample sizes are needed to enhance our results. Our sample consisted of pre-pubertal children only, and the results are not generalizable to other age groups. Application only uses extra-oral photograph, and thus it may not be generalizable to other types of imaging such as cephalometric radiographs.

In conclusion, our study demonstrates the feasibility and accuracy of using machine learning algorithms in a mobile application to pre-diagnose skeletal malocclusion in pre-pubertal children. Our application has the potential to improve access to early orthodontic treatment, particularly for Class III malocclusion. By leveraging mobile technology and AI, our approach offers a cost-effective and accessible solution to address orthodontic diagnosis and treatment in underserved populations before it is too late. Further research is needed to establish the clinical significance and impact of our approach on public health, but our results suggest that it holds promise as a tool for improving orthodontic care delivery.

Declaration of competing interests

All authors were partially supported by the company Kedi Mobil Uygulama Limited Şirketi, Muğla, Turkey during the study.

Acknowledgments

We revised the paragraph as; The Scientific and Technological Research Council of Turkey (Grant number 2210034) and Small and Medium Industry Development Organization of Turkey R&D and Innovation program funded this work. We thank Tülay Sevinç and Semanur Özer for their assistance in patient photo collection. The authors would like to extend their appreciation to the following faculty members of the Department of Orthodontics at Bezmialem Vakıf University: Prof. Dr. Gökmen Kurt, Assoc. Prof. Dr. Berza Yılmaz Şen, Assoc. Prof. Dr. Elif Dilara Şeker, Asst. Prof. Dr. Ezgi Sunal Aktürk, and Asst. Prof. Dr. Şerife Şahin for their valuable contributions in allowing the use of their archives during the collection of photographs of patients for the purpose of creating an independent test dataset.

References

1. Fleming P. Timing orthodontic treatment: early or late? *Aust Dent J* 2017;62:11—9.
2. Wahl N. Orthodontics in 3 millennia. Chapter 12: two controversies: early treatment and occlusion. *Am J Orthod Dentofacial Orthop* 2006;130:799—804.
3. American Association of Orthodontics. *Clinical practice guidelines for orthodontics and dentofacial orthopedics*. St. Louis, America: American Association of Orthodontics, 2017.
4. Baccetti T, Tollaro I. A retrospective comparison of functional appliance treatment of Class III malocclusions in the deciduous and mixed dentitions. *Eur J Orthod* 1998;20:309—17.
5. Sarangal H, Namdev R, Garg S, Saini N, Singhal P. Treatment modalities for early management of Class III skeletal malocclusion: a case series. *Contemp Clin Dent* 2020;11:91—6.
6. Begum Khan M, Karra A. Early treatment of Class III malocclusion: a boon or a burden? *Int J Clin Pediatr Dent* 2014;7:130—6.
7. Almeida MR, Almeida RR, Oltramari-Navarro PV, Conti AC, Navarro Rde L, Camacho JG. Early treatment of Class III malocclusion: 10-year clinical follow-up. *J Appl Oral Sci* 2011;19:431—9.
8. Leonardi R, Giordano D, Maiorana F. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. *J Biotechnol Biomed* 2009;2009:717102.
9. Mario MC, Abe JM, Ortega NR, Del Santo Jr M. Paraconsistent artificial neural network as auxiliary in cephalometric diagnosis. *Artif Organs* 2010;34:215—21.
10. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging* 2017;4:014501.

11. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: Part 1—comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod* 2019;89:903—9.
12. Hwang HW, Park JH, Moon JH, et al. Automated identification of cephalometric landmarks: Part 2-Might it be better than human? *Angle Orthod* 2020;90:69—76.
13. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop* 2016;149:127—33.
14. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod* 2019;20:41.
15. Makaremi M, Lacaule C, Mohammad-Djafari A. Deep learning and artificial intelligence for the determination of the cervical vertebra maturation degree from lateral radiography. *Entropy* 2019;21:1222.
16. Lu CH, Ko EWC, Liu L. Improving the video imaging prediction of postsurgical facial profiles with an artificial neural network. *J Dent Sci* 2009;4:118—29.
17. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, et al. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making-a systematic review. *J Dent Sci* 2021;16:482—92.
18. Aksoy S, Kılıç B, Süzek T. Comparative analysis of three machine learning models for early prediction of skeletal Class III malocclusion from profile photos. *Mugla J Sci Technol* 2022;8: 22—30.
19. Seo H, Hwang J, Jung YH, Lee E, Nam OH, Shin J. Deep focus approach for accurate bone age estimation from lateral cephalogram. *J Dent Sci* 2023;18:34—43.
20. Takeda S, Mine Y, Yoshimi Y, Ito S, Tanimoto K, Murayama T. Landmark annotation and mandibular lateral deviation analysis of posteroanterior cephalograms using a convolutional neural network. *J Dent Sci* 2021;16:957—63.
21. Thanathornwong B. Bayesian-based decision support system for assessing the needs for orthodontic treatment. *Healthc Inform Res* 2018;24:22—8.
22. Bichu YM, Hansa I, Bichu AY, Premjani P, Flores-Mir C, Vaid NR. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. *Prog Orthod* 2021;22:1—11.
23. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD. Health information seeking by parents in the internet age. *J Paediatr Child Health* 2008;44:419—23.
24. Ramgopal S, Heffernan ME, Bendelow A, et al. Parental perceptions on use of artificial intelligence in pediatric acute care. *Acad Pediatr* 2022;23:140—7.
25. Faber J, Faber C, Faber P. Artificial intelligence in orthodontics. *APOS trends orthod* 2019;9:201—5.
26. Khanagar SB, Al-ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry — a systematic review. *J Dent Sci* 2021;16:508—22.
27. Khosravi-Kamrani P, Qiao X, Zanardi G, Wiesen CA, Slade G, Frazier-Bowers SA. A machine learning approach to determine the prognosis of patients with Class III malocclusion. *Am J Orthod Dentofacial Orthop* 2022;161:1—11.
28. Baik CY, Ververidou M. A new approach of assessing sagittal discrepancies: the Beta angle. *Am J Orthod Dentofacial Orthop* 2004;126:100—5.
29. Pfeiffer J, Grobéty D. The Class II malocclusion: differential diagnosis and clinical application of activators, extraoral traction, and fixed appliances. *Am J Orthod* 1975;68:499—544.
30. Pfeiffer JP, Grobéty D. A philosophy of combined orthopedic-orthodontic treatment. *Am J Orthod* 1982;81:185—201.
31. Dodda KK, Prasad SE, Kanuru RK, Nalluri S, Mittapalli R. Diagnostic features of Angle's Class II div 2 malocclusion. *J Int Soc Prev Community Dent* 2015;5:513.
32. Knigge RP, McNulty KP, Oh H, et al. Geometric morphometric analysis of growth patterns among facial types. *Am J Orthod Dentofacial Orthop* 2021;160:430—41.
33. Tufekci E, Jahangiri A, Lindauer SJ. Perception of profile among laypeople, dental students and orthodontic patients. *Angle Orthod* 2008;78:983—7.
34. Takada K, Petdachai S, Sakuda M. Changes in dentofacial morphology in skeletal Class III children treated by a modified maxillary protraction headgear and a chin cup: a longitudinal cephalometric appraisal. *Eur J Orthod* 1993;15:211—21.
35. Schuster G, Lux CJ, Stellzig-Eisenhauer A. Children with Class III malocclusion: development of multivariate statistical models to predict future need for orthognathic surgery. *Angle Orthod* 2003;73:136—45.
36. Guyer EC, Ellis III EE, McNamara Jr JA, Behrents RG. Components of Class III malocclusion in juveniles and adolescents. *Angle Orthod* 1986;56:7—30.
37. Kurt G, Altug AT, Ustdal A. Comparison off soft tissue profile between subgroups of skeletal Class III malocclusion. *Turk J Orthod* 2010;23:30—40.
38. Lee BD. Correction of crossbite. *Dent Clin N Am* 1978;22: 647—68.