Original Article

Artificial intelligence for predicting orthodontic patient cooperation: Voice records versus frontal photographs

Farhad Salmanpour[1] , Hasan Camci[1]

[1]Department of Orthodontics, Afyonkarahisar Health Science University, Afyonkarahisar, Turkey.

*Corresponding author:
Farhad Salmanpour,
Department of Orthodontics,
Afyonkarahisar Health Science
University, Afyonkarahisar,
Turkey.

ferhad_selmanpur@yahoo.com
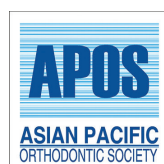
ABSTRACT

Objectives: The purpose of this study was to compare the predictive ability of different convolutional neural network (CNN) models and machine learning algorithms trained with frontal photographs and voice recordings.

Material and Methods: Two hundred and thirty-seven orthodontic patients (147 women, 90 men, mean age 14.94 ± 2.4 years) were included in the study. According to the orthodontic patient cooperation scale, patients were classified into two groups at the 12th month of treatment: Cooperative and non-cooperative. Afterward, frontal photographs and text-to-speech voice records of the participants were collected. CNN models and machine learning algorithms were employed to categorize the data into cooperative and non-cooperative groups. Nine different CNN models were employed to analyze images, while one CNN model and 13 machine learning models were utilized to analyze audio data. The accuracy, precision, recall, and F1-score values of these models were assessed.

Results: Xception (66%) and DenseNet121 (66%) were the two most effective CNN models in evaluating photographs. The model with the lowest success rate was ResNet101V2 (48.0%). The success rates of the other five models were similar. In the assessment of audio data, the most successful models were YAMNet, linear discriminant analysis, K-nearest neighbors, support vector machine, extra tree classifier, and stacking classifier (%58.7). The algorithm with the lowest success rate was the decision tree classifier (41.3%).

Conclusion: Some of the CNN models trained with photographs were successful in predicting cooperation, but voice data were not as useful as photographs in predicting cooperation.

Keywords: Artificial intelligence, Cooperation, Orthodontics

INTRODUCTION

Patient cooperation refers to the harmony between the patient and the doctor throughout the course of the treatment and the patient's compliance with the doctor's orders. In terms of orthodontics, a compatible patient is one who maintains good oral hygiene, continues to keep all of his/her appointments, does not break brackets, uses elastic correctly and properly, and pays attention to his/her diet.[1] If the patient fails to meet any of these requirements, the treatment will be prolonged, the targeted occlusion will not be achieved, or the treatment will be terminated before these goals are met. In other words, the orthodontist's knowledge, skills, and experience, as well as patient compliance and cooperation, are required to achieve an ideal occlusion at the end of treatment.[2-4]

In the present literature, orthodontists employ scales such as the orthodontic patient cooperation scale (OPCS) [4] or clinical cooperation evaluation[3] to estimate patient participation.[5] However,

in recent years, the use of artificial intelligence (AI) in dentistry and medicine has become widespread.[6,7] In this case, the following question comes to mind: Can an alternative method be created for the above-mentioned cooperation scales using AI in patient cooperation estimation?

The term "artificial intelligence (AI)" refers to computing systems that have human-like abilities such as learning and problem-solving. AI can complete some tasks more efficiently than humans because it is less prone to errors. Furthermore, AI works faster than human intelligence, with performance not expected of any human, and it can work accurately and effectively 24 h a day, seven days a week.

Making a choice or decision about any subject can be influenced by a person's emotions or prejudices. By eliminating human prejudices and emotions from the equation, AI, on the other hand, makes evaluations based on numerical data. When assessing a patient's cooperation at the start of treatment, the orthodontist's subjective criteria may result in a more error-prone evaluation. In this situation, AI's cooperation prediction can be more objective and consistent.

Photographs of a person's face can be used to learn about their psychological characteristics. This is called the physiognomy of the face.[8] Some psychologists believe that a person's face can provide more quantitative information about their psychological characteristics than questionnaires and neuropsychological tests.[9] Modern face analysis techniques allow for the assessment of not only psychological traits but also general health conditions. For instance, it is known that there are more than 700 genetic disorders that affect the structure and features of the face, and specialized software has been created to help diagnose these conditions.[10,11]

Voice is a crucial communication tool that enables the speaker to express himself/herself. The human brain analyzes a person's face, voice, and other physical characteristics using unidentified criteria when forming first impressions and making personality judgments.[12,13] These data points could allow an orthodontist to make a preliminary assessment of the patient's cooperation. This evaluation, however, cannot be confirmed due to its subjectivity and lack of reproducibility. According to the findings of numerous previous studies, researchers agree that personality traits can be successfully learned through voice.[14,15]

The primary objective of the study is to evaluate the performance of various AI algorithms using frontal photographs and voice recordings. The secondary purpose is to compare the effectiveness of voice recording data versus photographic data in training AI algorithms.

MATERIAL AND METHODS

The research protocols of this study were approved by the Afyonkarahisar Health Science University Clinical Research Ethics Committee (ID = 2022/30). The study included 237 (147 females and 90 males; mean age 14.94 ± 2.4) systemically and mentally healthy patients undergoing fixed

Figure 1: Orthodontic patient cooperation scale.

orthodontic treatment. The inclusion criteria were no prior orthodontic treatment history and no prior orthodontic treatment history, and undergoing at least 12 months of ongoing fixed orthodontic treatment were requirements for inclusion.

The doctors filled out an OPCS form at the 12th month of treatment to determine the patients' level of cooperation. This scale, created by Slakter *et al.*, has ten criteria that describe patient behaviors that are frequently taken into account when evaluating the level of patient cooperation [Figure 1].[4]

OPCS forms were filled out for 286 patients who had been pre-evaluated and met the study's eligibility criteria. Patients who received a total score of 60% or higher were considered cooperative (34 males and 98 females), while those who received a score of 40% or less were considered non-cooperative (56 males and 49 females). Patients with a score between 41% and 59% (25 females and 31 males) were not included in the study because they were on the borderline.

Routine extraoral photographs were taken with a standardized protocol in our orthodontic clinic. The images were taken with a Canon (EOS 60D, Japan-Tokyo) camera (shutter speed: 1/125, aperture: F22, ISO: 100) from a distance of 1.5 m and under identical lighting conditions. Resting frontal photographs of the patients, which were standardized as described above, were used in the present study.

Patients were asked to read aloud a standard text to obtain voice recordings that were taken in a quiet room using an IPhone (Apple 12 pro max, USA-California). Patients were not allowed to read the text aloud a second time to standardize the records, and only the recording from the first reading was analyzed.

The images were processed using nine different convolutional neural network (CNN) models. The architecture and number of layers of each of these models (VGG16, ResNet50 V2, ResNet101 V2, ResNet152 V2, InceptionResnetV2, Xception, MobileNet V2, NasNetMobile, and DenseNet) varies. CNN continuously analyzes the information in the data set using its mathematical and logical structures to generate a result that is similar to human conclusions.[16] Each layer tries to learn

some qualities by applying different filters to the image sample, and one layer transmits the image it processes to the next layer to learn other characteristics. The final layer divides the samples into two categories based on the retrieved features: Cooperative and non-cooperative.[17]

In the present study, the models were first trained on the ImageNet dataset using the transfer learning technique to get the best performance out of the dataset, and the knowledge they acquired was then applied to the existing dataset's training and classification.

For audio recording analysis, the machine learning algorithms K-nearest neighbors (KNN), logistic regression, random forest classifier, support vector machine (SVM), Decision Tree classifier, Gaussian Naive Bayes, linear discriminant analysis, AdaBoost classifier, Gradient Boosting classifier, XGBoost classifier, Extra tree classifier, voting classifier, Stacking classifier, and YAMNet (Yet Another Mobile Network), which are a CNN model, were used. YAMNet is a deep neural network that has been pre-trained with hundreds of sound spectrograms and is capable of predicting 521 different sound categories. Before processing audio input, YAMNet transforms it into spectrograms. It, then, analyzes the spectrogram images and identifies the sounds based on the information acquired from the analysis.

The models were created in the integrated development and learning environment using the Python programming language. A total of 80% of the sample was used for training, and 20% for testing. The models' success was determined by their classification accuracy. Analysis for precision, recall, and F1 score was also used to assess the findings. The receiver operating characteristic (ROC) curve and confusion matrix of each model of the AI models used for photographs were examined comparatively.

RESULTS

The accuracy, precision, recall, and F1-score values of nine different CNN models trained with photos are shown in [Table 1]. The ROC curves which graphically illustrated

Table 1: Findings of the CNN models for photographs

Convolutional neural network models	Accuracy	Precision	Recall	F1-Scores
Xception	%66.0	0.640	0.582	0.610
DenseNet121	%66.0	0.630	0.586	0.607
ResNet50V2	%62.0	0.636	0.467	0.538
MobilNetV2	%62.0	0.596	0.509	0.549
ResNet152V2	%61.0	0.607	0.548	0.576
NasNetMobil	%57.0	0.537	0.400	0.458
InceptionResNetV2	%53.0	0.519	0.452	0.483
VGG16	%53.0	0.500	0.367	0.423
ResNet101V2	%48.0	0.441	0.473	0.456

VGG16: Visual Geometry Group 16, CNN: Convolutional neural network

each model's efficacy, are also shown in [Figure 2]. Based on overall success, Xception (66.0%) and DenseNet121 (66.0%) were the two most effective CNN models. The success rate for ResNet101V2 was the lowest (48.0%). The success rate of the remaining five models' were similar.

The confusion matrix results of all photo-trained models are shown in [Figure 3]. The results of the confusion matrix showed that the ResNet50V2 (70.4%) model outperformed all others in predicting cooperative patients. The best models for predicting uncooperative patients (57.9%) were Xception and DenseNet121. The success rates of the remaining models for both cooperative and non-cooperative patients varied.

The accuracy rates of the voice-trained models are shown in [Table 2]. Machine learning algorithms with the highest success rate were as follows: Linear discriminant analysis, KNN, SVM, Extra tree classifier, Stacking classifier), and YAMNet. The

success rate of all of these five algorithms was 58.7%. The decision tree had the lowest prediction rate (41.3%).

In [Figure 4], the voice-model confusion matrix results were shown. Linear discriminant and AdaBoost classifier (56.2%) models, performed the best in terms of predicting cooperative patients, according to the results of the confusion matrix. The extra tree classifier and stacking classifier were the most successful models in predicting non-cooperative patients (70.0%). The prediction rates of other models' varied for both cooperative and non-cooperative patients.

DISCUSSION

OPCS or Clinical compliance evaluation (CCE) scales have been used in previous studies to assess patient cooperation. [3,4,18] However, these scales are composed of a series of clinically scored questions that assess some of the patients' behaviors.
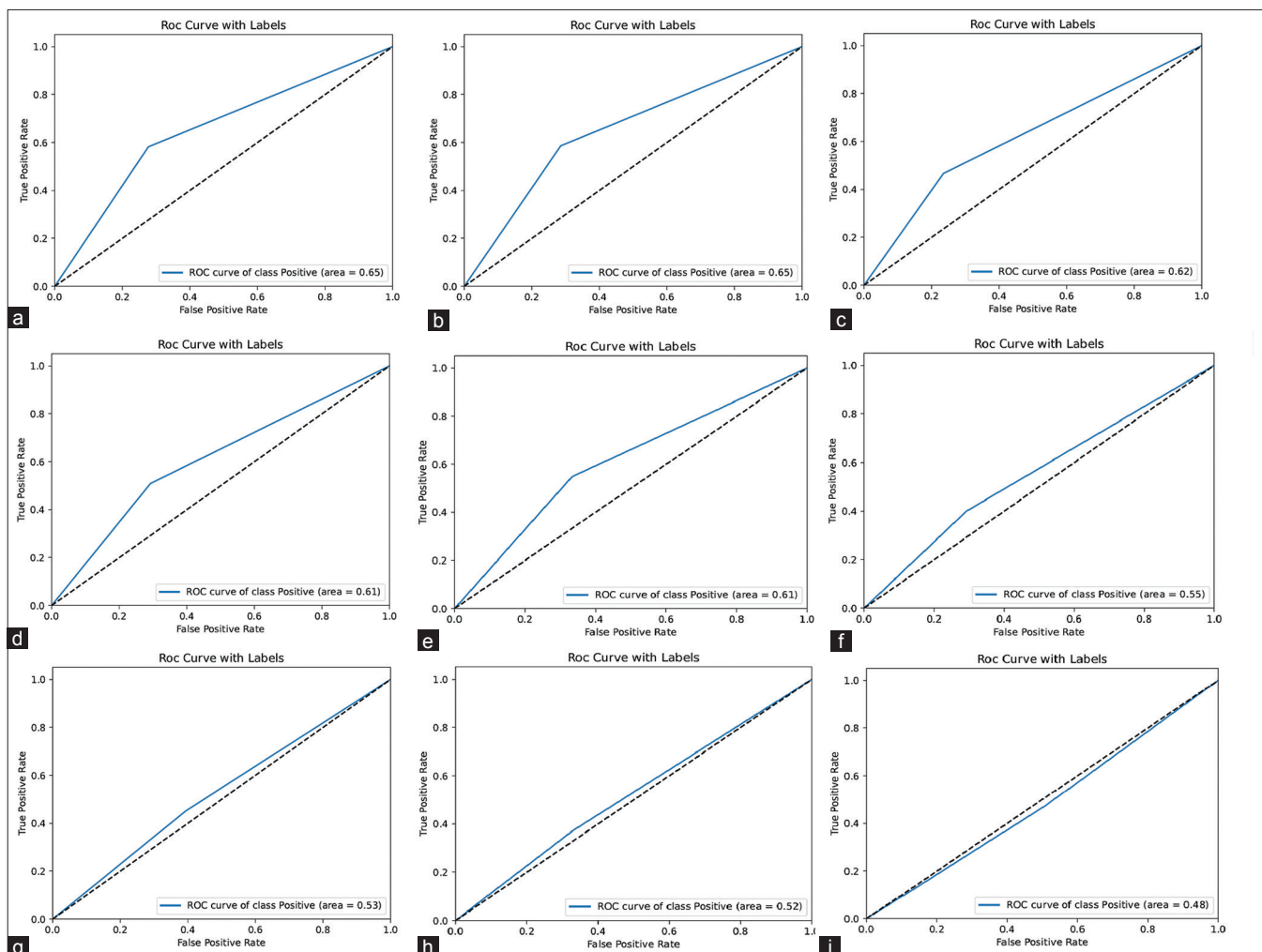

Figure 2: Receiver operating characteristic curves of the models for photographs (blue curves); angular dashed lines (black); (a) Xception, (b) DenseNet121, (c) ResNet50V2, (d) MobilNetV2, (e) ResNet152V2, (f) NasNetMobil, (g) InceptionResNetV2, (h) VGG16, and (i) ResNet101V2. ROC: Receiver operating characteristic; VGG16: Visual geometry group16

Figure 3: Confusion matrix results of models for photographs (a) Xception, (b) DenseNet121, (c) ResNet50V2, (d) MobilNetV2, (e) ResNet152V2, (f) NasNetMobil, (g) InceptionResNetV2, (h) VGG16, and (i) ResNet101V2. ROC: Receiver operating characteristic; VGG16: Visual geometry group16 (VGG16)
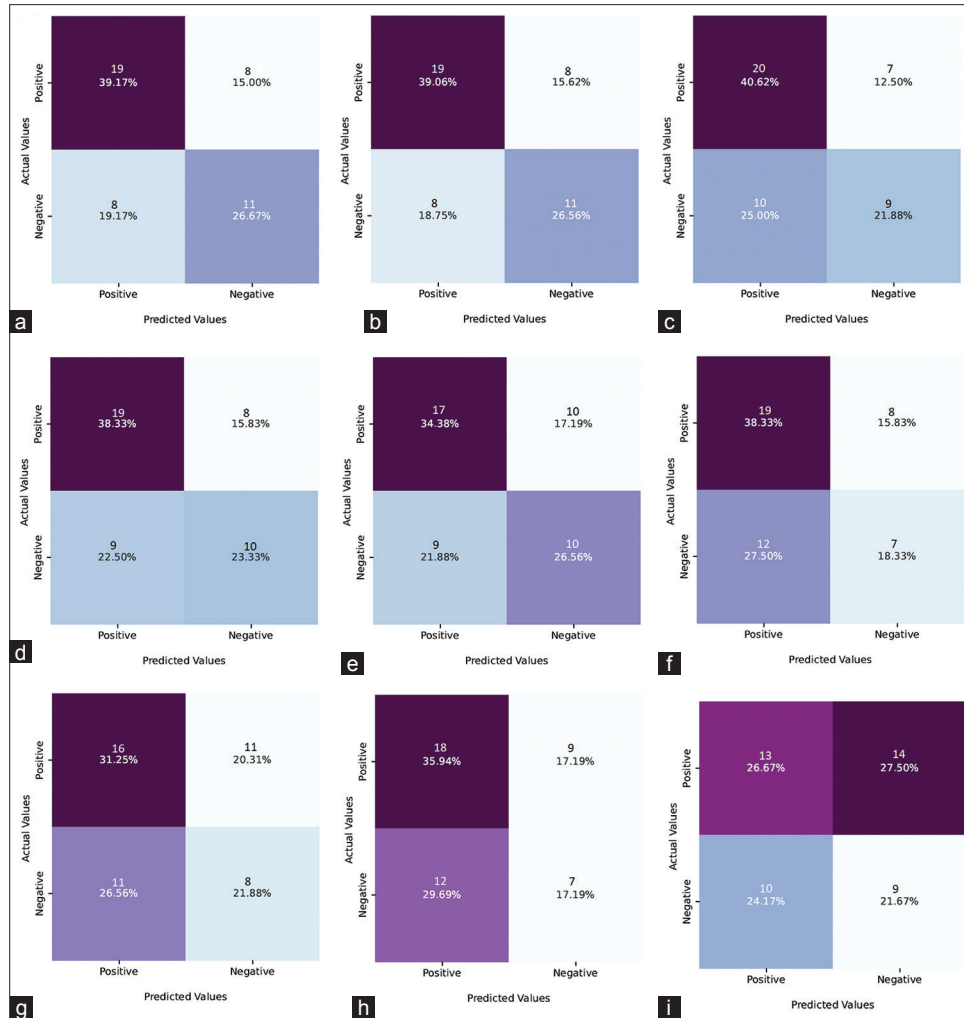
The clinician's qualitative impressions of the patient may influence the scale results.[19] Therefore, the present study was designed with the idea of developing a method that provides orthodontists with more quantitative and reproducible results in learning about patient cooperation.

Many research have shown that facial anatomy and voice carry important personal information.[20-23] Personality analysis methods may now be performed with greater precision and reproducibility because of the advancement of AI technology and its application in the medical industry.[7,24] In the literature, CNN models are commonly utilized in the processing of visual data for personality analysis, and successful results have been achieved. Mukhopadhyay *et al.* employed the CNN model in their study to characterize photographers based on their emotional state using facial images, and they had 76.6% success with

this model.[25] In another identical study, a CNN model with a different architecture was utilized for the same goal, and the CNN's success rate was 80.5%.[26]

One of the most critical factors influencing the outcome of deep learning studies is the architecture of CNN models. The variation in model architecture has a direct impact on model performance, and the success rate can alter. CNNs with three different structures, for example, were utilized in a study that predicts the most prominent personality qualities (extraversion, openness, agreeableness, and meticulousness) using 240 pictures. As a result, varying success rates of 77.2%, 81.5%, and 85.3% were achieved.[27] Nine alternative CNN models were utilized to strengthen the reliability of the present study's results and to determine the most effective CNN model.

The content and quality of the data set, like the model architecture, might influence the success rate. Allen-Zhu *et al.*, for instance,
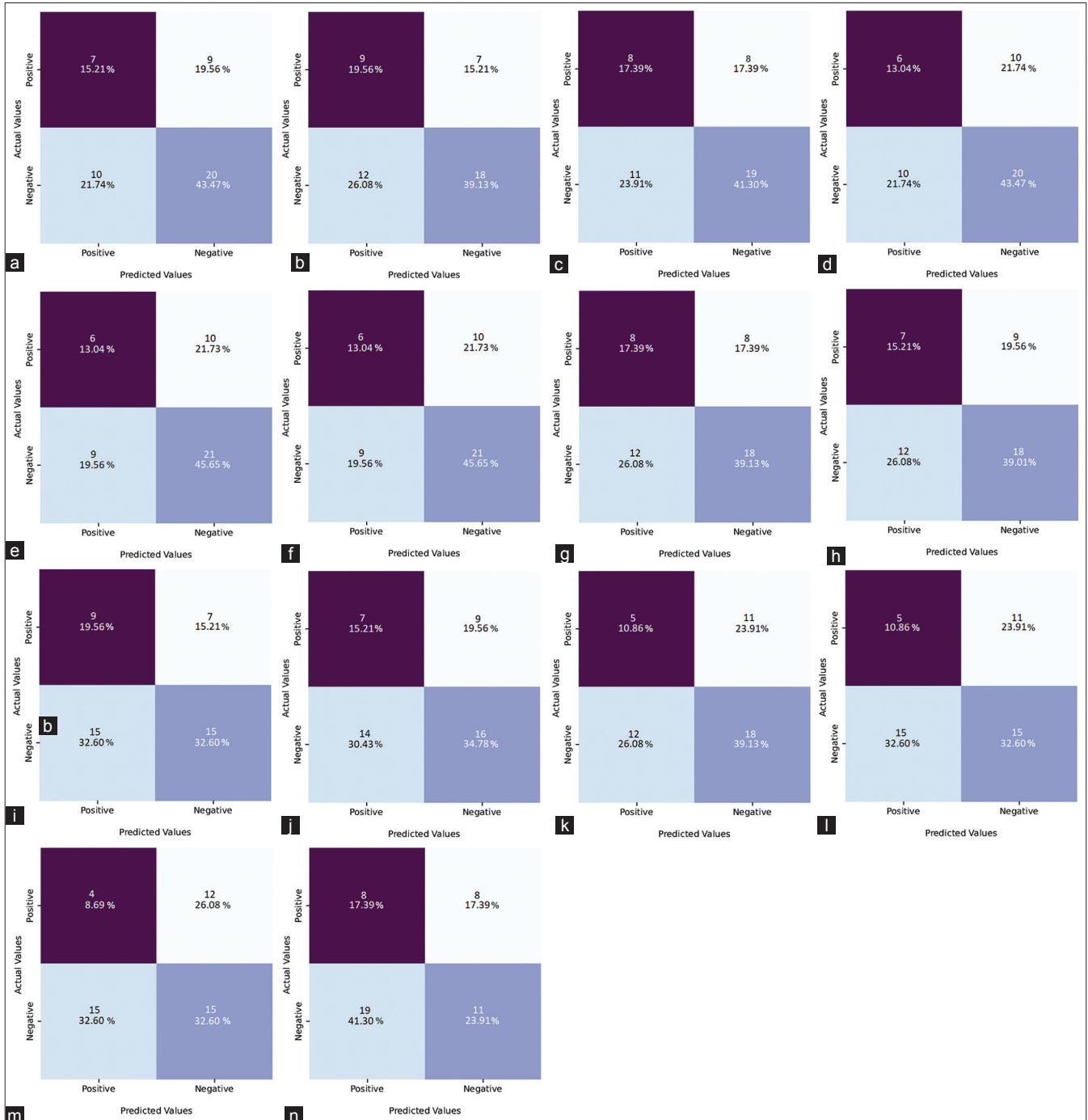
Figure 4: Confusion matrix results of the YAMNet models and machine learning algorithms (a) YAMNet, (b) Linear discriminant analysis, (c) K-nearest neighbors, (d) Support vector machine, (e) Extra tree classifier, (f) Stacking classifier, (g) Extra gradient boosting classifier, (h) Logistic regression, (i) AdaBoost classifier, (j) Random forest classifier, (k) Gaussian NB, (l) Gradient boosting classifier, (m) Voting classifier, and (n) Decision tree classifier. YAMNet: Yet Another Mobile Network, Gaussian NB: Gaussian naive bayes

suggested that when two separate data sets were processed by the same CNN model, different results were achieved.[28] Only rest frontal images and text-to-speech recordings that did not contain any visible emotion were preferred to establish standardization in the present study's data collection.

High success rates are attained while using YAMNet in a variety of sectors, including technology,[29] health,[30] and industry.[31] For example, the accuracy of YAMNet for the COVID-19 diagnosis[32] was >97%, while it was 82% for the Parkinson's disease diagnosis.[30]

Table 2: Findings of the YAMNET model and machine learning algorithms for voice recording

Convolutional neural network model and machine learning algorithms	Accuracy	Precision	Recall	F1-Scores
YamNet	%58.7	0.710	0.640	0.677
Linear Discriminant Analysis	%58.7	0.720	0.581	0.643
K-NN	%58.7	0.704	0.613	0.655
SVM	%58.7	0.677	0.677	0.677
Extra Tree Classifier	%58.7	0.677	0.677	0.677
Stacking Classifier	%58.7	0.677	0.677	0.677
Extra Gradient Boosting Classifier	%56.5	0.692	0.581	0.632
Logistic Regression	%54.2	0.667	0.581	0.621
AdaBoost Classifier	%52.1	0.682	0.484	0.566
Random Forest Classifier	%50.0	0.640	0.516	0.571
Gaussian NB	%50.0	0.621	0.581	0.600
Gradient Boosting Classifier	%43.4	0.577	0.484	0.526
Voting Classifier	%43.4	0.571	0.516	0.542
Decision Tree Classifier	%41.3	0.579	0.355	0.440

YAMNet: Yet Another Mobile Network, K-NN: K-Nearest Neighbors, SVM: Support Vector Machine

We had to generate our own dataset in the present study to estimate patient cooperation because there was no pre-prepared dataset. Nevertheless, it is not always possible to produce an extensive data set in a single center. Creating mega datasets, particularly in the health industry, takes time and money. However, the invention of the transfer learning technique has aided in the abolition of this difficulty. Using pre-trained and highly successful model knowledge eliminates the need for mega data sets in this approach.[33,34]

Maray *et al.* used ImageNet, a mega dataset, to perform transfer learning before employing AlexNet in the diagnosis of Alzheimer's disease. As a result, they reported that AlexNet powered by transfer learning outperformed models trained using other methods.[34] In another study, the classification of a small dataset using the transfer learning technique performed significantly better than the classification of big datasets using a standard CNN model.[35] In the present study, a transfer learning technique was used with the ImageNet data set to improve the efficiency of model training. As a result, the Xception and DenseNet121 models achieved a 66.0% success rate. When the confusion matrix was analyzed, ResNet50V2 performed best in predicting cooperative patients, with a success rate of 70.4%, whereas Xception and DenseNet121 performed best in predicting non-cooperative patients, with a success rate of 57.9%. YAMNet was employed for audio data and had a success rate of 58.7%. In addition to YAMNet, a CNN, 13 commonly used machine learning algorithms were tested in voice data processing. Some of these algorithms, however, produced results similar to YAMNet.[36-41] The present study's relatively low success rate may be attributed to the employment of just standardized models on a specific

issue, such as cooperation prediction. However, even better results may have been achieved by combining the models that were most successful in predicting cooperative and non-cooperative patients.

There are also studies in the literature that use voice recordings to evaluate personality or psychological traits. For instance, in the study by Liu *et al.*, the scale selection pyramid network dataset was used to assess the performances of SVM, KNN, and logistic regression algorithms in defining personality traits.[37] The success rate obtained using these algorithms was reported as 72.0% (SVM), 70.0% (KNN), and 71.0% (logistic regression). Various machine learning algorithms and an CNN model were used in the present study to determine whether patients were cooperative or non-cooperative using voice data, and acceptable results were obtained when compared to the results of the previous studies.

The present study aimed to investigate the performance of AI, particularly deep learning, on patient cooperation prediction and to pave the way for further researches on this topic. The present study's findings are intriguing, but more research using more complex models and data sets acquired from various centers is required.

CONCLUSION

The present study's findings demonstrate that cooperation prediction can be made using AI in orthodontics. Some CNN models trained using patient photographs predicted cooperation with acceptable accuracy. Voice recordings, on the other hand, were not as effective as photographs in predicting cooperation. However, it is clear that both types of data can help the orthodontist predict cooperation to varying degrees.

Availability of data and materials

Data and materials are available at the Orthodontic Department in the Faculty of Dentistry, Afyonkarahisar Health Science University.

Ethical approval

This study was approved by the Clinical Research Ethics Committee. Afyonkarahisar Health Science University (ID:2022/30).

Declaration of patient consent

The authors certify that they have obtained all appropriate patient consent.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Use of artificial intelligence (AI)-assisted technology for manuscript preparation

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

REFERENCES

1. Egolf RJ, BeGole EA, Upshaw HS. Factors associated with orthodontic patient compliance with intraoral elastic and headgear wear. Am J Orthod Dentofac Orthop 1990;97:336-48.
2. Albino JE, Lawrence SD, Lopes CE, Nash LB, Tedesco LA. Cooperation in orthodontic treatment. J Behav Med 1991;14:53-70.
3. Richter DD, Nanda RS, Sinha PK, Smith DW, Currier GF. Effect of behavior modification on patient compliance in orthodontics. Angle Orthod 1998;68:123-32.
4. Slakter MJ, Albino JE, Fox RN, Lewis EA. Reliability and stability of the orthodontic patient cooperation scale. Am J Orthod 1980;78:559-63.
5. Albino JE. Factors influencing adolescent cooperation inorthodontic treatment. Semin Orthod 2000;6:214-23.
6. Allareddy V, Venugopalan SR, Nalliah RP, Caplin JL, Lee MK, Allareddy V. Orthodontics in the era of big data analytics. Orthod Craniofac Res 2019;22(S1):8-13.
7. Mohammad-Rahimi H, Nadimi M, Rohban MH, Shamsoddin E, Lee VY, Motamedian SR. Machine learning and orthodontics, current trends and the future opportunities: A scoping review. Am J Orthod Dentofacial Orthop 2021;160:170-92.e4.
8. Kamenskaya E, Kukharev G. Recognition of psychological characteristics from face. Metod Inf Stosow 2008;1:59-73.
9. Gaussian NB, Kamenskaya E, Kukharev G. Recognition of psychological characteristics from face. Metod Inf Stosow 2008;1:59-73.
10. Boehringer S, Vollmar T, Tasse C, Wurtz RP, Gillessen-Kaesbach G, Horsthemke B, et al. Syndrome identification based on 2D analysis software. Eur J Hum Genet 2006;14:1082-9.
11. Loos HS, Wieczorek D, Würtz RP, von der Malsburg C, Horsthemke B. Computer-based recognition of dysmorphic faces. Eur J Hum Genet 2003;11:555-60.
12. Leongómez JD, Binter J, Kubicová L, Stolařová P, Klapilová K, Havlíček J, et al. Vocal modulation during courtship increases proceptivity even in naive listeners. Evol Hum Behav 2014;35:489-96.
13. Belin P, Fecteau S, Bédard C. Thinking the voice: Neural correlates of voice perception. Trends Cogn Sci 2004;8:129-35.
14. Addington DW. The relationship of selected vocal characteristics to personality perception. Speech Monogr 1968;35:492-503.
15. Cantril H, Allport G. Judging personality from voice. J. Soc. Psychol. 1934;5:37-55.
16. Goodfellow I, Bengio Y, Courville A. Gaussian NB: Gaussian Naive Bayes Deep Learning? Nature 2016;29:1-73.
17. Aghdam H, Heravi EJ. Guide to convolutional neural networks. Guid to convolutional neural networks. New York, NY: Springer; 2017.
18. Sergl HG, Klages U, Pempera J. On the prediction of dentist-evaluated patient compliance in orthodontics. Eur J Orthod 1992;14:463-8.
19. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: A review of classification and combining techniques. Artif Intell Rev 2006;26:159-90.
20. Zhang T, Qin RZ, Dong QL, Gao W, Xu HR, Hu ZY. Physiognomy: Personality traits prediction by learning. Int J Autom Comput 2017;14:386-95.
21. Samizadeh S. Facial physiognomy. In: Non-surgical rejuvenation of Asian faces. Cham: Springer; 2022. p. 33-9.
22. Kröger JL, Lutz OH, Raschke P. Privacy implications of voice and speech analysis - information disclosure by inference. IFIP Adv Inf Commun Technol 2020;576:242-58.
23. Zare M, Flinchbaugh C. Voice, creativity, and big five personality traits: A meta-analysis. Hum Perform 2019;32:30-51.
24. Ilmini K, Fernando TG. Persons' personality traits recognition using machine learning algorithms and ımage processing techniques. Adv Comput Sci Res 2016;5:40-4.
25. Mukhopadhyay M, Dey A, Shaw RN, Ghosh A. Facial emotion recognition based on Textural pattern and Convolutional Neural Network. 2021 IEEE 4th 4th International Conference on computing, power and communication technologies (GUCON); 2021.
26. Gao Y, Gao L, Li X. A generative adversarial network based deep learning method for low-quality defect ımage reconstruction and recognition. IEEE Trans Ind Inform 2021;17:3231-40.
27. Fu JF, Zhang H. Personality trait detection based on ASM localization and deep learning. Sci Program 2021;2021:5675917.
28. Allen-Zhu Z, Li Y, Song Z. On the convergence rate of training

recurrent neural networks. Adv Neural Inf Process Syst 2019;32:1-20.

29. Lee E, Kim D, Min C, Kang S. Automatic recognition of vocal reactions in music listening using smart earbuds. In: SenSys '20: The 18th ACM Conference on Embedded Networked Sensor Systems; 2020.

30. Kumar SA, Sasikala S, Arthiya KB, Sathika J, Karishma V. Parkinson's speech detection using YAMNet; 2023. Available from: https://ieeexplore.ieee.org/abstract/document/10200704 [Last accessed on 2023 Aug 18].

31. Patil S, Wani K. Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network. Mater Today Proc 2023;72:1322-7.

32. Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. Biocybern Biomed Eng 2022;42:1066-80.

33. Brusa E, Delprete C, Di Maggio LG. Deep transfer learning for machine diagnosis: From sound and music recognition to bearing fault detection. Appl Sci 2021;11:11663.

34. Maray N, Ngu AH, Ni J, Debnath M, Wang L. Transfer learning on small datasets for ımproved fall detection. Sensors 2023;23:1105.

35. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Lecture notes computer science (including subseries Lecture notes in artificial ıntelligence and lecture notes bioinformatics). LNCS, vol. 11141. Cham: Springer; 2018. p. 270-9.

36. Jawarkar NP, Holambe RS, Basu TK. Use of fuzzy min-max neural network for speaker identification. In: IEEE-International Conference on Recent Trends in Information Technology, ICRTIT; 2011. p. 178-82.

37. Liu ZT, Rehman A, Wu M, Cao WH, Hao M. Speech personality recognition based on annotation classification using log-likelihood distance and extraction of essential audio features. IEEE Transac Multimed 2020;23:3414-26.

38. Tulics MG, Szaszák G, Mészáros K, Vicsi K. Artificial neural network and svm based voice disorder classification. In: 2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Naples, Italy. Available from: https://ieeexplore.ieee.org/abstract/document/9089908 [Last accessed on 2023 Aug 18].

39. Liu Z, Wang D, Zhang L, Hu B. A novel decision tree for depression recognition in speech; 2020. Available from: http://arxiv.org/abs/2002.12759 [Last accessed on 2023 Aug 18].

40. Hassani R. Predicting Parkinson's disease using different features based on Xgboost of voice data; 2022. Available from: https://ieeexplore.ieee.org/abstract/document/9988089 [Last accessed on 2023 Aug 18].

41. Koduru A, Valiveti HB, Budati AK. Feature extraction algorithms to improve the speech emotion recognition rate. Int J Speech Technol 2020;23:45-55.