AJO-DO

Estimating the size of unerupted teeth: Moyers vs deep learning

Hasan Camcı and Farhad Salmanpour
Afyonkarahisar, Turkey

Introduction: This study aimed to design a deep learning (DL) system for estimating the sum of the mesiodistal widths (MDWs) of unerupted mandibular canines and premolars in the mixed dentition period and to clarify its performance by comparing DL estimates with Moyers' table (MT) results. **Methods:** The training dataset was obtained from 974 patients with permanent dentition. On the 3-dimensional digital models, MDWs of the mandibular right teeth were measured using Ortho Analyzer software (3Shape, Copenhagen, Denmark). A system was designed that could predict the total width of the mandibular canines and premolars using the mandibular central, lateral incisor, and first molar MDWs. This artificial neural system had 5 layers (4 hidden and 1 output) and 886 neurons. The MDWs of the mandibular teeth were introduced to the DL system in the form of datasets. The DL system's predicted results for 100 randomly selected patients were compared with the probability values obtained from the MT. **Results:** The estimation performance of the DL system for the unerupted mandibular canines and premolars was acceptable, with 49.5% accuracy. The success rate for the MT, in comparison, was 45.0%, with an error margin of 1.00 mm. **Conclusions:** The DL system offers a potential alternative to current methods in estimating unerupted tooth size. The results of the DL system are expected to provide diagnostic support for mixed dentition analysis on dental casts. (Am J Orthod Dentofacial Orthop 2022;161:451-6)

The basic requirement for good orthodontic treatment is proper diagnosis.[1] Correct and effective diagnostic analysis directly affects the success of treatment. In the early diagnosis of malocclusions, the existence of unerupted permanent teeth limits the efficacy of model analysis. Accurate estimation of the mesiodistal widths (MDWs) of unerupted canines and premolars (CP_1P_2) in mixed dentition is critical for determining tooth size-arch length discrepancies.[2] In the existing literature, 3 basic methods for estimating the MDWs of unerupted teeth are identified[3,4]: (1) radiographic method: the MDWs of the unerupted teeth are determined directly from x-ray images, taking into account the magnification rate; (2) prediction equations: the estimation is carried out with the help of a regression or correlation analysis; and (3) combined method: both the radiographic method and prediction equations are used together.

The Moyers' table (MT) estimates the sum of the unerupted CP_1P_2 widths using the sum of the mandibular incisors' MDW, with no requirement for radiographic measurement. Because of its simple applicability, this regression scheme, developed by Moyers, is often preferred by orthodontists.[5] However, the reliability of the MT for different ethnic groups remains questionable, as the MT was created using data from a single population, and there are different levels of estimation rate, ranging from 5.0% to 95.0%. Therefore, many researchers have investigated which prediction rate is most applicable to their society.[6-9]

In recent years, artificial intelligence (AI) has become widely used in medicine and dentistry.[10-12] Machine learning is a type of AI that educates itself using specific datasets and draws conclusions through observation and analysis. Throughout the evolutionary development of AI, machine learning has evolved into deep learning (DL).[13] DL systems imitate human learning that can not be formulated or standardized.

Model architecture is an important concept in DL systems, and the proper construction of an artificial neural network (ANN) model is mandatory for the correct and effective working of the system. Artificial nerve cells

From the Department of Orthodontics, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkey.
All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.
Approval was obtained from the Afyonkarahisar Health Sciences University Clinical Research Ethics Committee.
Address correspondence to: Hasan Camcı, Faculty of Dentistry, Department of Orthodontics, Afyonkarahisar Health Sciences University, Güvenevler mah., İsmet İnönü Cad. No:4, Afyonkarahisar 03030, Turkey; e-mail, dt.hasan@hotmail.com.
Submitted, October 2020; revised, February 2021; accepted, March 2021.
0889-5406/$36.00
© 2021 by the American Association of Orthodontists. All rights reserved.
https://doi.org/10.1016/j.ajodo.2021.03.015

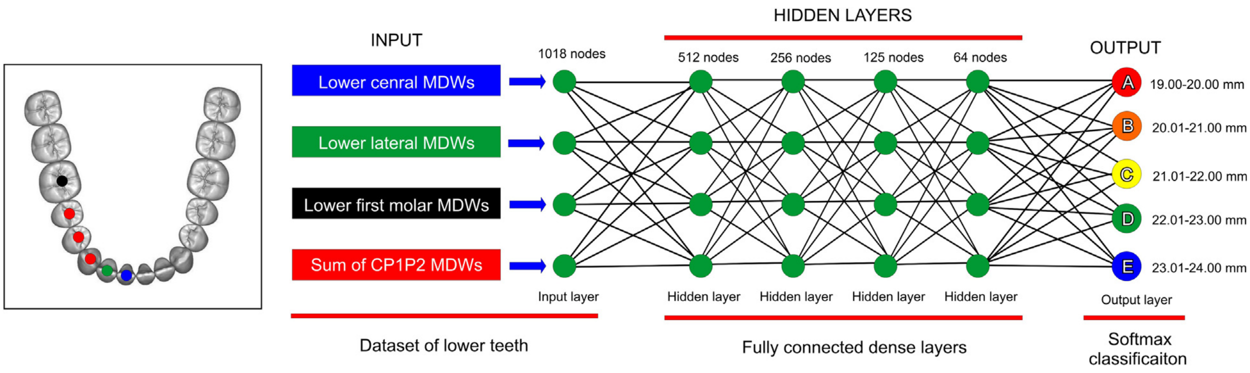

Fig 1. The general architecture of the present ANN model. The dataset for the mandibular incisor was labeled as input. Each of the convolutional layers was followed by a rectified linear unit activation function. In the final output layer, the dataset classification was performed using the Softmax function. The subgroups were classified as follows: A, 19.00-20.00 mm; B, 20.01-21.00 mm; C, 21.01-22.00 mm; D, 22.01-23.00 mm; and E, 23.01-24.00 mm. CP1P2, canines and premolars.

create a neural network by establishing links between themselves. This ANN consists of 3 layers: input, hidden, and output (Fig 1).[14] The data input is performed on the first layer. The AI trains itself in the hidden layers. The hidden nodes in these layers work as ANN system interneurons. In the output layer, the AI makes predictions on the basis of the input dataset. Not all data are used simultaneously during model training; datasets are gradually included as small groups in the AI training. Training starts with the first minor group, and the performance of the model is tested. Based on the success rate, the neurons' weights are updated from previous iterations. Subsequently, the model is retrained, and the weights are updated again using the new training set. This process is repeated at each training step, and the most appropriate weight values are calculated for the model. Each training step is called an epoch, which refers to 1 cycle through the full training dataset. Because these back-and-forth movements are repeated dozens or hundreds of times, the ANN trains itself on the basis of the designed model. A certain number of learning iterations (epoch) is another key element in the high success rate of training.

The present research study aimed to create an ANN model that could predict the MDW of the mandibular CP_1P_2 teeth and test its success after training. The accuracy and validity of the DL system were assessed by comparing it with the MT results.

MATERIAL AND METHODS

The experimental protocols of this study were approved by the Afyonkarahisar Health Sciences University Clinical Research Ethics Committee. The research was conducted using digital models acquired by the TRIOS (3Shape, Copenhagen, Denmark) intraoral scanner in the orthodontics archive of our faculty. Mandible digital models of 974 patients with permanent dentition were used. Because of the possibility of congenitally missing lateral incisors in the maxilla, only lower models were preferred. The exclusion criteria were as follows: incomplete permanent dentition, history of previous orthodontic treatment, presence of proximal restoration, caries, attrition, and dental anomalies. The mesiodistal crown widths of the mandibular right teeth were measured digitally by a single researcher (H.C.) using Ortho Analyzer software (Copenhagen, Denmark).

The ANN model, which can predict the MDW of unerupted CP_1P_2 teeth, was designed by a software developer (F.S.) (Fig 1). The Google Colab AI Laboratory was used in the ANN construction (https://github.com/farhadsalmanpour/cp1p2-class). Python was the preferred programming language. MDW measurements of the mandibular teeth were introduced to AI in the form of datasets. The same tooth datasets were written in a single column 1-by-1. Four columns were used to enter the dataset: column 1, mandibular central MDWs; column 2, mandibular lateral MDWs; column 3, mandibular first molar MDWs; and column 4 (C4), the sum of the CP_1P_2 MDWs. The C4 data were classified into 5 subgroups according to the MDWs of the teeth to train the AI more precisely and effectively. The subgroups were categorized as follows: A, 19.00-20.00 mm; B, 20.01-21.00 mm; C, 21.01-22.00 mm; D, 22.01-23.00 mm; and E, 23.01-24.00 mm. All 4 columns were used to train the AI, and then the AI was asked to estimate the C4 data using the column 1-column 3 datasets. Supervised learning algorithms

Table I. Current DL system characteristics

Layer (type)	*No. of nodes*	*Parameter*
Input 1	512	2048
Hidden 1	256	131328
Hidden 2	64	16448
Hidden 3	32	2080
Hidden 4	16	528
Output 1	6	102

Note. Total parameters, 152,534; trainable parameters, 152,534; nontrainable parameters, 0.

were used to train the model.[15,16] As a consequence, an ANN model consisting of 5 layers (4 hidden layers and 1 output layer) was obtained (Table I). The model had 512 input layer nodes and 6 output layer nodes. There were also 368 hidden nodes in the 4 hidden layers that served as interneurons.

There are many activation functions for DL systems.[17] Maximizing the success rate of DL learning can only be achieved by selecting a sufficient number of layers and neurons and an accurate activation function. In the present study, the rectified linear unit activation function and the Softmax classifier function were used to create the current ANN model.[18,19] The main advantage of using the rectified linear unit activation function over other activation functions is that it does not activate all the neurons simultaneously. The active use of some but not all of the neurons in the network improves the calculation efficiency.[20] The Softmax function was used as the final activation function in the neural networks to classify the output dataset.

In the current ANN model, the success of the DL system reached its highest level after 200 epochs. The AI tested the validation of the model using a backpropagation algorithm. The algorithm achieved this by sharing errors with the nodes in each layer from the beginning to the end. The error function was therefore reduced as much as possible.

The first 70.00% of the data were assigned as the training set, and another 20.00% were assigned to the validation set.[18,21,22] The remaining 10.00% of the data was used to test the success of the DL system. For 100 randomly selected patients, the MT prediction values were compared with the DL system estimates. The testing dataset (10.00%) was the same as the 100 patients' data used to compare the MT and DL results. An error margin of 1.00 mm was created for the MT prediction values by adding $\pm$ 0.50 mm to these values. In other words, the estimation results of patients with real CP_1P_2 values between 18.90 mm and 19.90 mm were considered correct. The purpose of this 1.00 mm error margin was to create 1.00 mm intervals for the MT results, similar to the output layer subgroups. Otherwise, the results of the MT and DL could not be compared.

Statistical analysis

SPSS software (version 22.0; IBM Corp, Armonk, NY) was used for the analysis. Ten randomly selected patient models were remeasured 10 days later by the same researcher (H.C) to assess the error rate in the MDW measurements. Repeated measurements were compared using the intraclass correlation coefficient.

RESULTS

The training and validation test graphics that represent the success rate of the AI training are shown in (Fig 2). The validation test results reached 49.50% during model training. The current ANN model was tested using 100 randomly selected patients, and the model made the correct prediction for 50.00% of the patients. A patient prediction table is shown as subgroups in (Fig 3, A and E). In the ANN model, the output layer dataset was categorized at an interval of 1.00 mm. If the second-highest column is accepted as a clinically acceptable prediction, the success rate could increase to 75.00%. The intraclass correlation coefficient results of repeated MDW measurements are shown in Table II.

The success rate of the DL system was 49.50%, whereas the MT correctly predicted 45.00% with a 1.00 mm error margin. This 45.00% success rate was valid for the 50.00% probability value. The success values for other MT probability values are shown in (Table III). The results of this table revealed that success rates could vary by different MT probability values. The authors considered that this situation was associated with the use of data from a single community. These percentages may vary if the DL system is trained with data from another society.

DISCUSSION

DL is a subset of machine learning that mimics the process of human brain learning. DL systems analyze datasets using ANN models specially designed to perform certain tasks. DL evaluates several datasets at the same time in different hidden layers. In the last 2 decades, DL model architectures have been created for many purposes, such as identifying cephalometric landmarks,[23] determining the need for orthodontic extraction,[24] evaluating cervical vertebral maturation,[25] and predicting facial attractiveness after orthognathic surgery.[26] The efficiency of some of these algorithms goes beyond what humans can do.[27]

In the existing literature, there is no precise method for determining the MDWs of unerupted teeth. This

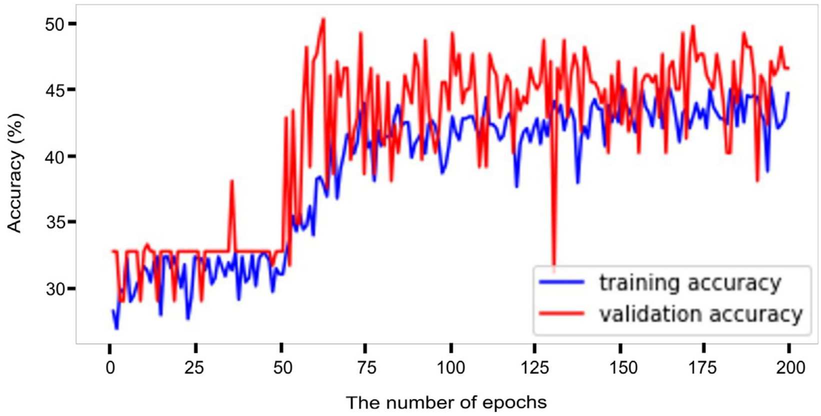

Fig 2. The training and validation test graphics.

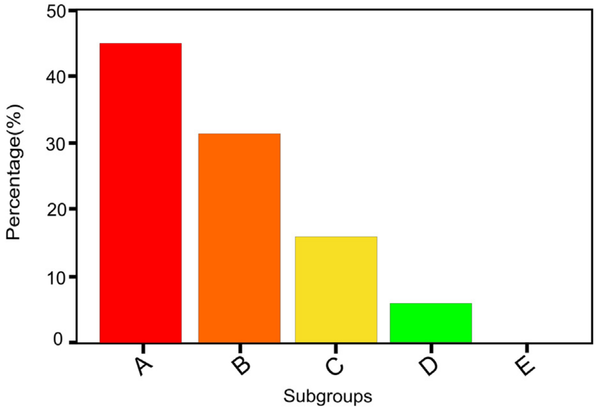

Fig 3. This bar chart illustrates the results of a randomly selected patient and explains how the ANN system works. The patient's MDWs of mandibular first molars and central and lateral incisors were entered into the system, and the DL system was asked to estimate unerupted CP_1P_2 widths. The software gave us the patient's probability values in 5 subgroups. According to this bar graph, the patient's unerupted CP_1P_2 widths most likely belonged to subgroup A, and the second most likely prediction was in subgroup B. The probability value decreased in other subgroups consecutively for this randomly selected patient. In addition, the possibility of the patient's unerupted CP_1P_2 widths belonging to subgroup E was zero.

research study was not intended to develop a precise method using the DL system but to provide an alternative to the currently available methods. Similar to the present study, an ANN system was developed by Moghimi et al[28] to estimate the mesiodistal width of unerupted canines and premolars. In their study, only the error rate of the linear regression analysis and their ANN system were compared. It was reported that the highest correlation in the mandible was between the sum of the CP_1P_2 MDWs and the sum of the MDWs of the mandibular first molars and central and lateral incisors ($r = 0.697$). This finding was why mandibular first molars and central and lateral incisors teeth were used as a reference for estimating the width of unerupted canines and premolars in the current DL model. However, the present study was conducted using a larger sample size, and the evaluation was performed with a success rate rather than an ANN error rate. In comparing the MT and DL predictions, the implementation of the DL system in clinical use was targeted. In addition, the training dataset was categorized into subgroups, and the training success of the DL was increased.

The first molar measurement, which is not used for Moyers, was added to the research protocol. The findings of the current study revealed its potential correlation with the unerupted canines or premolars. This is an interesting finding, according to the authors, because it could enable orthodontists to look at the mixed dentition analysis in a new way.

Table II. The ICC results for the mandibular teeth measurements

	Teeth					
Variable	*Central*	*Lateral*	*Canine*	*First premolar*	*Second premolar*	*First molar*
ICC	0.957	0.974	0.957	0.982	0.865	0.960

ICC, intraclass correlation coefficient.

Table III. The MT probability values and their success rates

Variable			%		
Probability values	50	65	75	85	95
Success rate	45	46	43	29	12

The size of the teeth, like facial characteristics, varies among different ethnic groups.[29] The MT was derived from Caucasian population data.[30] The applicability of the table to different populations is therefore still debated. The effectiveness of the table in each society has been investigated by researchers to obtain more accurate prediction results.[31-33] For example, it was suggested by Güner and Ülgen[34] that the 50.00% MT probability value was suitable for Turkish children. Therefore, 50.00% probability values were used in the comparisons in this study.

In previous studies, an average of 100 samples was sufficient for estimation using linear[28] or multiple regression[35] analyzes. The present study was conducted using 974 samples because the size of the dataset has a direct impact on the success or efficacy of the DL model.[36,37] Generally, it is common knowledge that too little training data results in poor approximation and, not enough test data will lead to a high variance in the model performance estimation. When the training dataset is large enough, the network has a stronger estimation capability and self-learning function. A large number of samples was therefore introduced into the DL model for proper training.

In the current ANN model, the output dataset was classified into 5 subgroups (from A to E) at 1.00 mm intervals to increase the success rate of the AI prediction. However, there was no homogeneous distribution in the output layer datasets. The distribution of the samples in the subgroups was as follows: A, 147; B, 273; C, 314; D, 181; and E, 59. The authors consider that the success of the DL was negatively affected by this situation. The DL could have generated more successful results by using more homogeneous datasets. In other words, if there had been an equal number of samples in all of the subgroups, the success of the DL model could have been more than 49.50%. In addition, the current model is improvable. In the long term, the success rate could be increased using better-designed ANN models.

The parallelism between the validation and the training line was not distorted as more datasets were entered, so early stop was not used (Fig 2). However, it was stated in a previous study that the researchers had limited the number of references to reduce the error rate.[28]

In the present ANN model, the output layer results were classified into 5 subgroups with 1.00 mm intervals. If they had been categorized at intervals of 0.50 mm, 10 subgroups would have appeared, and the prediction ability of the ANN model would have decreased. The prediction range of intervals greater than 2.00 mm would prevent the authors from achieving clinically acceptable results. Wide estimation intervals could have a negative impact on orthodontic treatment plans, as both overestimation and underestimation of the crown diameters of unerupted canines and premolars could change a plan, especially extraction decisions. When the predicted values underestimate the real values, a space will be required to properly align the teeth. However, an error margin of 1.00 mm or 2.00 mm in tooth size estimation is tolerable for clinical applications such as incisor protrusion or interproximal enamel reduction during treatment.

An error margin of 1.00 mm was generated in the MT results for comparison with the DL prediction. However, this artificial 1.00 mm error margin could have an impact on the fact that the prediction success rates for the 50.00%, 65.00%, and 75.00% probability values were close to each other (Table III).

This study was conducted with a limited number of patients from a single ethnic group. AI trained using datasets from only 1 ethnic group from a specific region of a country cannot be expected to maintain the same level of success in different regions of the world because the width of the teeth and the ratio of the mandibular teeth could show racial or regional differences.[38-40] However, the current DL system could be trained to estimate the MDWs of unerupted teeth belonging to various ethnic groups.

CONCLUSIONS

The current study findings were encouraging and promising, and the results suggest that the performance of the DL system was 1 step ahead of the MT. It appears to be possible that even more successful results could be achieved in further studies using larger samples collected from various ethnic groups and more advanced DL systems.

ACKNOWLEDGMENTS

The authors thank Aytekin Yıldızhan for his contribution to the improvement of this article.

REFERENCES

1. Carey CW. Diagnosis and case analysis in orthodontics. Am J Orthod 1952;38:149-61.
2. Paredes V, Gandia JL, Cibrian R. A new, accurate and fast digital method to predict unerupted tooth size. Angle Orthod 2006;76:14-9.
3. Pancherz H, Schäffer C. Individual-based prediction of the size of the supporting zones in the permanent dentition. A comparison of the Moyers method with a unitary prediction value. J Orofac Orthop 1999;60:227-35.
4. Tanaka MM, Johnston LE. The prediction of the size of unerupted canines and premolars in a contemporary orthodontic population. J Am Dent Assoc 1974;88:798-801.

5. Al-Khadra BH. Prediction of the size of unerupted canines and premolars in a Saudi Arab population. Am J Orthod Dentofacial Orthop 1993;104:369-72.
6. Bherwani AK, Fida M. Development of a prediction equation for the mixed dentition in a Pakistani sample. Am J Orthod Dentofacial Orthop 2011;140:626-32.
7. Sherpa J, Sah G, Rong Z, Wu L. Applicability of the Tanaka-Johnston and Moyers mixed dentition analyses in Northeast Han Chinese. J Orthod 2015;42:95-102.
8. Burhan ASH, Nawaya FR. Prediction of unerupted canines and premolars in a Syrian sample. Prog Orthod 2014;15:4.
9. Hammad SM, Abdellatif AM. Mixed dentition space analysis in Egyptian children. Pediatr Dent J 2010;20:115-21.
10. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73-81.
11. Gilvary C, Madhukar N, Elkhader J, Elemento O. The missing pieces of artificial ıntelligence in medicine. Trends Pharmacol Sci 2019;40:555-64.
12. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res 2020;99:769-74.
13. Jakhar D, Kaur I. Artificial intelligence, machine learning and deep learning: definitions and differences. Clin Exp Dermatol 2020;45:131-2.
14. Pyo S, Lee J, Cha M, Jang H. Predictability of machine learning techniques to forecast the trends of market index prices: hypothesis testing for the Korean stock markets. PLoS One 2017;12:e0188107.
15. Barto AG, Dietterich TG. Reinforcement learning and its relationship to supervised learning. In: Handbook of Learning and Approximate Dynamic Programming. Hoboken: Wiley-IEEE Press; 2004. p. 47-63.
16. Hayashi K, Sachdeva AU, Saitoh S, Lee SP, Kubota T, Mizoguchi I. Assessment of the accuracy and reliability of new 3-dimensional scanning devices. Am J Orthod Dentofacial Orthop 2013;144:619-25.
17. Rasamoelina AD, Adjailia F, Sincak P. A review of activation function for artificial neural network. In: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics. SAMI 2020: Proceedings of the 18th World Symposium on Applied Machine Intelligence and Informatics; 2020 Jan 23-25; Herl'any, Slovakia. Piscataway: IEEE; 2020. p. 281-6.
18. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci 2018;48:114-23.
19. Ngoc V, Agwu AC, Son LH, Tuan TM, Nguyen Giap C, Thanh M, et al. The combination of adaptive convolutional neural network and bag of visual words in automatic diagnosis of third molar complications on dental x-ray images. Diagnostics (Basel) 2020;10:209.
20. Schmidt-Hieber J. Nonparametric regression using deep neural networks with ReLU activation function. Ann Stat 2020;48:1875-97.
21. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent 2018;77:106-11.
22. Hegde H, Shimpi N, Panny A, Glurich I, Christie P, Acharya A. Development of non-invasive diabetes risk prediction models as decision support tools designed for application in the dental clinical environment. Inform Med Unlocked 2019;17:100254.
23. Park JH, Hwang HW, Moon JH, Yu Y, Kim H, Her SB, 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.
24. Xie X, Wang L, Wang A. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. Angle Orthod 2010;80:262-6.
25. 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.
26. Patcas R, Bernini D, Volokitin A, Agustsson E, Rothe R, Timofte R. Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. Int J Oral Maxillofac Surg 2019;48:77-83.
27. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging (Bellingham) 2017;4:014501.
28. Moghimi S, Talebi M, Parisay I. Design and implementation of a hybrid genetic algorithm and artificial neural network system for predicting the sizes of unerupted canines and premolars. Eur J Orthod 2012;34:480-6.
29. Warren JJ, Bishara SE. Comparison of dental arch measurements in the primary dentition between contemporary and historic samples. Am J Orthod Dentofacial Orthop 2001;119:211-5.
30. Buwembo W, Luboga S. Moyer's method of mixed dentition analysis: a meta-analysis. Afr Health Sci 2004;4:63-6.
31. Durgekar SG, Naik V. Evaluation of Moyers mixed dentition analysis in school children. Indian J Dent Res 2009;20:26-30.
32. Carrillo JJP, Rubial MC, Albornoz C, Villalba S, Damiani P, de Cravero MR. Applicability of the Moyers' probability tables in adolescents with different facial biotypes. Open Dent J 2017;11:213-20.
33. Rojo JFG, Sandoval DD, Minjarez AM, García ARR. Gender-adjustment of Moyers dentition analysis for the Nayarit (Mexico) population. Rev Odontol Mex 2015;19:e224-7.
34. Güner D, Ülgen M. Adaptations of the Moyers tables for Turkish children (Determination of the mesio-distal widths of the canines and the premolars). Turk J Orthod 2000;13:102-7.
35. Boitor CG, Stoica F, Nasser H. Prediction of the mesiodistal size of unerupted canines and premolars for a group of Romanian children: a comparative study. J Appl Oral Sci 2013;21:225-30.
36. Barbedo JGA. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput Electron Agric 2018;153:46-53.
37. Caluña G, Guachi-Guachi L, Brito R. Convolutional neural networks for automatic classification of diseased leaves: the impact of dataset size and fine-tuning. In: Computational Science and Its Applications – ICCSA 2020. ICCSA 2020: Proceedings of the 2020 International Conference on Computational Science and Its Applications; 2020 Jul 1-4; Cagliari, Italy. Cham: Springer; 2020. p. 951-66.
38. Schirmer UR, Wiltshire WA. Orthodontic probability tables for black patients of African descent: mixed dentition analysis. Am J Orthod Dentofacial Orthop 1997;112:545-51.
39. Niwa T. Prediction of biological targets using probabilistic neural networks and atom-type descriptors. J Med Chem 2004;47:2645-50.
40. Yang ZR, Thomson R. Bio-basis function neural network for prediction of protease cleavage sites in proteins. IEEE Trans Neural Netw 2005;16:263-74.