scientific reports

OPEN

Limitations of panoramic radiographs in predicting mandibular wisdom tooth extraction and the potential of deep learning models to overcome them

Atsushi Danjo[1,4✉], Chiaki Kuwada[2,4], Reona Aijima[1], Asana Kamohara[1], Motoki Fukuda[3], Yoshiko Ariji[3], Eiichiro Ariji[2] & Yoshio Yamashita[1]

Surgeons routinely interpret preoperative radiographic images for estimating the shape and position of the tooth prior to performing tooth extraction. In this study, we aimed to predict the difficulty of lower wisdom tooth extraction using only panoramic radiographs. Difficulty was evaluated using the modified Parant score. Two oral surgeons (a specialist and a clinical resident) predicted the difficulty level of the test data. This study also aimed to evaluate the performance of a deep learning model in predicting the necessity for tooth separation or bone removal during wisdom tooth extraction. Two convolutional neural networks (AlexNet and VGG-16) were created and trained using panoramic X-ray images. Both surgeons interpreted the same images and classified them into three groups. The accuracies for humans were 54.4% for both surgeons, 57.7% for AlexNet, and 54.4% for VGG-16. These results indicate that accurately predict the difficulty of wisdom teeth extraction using panoramic radiographs alone is challenging. However, AlexNet and VGG-16 had sensitivities of more than 90% for crown and root separation. The predictive ability of our proposed model is equivalent to that of an oral surgery specialist, and a recall value > 90% makes it suitable for screening in clinical settings.

Keywords Panoramic Radiographs, Mandibular wisdom Tooth Extraction, Deep learning Models, Screening

In the field of oral surgery, extraction of impacted mandibular wisdom teeth are relatively common. They often cause surgical complications, such as bleeding, damage to the inferior alveolar nerve, postoperative pain, swelling, and infection. Oral surgeons often encounter cases of interrupted extraction or severe surgical complications referred by unfamiliar oral surgeons or general dentists. This may be due to a lack of preoperative predictions of extraction difficulties. However, it is difficult for novice oral surgeons and general dentists to predict the difficulty of wisdom tooth extractions.

Factors that increase the difficulty of a wisdom tooth include proximal–distal inclination of the tooth on radiographs[1], distance to the mandibular ramus and second molar, depth of impaction[2], root shape and morphology (curvature and enlargement), width of the periodontal ligament space, thin dental sac[3], bone quality, contact between the second molar and wisdom tooth, proximity to the inferior alveolar nerve, and a fully bony impacted wisdom tooth[4]. Several classifications have been published as indicators for predicting the degree of difficulty in using radiographic images. These methods include evaluation of the depth of impaction, distance between the mandibular ramus and the second molar, and proximal–distal tilt (Pell and Gregory[2], Pederson[5]), as well as the WHARFE classification system, which considers the root morphology and size of the dental sac[3]. Although these methods are good indicators of anatomical difficulty, some reports have indicated that they do not correlate with the actual difficulty of the surgery. Therefore, we used the modified Parant score as a method

[1]Department of Oral and Maxillofacial Surgery, Faculty of Medicine, Saga University, 5-1-1 Nabeshima, Saga 849-8501, Japan. [2]Department of Oral and Maxillofacial Radiology, Aichi Gakuin University School of Dentistry, Nagoya, Japan. [3]Department of Oral Radiology, School of Dentistry, Osaka Dental University, Osaka, Japan. [4]Atsushi Danjo and Chiaki Kuwada contributed equally to this work. ✉email: danjoat@cc.saga-u.ac.jp

for predicting the difficulty of simple extraction (forceps extraction), bone removal, crown division, crown-root division, and other procedures as an index of difficulty[6]. This index is easily understandable for both general dentists and patients during preoperative explanations. Experienced oral specialists and residents performed the predictions, and their accuracy and sensitivity were verified.

Machine learning methods that use deep learning (DL) with convolutional neural networks have been used in various medical fields. In dentistry, it is used to detect impacted supernumerary teeth, cleft lip and palate, dental caries, periodontal disease, and oral cancer[7891011]. Studies have been conducted to identify the mandibular canal[12], classify the depth of mandibular wisdom teeth[13] and predict the position of wisdom teeth in relation to nearby anatomical structures[14]. Although these previous studies provided important information regarding tooth extraction, they did not directly investigate the difficulty of tooth extraction. In this study, we developed a DL model to predict the difficulty of wisdom tooth extraction. We used the modified Parant score as the gold standard and tested whether DL could predict the required procedure for lower wisdom teeth extraction and whether it was better than the imaging ability of oral surgeons.

This is the first study to validate the results of DL-based image analysis using actual tooth extraction procedures. The null hypothesis was that there was no difference in the predictive ability of the DL model versus human observers.

Methods

Patients

A total of 1,095 panoramic images were obtained retrospectively from 695 patients with mandibular wisdom teeth (253 males and 442 females; average age: 31.2 years) between October 1, 2018, and September 30, 2021. All patients underwent unilateral extraction of the lower wisdom teeth under local anesthesia in an outpatient setting at the Department of Oral and Maxillofacial Surgery, Saga University Hospital. The patients were randomly selected, and panoramic images were obtained from the maximum number of patients available during the experimental period. Only images from patient who provided consent were included. For our analyses, we included lower wisdom teeth with the adjacent second molars and excluded those with severe periodontitis or mandibular lesions, such as odontogenic cysts or tumors. The panoramic radiographs obtained immediately before the tooth extraction were used for image analysis. The images were exposed to a tube voltage of 75 kV, tube current of 8 mA, and an exposure time of 10 s (VeraView IC5; Morita Manufacturing, Kyoto, Japan).

This study was conducted in accordance with the guidelines of the World Medical Association's Helsinki Declaration for Biomedical Research Involving Human Subjects, and the protocol was approved by the Institutional Review Board (IRB) of Saga University (Approval No. 2021–02-R-03). Due to the retrospective nature of the study, IRB of Saga University waived the need of obtaining informed consent. Additional information, including the opt-out format, was posted on the IRB website. All data were analyzed anonymously.

Diagnostic performance of oral surgeons

We retrospectively extracted data on wisdom tooth extraction from the surgical and medical records of the patients. All surgeries were performed by five oral and maxillofacial surgery specialists. We used the modified Parant score [6] (an indicator of post-extraction difficulty) to classify tooth extraction procedures into four grades: Grade I, extraction with forceps only; Grade II, extraction by ostectomy; Grade III, extraction by ostectomy and coronal section; and Grade IV, complex extraction (with bone removal, crown, and root division). For our analysis, we combined these grades into three groups: Group 1, easy (Grades I and II); Group 2, moderately difficult (Grade III); Group 3, highly difficult (Grade IV), and evaluated 30 images that were randomly assigned to the dataset (Table 1) by 10 oral surgeons (five specialists and five trainees). Of these, a doctor with the median sensitivity was selected from among the specialists and trainees as a human observer: one primary resident and one specialist with > 10 years of experience in clinical oral surgery and interpretation of radiographs. These analyses were performed separately for the results reported by the specialists and trainee surgeons.

DL architecture

The DL system was created using the Linus Ubuntu operating system (version 16.04.2). The network station was equipped with a GeForce 1080Ti graphics processing unit with 11 GB of memory (version 5.0; NVIDIA, Santa Clara, CA; https://developer.nvidia.com/digits). DL was performed using the AlexNet and VGG-16 architectures—both of which are available in the DIGITS library (version 5.0; NVIDIA; https://developer.nvidia.com/digits)—and the Caffé framework. The adaptive moment-estimation (Adam) solver was used with a base learning rate of 0.0001.

Group	Dataset			
	Training data	Validating data	Testing data	Total
1	120	31	30	181
2	232	58	30	320
3	451	113	30	594

Table 1. Total number of panoramic images in each dataset per group.

Development of learning models

The panoramic images from each group were downloaded from the image database in JPEG format, and each image was cropped to set a rectangular region of interest depending on the position of the wisdom tooth. If the wisdom tooth was impacted horizontally or angulated (mesial/distal), the positioning was defined as follows: the mesial end was set as the mesial end of the crown of the second molar, the distal end as the apex of the wisdom tooth, the upper end as the cusp of the second molar or the wisdom tooth (whichever was higher), and the lower end as the apex of the second molar. If the wisdom tooth was in the vertical (normal/inverted) position, the mesial end was set as the mesial end of the crown of the second molar, the distal end as the distal end of the crown of the wisdom tooth, the upper end as the second molar or cusp of the wisdom tooth (whichever was higher), and the lower end as the apex of the second molar (Fig. 1). Subsequently, we created two classification models, AlexNet and VGG-16. In each group, 30 images were randomly assigned to the test dataset, and the remaining images were assigned to the training and validation datasets to create the learning models (Table 1 and Fig. 1). The number of test data was determined based on the amount of data required for the chi-square test. The training data were augmented to create 7,000 images using the IrfanView software (Irfan Škiljan, Austria; https://www.irfanview.com/). Each model was created with a total of 100 learning epochs. The training and validation were conducted by C.K. Confusion matrices were created for each model (Fig. 2) and the recall, precision, and F-measure parameters were calculated. The F-measure was calculated as the harmonic average of recall and precision and served as an indicator of poor balance between the two parameters.
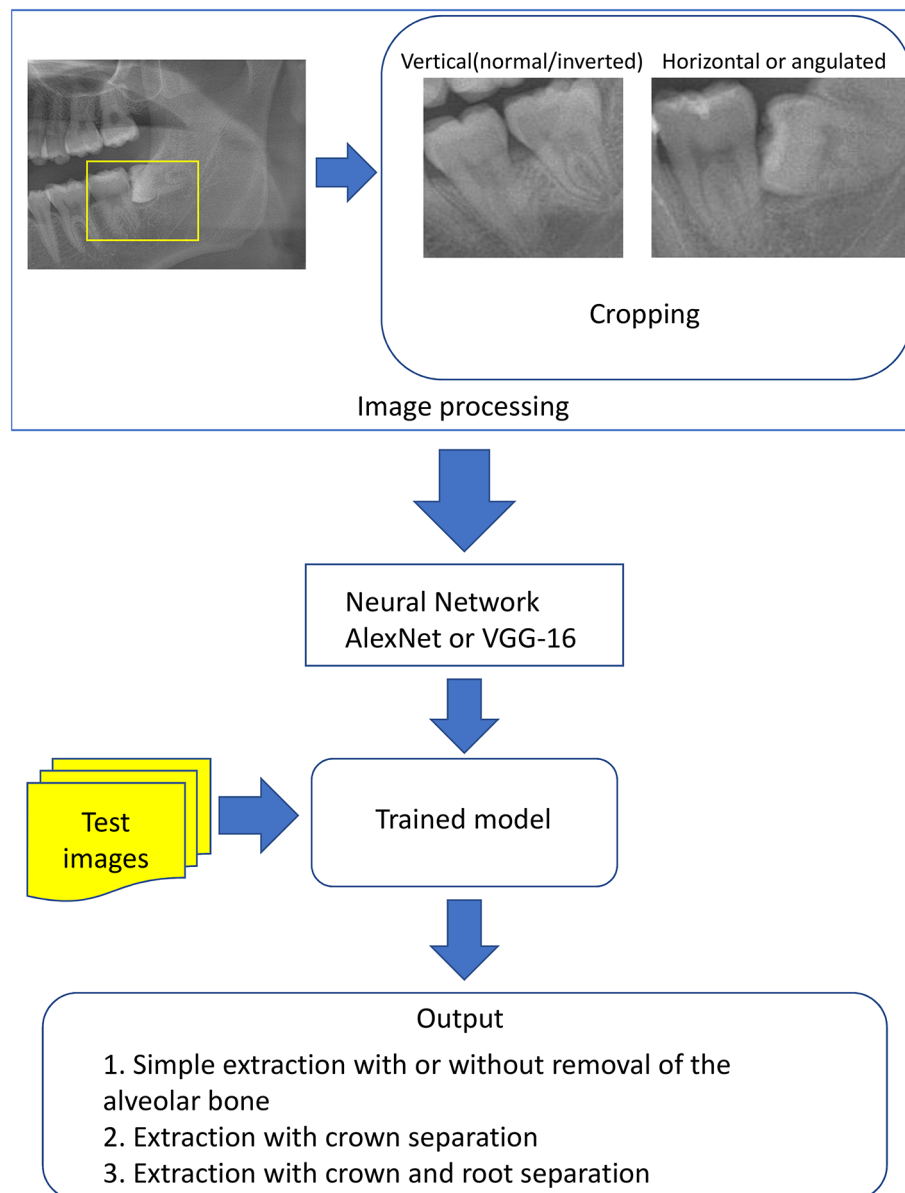

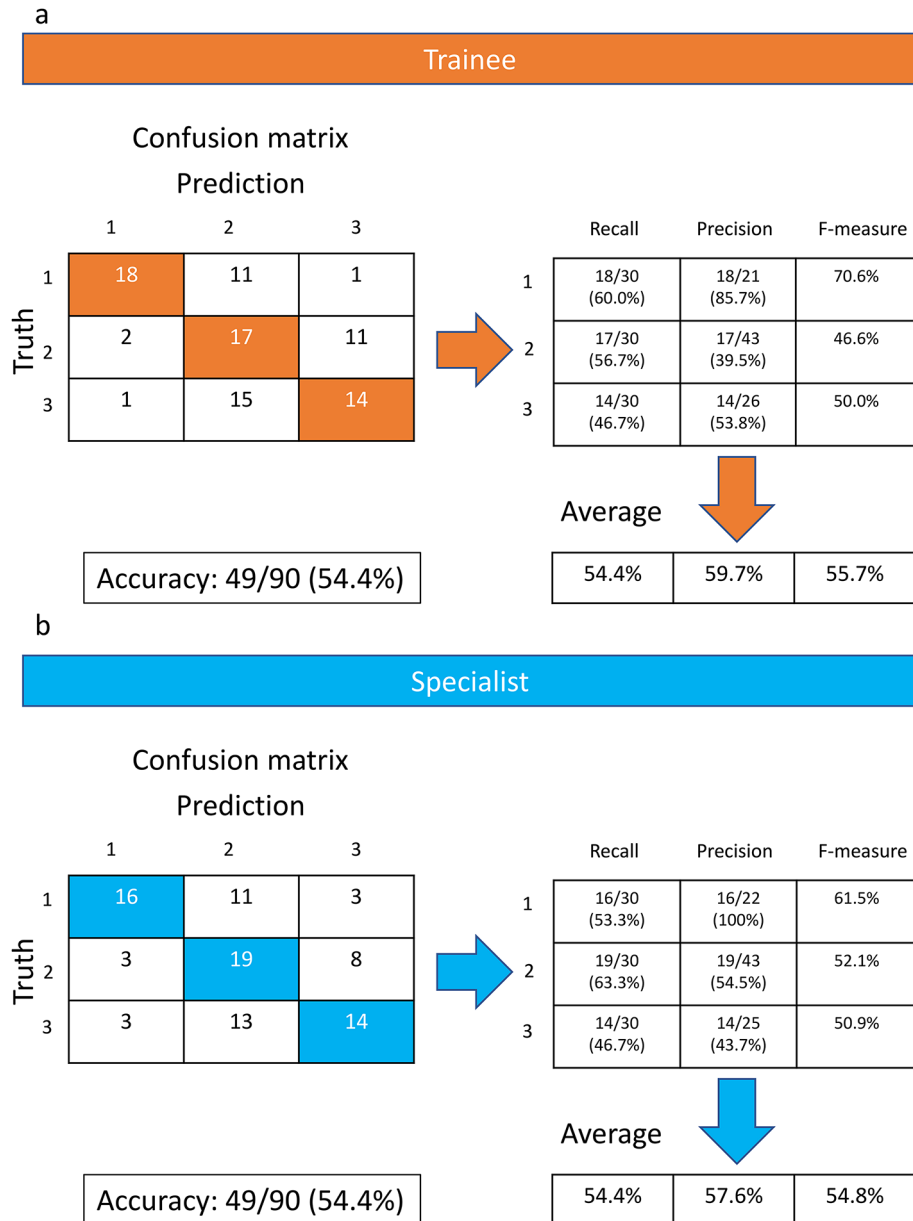**Fig. 1**. Processing of panoramic images and the diagnostic process used in this study.

Fig. 2. Confusion matrices used to analyze the performance of human observers: trainee surgeon (a, top) and specialist with > 10 years of experience in oral surgery (b, bottom).

Comparison of model performance with the diagnostic performance of human observers

The surgeons and DLs classified the same set of images as those included in the test dataset. They classified the teeth into three groups using the modified Parant score [6] as above and compared the findings of our DL-based image analysis with those of human observers. The results from AlexNet, VGG-16, and the two human observers were divided into two groups comprising teeth extracted with and without crown separation (Group 1 vs. Groups 2 and 3). We calculated the sensitivity, specificity, accuracy, and AUC (area under the curve) considering Groups 2 and 3 as positive (Table 2 top). The results were divided into two groups comprising teeth extracted with and without root separation (Groups 1 and 2 vs. Group 3). In addition, we calculated the sensitivity, specificity, accuracy, and AUC for Group 3 as positive (Table 2, bottom panel).

Statistical analysis

Comparisons between two groups were statistically evaluated using the chi-square test. The threshold for significance was set at $p < 0.05$.

Group 1 vs. Groups 2 and 3	Sensitivity	Specificity	Accuracy	AUC
AlexNet	0.93	0.66	0.84	0.78[a,b]
VGG-16	1.00	0.50	0.83	0.75[c]
Trainee	0.95	0.60	0.83	0.65[a,c]
Specialist	0.90	0.53	0.77	0.71[b]
Groups 1 and 2 vs. Group 3	**Sensitivity**	**Specificity**	**Accuracy**	**AUC**
AlexNet	0.90[d,e]	0.50	0.63	0.69
VGG-16	0.93f.[g]	0.40	0.57	0.66
Trainee	0.46[d,f]	0.80	0.68	0.63
Specialist	0.46[e,g]	0.81	0.70	0.64

Table 2. Summary of the sensitivity, specificity, accuracy, and AUC in Group 1 vs. Groups 2 and 3 (Top) and Groups 1 and 2 vs. Group 3 (bottom). AUC: Area under the receiver operating characteristic curve. a, b, c, d, e, f, and g: Values with the same character denote significant differences between them by the chi-square test with $p < 0.05$.

Results

As described in the method section, one doctor was selected from the five primary residents and one from the five specialists with over 10 years of experience, based on their median sensitivity. The sensitivities of the primary residents were 0.433, 0.444, 0.488, 0.522, and 0.544, while the sensitivities of the specialists were 0.4, 0.433, 0.544, 0.555, and 0.566. The results of the image analysis were interpreted and compared for all three groups by two oral surgeons (groups 1 vs. 2 vs. 3; Fig. 2). The average recall, average precision, average F-measure, and overall accuracy were 54.4%, 59.7%, 55.7%, and 54.4%, respectively, for the trainee surgeon (Fig. 2a) and 54.4%, 57.6%, 54.8%, and 54.4%, respectively, for the specialist with > 10 years of experience in oral surgery (Fig. 2b). No significant difference in accuracy was found between the specialist and trainee ($p = 1.000$).

Next, the image results were interpreted using both models (Groups 1, 2, and 3 for the AlexNet and VCG-16 models; Fig. 3). The AlexNet model had an average recall (sensitivity) of 57.7%, an average precision of 62.0%, an average F-measure of 53.8%, and an overall accuracy of 57.7% (Fig. 3a). The VGG-16 model had an average recall of 54.4%, an average precision of 66.0%, an average F-measure of 51.7%, and an overall accuracy of 54.4% (Fig. 3b). No significant difference in accuracy was observed between the two models ($p = 0.6523$).

When comparing the three groups in terms of predicting the need for crown separation during tooth extraction (Group 1 vs. Groups 2 and 3; Fig. 4a, Table 2 top), we found that AlexNet could significantly predict this need better than the trainee could ($p = 0.0016$) and the specialist ($p = 0.0347$). Although the VGG-16 model significantly outperformed the trainee in predicting the need for crown separation ($p = 0.0143$), it did not predict it better than the specialist ($p = 0.1942$). Finally, no significant difference was noted between the three groups in predicting the need for root separation during tooth extraction (groups 1 and 2 vs. group 3; Fig. 4b, Table 2 bottom). However, the sensitivities of both AlexNet and VGG-16 were significantly higher than those of the human observers (AlexNet vs. Trainee and Specialist: $p = 0.0003$, VGG vs. Trainee and Specialist: $p < 0.0001$) (Table 2 bottom).

Discussion

The null hypothesis of this study was that there would be no difference between the DL and human observers in predicting the difficulty of wisdom tooth extraction.

Prior to the extraction of the lower wisdom tooth, the surgeon estimated the degree of difficulty. Different factors affect the difficulty of the procedure, including tooth morphology, depth, patient age, gender, physique, and the surgeon's skill and experience[15,16]. Computed tomography can be used to examine the three-dimensional morphology of the wisdom tooth and its position in relation to the surrounding tissues to estimate the degree of difficulty. Panoramic radiography is a very useful imaging test in terms of its simplicity and popularity for screening.

This study examined whether panoramic radiography could predict the procedure required for mandibular wisdom tooth extraction. We also examined whether DL could outperform human diagnostic ability. The results of this study showed that it is difficult to accurately predict the necessary surgical procedure from panoramic radiography because the prediction by both oral surgeons and DL was in the 50% range. When comparisons were made under narrowly defined conditions (Table 2), the predictive accuracy of root separation was low but nearly identical between the specialist, resident, and DL-based procedures. This suggests that even surgeons highly skilled in surgical techniques and radiographic interpretation (including factors, such as root hypertrophy and thickening) have poor predictive accuracy. It was difficult to predict root separation from images alone, including the number of roots and their adhesion to the surrounding alveolar bone, which is consistent with previous reports[17].

In addition to imaging findings, several factors, such as an individual's age[15,16,18,19], gender[15], and body size[16,20] can also affect the difficulty of wisdom tooth extraction. Root blurring in dental imaging may indicate root adhesions or buccolingual curvature, potentially affecting the complexity of root separation during tooth extraction[15]. Several studies, including the present study, have reported that root separation and bone removal affect the surgical difficulty of wisdom tooth extraction[20,21].
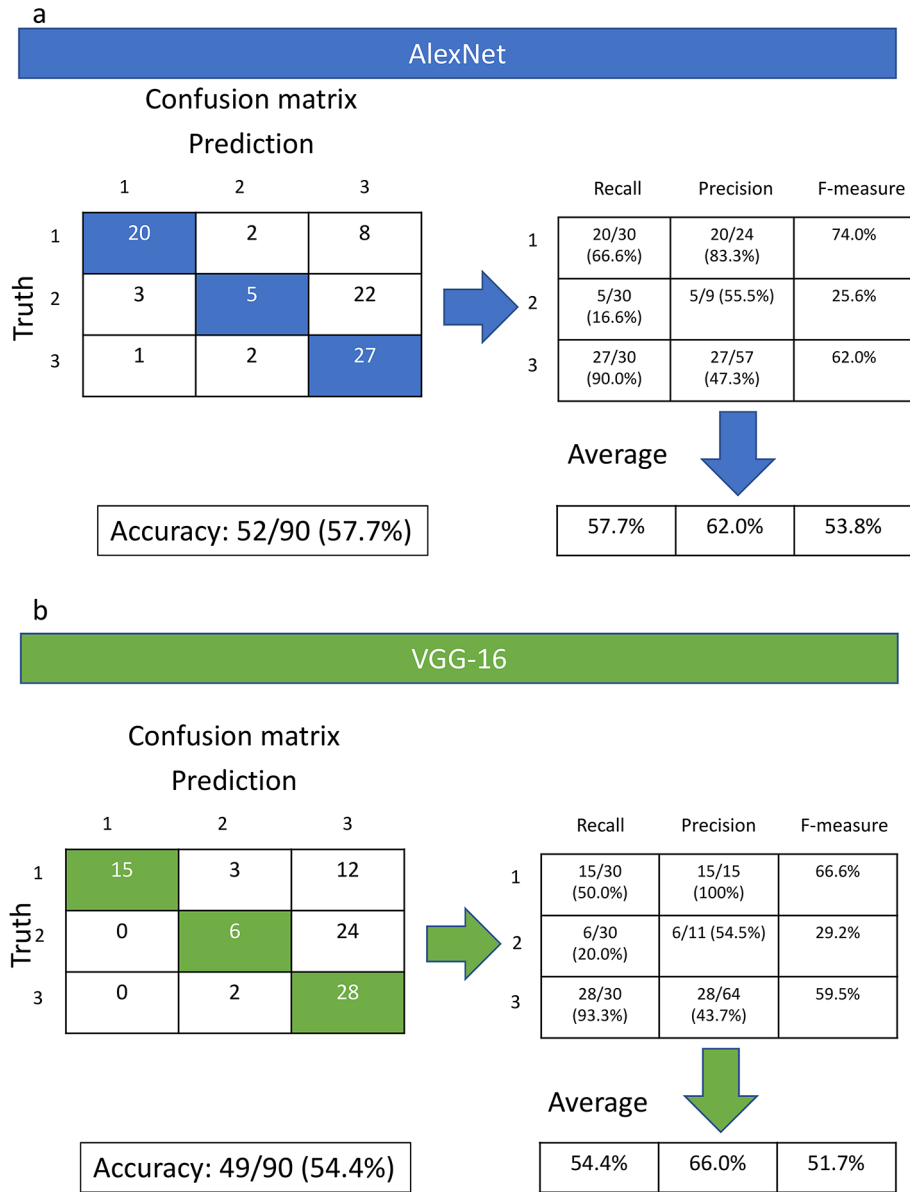
Fig. 3. Confusion matrices used to analyze the performance of the AlexNet (**a**) and VGG-16 (**b**) models.

These findings suggest that it is difficult to accurately predict the surgical difficulty of wisdom tooth extraction using panoramic images alone[17]. Komerik et al. reported that both residents and experienced surgeons have difficulty predicting the time required for wisdom tooth extraction solely based on information from panoramic images[22]. Therefore, without considering additional factors beyond imaging findings, neither DLs nor oral surgeons can accurately predict the surgical difficulty of wisdom tooth extraction.

According to a systematic review of methods for predicting the difficulty of mandibular wisdom tooth extraction, the Parant score was the major evaluation method, along with operative time and surgeon subjectivity[18,21,22]. All methods have the disadvantage of being influenced by the surgeon's experience, skill, instruments, and technique. However, in this study, the Parant score was used because the surgery was performed at a single institution with the same instruments and technique (separation method), which was considered to have relatively few unstable factors. Other indices include the Pell-Gregory and Pederson indices; however, these classifications have been reported to poorly express the actual difficulty[6,23]. We agree with these reports because the classification of the depth of impaction and the position of the teeth and mandibular ramus do not reflect the curvature of the root, bony attachments, or the position of the tooth in relation to the inferior alveolar nerve, which increases the difficulty of tooth extraction. In fact, a previous study also supports our opinion, as it found that the position of the wisdom tooth (horizontal, upright, tilted, and so on) does not correlate with the difficulty of wisdom tooth extraction; rather, it is influenced by the depth and root morphology. The Pell and Gregory and Pederson indices, which do not reflect root morphology, are considered to be incomplete evaluation methods[24].
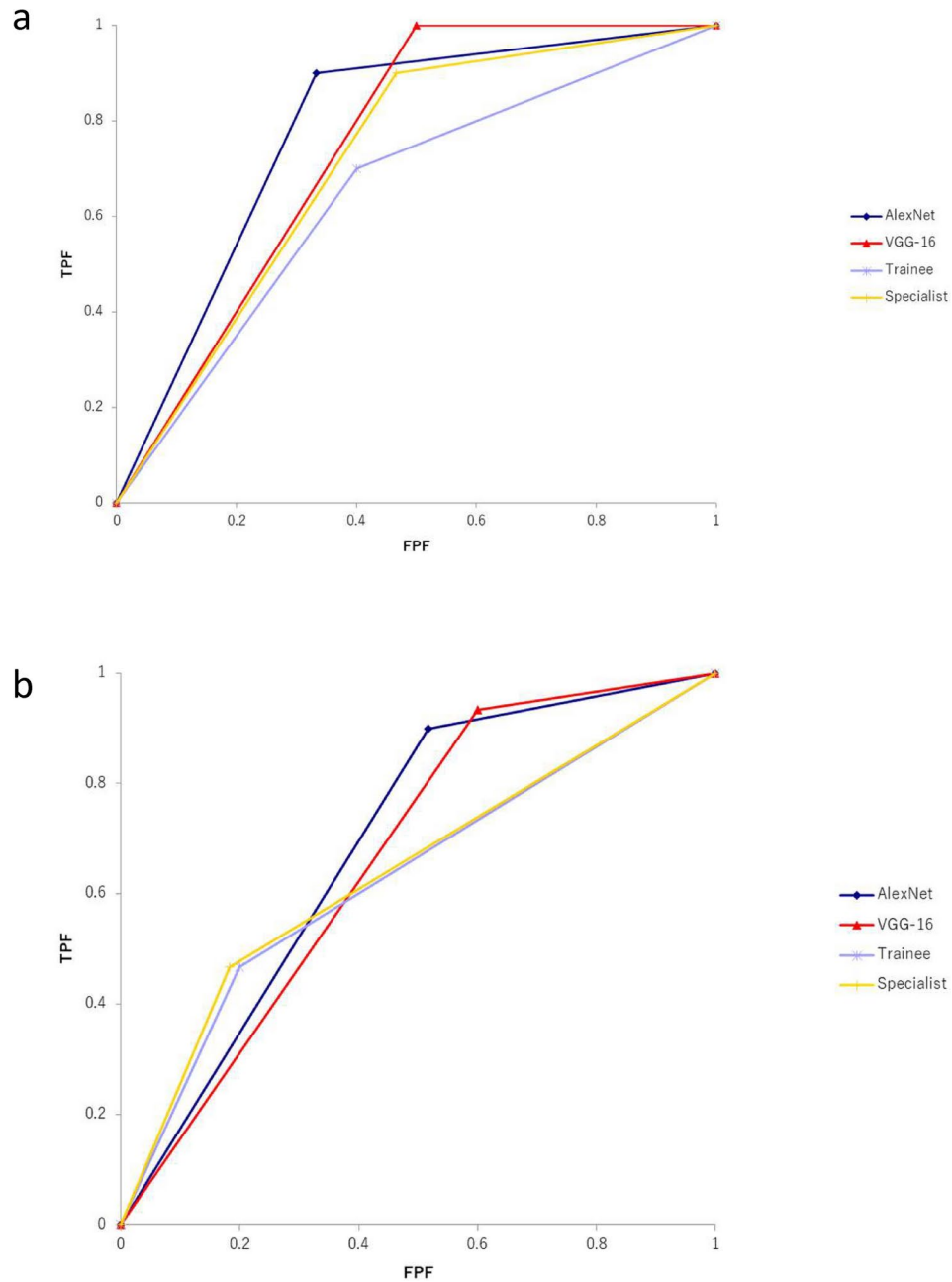
Fig. 4. Receiver operating characteristic (ROC) curve in Group 1 vs. Groups 2 and 3 (**a**, Top) and ROC curve in Groups 1 and 2 vs. Group 3 (**b**, bottom). The blue line indicates AlexNet, the red line indicates VGG-16, the purple line indicates a trainee, and the yellow line indicates a specialist.

Although we know that there is no perfect method to predict the difficulty based on imaging alone, the evaluation by panoramic radiographs is necessary for dentists who perform wisdom tooth extraction. It is also difficult for novice dentists to understand which elements of panoramic radiographs increase the difficulty of extracting a wisdom tooth. In such cases, high sensitivity is required for reading panoramic radiographs.

In this study, both DL models demonstrated a sensitivity of over 90% in the most complex procedures (Groups 3). For crown separation, the AUC of the DL models was significantly higher than that of human observers, suggesting that these DL models are effective as screening methods. Surgeons determine the need for crown removal based on the presence of obstructions (e.g., lower second molars or mandibular ramus) in the direction of wisdom tooth extraction. Thus, we expected humans to outperform the DL-based analysis in this regard. Contrary to our expectations, the DL-based analysis was more accurate than the human observers and had similar sensitivity (Table 2).

Previous studies on wisdom tooth extraction used the DL framework to detect and classify impacted teeth and determine the positional relationship between the root apex of the tooth and the mandibular canal[14,25]. These studies primarily used the DL algorithm to identify the object and its positional relationship on the

radiographic image and reported that their results showed high accuracy[13,14,25]. Yoo et al. developed a model to predict the Pederson difficulty index of mandibular wisdom teeth using panoramic radiographs and reported high accuracy (78.9–90.23%)[26]. These reports indicate that an artificial intelligence framework can accurately predict the classification of wisdom teeth and their positional relationship with the mandibular canal.

In addition to evaluating anatomical positioning, there have also been reports of using DL technology to analyze preoperative panoramic images to predict different surgical outcomes. These include sensory disturbance after lower wisdom teeth extraction[27] and maxillary sinus perforation after upper wisdom teeth extraction[28]. However, in the case of these studies, their findings tend to be inaccurate because the gold standard differs from the object in the image. The results of these studies could also have been influenced by several factors, such as the surgeon's level of surgical experience and the patient's condition. This trend is the same as that observed in the present study. This is one of the limitations of the present study.

Another limitation is that this study is based on the interpretation of panoramic images and that it is a single-institution study. The institution where this study was conducted is a university hospital, to which many patients are referred by general dentists. Therefore, there were very few patients in group 1 (simple extractions), which may have introduced bias due to variations in sample size between groups. The retrospective design of this study is another limitation. Future multicenter prospective studies with larger and more uniform sample sizes are required to validate our results. However, a multicenter study includes the possibility of a larger bias in the surgical technique. Despite these limitations, this is the first study to validate the results of a DL-based image analysis using actual extraction procedures.

Conclusion

This study demonstrated that it is difficult to predict the extraction difficulty with high accuracy using panoramic radiographs alone. The accuracy of the reading extraction difficulty from panoramic radiographs did not differ between residents and specialists, and the accuracy of the two DLs was similar to that of humans. Panoramic radiograph is a 2D imaging modality, making it challenging to accurately assess the three-dimensional morphology and spatial relationships of third molars. The findings of this study suggest that more advanced imaging modalities, such as cone-beam computed tomography (CBCT) or supplementary diagnostic tools, are necessary for accurately diagnosing the difficulty of mandibular third molar extraction. However, the information obtained from panoramic radiographs remains essential for extraction in current clinical practice. The AUC and sensitivity of the DLs for crown and root separation exceeded 90%, suggesting that DLs may be a good method for screening tooth extraction difficulty. Based on these results, the null hypothesis is partially rejected.

This is the first study to develop a DL model capable of predicting the difficulty of wisdom tooth extraction using surgical techniques. We believe that this is a practical and useful educational tool for inexperienced dentists and students. We anticipate that integrating three-dimensional image data, such as CBCT, in the future could further improve the performance of DL models, contributing to advancements in diagnosing the difficulty of mandibular wisdom teeth extractions.

Data availability

The datasets generated and/or analyzed in the current study are available from the corresponding author upon reasonable request.

References

1. Winter, G. B. *Impacted Mandibular Third Molar* 41–100 (American Medical Book Co, St. Louis, 1926).
2. Pell, G. J. & Gregory, B. T. Impacted mandibular third molars: Classification and modified techniques for removal. *Dent. Dig.* **39**, 330–338 (1933).
3. MacGregor, A. J. *The Impacted Lower Wisdom Tooth* (Oxford Univ, 1985).
4. Sammartino, G. et al. Extraction of mandibular third molars: Proposal of a new scale of difficulty. *Br. J. Oral Maxillofac. Surg.* **55**, 952–957 (2017).
5. Pederson, G. W. Oral surgery (W B Saunders, Philadelphia, 1994), 1988. (Cited in: Koerner, K. R. The removal of impacted third molars. Principles and procedures. *Dent. Clin. North Am.* **38**, 255–278 (1994).
6. García, A. G., Sampedro, F. G., Rey, J. G., Vila, P. G. & Martin, M. S. Pell-Gregory classification is unreliable as a predictor of difficulty in extracting impacted lower third molars. *Br. J. Oral Maxillofac. Surg.* **38**, 585–587 (2000).
7. Kuwada, C. *et al.* Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **130**, 464–469 (2020).
8. Kuwada, C. et al. Detection and classification of unilateral cleft alveolus with and without cleft palate on panoramic radiographs using a deep learning system. *Sci. Rep.* **11**, 16044 (2021).
9. Mohammad-Rahimi, H. et al. Deep learning for caries detection: A systematic review. *J. Dent.* **122**, 104115 (2022).
10. Krois, J. et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci. Rep.* **9**, 8495 (2019).
11. Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S. & Jantana, P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J. Oral Pathol. Med.* **50**, 911–918 (2021).
12. Choi, E. et al. Artificial intelligence in positioning between mandibular third molar and inferior alveolar nerve on panoramic radiography. *Sci. Rep.* **12**, 2456 (2022).
13. Sukegawa, S. et al. Deep learning model for analyzing the relationship between mandibular third molar and inferior alveolar nerve in panoramic radiography. *Sci. Rep.* **12**, 16925 (2022).
14. Vinayahalingam, S., Xi, T., Bergé, S., Maal, T. & de Jong, G. Automated detection of third molars and mandibular nerve by deep learning. *Sci. Rep.* **9**, 9007 (2019).
15. Park, K. L. Which factors are associated with difficult surgical extraction of impacted lower third molars?. *J. Korean Assoc. Oral Maxillofac. Surg.* **42**, 251–258 (2016).

16. Gbotolorun, O. M., Arotiba, G. T. & Ladeinde, A. L. Assessment of factors associated with surgical difficulty in impacted mandibular third molar extraction. *J. Oral Maxillofac. Surg.* **65**, 1977–1983 (2007).
17. Barreiro-Torres, J. et al. Evaluation of the surgical difficulty in lower third molar extraction. *Med. Oral Patol. Oral Cir. Bucal* **15**, e869–e874 (2010).
18. Sánchez-Torres, A. et al. Does mandibular gonial angle predict difficulty of mandibular third molar removal?. *J. Oral Maxillofac. Surg.* **77**, 1745–1751 (2019).
19. Obimakinde, O. S., Okoje, V., Ijarogbe, O. A. & Obimakinde, A. Role of patients' demographic characteristics and spatial orientation in predicting operative difficulty of impacted mandibular third molar. *Ann. Med. Health Sci. Res.* **3**, 81–84 (2013).
20. Alvira-González, J., Figueiredo, R., Valmaseda-Castellón, E., Quesada-Gómez, C. & Gay-Escoda, C. Predictive factors of difficulty in lower third molar extraction: A prospective cohort study. *Med. Oral Patol. Oral Cir. Bucal* **22**, e108–e114 (2017).
21. Aznar-Arasa, L., Figueiredo, R., Valmaseda-Castellón, E. & Gay-Escoda, C. Patient anxiety and surgical difficulty in impacted lower third molar extractions: A prospective cohort study. *Int. J. Oral Maxillofac. Surg.* **43**, 1131–1136 (2014).
22. Komerik, N., Muglali, M., Tas, B. & Selcuk, U. Difficulty of impacted mandibular third molar tooth removal: Predictive ability of senior surgeons and residents. *J. Oral Maxillofac. Surg.* **72**, e1–e6 (2014).
23. Diniz-Freitas, M. et al. Pederson scale fails to predict how difficult it will be to extract lower third molars. *Br. J. Oral Maxillofac. Surg.* **45**, 23–26 (2007).
24. Vranckx, M., Lauwens, L., Moreno Rabie, C. M., Politis, C. & Jacobs, R. Radiological risk indicators for persistent postoperative morbidity after third molar removal. *Clin. Oral Investig.* **25**, 4471–4480 (2021).
25. Fukuda, M. *et al.* Comparison of 3 deep learning neural networks for classifying the relationship between the mandibular third molar and the mandibular canal on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **130**, 336–343 (2020).
26. Yoo, J. H. et al. Deep learning based prediction of extraction difficulty for mandibular third molars. *Sci. Rep.* **11**, 1954 (2021).
27. Kim, B. S. *et al.* Deep learning-based prediction of paresthesia after third molar extraction: A preliminary study. *Diagnostics (Basel)* **11**, 1572 (2021).
28. Vollmer, A. *et al.* Artificial intelligence-based prediction of oroantral communication after tooth extraction utilizing preoperative panoramic radiography. *Diagnostics (Basel)* **12**, 1406 (2022).

Acknowledgements

Author contributions

A.D. conceived the study, compiled the results, interpreted the data, and wrote the manuscript. C.K. conducted the experiments involving the deep learning technique and analyzed the data. R.A., A.K., and S.T. contributed to data collection. M.F. contributed to data analysis and interpretation. Y.A. contributed to data interpretation and provided expertise. E. A. and Y. Y. supervised the experiments and provided instructions and advice.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.