scientific reports

Check for updates

OPEN

Deep learning-based ordinal classification overcomes subjective assessment limitations in intraoral free flap monitoring

Hyounmin Kim[1], Dongwook Kim[1,2,4] & Juho Bai[3,4]

While free flap reconstruction is essential for repairing post-surgical defects or oncologic defects, current monitoring methods lack standardization and rely on subjective assessments. This study aims to introduce an artificial intelligence model that quantitatively analyzes and classifies the status of intraoral free flaps to enhance clinical monitoring capabilities. In this study, a total of 1862 clinical photographs from 131 patients who underwent intraoral free flap reconstruction were analyzed. Based on final flap outcomes and expert evaluation, images were classified into three ordinal categories: "Normal," "Suspicious," and "Compromised." The ordinal classification model was developed using a novel approach using dynamic triplet margin loss technique. Using 1489 images for training and 373 for testing, our model achieved 0.9571 accuracy, significantly outperforming the baseline model (0.8847). The model showed F1 scores of 0.98, 0.85, and 0.73 for normal, suspicious, and compromised classes respectively, with AUC values exceeding 0.97 for all classes.

Keywords Artificial intelligence, Deep learning, Free flap monitoring, Oral surgery

Free flap reconstruction is an essential surgical procedure that can restore both function and aesthetics in patients with extensive defects caused by cancer or trauma[1,2]. Particularly in the head and neck region, defects beyond a certain size significantly diminish patients' quality of life, making successful free flap reconstruction crucial[2].

While successful free flap reconstruction requires meticulous monitoring over a specific period, current monitoring methods vary considerably without standardized protocols. Some practitioners detect the condition of flaps through visual observation, while others utilize methods such as laser doppler, photoplethysmography, or diffuse reflectance spectroscopy[3–5]. However, the qualitative nature of these assessments can lead to divergent opinions among observers, potentially delaying decisions regarding salvage procedures. Such delays can leave patients with critical and fatal postoperative complications[1,6].

Flap deterioration occurs progressively rather than abruptly, involving multiple transitional stages. Identifying these intermediate stages is critical for determining salvage timing by the time compromise is obvious, intervention may be too late. While previous AI models have addressed extraoral flap monitoring[7] or employed binary classification for intraoral flaps[8], we developed a three-tier ordinal classification system (normal, suspicious, and compromised) to better capture the gradual nature of tissue changes and provide more actionable clinical information.

Some researchers have considered this aspect and developed artificial intelligence (AI) models for the gradual changes of flaps; however, these models are still tailored to extraoral settings[7].

The aim of this study is to develop an AI model that can detect gradual changes in flaps reconstructing the oral cavity and assist in diagnosis.

[1]Department of Oral and Maxillofacial Surgery, Yonsei University College of Dentistry, Seoul 03722, Republic of Korea. [2]Oral Science Research Center, Yonsei University College of Dentistry, Seoul 03722, Republic of Korea. [3]Department of Global Business and Technology, Hankuk University of Foreign Studies, Yongin 17035, Republic of Korea. [4]Dongwook Kim and Juho Bai contributed equally to this work. ✉email: dwkimomfs@yuhs.ac; juho@hufs.ac.kr

Results

Patients characteristics

Medical records, including clinical photographs, were reviewed for 131 patients. Free flap tissue reconstruction was performed on various sites including the tongue and mandible. Vascular compromise occurred in 10 patients (7.6%). Among these, four patients presented with signs consistent with insufficient arterial flow, and four exhibited signs of venous congestion, necessitating salvage procedures. The remaining two patients experienced delayed total flap tissue loss of undetermined vascular etiology. Detailed characteristics are summarized in Table 1.

Model performance

Our model utilized dynamic margin triplet loss as a type of loss function, and its performance was compared against a baseline model trained solely with cross-entropy loss. Model performance was evaluated using precision, recall, and F1 score metrics, from which classification accuracy was calculated. The proposed model achieved an overall accuracy of 0.9571, surpassing the baseline model's accuracy of 0.8847. Detailed performance metrics for each class across both models are presented in Table 2.

Of the 316 images classified as 'normal', our proposed model misclassified only 6 images (5 as 'suspicious' and 1 as 'compromised'). Among the 52 images in the suspicious class, 8 were diagnosed as normal, and 1 was diagnosed as the compromised class. Of the 5 images in the compromised class, only 1 was misdiagnosed as the suspicious class, while all others were correctly classified as 'compromised' class, indicating flap failure.

In contrast, the baseline model using only cross-entropy loss showed significantly different error patterns. Of the 52 images in the suspicious class, 37 were misclassified as the normal class. Among the 5 images in the compromised class, 1 was misdiagnosed as the normal class and another as the suspicious class. This can be readily observed in the confusion matrix presented in Fig. 1, while the model's performance, as depicted by the receiver operating characteristics (ROC) curve, is shown in Fig. 2.

Characteristics	Patients (N = 131)
Age (mean)	58.1
Sex	
Male	81 (61.8%)
Female	50 (38.2%)
Primary site	
Tongue	47 (35.9%)
Mandible*	39 (29.8%)
Maxilla	11 (8.4%)
Floor of mouth	10 (7.6%)
Buccal mucosa	7 (5.3%)
Soft palate	2 (1.5%)
Lower gingiva	1 (0.7%)
Retromolar trigone	1 (0.7%)
Composite	13 (9.9%)
Mandible + Skin	4
Mandible + oropharynx	2
Maxilla + Skin	2
Parotidectomy defect	2
Mandible + Base of the tongue	1
Buccal mucosa + Skin	1
Hypopharynx + Skin	1
Type of free flap	
Anterolateral thigh free flap	64 (48.9%)
Fibula free flap	45 (34.3%)
Radial forearm free flap	22 (16.8%)
Occurrence of vascular compromise	10 (7.6%)
Arterial insufficiency	4
Venous congestion	4
Anonymous late failure	2
Total flap loss	2 (1.5%)

Table 1. Clinical characteristics of the study patients. * Include all cases that involve the mandible, not just the lower gingiva.

Model configuration	Class	Accuracy	Precision	Recall	F1 Score
Proposed (triplet loss + functional loss)	Normal	0.981	0.97	0.98	0.98
	Suspicious	0.827	0.88	0.83	0.85
	Compromised	0.800	0.67	0.80	0.73
Cross-entropy loss	Normal	0.991	0.89	0.99	0.94
	Suspicious	0.269	0.78	0.27	0.40
	Compromised	0.600	0.75	0.60	0.67

Table 2. Compare model performance between control and experimental groups.
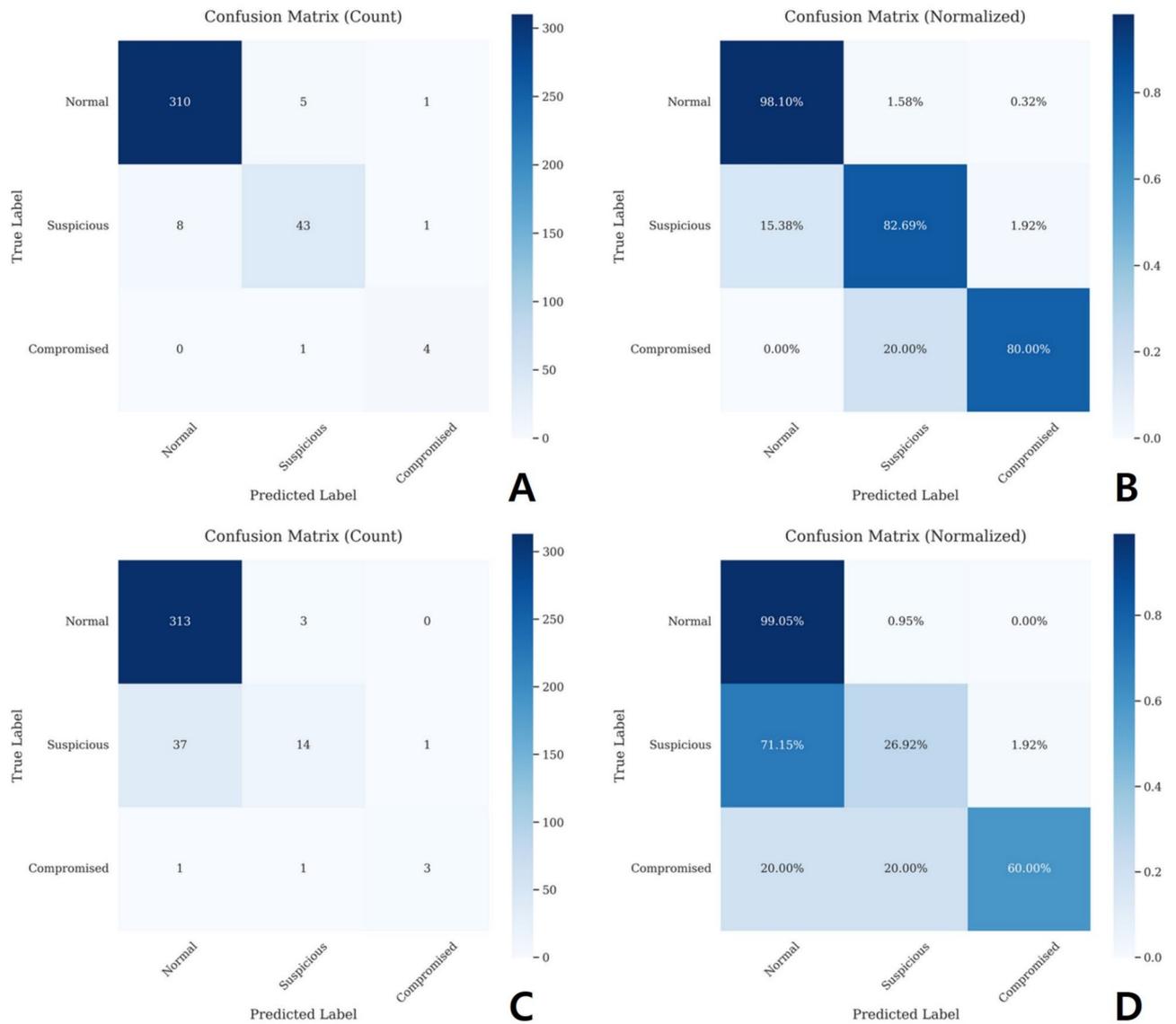
Fig. 1. Confusion matrices for both models. (**A**, **B**) Confusion matrix for the proposed model. (**A**) Displays the number of misclassified images, while (**B**) shows the corresponding probabilities. (**C**, **D**) Confusion matrix for the control model. (**C**) Shows the number of images, while (**D**) displays the corresponding probabilities.

Cross-validation analysis

To validate our model's stability and performance, we implemented a patient-level stratified 5-fold cross-validation strategy. Patients were first grouped by their clinical outcomes, and then randomly assigned to five folds while maintaining the proportion of each class (normal, suspicious, compromised) across all folds. The performance metrics for each fold and their average values are presented in Table 3. This model demonstrated
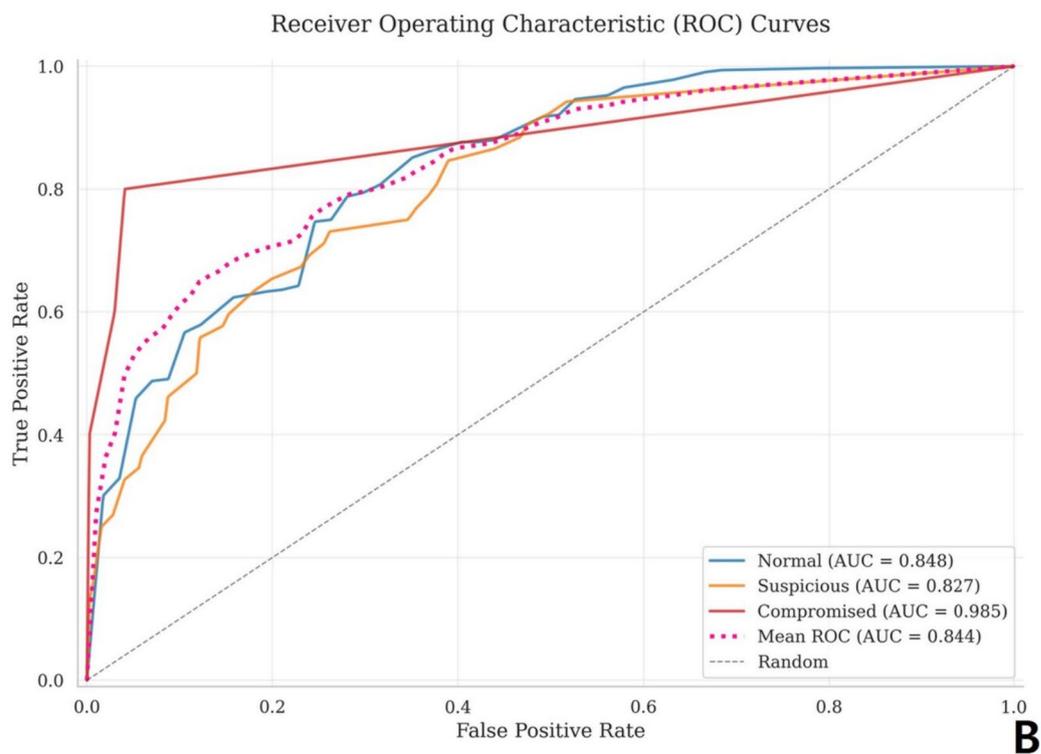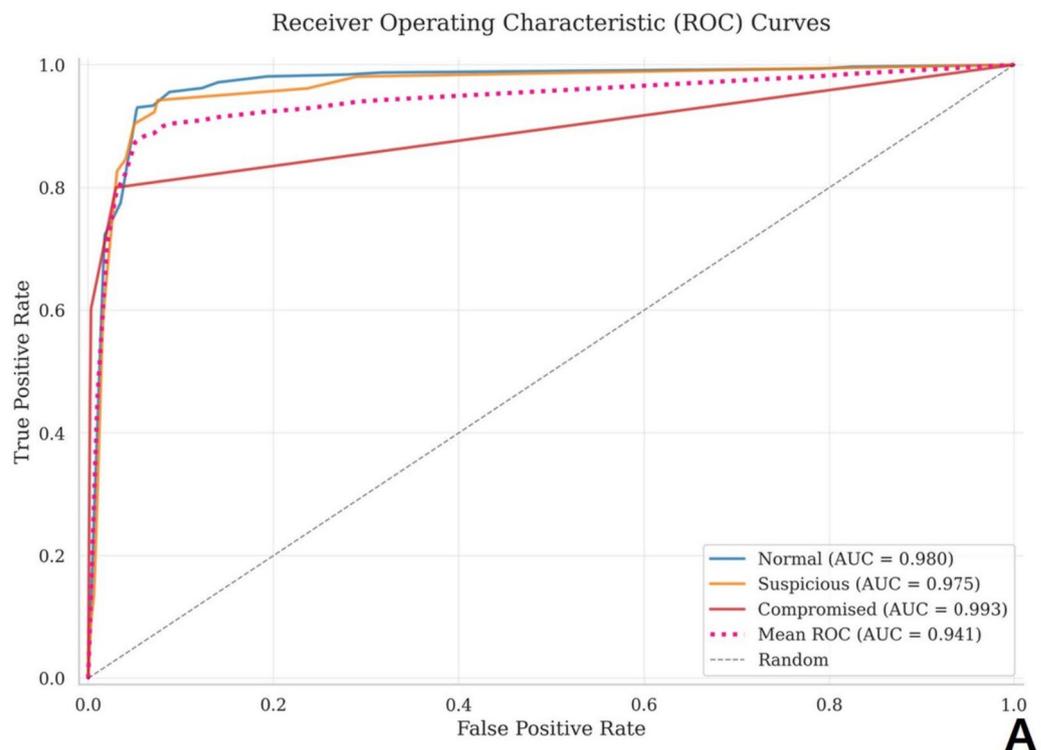
Fig. 2. ROC curves for model comparison. Model discriminative ability can be validated through the area under the curve (AUC) measurements. (**A**) ROC curves for the proposed model. AUC values for each class: Normal: 0.980, Suspicious: 0.975, Compromised: 0.993 (**B**) ROC curves for the cross-entropy loss only model. AUC values for each class: Normal = 0.848, Suspicious = 0.827, Compromised = 0.985.

Fold	Acc.	Normal			Suspicious			Compromised		
		Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Fold 1	0.954	0.971	0.988	0.980	0.667	0.421	0.516	0.750	0.818	0.783
Fold 2	0.970	0.974	1.000	0.987	0.909	0.526	0.667	0.900	0.818	0.857
Fold 3	0.970	0.985	0.983	0.984	0.737	0.737	0.737	0.917	1.000	0.957
Fold 4	0.968	0.972	1.000	0.986	0.833	0.526	0.645	1.000	0.727	0.842
Fold 5	0.989	0.994	0.994	0.994	0.895	0.895	0.895	1.000	1.000	1.000
Avg.	0.970 ± 0.011	0.979 ± 0.009	0.993 ± 0.007	0.986 ± 0.005	0.808 ± 0.093	0.621 ± 0.171	0.692 ± 0.124	0.913 ± 0.092	0.873 ± 0.109	0.888 ± 0.079

Table 3. Performance metrics across 5-fold cross-validation.

consistent performance across all folds (average accuracy: 0.9704 ± 0.01111) and showed high sensitivity in the compromised class (average recall: 0.8727 ± 0.1091). However, it exhibited relatively low performance metrics for images in the suspicious class (average F1 score: 0.6919 ± 0.1240).

Discussion

Regular monitoring at short intervals is essential after free flap reconstruction in the head and neck region. Frequent monitoring that enables early salvage procedures, particularly within 48 h, results in higher success rates compared to cases without such early intervention[9]. Given this critical time window, current monitoring approaches face several limitations. Current widely adopted monitoring methods are traditionally based on ultrasound, specialized imaging, or Doppler sound, offering limited qualitative assessment of flap changes10. Certainly, there have been studies that have incorporated quantitative measures such as surface temperature into flap monitoring. While Kraemer et al. demonstrated in 2011 that flap surface temperature measurements could detect changes, and Lee et al. developed a convolutional neural network(CNN) model using infrared camera images to diagnose surface temperature variations[11],[12].

The severe class imbalance inherent in our dataset—with compromised cases comprising only 2.7% (51/1862) of images—accurately reflects the clinical reality where free flap success rates exceed 95%[13]. While this distribution could potentially raise concerns about statistical robustness, our methodology was specifically designed to address this challenge. The dynamic triplet margin loss mechanism prioritizes learning from rare compromised cases by enforcing larger separation margins between normal and compromised states in the embedding space. This approach proved effective, as evidenced by our model achieving an F1 score of 0.73 and recall of 0.80 for the compromised class despite the severe imbalance. The clinical reliability is further demonstrated by the model's ability to detect 4 out of 5 compromised cases in the test set, comparable to human expert performance in similarly challenging scenarios.

Recent advances have explored quantitative approaches, including color tone measurement for extraoral flaps[7]. However, all of these approaches have focused on extraoral flaps. Unlike extraoral flaps, intraoral flaps are susceptible to contamination by saliva and blood, provide limited visibility making observation difficult, and are located within the confined space of the oral cavity, making it challenging to accurately measure even the surface temperature of the flap[14–16]. Additionally, the use of various capture devices by different residential staff members presents another challenge in maintaining consistency of image quality. In clinical settings, photographs are typically captured using multiple smartphone models with varying camera specifications and under inconsistent lighting conditions within the oral cavity. To address these technical limitations, our model incorporated comprehensive data augmentation techniques during the preprocessing stage, including random adjustments to brightness, contrast, saturation, and color jittering. Furthermore, normalization using the ImageNet dataset enhanced the model's resilience to variations in image acquisition conditions, enabling robust performance despite the heterogeneity of capture devices and lighting environments. In this study, we aimed to quantitatively analyze and characterize flaps using images of intraoral flaps.

Based on our previous binary classification model for intraoral free flaps, the current study addresses the clinical need for a more granular assessment of flap status[8]. Binary classification, while effective in distinguishing compromised from normal flaps, does not capture the gradual nature of tissue deterioration. In clinical practice, identifying the intermediate suspicious stage is critical for determining the optimal timing of salvage procedures. Early intervention during this transitional phase can prevent progression to complete compromise, whereas delayed action after obvious compromise may result in irreversible tissue loss. Therefore, we developed a three-tier ordinal classification system to provide clinicians with actionable information about progressive flap changes.

To address this clinical need for progressive assessment, we imployed ordinal classification techniques. In machine learning, such ordinal classification has evolved to its current performance level through neural network approaches and complex architectures since Frank and Hall established the fundamental techniques in 2001[17,18]. Our model learned the progressive changes in flap status by combining dynamic margin loss and classification loss functions. This approach enhanced predictive performance in clinical settings compared to conventional independent class classification methods by reflecting the continuous state transitions from normal to suspicious to compromised states. This can be intuitively observed in the data visualized in two-dimensional space, with the uniform manifold approximation and projection (UMAP) of the control and experimental models shown in Fig. 3.

This model misclassified 16 images out of 373 test images. Of these, 9 errors could potentially delay salvage procedures in actual clinical settings by overestimating the flap's condition relative to its true state. Specifically,
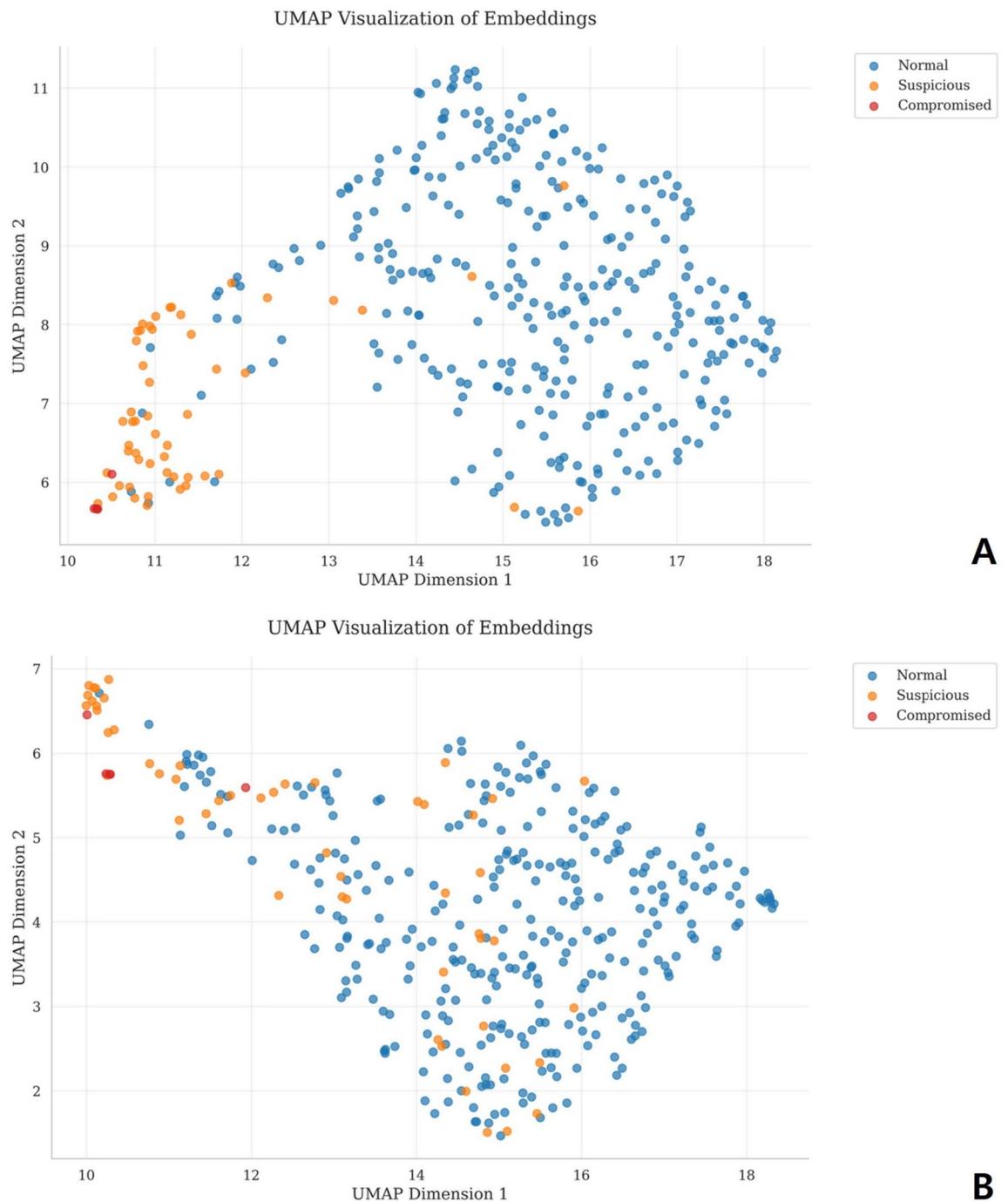
Fig. 3. The UMAP of each model. Test set data points are represented as dots, with blue indicating normal class, orange indicating suspicious class, and red indicating the compromised class. (**A**) The UMAP of our proposed model. Data points are clustered according to their ordinal relationship by class, and the distance between disparate classes (e.g., the healthy and the compromised class images) is appropriately represented. (**B**) The UMAP of the control model using only cross-entropy loss. The data points were not clustered in distinct stages; instead, they exhibited ambiguous mixing patterns.

the 8 images that the model classified as normal were actually in a state where vascular compromise was "suspicious" according to the expert, and 1 image that the model identified as suspicious was actually an image of a necrotized free flap tissue.

During the 5-fold cross-validation process, this model showed notably lower performance for the suspicious class compared to the normal and compromised classes (average F1 score: 0.6919 ± 0.1240). This is likely attributed to the difficulty in distinguishing borderline cases. The suspicious class encompasses both early-stage clinical compromise and cases where diagnostic uncertainty arises from incomplete visualization (e.g., obscured

margins), though images with fundamentally poor technical quality were excluded during preprocessing as described in Methods. Images in the suspicious class do not indicate total flap loss but are not completely healthy either, representing a transitional state that is challenging to characterize even for experienced clinicians. The accurate differentiation of these borderline cases appears to reflect the reality that this is challenging even for experienced clinicians.

Our model displays the probability of belonging to each class on the final result screen (Fig. 4). The example image shows an anterolateral free-flap reconstruction after mandibulectomy for squamous cell carcinoma in the lower gingiva, taken 48 hours after the operation. The flap's pinkish and yellowish appearance and adequate turgor led both clinical experts and our model to classify it as normal. The current implementation operates efficiently on mobile devices through external server processing, completing analysis within seconds. This represents a significant advantage in terms of convenience compared to both traditional monitoring methods requiring substantial equipment and existing deep learning approaches in flap monitoring[12,19].

Although the current model has limitations, including the use of previously captured images with inconsistent angles and lighting conditions, classification by a single expert which may introduce inter-rater variability, and restricting to Asian skin complexion, its fundamental strength lies in providing quantitative ordinal assessment of intraoral flaps using only smartphone photographs obtained during routine monitoring. While additional expert annotations could provide inter-rater reliability estimates, recent evidence suggest that expert consensus does not always outperform well trained AI models[20]. Whether future development should prioritize multi-expert labeling or model self-improvement through active learning remains an open research question. This work establishes a foundation for accessible, quantitative monitoring of intraoral flaps and highlights the potential of ordinal classification approaches in addressing real-world clinical challenges.
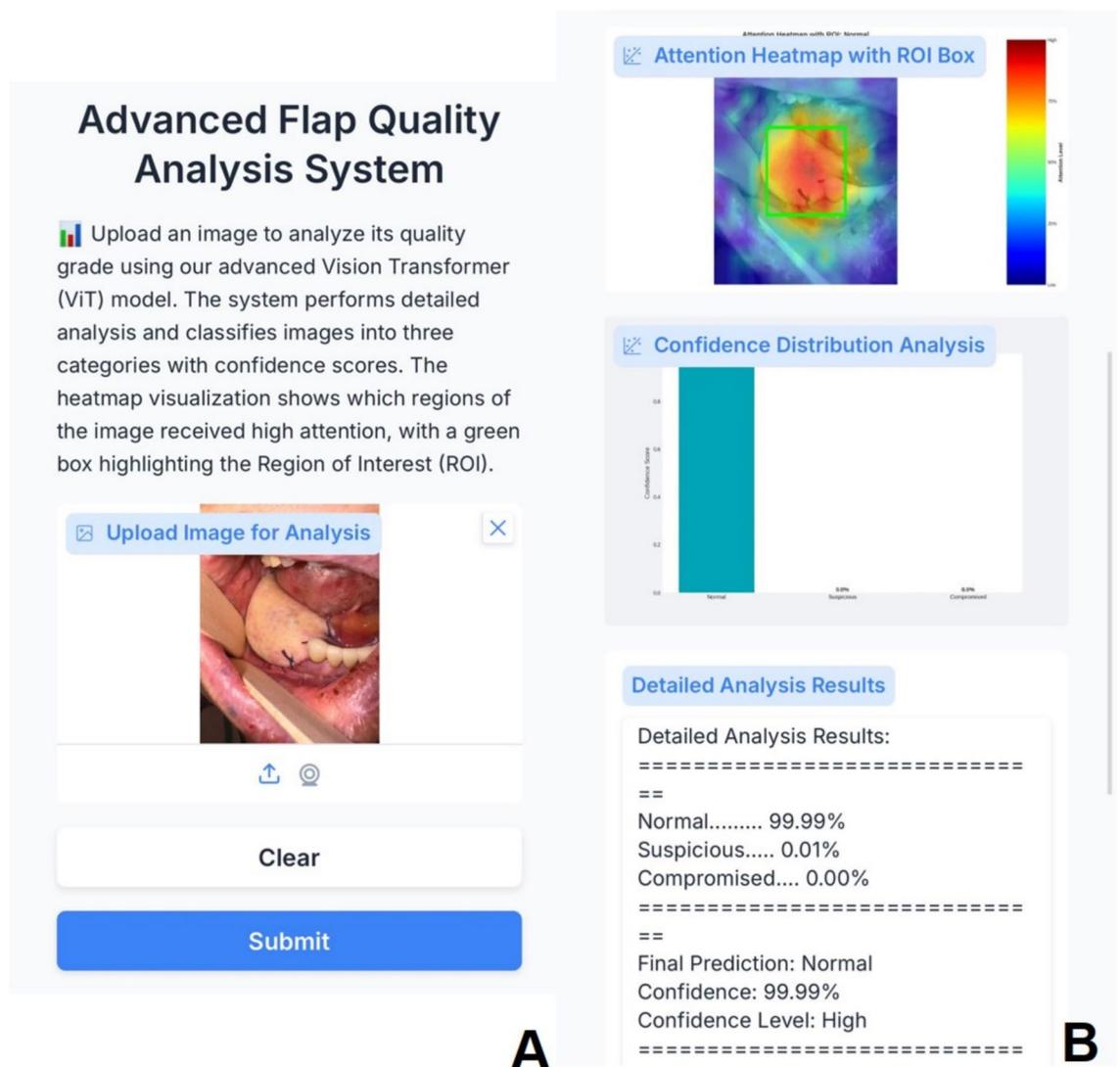

Fig. 4. Model Interface. When an image is submitted to the server as shown in (**A**), the model segments the flap within the photograph as illustrated in (**B**), and displays the probability of vascular compromise along with the most appropriate classification grade accordingly.

Criteria	Score
Color	
Flap color matches patient's tissue color or appears pinkish	0
Partial congestion or pallor of the flap	1
Complete necrosis (black appearance) or total pallor of the flap	2
Margin	
Complete visualization of flap margins with satisfactory appearance	0
Partial margin obscurement with satisfactory appearance of visible margins	1
Partial margin obscurement with congestion of necrosis in visible margins	2
Uniform necrotic appearance of flap and margins with identical coloration	3
Margins not visible for assessment	R
Turgidity	
Elasticity comparable to or greater than surrounding mucosa	0
General wrinkling or focal areas of decreased elasticity	1
Global loss of elasticity with sloughy appearance	2
Texture assessment impossible due to presence of foreign matter (blood, saliva, food debris)	R

Table 4. Flap evaluation criteria and scoring system.

Dataset	Samples	Normal	Suspicious*	Compromised†	Images that need to be reshot
Training	1,489	1,270	173	46	Not divided
Testing	373	316	52	5	Not divided
Total	1,862	1,586	225	51	

Table 5. Distribution of samples by sets and classes. * Group suspected vascular compromise based on assessment of flap color, texture, and margins. † Confirmed vascular compromise or determined as failure due to impossible salvage procedure, based on evaluation of flap color, texture, margins, and elapsed time.

Methods

Study population and materials

Between June 2021 and March 2024, a total of 207 consecutive patients underwent surgical reconstruction of intraoral postoperative defects using various free flap tissues. Among these patients, 131 provided postoperative photographs of sufficient quantity and quality, resulting in a comprehensive dataset of 1,883 sequential images for detailed analysis. Clinical photographs were obtained from the Department of Oral and Maxillofacial Surgery, Yonsei University Dental Hospital. All photographs were taken by residential staff who performed immediate flap monitoring, using various capture devices (iPhone 13 mini (Apple Inc.), iPhone 13 Pro (Apple Inc.), iPhone 15 (Apple Inc.), Galaxy S21 (Samsung Group), and Galaxy Z Flip4 (Samsung Group)). During the first 48 hours, when there is a higher probability of successful salvage procedures, images were captured at 2-hour or 3-hour intervals[21]. On the third day after surgery, the flap was observed once every 6 h, and from the fourth day post-surgery onwards, one photograph was taken daily. We collected patient data including surgery date, age at diagnosis, sex, primary site, and type of free flap. All collected photographs and corresponding patient information were anonymized before storage.

For the dataset compilation, a total of 1862 images were utilized for training after excluding 21 images that required second take due to inadequate visualization in two or more categories as outlined in Table 4. Of these, 372 images (20%) were allocated to the test set and used exclusively for external validation. The class distribution within the dataset is detailed in Table 5. To prevent data leakage and ensure robust generalization, we performed patient-level stratified splitting, ensuring that all images from the same patient were exclusively assigned to either the training set or test set, never both. The class distribution was maintained consistently between sets.

Data preparation

Flaps were evaluated and analyzed by a single specialist based on color, margin, and turgidity, and classified into three classes based on the sum of scores for each parameter (Table 4)[22–24].

Among the three categories, classification into the suspicious category required particular consideration regarding flap margin visibility. Since flap margins serve as early indicators of venous congestion immediately after surgery[24], images where margins could not be clearly visualized were classified as suspicious, even when the visible portions of the flap displayed healthy color and adequate turgidity without foreign material contamination (e.g., saliva, blood). This classification approach reflects the clinical reality that incomplete visualization prevents definitive assessment of flap viability, rather than indicating actual vascular compromise. Thus, the suspicious
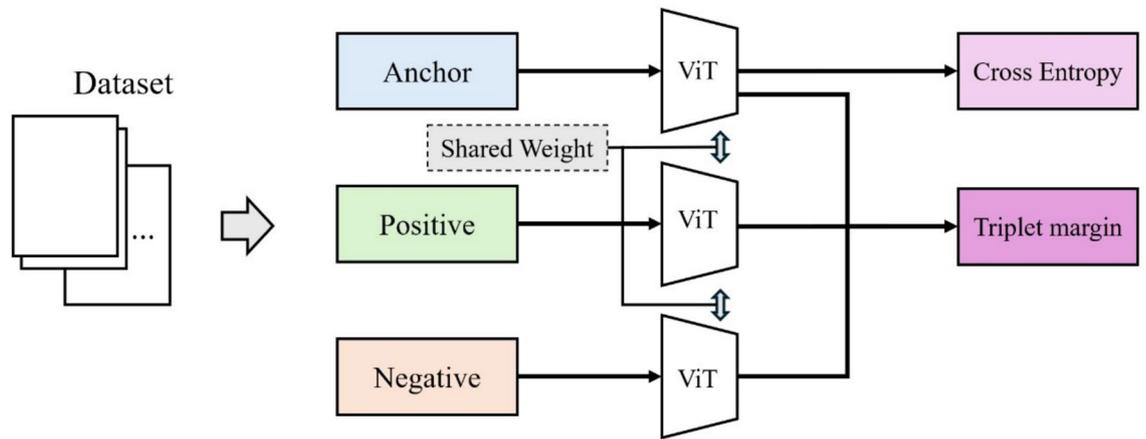
Fig. 5. Our architecture implements a Siamese Vision Transformer (ViT) that utilizes weight sharing across three parallel branches to process anchor, positive, and negative samples. The learning framework incorporates a dual-objective strategy, combining cross-entropy loss for anchor sample classification with a dynamic triplet margin loss applied to embeddings. Through weight sharing, the network ensures consistent feature extraction while maintaining computational efficiency. This unified approach facilitates the simultaneous learning of discriminative classification features and metric relationships between samples of varying quality grades, while the dynamic triplet margin mechanism adaptively adjusts separation requirements to address class imbalance and maintain optimal embedding space distributions.

category includes cases with actual or potential changes, as well as those reflecting diagnostic uncertainty due to restricted visibility.

Following classification, the image preparation process was implemented through four approaches. All images were resized to 224×224 pixels[25], randomly flipped with a probability of 0.5 to introduce reflection invariance[26], and randomly rotated by 10 degrees to enhance rotational invariance[27]. Since various devices were used to capture the images and lighting conditions varied due to the intraoral photography, these conditions were randomly adjusted[28]. This approach prevents overfitting, improves model generalization, and enhances the model's resilience when analyzing photographs of free flap tissue located within the oral cavity.

Conventional transformations, such as random rotations, flips, and color jitter, were combined with domain-specific techniques. These augmentations effectively expanded our training set, enhancing the model's generalization capability.

Model development

Conventional multiclass grading systems typically treat each class as an independent category. In our approach, we incorporate a contrastive classification learning framework to simultaneously learn both ordinal relationships between grades and discrete class boundaries. The model was implemented using PyTorch and trained on a single NVIDIA RTX 4090 GPU.

A dynamic margin mechanism and multi-task learning framework were employed to ensure that the inter-sample distances in the embedding space are proportional to grade differences, enabling the model to learn both continuous quality spectra and discrete decision boundaries. Furthermore, a Siamese network architecture based on Vision Transformers was developed that effectively processes image triplets while maintaining parameter efficiency through weight sharing (Fig. 5).

Transfer learning was utilized by initializing the model with weights from large-scale pretrained networks (ViT-large, patch16-224 model). Furthermore, the model architecture and loss functions were refined to address the ordinal nature of the classification task. By incorporating a triplet margin loss with a dynamic margin mechanism, clear class boundaries were maintained while capturing progressive transitions between grades. This specialized loss function design can robustly handle the class imbalance problem that commonly occurs in medical data. We have previously demonstrated this in our binary classification study[8].

The final loss function (L_{total}) is formulated as a combination of dynamic triplet loss and cross-entropy classification loss as defined below:

$$L_{total} = L_{triplet}(e_a, e_p, e_n) + \lambda L_{CE}(y_a, l_a) \tag{1}$$

The ($L_{triplet}$) and y_a, l_a are defined as follows:

$$L_{triplet}(e_a, e_p, e_n) = \max(0, \|e_a - e_p\|_2 - \|e_a - e_n\|_2 + m(l_a, l_n)) \tag{2}$$

$$m(l_a, l_n) = \begin{cases} m_b \cdot m_m & \text{if } |l_a - l_n| \geq 2 \\ m_b & \text{otherwise} \end{cases} \tag{3}$$

$$y_a = c_\psi(h_a) \tag{4}$$

Further, where y_a represents the class logits for the anchor sample, and e_i and h_i are defined as follows:

$$e_i = g_\phi(h_i) \text{ where } i \in \{a, p, n\} \tag{5}$$

$$h_i = f_\theta(x_i) \text{ where } i \in \{a, p, n\} \tag{6}$$

Data availability

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

References

1. Spindler, N. et al. Free flap reconstruction of the extremities in patients who are? 65 years old: a single-center retrospective 1-to-1 matched analysis. *Clin. Intervent. Aging* **2021**, 497–503 (2021).
2. Lee, T.-Y., Lee, S. & Eun, S. The free flap reconstruction of facial defects after squamous cell carcinoma excision. *Medicina* **60**, 1432 (2024).
3. Mücke, T. et al. A comparative analysis using flowmeter, laser-doppler spectrophotometry, and indocyanine green-videoangiography for detection of vascular stenosis in free flaps. *Sci. Rep.* **10**, 939 (2020).
4. Schraven, S. P. et al. Continuous intraoperative perfusion monitoring of free microvascular anastomosed fasciocutaneous flaps using remote photoplethysmography. *Sci. Rep.* **13**, 1532 (2023).
5. Moreno-Oyervides, A. et al. Design and testing of an optical instrument for skin flap monitoring. *Sci. Rep.* **13**, 16778 (2023).
6. Patel, U. A. et al. Free flap reconstruction monitoring techniques and frequency in the era of restricted resident work hours. *JAMA Otolaryngol. Head Neck Surg.* **143**, 803–809 (2017).
7. Huang, R.-W. et al. Reliability of postoperative free flap monitoring with a novel prediction model based on supervised machine learning. *Plast. Reconstr. Surg.* **152**, 943e–952e (2023).
8. Kim, H., Kim, D. & Bai, J. Machine learning approaches overcome imbalanced clinical data for intraoral free flap monitoring. *Sci. Rep.* **15**, 34849 (2025).
9. Novakovic, D., Patel, R. S., Goldstein, D. P. & Gullane, P. J. Salvage of failed free flaps used in head and neck reconstruction. *Head Neck Oncol.* **1**, 1–5 (2009).
10. Rogoń, I. et al. Flap monitoring techniques: a review. *J. Clin. Med.* **13**, 5467 (2024).
11. Kraemer, R. et al. Free flap microcirculatory monitoring correlates to free flap temperature assessment. *J. Plastic Reconstruct. Aesthetic Surg.* **64**, 1353–1358 (2011).
12. Lee, C.-E., Chen, C.-M., Wang, H.-X., Chen, L.-W. & Perng, C.-K. Utilizing mask rcnn for monitoring postoperative free flap: circulatory compromise detection based on visible-light and infrared images. *IEEE Access* **10**, 109510–109525 (2022).
13. Pohlenz, P. et al. Microvascular free flaps in head and neck surgery: complications and outcome of 1000 flaps. *Int. J. Oral Maxillofac. Surg.* **41**, 739–743 (2012).
14. Nielsen, H. T., Gutberg, N. & Birke-Sorensen, H. Monitoring of intraoral free flaps with microdialysis. *Br. J. Oral Maxillofac. Surg.* **49**, 521–526 (2011).
15. Lovětínská, V., Sukop, A., Klein, L. & Brandejsova, A. Free-flap monitoring: review and clinical approach. *Acta Chir. Plast.* **61**, 16–23 (2020).
16. Felicio-Briegel, A. et al. Hyperspectral imaging for monitoring of free flaps of the oral cavity: a feasibility study. *Lasers Surg. Med.* **56**, 165–174 (2024).
17. Frank, E. & Hall, M. A simple approach to ordinal classification. In *Machine Learning: ECML 2001: 12th European Conference on Machine Learning Freiburg, Germany, September 5–7, 2001 Proceedings 12* 145–156 (Springer, 2001).
18. Liu, J., Chang, W.-C., Wu, Y. & Yang, Y. Deep learning for extreme multi-label text classification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* 115–124 (2017).
19. Chae, M. P. et al. Current evidence for postoperative monitoring of microvascular free flaps: a systematic review. *Ann. Plast. Surg.* **74**, 621–632 (2015).
20. Nori, H. et al. Sequential diagnosis with language models. arXiv preprint arXiv:2506.22405 (2025).
21. Shen, A. Y. et al. Free flap monitoring, salvage, and failure timing: a systematic review. *J. Reconstr. Microsurg.* **37**, 300–308 (2021).
22. Knoedler, S. et al. Postoperative free flap monitoring in reconstructive surgery—man or machine?. *Front. Surg.* **10**, 1130566 (2023).
23. Smit, J. M., Zeebregts, C. J., Acosta, R. & Werker, P. M. Advancements in free flap monitoring in the last decade: a critical review. *Plast. Reconstr. Surg.* **125**, 177–185 (2010).
24. Liu, Y. et al. Analysis of 13 cases of venous compromise in 178 radial forearm free flaps for intraoral reconstruction. *Int. J. Oral Maxillofac. Surg.* **41**, 448–452 (2012).
25. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 256 (2012).
26. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
27. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. & Le, Q. V. Autoaugment: learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 113–123 (2019).
28. Berggren, A., Weiland, A. J. & Dorfman, H. The effect of prolonged ischemia time on osteocyte and osteoblast survival in composite bone grafts revascularized by microvascular anastomoses. *Plast. Reconstr. Surg.* **69**, 290–298 (1982).

Author contributions

H.K. and D.K. conceived and designed the experiments. H.K. and J.B. interpreted the data. H.K. drafted the work, and D.K. and J.B. substantively revised the manuscript. All authors approved the submitted version and agreed to be personally accountable for their own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even those in which they were not personally involved, were appropriately investigated, resolved, and documented in the literature.

Funding

Competing interests

The authors declare no competing interests.

Ethics approval

This retrospective study was approved by the Institutional Review Board of Yonsei University College of Dentistry, Dental Hospital (IRB No.:2-2024-0008) with a waiver of informed consent. The study was registered with the Clinical Research Information Service (CRiS, Registration No.: KCT0009962). All methods were carried out in accordance with relevant guidelines and regulations and the Declaration of Helsinki.

Additional information

Correspondence and requests for materials should be addressed to D.K. or J.B.

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.