



# OPEN Machine learning approaches overcome imbalanced clinical data for intraoral free flap monitoring

Hyounmin Kim<sup>1,4</sup>, Dongwook Kim<sup>1,2,4</sup>✉ & Juho Bai<sup>3</sup>✉

Free flap reconstruction is essential for treating intraoral defects; however, failure can lead to complex and prolonged complications. While various monitoring methods have been employed to prevent such situations, they are qualitative and sometimes unfamiliar to novices. The purpose of this study was to develop a user-friendly model using artificial intelligence that quantitatively represents flap status. We analyzed 1877 images from 131 patients who underwent free flap reconstruction for intraoral defects between June 2021 and March 2024. Since patients with vascular damage were very few in number, class weighting and focal loss techniques were used to address this imbalance. The proposed model achieved high overall accuracy and F1 scores of 0.9867 and 0.9863, respectively. This study introduces the first deep learning model for intraoral flaps and demonstrates the possibility of quantitative measurement of flap changes. This tool can assist surgeons in making timely decisions regarding salvage procedures and facilitate easier monitoring for resident care-givers.

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

Free flap reconstruction is a common procedure used to reconstruct large defects resulting from tumor resection or trauma<sup>1</sup>. In the oral and maxillofacial region, flaps play a crucial role both functionally and aesthetically; consequently, flap necrosis can lead to challenging medical situations<sup>2,3</sup>. To prevent such situations, postoperative flap monitoring is essential and various device-based monitoring methods have been introduced<sup>4–6</sup>. However, these monitoring methods still rely on multiple observations from personnel who are not constantly present at the patient's bedside, such as residents<sup>7,8</sup>. Moreover, the resulting data, often qualitative images, can be difficult to interpret and may not provide conclusive evidence to the operating surgeon.

To resolve these difficulties, artificial intelligence (AI) models have been applied to flap monitoring in extraoral region<sup>9–12</sup>. However, these approaches often struggled to achieve robust performance while addressing the class imbalance inherent in clinical settings, or they required specialized imaging equipment. Our research utilizes photographs obtained from patients already under observation, incorporating images from diverse anatomical sites and under varying conditions, with the aim of developing a model that can be broadly implemented in real-world clinical scenarios.

## Results

### Patients characteristics

Medical records, including clinical photographs, of 131 patients were reviewed. The most common case was the reconstruction of glossectomy defects using anterolateral thigh (ALT) free flaps. Among the 131 patients, 8 underwent salvage procedures due to vascular compromise and late failure leading to total flap loss was observed in 2 patients. Details are presented in Table 1.

### Dataset details

A total of 1877 images obtained from these patients were divided into training and test sets. Only the training set was used for model learning, while the test set was utilized solely for validation. The data were split in an 8: 2 ratio, resulting in 1501 samples in the training set and 376 samples in the test set.

Images were classified into two classes based on the presence or absence of vascular compromise in the flap. Class 1 represented a set of images with confirmed vascular compromise, which was verified either through

<sup>1</sup>Department of Oral and Maxillofacial Surgery, Yonsei University College of Dentistry, Seoul 03722, Republic of Korea. <sup>2</sup>Oral Science Research Center, Yonsei University College of Dentistry, Seoul 03722, Republic of Korea.

<sup>3</sup>Department of Global Business and Technology, Hankuk University of Foreign Studies, Yongin 17035, Republic of Korea. <sup>4</sup>Hyounmin Kim and Dongwook Kim contributed equally to this work. ✉email: dwkimomfs@yuhs.ac; juho@hufs.ac.kr

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
Free flap	
Anterolateral thigh free flap	64 (48.9%)
Fibula free flap	45 (34.3%)
Radial forearm free flap	22 (16.8%)
Compromised patients	10 (7.6%)
Success in flap salvage	8
Arterial insufficiency	4
Venous thrombosis <sup>†</sup>	4
Total flap loss	2
Arterial insufficiency	2

**Table 1.** Clinical characteristic of the study patients. \*This includes all cases where both the lower gingiva and mandible bone were reconstructed. <sup>†</sup>This includes both cases where there was a problem with the anastomosis as well as cases where there was a venous compromise due to kinking

Data set	Samples	Classification	
		Class 0 (Absence of vascular compromise)	Class 1 (Presence of vascular compromise)
Training	1501	1404 (93.5%)	97 (6.4%)
Testing	376	352 (93.6%)	24 (6.4%)
Total	1877	1756 (93.6%)	121 (6.4%)

**Table 2.** Distribution of the images used for training and tests.

the patient undergoing a repeat salvage procedure or through confirmation of flap necrosis as evidenced by comprehensive color change and complete alteration in turgidity. All other flaps that remained normal and viable were categorized as Class 0. In patients who underwent re-entry for salvage procedures due to detected vascular compromise, images taken after the second vascular anastomosis period were not collected.

The samples were randomly selected for each class, ensuring similar class distributions in both the test and training sets. This information is summarized in Table 2.

### Clinical application

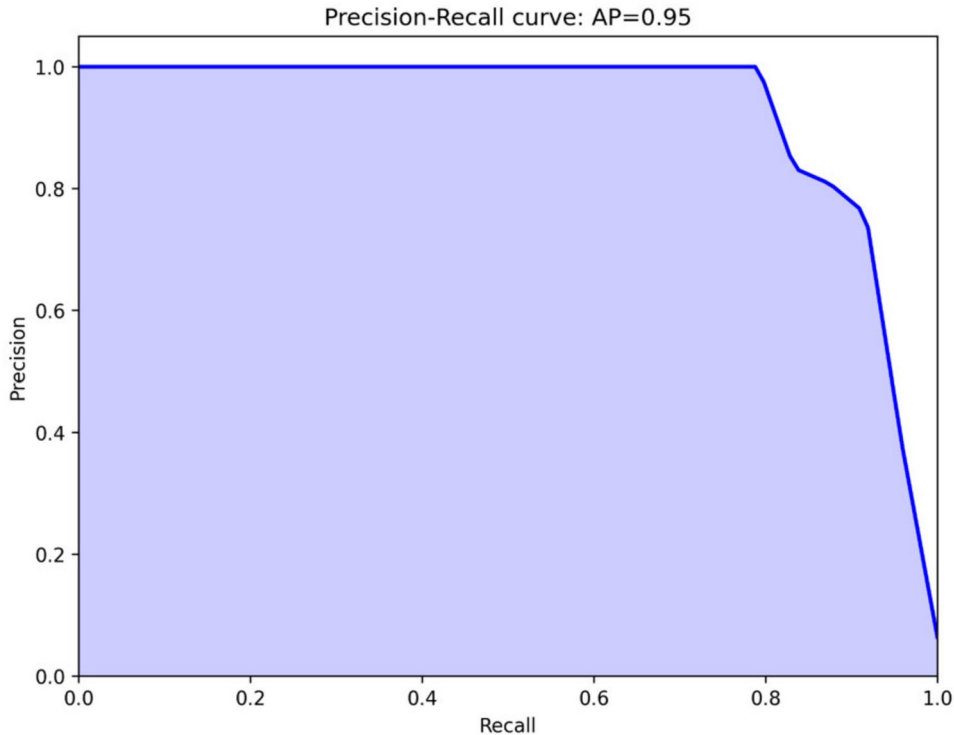
Of the 376 images, 352 represented healthy flaps without any vascular compromise, partial flap loss. The remaining 24 images depicted flaps that had experienced venous congestion, arterial insufficiency. The model correctly classified 351 of 352 normal flap images as class 0, and identified 20 of 24 unhealthy flap images that required salvage procedures.

Model name	Class weighting	Loss function
CE loss without weight	No	Cross-Entropy
Weighted CE loss	Yes	Cross-Entropy
Focal loss without weight	No	Focal Loss ( $\gamma = 2$ )
Weighted focal loss (proposed)	Yes	Focal Loss ( $\gamma = 2$ )

**Table 3.** The four experimental configurations.

Model configuration	Accuracy	F1 score
Proposed (weighted focal loss)	0.9867	0.9863
Focal loss without weight	0.9761	0.9741
Weighted cross-entropy loss	0.9787	0.9783
CE loss without weight	0.9814	0.9804

**Table 4.** Overall performance.



**Fig. 1.** The precision-recall curve maintains high precision (close to 1.0) for a wide range of recall values, only dropping sharply at very high recall levels. This demonstrates the model’s ability to effectively identify class 1 flaps (those requiring salvage procedure) with minimal errors. The combination of class weighting and Focal loss appears to effectively address the class imbalance problem.

**Model performance**

We previously explored four distinct model configurations, as shown in Table 3, considering the application of class weighting and the choice of loss function. All configurations utilized the Vision Transformer (ViT) large model as the common architecture. The learning rate was set to 0.0001 across all experiments, with other hyperparameters following the default recommendations for Vision Transformers.

The models were assessed using accuracy, F1 score, precision, recall, and class-specific metrics. Table 4 summarizes the overall performance metrics for each model.

The overall accuracy of the proposed model across all classes was 0.9867, with an F1 score of 0.9863, indicating exceptionally high performance among the four models. The corresponding precision-recall curve is presented in Fig. 1.

Model configuration	Class	Precision	Recall	F1 score
Proposed (weighted Focal loss)	0	0.99	1.00	0.99
	1	<b>0.95</b>	<b>0.83</b>	<b>0.89</b>
Focal loss without Weight	0	0.98	1.00	0.99
	1	0.94	0.67	0.78
Weighted Cross-entropy loss	0	0.99	0.99	0.99
	1	0.86	0.79	0.83
CE loss without Weight	0	0.98	1.00	0.99
	1	0.95	0.75	0.84

**Table 5.** Class-specific performance-metrics table for all models. The results of the proposed model are shown in bold.

Model configuration	Accuracy	Precision	Recall	F1 score
Proposed (weighted focal loss)	<b>0.9379 ± 0.0173</b>	<b>0.9421 ± 0.0152</b>	<b>0.9379 ± 0.0173</b>	<b>0.9080 ± 0.0253</b>
Focal loss without weight	0.8956 ± 0.0229	0.8987 ± 0.0185	0.8956 ± 0.0229	0.8962 ± 0.0099
Weighted cross-entropy loss	0.8634 ± 0.0748	0.9039 ± 0.0221	0.8634 ± 0.0748	0.8789 ± 0.0524
CE loss without weight	0.9339 ± 0.0197	0.9187 ± 0.0380	0.9339 ± 0.0197	0.9060 ± 0.0263

**Table 6.** Patient-level 5-fold cross-validation results for all model configurations. The results of the proposed model are shown in bold.

As expected, the proposed model among 4 models performed exceptionally well on the majority class 0 samples. Notably, it also achieved a precision of 0.95 for the minority class 1 samples, indicating superior performance compared to models without weighting or those using simple cross-entropy (CE) loss. The class-specific performance metrics of all models are presented in Table 5.

Patient-level cross-validation for model validation

We initially evaluated our model using an 80:20 train-test split at the image level. To more rigorously validate the model’s clinical applicability, we additionally performed 5-fold cross-validation at the patient level. In this approach, 131 patients were randomly divided into 5 groups, ensuring that all images from the same patient were placed exclusively in either the training or testing set.

The results of patient-level cross-validation are summarized in Table 6. The weighted focal loss model demonstrated superior performance with an average accuracy of 0.9379 ± 0.0173 and precision of 0.9421 ± 0.0152. For comparison, the CE loss without weight model achieved an accuracy of 0.9339 ± 0.0197, while focal loss without weight and weighted CE loss models showed lower performance with accuracies of 0.8956 ± 0.0229 and 0.8634 ± 0.0748, respectively.

Furthermore, our proposed weighted focal loss model achieved the highest F1 score (0.9080 ± 0.0253) among all configurations, confirming its effectiveness in handling the imbalanced class distribution present in this clinical dataset.

Discussion

Artificial intelligence has revolutionized medical imaging, particularly in diagnosis and segmentation. Numerous studies have employed artificial neural networks to identify diseases and predict prognosis in various fields, including neurosurgery and plastic surgery<sup>13–16</sup>. Subsequently, convolutional neural network (CNN) models were developed to recognize abnormal regions, diagnose conditions, and predict outcomes from clinical photographs<sup>17,18</sup>.

Mantelakis et al. provided a comprehensive review of artificial intelligence applications in plastic surgery, ranging from early non-neural network models to contemporary CNN approaches. They reported that AI systems analyzing visual images for lesion assessment and treatment planning demonstrated remarkably high accuracy<sup>19</sup>.

Free flap reconstruction is a crucial treatment modality for both functional and aesthetic perspective following tumor resection or trauma<sup>20</sup>. Although this procedure has a success rate that exceeds 95%, rare instances of vascular compromise can lead to complicated results<sup>2,3</sup>. Such complications can often be averted through frequent monitoring and prompt salvage procedures<sup>4–6</sup>.

Current monitoring methods include traditional observation, Doppler systems, color duplex sonography, and near-infrared spectroscopy. Despite their popularity, each of these methods has critical limitations: they may produce meaningless signals, prove challenging for inexperienced resident staff to interpret, or involve prohibitively large observational devices that preclude frequent use<sup>21,22</sup>. Moreover, none of the existing monitoring methods provide quantifiable measurements of flap changes<sup>6</sup>.

In response to these limitations, Hsu et al. developed a supervised learning approach to generate quantified results for extraoral flap changes<sup>12</sup>. However, to date, no such attempt has been made for monitoring free

flaps inserted inside the oral cavity. Monitoring flaps located within the oral cavity presents greater challenges compared to observing flaps positioned in external sites. Despite regular oral cleaning procedures, the flap is often covered by blood mixed with saliva, making complete visualization difficult. Additionally, the complex anatomical structure of the oral cavity can make the photographic documentation itself challenging<sup>23,24</sup>. Furthermore, some of the aforementioned monitoring devices are entirely inapplicable within the oral cavity.

Our model utilizes only 2D images of the flap, eliminating the discomfort associated with inserting large probes into the oral cavity. While the cave-like structure of the oral cavity inevitably leads to significant lighting variations between images, this challenge was overcome through random adjustments within the model. Consequently, only a smartphone flash was required for image capture.

Given that the success rate of vascular anastomosis exceeds 95%, only 10 out of 131 patients in our study underwent salvage procedures or experienced total flap loss. We addressed the inherent dataset imbalance by applying class weighting and incorporating focal loss during the training process. The combination of class weighting and focal loss presents a synergistic approach to addressing class imbalance in machine learning models. Class weighting mitigates the under representation of minority classes by assigning them higher importance, while focal loss dynamically adjusts the learning process to focus on hard-to-classify samples. This integrated method enhances the model's ability to learn discriminative features from underrepresented classes and challenging instances, thereby improving overall classification performance. Such an approach is particularly effective in highly skewed datasets, where it can significantly boost the model's generalization capabilities and increase the recognition rate of critical minority classes encountered in real-world applications.

Despite the presence of complex structures such as teeth, lips, and tongue in the images, often obscured by blood and saliva, our model successfully recognizes the flap and detects changes in its viability. This capability persists even when these structures deviate from their normal appearance due to surgical intervention and postoperative conditions.

The model's ability to quantify changes allows for more sensitive detection of alterations compared to human observers. As illustrated in Fig. 2, two input images were captured with a 3-h interval (A: 2 h post-initial salvage procedure, B: 5 h post-initial salvage procedure). While the second image shows a slight increase in the congestive margin compared to the first, the difference was subtle enough that clinicians deemed the flaps similar and deferred further salvage procedures. However, the model indicated a significant change, estimating a 1.3% probability of vascular compromise in the flap in the first image, which increased nearly tenfold to 13% in the second image. Subsequently, this patient underwent a second salvage procedure 15 h after the initial one. Despite these interventions, the vascular compromise remained unresolved, ultimately necessitating flap removal.

Despite the model's overall fine performance (F1 score = 0.9863), the relatively low recall value (0.83) observed for class 1 requires improvement. This indicates that the model may incorrectly classify actual cases of vascular compromise as normal with approximately 17% probability, which could potentially lead to significant clinical risks in practice. These errors primarily stem from the limited dataset, with only a few hundred images representing vascular compromise, as well as suboptimal configuration of the confidence level threshold for



**Fig. 2.** Input images from a 58-year-old male patient who underwent glossectomy and pharyngectomy for oropharyngeal cancer, with the defect repaired using an anterolateral thigh flap. The patient exhibited congestive symptoms following the initial surgery, prompting vein re-anastomosis after 11 h. The image on the left shows the model's analysis 2 h after re-anastomosis, while the image on the right represents the analysis 5 h after re-anastomosis.

determining vascular compromise. As previously described, the current model does not directly print class 1 when the confidence level falls below 50%; however, it does indicate potential vascular disorders through increased confidence levels. Therefore, prior to clinical deployment, it appears necessary to recalibrate the confidence level thresholds based on images where blood flow disorders are suspected. In clinical settings, where missing a compromised flap (false negative) could lead to irreversible tissue loss, we would recommend setting a lower threshold for flagging potential compromise, accepting a higher false positive rate as a reasonable trade-off for enhanced patient safety. A clinically acceptable recall should approach 0.95 or higher, given the severe consequences of missed vascular compromise.

In the test set of 376 images, the model misclassified only five. One healthy flap was incorrectly classified as necrotic (false negative), while four flaps with vascular compromise were misclassified as healthy (false positives). The false-negative case involved a glossectomy defect repaired with an ALT free flap, where the contralateral tongue appeared necrotic. The model mistakenly identified this darkened tongue as the flap due to similar proportions in the image. All four false positive cases exhibited vascular compromise. In three of these cases, the flap occupied only a small portion of the image, and the majority of the photograph consisted of normal tissue, which likely contributed to the model's misclassification. The remaining case involved an image taken 12 h after a forearm flap was transplanted to the buccal cheek. In this image, the color and margins of the flap closely resembled those of a healthy flap, leading the model to diagnose it as having no vascular compromise. However, by postoperative day four, the flap showed signs of necrosis, including partial skin sloughing.

This model, initially developed for diagnostic purposes, demonstrates potential as a sensitive detector of vascular compromise in flaps, serving as an assistant to surgeons. With the capability to analyze images taken under various conditions and lighting, it could provide round-the-clock, location-independent results, supporting decision-making processes for post-operative surgeons.

However, it is widely acknowledged that AI should augment, not replace, the surgeon's decision-making process<sup>20</sup>. As the model quantifies the probability of vascular compromise rather than dictating treatment timing, clinical judgment remains crucial in determining the threshold for initiating salvage procedures.

This study presents several areas for potential improvement. While the model currently provides a binary classification based on the probability of vascular compromise, future iterations could incorporate expert annotations to offer more nuanced explanations for the flap's apparent condition, potentially providing clinicians with additional insights for treatment decisions. Moreover, this study was conducted at a single institution with patients of a single nationality, limiting the patient cohort to Asian ethnicity. To expand the model's applicability, further training on more diverse skin complexions would be necessary.

Objectively, our model demonstrates high performance in identifying flaps and assessing their viability. Notably, this represents the first deep learning model developed for intraoral flap monitoring.

## Methods

### Ethics declaration

This retrospective study was waived for the requirement of informed consent from all subjects by the Institutional Review Board (IRB) of Dental Hospital of Yonsei University, College of Dentistry, and the experimental protocol of this study was approved by the IRB of Dental Hospital of Yonsei University, College of Dentistry (Approval number: 2-2024-0008). All methods were carried out in accordance with relevant guidelines and regulations of Dental Hospital of Yonsei University, College of Dentistry.

### Study population

From June 2021 to March 2024, 207 consecutive patients underwent surgery to reconstruct intraoral postoperative defects using various free flaps. Among these, only 131 patients provided a sufficient quantity and appropriate quality of sequential clinical photographs, yielding a total of 1877 images for analysis. Following established protocols, clinical observations focused on flap color, temperature, capillary refill, and turgidity. In cases of suspected arterial insufficiency, a pin-prick test was performed. Indications for salvage procedure included noticeable changes in flap color compared to previous observations, widening of congestive margins, absence of blood flow in the pin-prick test, or alterations in flap turgidity.

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 h, when there is a higher probability of successful salvage procedures, images were captured at 2-h or 3-h intervals<sup>6</sup>. 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. The photographs were categorized by case, date, and time of capture.

### Data preparation

Due to the variety of devices used to take the photographs, a standardized pre-processing pipeline was implemented: resizing, random horizontal flipping, random rotation, color jittering, and normalization. These processes help overcome limitations in color tonality that might arise from restricting the dataset to a single population, as well as the lighting inconsistencies characteristic of the oral cavity when flaps are located in deeper regions. Detailed explanations of each method are provided below.

1. Resizing: All images were maintained at  $224 \times 224$  pixels while preserving aspect ratios<sup>25</sup>.
2. Random horizontal flipping: Images were horizontally flipped with a 0.5 probability to introduce reflection invariance<sup>26</sup>.
3. Random rotation: Images were rotated by a random angle within  $\pm 10$  degrees to enhance rotational invariance<sup>27</sup>.

4. Color jittering: Random adjustments to brightness, contrast, saturation, and hue were applied to improve model resilience to variations in lighting conditions and color distribution<sup>28</sup>.
5. Normalization: Recently, Huang et al. pointed out in their 2023 study that even within a single image capturing a flap, there can be differences in RGB values between the flap tissue and the original skin. The research team therefore underwent a normalization process during pre-processing, utilizing images from the ImageNet dataset, which can enhance the diversity of skin complexion that would otherwise be limited by including only a single ethnicity<sup>29,30</sup>.

### Model development

We employed an AMD 7500F CPU and an RTX 4090 with 24GB RAM for training. The pre-trained model utilized was the vit\_l\_16 model with default parameter values.

To address the class imbalance in our prepared dataset, we implemented a sample weighting strategy<sup>31</sup>. We define the weight  $w_i$  for each class using the total number of samples (N), the number of classes (C), and the number of samples in class  $i$   $n_i$  as follows<sup>32</sup>.

$$w_i = N / (C * n_i) \quad (1)$$

Using this, we defined an adjusted weight  $w'_1$  for the minority class. In our implementation,  $\beta = 2$ .

$$w'_1 = \beta \cdot w_1 \quad (2)$$

This additional boost emphasizes the importance of minority class samples during training<sup>33</sup>.

To further address imbalance during training, we employed the Focal Loss technique at the loss function level. Using a balance coefficient  $\alpha_i$  for class  $i$  and a focusing parameter  $\gamma$ , we defined the Focal Loss (FL) as:

$$FL(y, \hat{y}) = -\alpha_i(1 - \hat{y})^\gamma y \log(\hat{y}) - (1 - \alpha_i)(\hat{y})^\gamma (1 - y) \log(1 - \hat{y}) \quad (3)$$

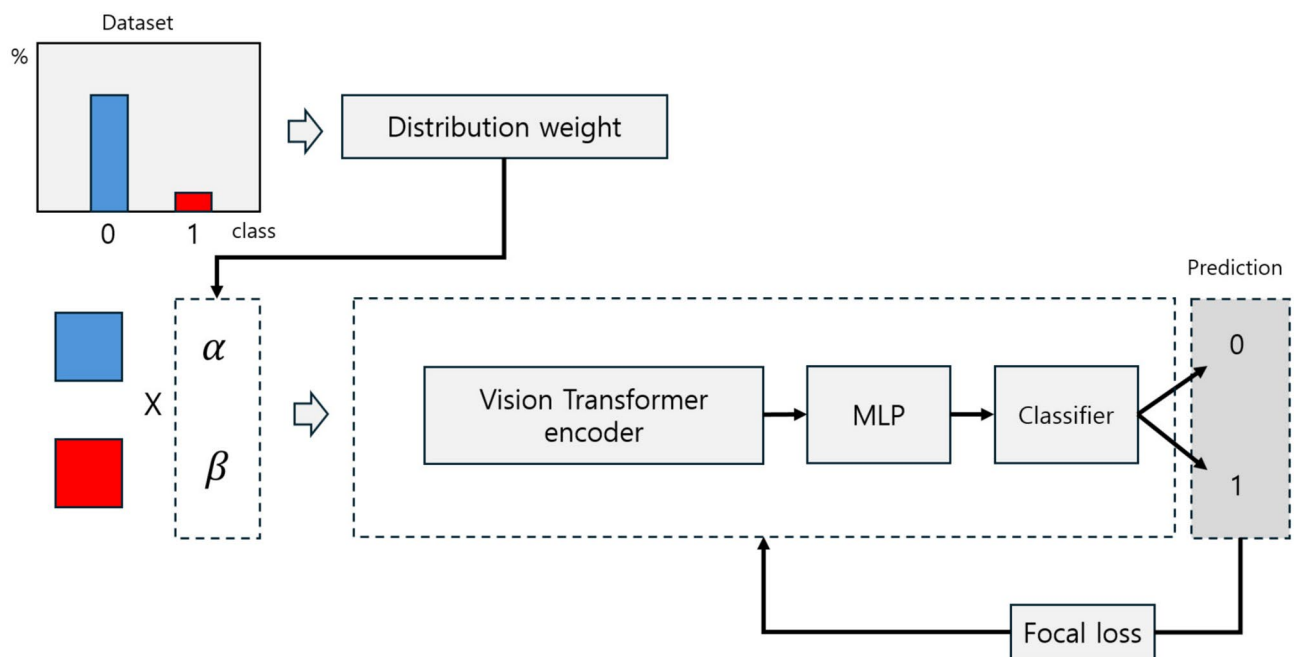
$\alpha_i$  is 0.25 and  $\gamma$  is 2 in our implementation<sup>34</sup>.

This approach helps to “reduce the loss contribution from easy examples (where  $y$  is close to  $\hat{y}$ ) and increase the importance of correcting misclassified examples”. The overall training architecture is briefly illustrated in Fig. 3.

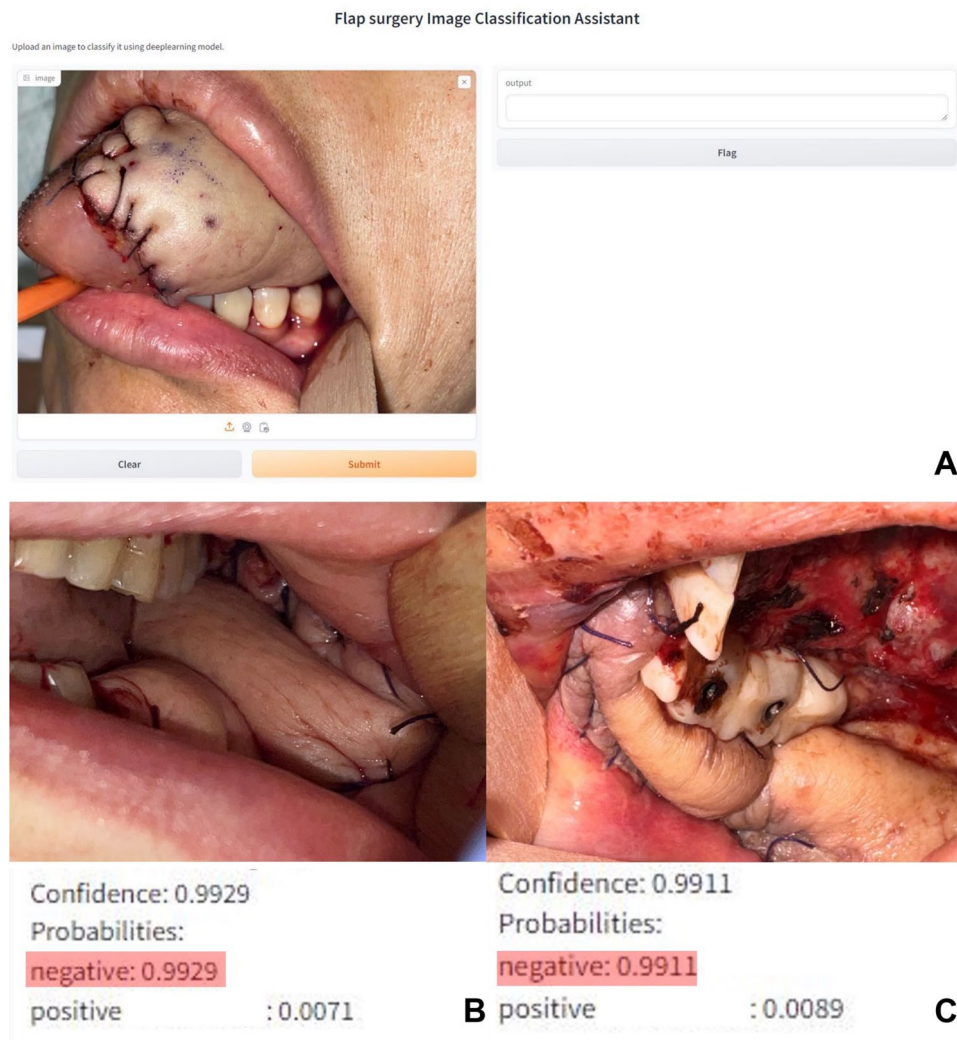
### Clinical application

The 376 images used for testing were input into the model, which output a confidence level representing the probability of vascular compromise (Fig. 4). When this probability fell below 0.5, the image was classified as class 1 (vascular compromise is likely present).

The inference process of our proposed method, which utilizes the Vision Transformer (ViT) architecture, can be executed on hardware with less than 8 GB of GPU memory, based on the ViT-Large model configuration. Upon quantization, the proposed method can be effectively used in mobile environments, such as smartphones.



**Fig. 3.** Schematic diagram of a model that uses class weighting and focal loss to overcome class imbalance. MLP = Multilayer perceptron.



**Fig. 4.** (A) Model interface. When a postoperative image is submitted to the model, it recognizes the flap among the structures in the image and outputs the probability of vascular compromise. (B) A 3-h postoperative photograph of a 64-year-old female patient who underwent resection of squamous cell carcinoma in the left lower gingiva, covered with an anterolateral thigh free flap. Considering the color and turgidity of the flap, it appears very healthy, and the model outputs a probability of less than 1% for vascular compromise. This flap has remained healthy to date. (C) A 48-h postoperative photograph of a 43-year-old male patient who underwent resection of ameloblastoma in the left mandible, reconstructed using the jaw-in-a-day technique. Although the anterior margin appears slightly dark, the overall appearance of the flap is considered healthy, and the model calculated a low probability of vascular compromise. This flap has also remained healthy to date.

equipped with Neural Processing Units (NPUs). This optimization enables the implementation of our method as a smartphone application, facilitating on-device inference capabilities.

### Data availability

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

Received: 20 February 2025; Accepted: 6 August 2025

Published online: 07 October 2025

### References

1. Wong, C.-H. & Wei, F.-C. Microsurgical free flap in head and neck reconstruction. *Head Neck* **32**, 1236–1245 (2010).
2. Lutz, B. S. & Wei, F.-C. Microsurgical workhorse flaps in head and neck reconstruction. *Clin. Plast. Surg.* **32**, 421–430 (2005).
3. 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).
4. Kohlert, S., Quimby, A. E., Saman, M. & Ducic, Y. Postoperative free-flap monitoring techniques. In *Seminars in plastic surgery*, vol. 33, 013–016 (Thieme Medical Publishers, 2019).

5. Chae, M. P. et al. Current evidence for postoperative monitoring of microvascular free flaps: A systematic review. *Ann. Plast. Surg.* **74**, 621–632 (2015).
6. Shen, A. Y. et al. Free flap monitoring, salvage, and failure timing: A systematic review. *J. Reconstr. Microsurg.* **37**, 300–308 (2021).
7. Jackson, R. S., Walker, R. J., Varvares, M. A. & Odell, M. J. Postoperative monitoring in free tissue transfer patients: Effective use of nursing and resident staff. *Otolaryngol. Head Neck Surg.* **141**, 621–625 (2009).
8. 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).
9. Kiranantawat, K. et al. The first smartphone application for microsurgery monitoring: Silparamanitor. *Plast. Reconstr. Surg.* **134**, 130–139 (2014).
10. Jarvis, T., Thornburg, D., Rebecca, A. M. & Teven, C. M. Artificial intelligence in plastic surgery: Current applications, future directions, and ethical implications. *Plast. Reconstr. Surg. Glob. Open* **8**, e3200 (2020).
11. Pettit, R. W., Fullem, R., Cheng, C. & Amos, C. I. Artificial intelligence, machine learning, and deep learning for clinical outcome prediction. *Emerg. Top. Life Sci.* **5**, 729–745 (2021).
12. Hsu, S.-Y. et al. Quantization of extraoral free flap monitoring for venous congestion with deep learning integrated ios applications on smartphones: A diagnostic study. *Int. J. Surg.* **109**, 1584–1593 (2023).
13. Yeong, E.-K., Hsiao, T.-C., Chiang, H. K. & Lin, C.-W. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. *Burns* **31**, 415–420 (2005).
14. Shi, H.-Y., Hwang, S.-L., Lee, K.-T. & Lin, C.-L. In-hospital mortality after traumatic brain injury surgery: A nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J. Neurosurg.* **118**, 746–752 (2013).
15. Azimi, P. & Mohammadi, H. R. Predicting endoscopic third ventriculostomy success in childhood hydrocephalus: An artificial neural network analysis. *J. Neurosurg. Pediatr.* **13**, 426–432 (2014).
16. Azimi, P., Benzel, E. C., Shahzadi, S., Azhari, S. & Mohammadi, H. R. Use of artificial neural networks to predict surgical satisfaction in patients with lumbar spinal canal stenosis. *J. Neurosurg. Spine* **20**, 300–305 (2014).
17. Hallac, R. R., Lee, J., Pressler, M., Seaward, J. R. & Kane, A. A. Identifying ear abnormality from 2d photographs using convolutional neural networks. *Sci. Rep.* **9**, 18198 (2019).
18. Ohura, N. et al. Convolutional neural networks for wound detection: The role of artificial intelligence in wound care. *J. Wound Care* **28**, S13–S24 (2019).
19. Mantelakis, A., Assael, Y., Sorooshian, P. & Khajuria, A. Machine learning demonstrates high accuracy for disease diagnosis and prognosis in plastic surgery. *Plast. Reconstr. Surg. Glob. Open* **9**, e3638 (2021).
20. U.S. Food and Drug Administration. Artificial intelligence and machine learning in software as a medical device. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> (2020). Accessed: 2025-02-19.
21. Wong, C.-H., Wei, F.-C., Fu, B., Chen, Y.-A. & Lin, J.-Y. Alternative vascular pedicle of the anterolateral thigh flap: The oblique branch of the lateral circumflex femoral artery. *Plast. Reconstr. Surg.* **123**, 571–577 (2009).
22. 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).
23. Lovčinská, V., Sukop, A., Klein, L. & Brandejsova, A. Free-flap monitoring: Review and clinical approach. *Acta Chir. Plast.* **61**, 16–23 (2020).
24. 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).
25. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (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).
29. Ioffe, S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
30. 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).
31. Cui, Y., Jia, M., Lin, T.-Y., Song, Y. & Belongie, S. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9268–9277 (2019).
32. Shen, L., Lin, Z. & Huang, Q. Relay backpropagation for effective learning of deep convolutional neural networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII* **14**, 467–482 (Springer, 2016).
33. Zhou, Z.-H. & Liu, X.-Y. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* **18**, 63–77 (2005).
34. Ross, T.-Y. & Dollár, G. Focal loss for dense object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2980–2988 (2017).

## Acknowledgements

Financial support and sponsorship: This work was supported by Hankuk University of Foreign Studies Research Fund of 2025.

## 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.

## Declarations

## Competing interests

The authors declare no competing interests.

### 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](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025