Utilizing Deep Learning to Detect the Perforators of Anterolateral Thigh Free Flap in Computed Tomography Images for Maxillofacial Reconstruction

Hyunyoung Kim

The Graduate School
Yonsei University
Department of Dentistry

Utilizing Deep Learning to Detect the Perforators of Anterolateral Thigh Free Flap in Computed Tomography Images for Maxillofacial Reconstruction

A Dissertation Submitted
to the Department of Dentistry
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Dentistry

Hyunyoung Kim

December 2024

This certifies that the Dissertation of Hyunyoung Kim is approved.

Thesis Supervisor	Dongwook Kim
Thesis Committee Member	Hyung Jun Kim
Thesis Committee Member	Woong Nam
Thesis Committee Member	Jun-Young Kim
Thesis Committee Member	Jong-Eun Kim

The Graduate School
Yonsei University
December 2024

Acknowledgements

먼저 이 논문을 이루기까지 지도와 격려를 베풀어 주신 지도교수 김동욱 교수님께 감사드립니다. 또한 연구가 결과로 도출되고, 결과가 논문으로 편찬되기까지 물심양면으로 지도해주신 김형준, 남웅, 김준영, 김종은 교수님께 감사드립니다.

인공지능 학습과 모델의 개발을 도움을 주신 고려대학교 컴퓨터학과 나지혜, 김채현 연구원 선생님들께도 깊은 감사를 드립니다.

로컬 치과원장으로서 박사학위 논문을 작성하면서 많은 소회가 있었습니다. 바쁜 진료를 마치고 연구를 시작하는 저녁 시간은 새로운 지식에 대한 열망과 아이디어로 가득했던 시간이었습니다. 이런 열정으로 시작한 이번 연구가 인류의 발전에 도움이 되는 연구가 되길 바랍니다. 나아가 박사학위 논문을 마무리하며 학업의 끝이 아닌 새로운 시작이 되기를 기대합니다.

연구를 위한 야근을 하면서 제가 연구활동을 영위할 수 있도록 곁에서 도움을 주신 부모님과 사랑하는 아내와 딸 지유 그리고 주위의 많은 분들께 감사드립니다.

마지막으로 본 연구를 가능케 한 환자분들의 쾌유와 안녕이 있기를 기도합니다. 이 연구가 향후 다른 환자분들을 도울 수 있기를 바랍니다.

TABLE OF CONTENTS

LIST OF FIGURES ii
LIST OF TABLES iii
ABSTRACT IN ENGLISH iv
1. INTRODUCTION 1
2. MATERIALS AND METHODS 4
2.1. DATA SELECTION 4
2.2. MANUAL ANNOTATION 6
2.3. DEVELOPMENT OF ARTIFICIAL INTELLIGENCE 8
2.3.1. nnUNet one 9
2.3.2. nnUNet adv 9
2.4. ASSESSMENT OF ARTIFICIAL INTELLIGENCE MODEL PERFORMANCE 10
2.4.1. GROUND TRUTH 10
2.4.2. INTERNAL AND EXTERNAL VALIDATION 10
2.4.3. PERFORATOR-SPECIFIC PERFORMANCE 11
2.4.4. SLICE-SPECIFIC PERFORMANCE 11
2.4.5. STATISTIC ANALYSIS 15
3. RESULTS 16
3.1. PERFORATOR-SPECIFIC PERFORMANCE 16
3.2. SLICE-SPECIFIC PERFORMANCE 18
4. DISCUSSION 19
5. CONCLUSION 22
REFERENCE 23
ABSTRACT IN KOREAN 25

LIST OF FIGURES

<Fig 1> A schematic diagram of a flap with perforators 1
<Fig 2> Cutting a perforator in harvesting free ALT flap 2
<Fig 3> Scheme of a perforator from lateral circumflex femoral artery for ALT flap 3
<Fig 4> Manual brush annotation tool to mark the perforator on Supervisely application 7
<Fig 5> A coronal-view perforator and brush annotation region 8
<Fig 6> True positive area 12
<Fig 7> False positive area 13
<Fig 8> False negative area 14
<Fig 9> Inaccuracy between the perforator and brush annotation 21

LIST OF TABLES

<Table 1> The CTA series and axial slices utilized in this study 6
<Table 2> Numbers of perforators validated by surgery (surgically confirmed perforators) and two experts (radiologically confirmed perforators) 16
<Table 3> The numbers of perforators detected by two experts and nnUNet adv 17
<Table 4> Perforator-specific performance of nnUNet adv 17
<Table 5> Slice-specific performance of nnUNet one 18
<Table 6> Slice-specific performance of nnUNet adv 19

ABSTRACT

Utilizing Deep Learning to Detect Perforators of Anterolateral Thigh Free Flap in Computed Tomography Images for Maxillofacial Reconstruction

The anterolateral thigh (ALT) free flap is widely utilized for reconstructive surgery in cases of anatomical defects caused by trauma, disease, or congenital malformations. Perforator vessels, which supply oxygen and blood, are essential for the successful transplantation of the ALT free flap. This study aimed to develop an artificial intelligence model to accurately detect perforator vessels of the ALT free flap in computed tomography angiography (CTA) images. For the image data, 53 CTA series from patients who underwent ALT free flap surgeries were used. Of these, 48 CTA series were used for training, and 5 CTA series were used for testing. The nnUNet model performed deep learning using manually annotated data provided by two oral and maxillofacial surgeons. Two models were developed in this study. nnUNet one trained on perforator vessels confirmed through surgery and nnUNet adv trained additionally on perforator vessels confirmed radiologically. The nnUNet adv demonstrated superior performance compared to nnUNet one in both internal and external validation, achieving higher scores in Dice, IoU, Precision, and Recall metrics. In external validation, nnUNet adv achieved a Dice score of 0.682 and a detection rate of 71.4% for identifying perforator vessels. Training with radiologically verified perforator vessels, in addition to surgically confirmed vessels, resulted in better performance than using only surgical confirmation. These results suggest that the nnUNet adv developed in this study can reduce the time surgeons spend interpreting CTA images and assist in preoperative planning by enabling more accurate design of various free flap configurations. Furthermore, nnUNet adv demonstrated potential in detecting perforator vessels that might be missed by specialists during CTA interpretation. Adding more training datasets in the future is expected to further enhance the model's accuracy and reliability.

Key words: Artificial intelligence, Deep learning, Computed tomography, Anterolateral thigh flap, Perforators, Reconstruction, Oral and maxillofacial area

1. Introduction

Defects in the body can occur due to accidents, diseases, or deformities. When a part of the body is missing, reconstructive surgery is needed with tissue from other parts of the body. The tissue harvested for this purpose is called a flap. Flaps can be in the form of composite flaps, consisting of multiple tissues such as skin, muscle, and bone, or simple flaps that consist of just one type of tissue (Mittal et al., 2018). Tissues need a vessel for oxygen and nutrients to survive in a recipient site. Such a vessel is called a perforator. When transplanting a flap, the blood vessels that supply it with nutrients and oxygen may also be transplanted. Depending on the cutting of a vessel, flaps are classified into pedicled flaps and free flaps. Pedicled flaps are transferred to the defect area without cutting the blood vessels, keeping the original blood supply intact (Kim et al., 2024). While this ensures sufficient blood supply, it limits the range of reconstruction to areas near the donor site due to the restricted length of the blood vessels. (Figure 1)

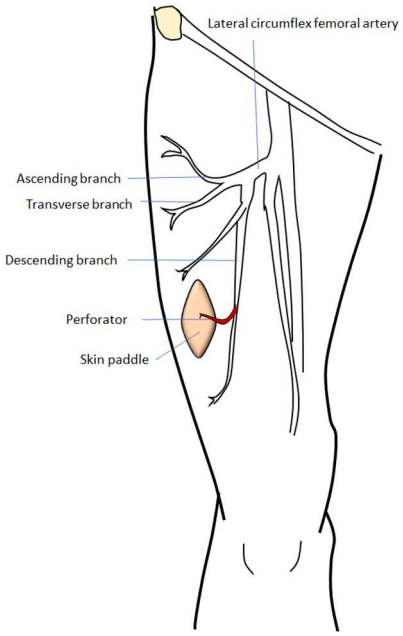

Figure 1. A schematic diagram of a flap with perforators. A flap for a transplantation needs a perforator vessel for oxygen and nutrients.

In contrast, free flaps involve cutting the blood vessels that supply the flap with nutrients and oxygen and anastomosing them to vessels at the recipient site. (Figure 2) This allows for reconstructive surgery regardless of the physical distance between the defect site and the donor site. Furthermore, the variety of sites from which free flaps can be harvested can allow for more diverse tissue compositions from donor sites, making it applicable to a wide range of defects. Due to these advantages, research on free flaps has been very active (Nghija and Son, 2024).

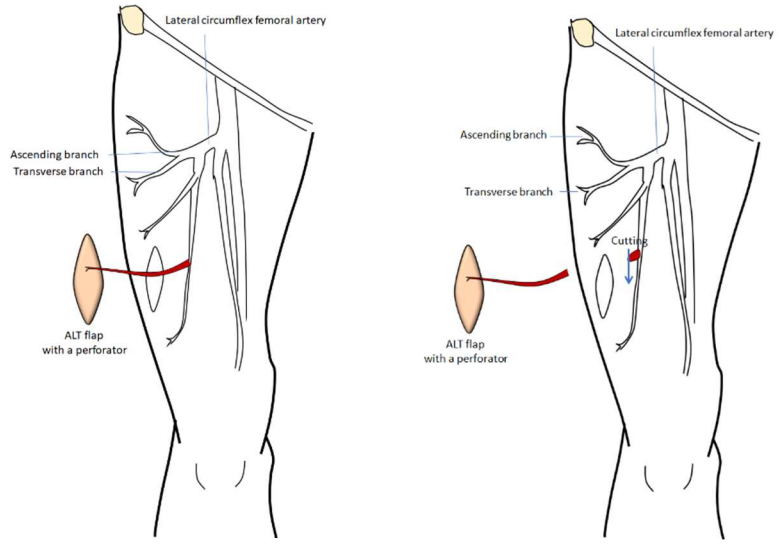

Figure 2. Cutting a perforator in harvesting free ALT flap.

Free flaps can be harvested from various parts of the body, such as the fibula flap, radial forearm flap, and anterolateral thigh (ALT) flap, which are commonly used for oral and maxillofacial defects. Among these, the ALT flap was first reported by Song in 1984 (Song et al., 1984). Since then, the ALT flap has been used for various reconstructive surgeries, including those for head and neck as well as ankle defects (Çiçek, 2023).

The ALT flap is a free flap with many advantages. It has a long vascular pedicle, providing sufficient length and flexibility for anastomosis with the recipient site's vessels. It can be very useful for reconstructing complex oral and maxillofacial defects. Due to these advantages, the ALT flap is considered a preferred choice among many surgeons in the field

of head and neck reconstruction (Besharah et al., 2020; Ranganath et al., 2022; Vijayasekaran et al., 2020).

The most important step in harvesting a flap is identifying and keeping the perforator intact during dissection. It can make the perforator vessel supply the flap with nutrients and oxygen. The perforators of the ALT flap mainly branch from the descending branch of the lateral circumflex femoral artery. However, the number, location, and pathway of these perforators vary significantly from a patient to a patient (Figure 3) (De Beule et al., 2016).

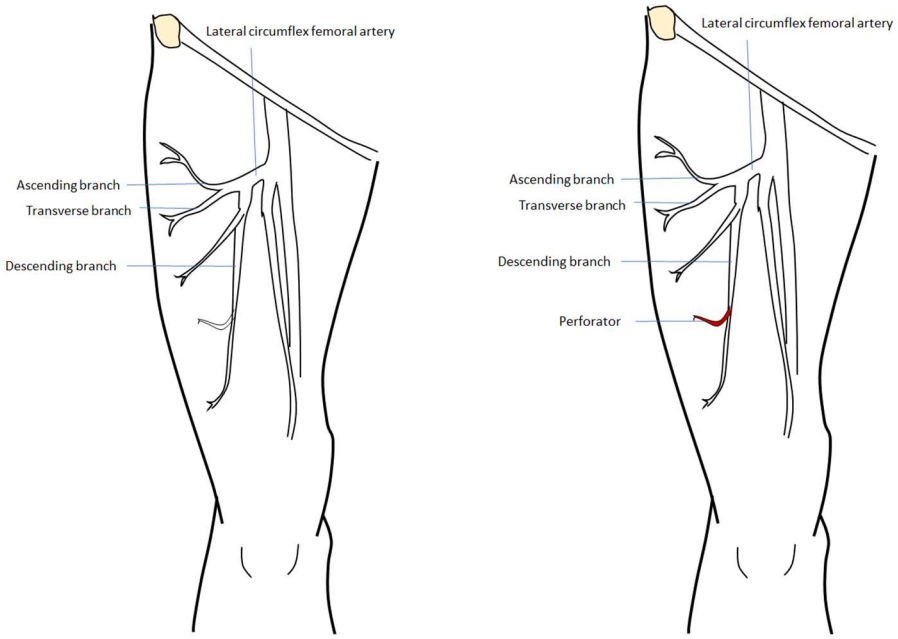

Figure 3. Scheme of a perforator from lateral circumflex femoral artery for ALT flap.

These anatomical variations in the flap can sometimes lead to inconsistent results. In some cases, the variation in the perforators connected to the ALT flap presents a challenge to the surgeon. When the perforator cannot be found or is encountered in an unexpected location, the surgeon can cut the perforator and have to resort to using a second optional flap. Furthermore, if the perforator is damaged during the transplant of a free flap, resulting in a lack of blood flow, the reconstructive surgery may fail (Smith et al., 2017). In such

cases, the patient may need a second surgery, which could also affect the morbidity of the donor site. Therefore, it is essential to control these variables before surgery (Kalra et al., 2022).

Traditionally, Doppler ultrasonography and computed tomography angiography (CTA) are used to locate perforators preoperatively (Hashimoto et al., 2012). Of these, CTA offers a higher predictive accuracy. However, interpreting CTA is time-consuming and requires a high level of expertise, which can pose challenges for the surgeon. Failure to predict the number, location, length, and diameter of the perforators can prolong the time required to locate them during surgery and lead to complications such as mismatches in the diameter of vessels between the donor and the recipient site. Therefore, accurate interpretation of CTA is crucial.

There is growing interest in artificial intelligence (AI) and deep learning technologies to analyze CT images. Khan et al. successfully reported a deep learning-based technique for intracranial hemorrhage detection using CT images from 75 patients (Khan et al., 2023). And Cevik et al. reported artificial intelligence for identification of the deep inferior epigastric artery perforator (DIEP) using preoperative CTA. Automatic detection of the deep inferior epigastric perforator is reported by Mavioso et al. They developed an automatic detection tool with 20 CTA from patients who undergone a DIEAP flap for breast reconstruction (Mavioso et al., 2018). And the automatic detection tool was evaluated in a prospective way with 40 patients. The reduction of the time is achieved during analyzing CTA and surgery (Mavioso et al., 2020).

Due to growing interest in artificial intelligence, the purpose of this study is to develop an artificial intelligence to detect a perforator of ALT flap using CTA images.

2. Materials and Methods

2.1. Data selection

Patients who visited the Department of Oral and Maxillofacial Surgery at Yonsei University Dental Hospital and underwent anterolateral thigh (ALT) flap surgery between March 2021 and July 2022 were selected. A total of 52 patients underwent the procedure, and all of them underwent a lower extremity computed tomography angiogram (CTA) prior

to surgery. Before utilizing the lower extremity CTA data, approval was obtained from the Institutional Review Board (IRB) of Yonsei University Dental Hospital (IRB 2-2024-0009).

The lower extremity CTA was used to explore the ALT perforator and provide the necessary images for deep learning in the AI model. Based on these criteria, lower extremity CTA data were requested from the Imaging Data Service Team at Yonsei University for all 52 selected patients. Since one patient had two lower extremity CTA scans on different dates, a total of 53 series of lower extremity CTA data were collected. Out of these, 48 CTA series were categorized into the learning group to be used for deep learning, while 5 CTA series were categorized into the test group to measure the performance of the artificial intelligence model developed through deep learning.

For lower extremity CTA series imaging, all patients are required to fast for 4–6 hours beforehand. The imaging range is set to include the lower extremities. A preliminary scan is performed before the main scan to confirm that the lower extremities, including the anterior superior iliac spine (ASIS) and patella, are properly captured. Afterward, a contrast agent is administered. Following the injection of the contrast agent, the main scan is conducted, and thinner axial slice thickness results in higher resolution. All patients were imaged with an axial slice thickness of 2 mm. The specifications of the equipment used for CTA imaging are represented by the model name: Siemens Somatom Definition Flash (FLASH), FORCE, or X.cite (Siemens Healthineers), Revolution CT (GE Healthcare, Chicago, IL, USA)

The 2mm-thickness axial data from the CTA series was extracted as DICOM files, and all information was anonymized. Due to variations in patient height, the number of axial slices in lower extremity angiography ranged from 500 to 800 slices per patient. To ensure a consistent number of axial slices across patients, 250 axial slices were selected from the anterior superior iliac spine (ASIS) toward the knee. At a 2mm thickness, the 250 axial slices cover approximately 500mm, which includes the region from the ASIS to the patella in most patients. This region is considered sufficient for exploring the ALT perforator. Consequently, 250 axial slices per CTA series were used across 53 series, resulting in a total of 13,250 axial slices. The data is shown in table 1.

	CTA series	Axial slices
Learning group	48	12000
Test group	5	1250

Table 1. The CTA series and axial slices utilized in this study. The Learning Group was used both for deep learning in the AI model and for internal validation. The Test Group, on the other hand, was not used in the deep learning process and was solely utilized for external validation.

2.2. Manual annotation

For deep learning in the artificial intelligence model, it was necessary to annotate perforators manually in the lower extremity angiography images by two experts(DW/HY). Perforators identified during surgery were manually annotated and classified as "surgically confirmed perforators." Additionally, perforators confirmed by two oral and maxillofacial surgery specialists in the CTA series were classified as "radiologically confirmed perforators." The annotation process was conducted across all 13,250 axial slices.

The program used for manual annotation was the Supervisely App, which provides a brush annotation tool to mark perforators in DICOM labeling online (Figure 4). Data manually annotated with the brush tool on axial slices can be reconstructed into 3D models, allowing for visualization in three dimensions and examination from coronal-sectional views (Figure 5). Access to the Supervisely app is restricted to securely encrypted accounts, minimizing the risk of personal information exposure, and data can be immediately deleted after processing. Additionally, both experts can access the same data, allowing for double-checking. For perforators in the expert-validated group, the consistency of brush annotations between the two experts was calculated using Cohen’s Kappa value.

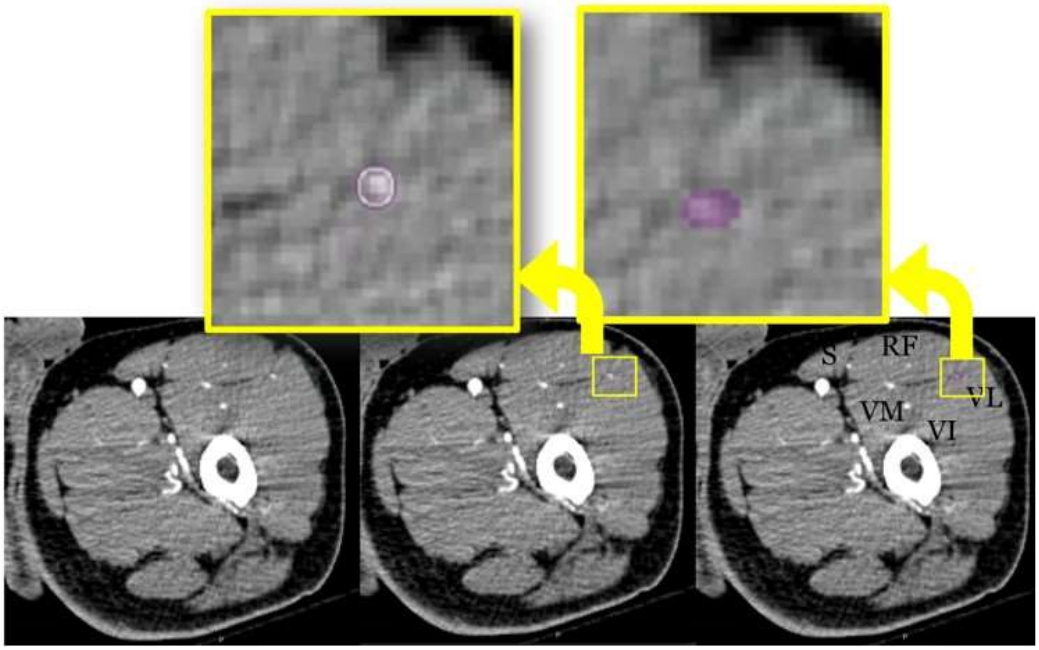

Figure 4. Manual brush annotation tool to mark the perforator on Supervisely application.

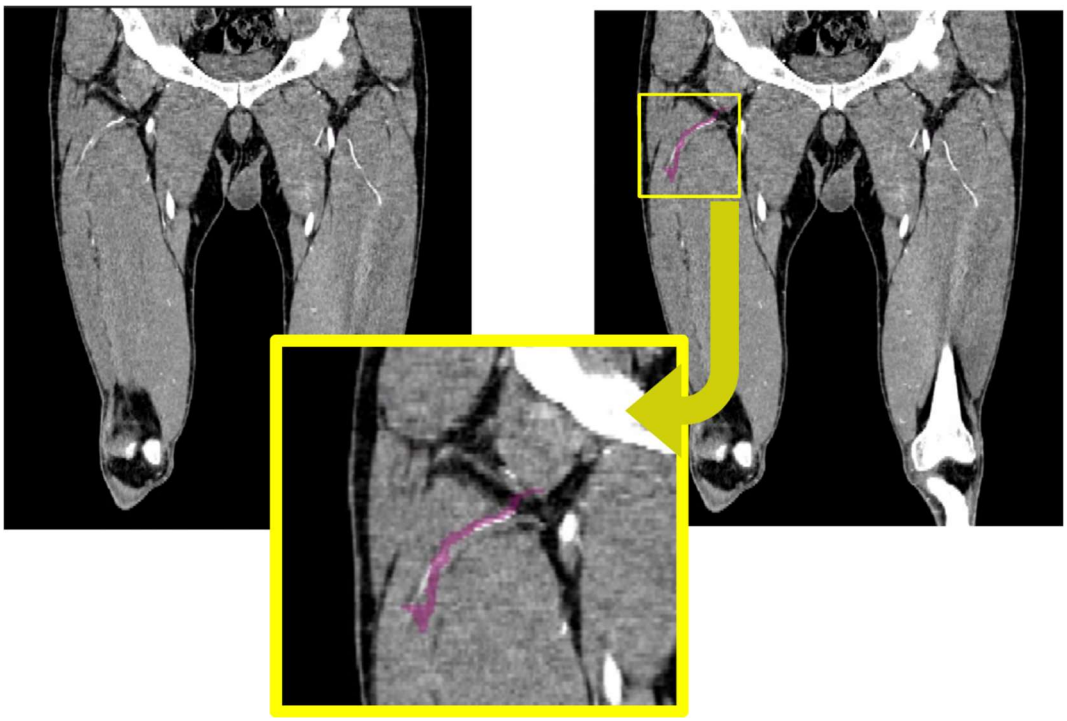

Figure 5. A coronal-view perforator and brush annotation region.

2.3. Development of artificial intelligence

nnUNet was chosen as an AI model for deep learning. This engine can support 3D image data to reflect the characteristic of perforators continuously existing between adjacent slices. the nnUNet is known for its high segmentation performance in medical data (Isensee et al., 2021).

The brush-annotated data were extracted in the .json file format. The CTA slides used for the analysis were extracted in the .nrrd file format instead of the original .dicom files. Then these file formats were sent to the Department of Computer Science at Korea University for AI model development. The data transfer was conducted after obtaining approval from the Severance Institutional Review Board.

The brush annotation extracted from the .json file was matched with the CTA image extracted from the .nrrd file. Registration is the process of aligning the coordinates of the

brush annotation with the coordinates in the CTA images, allowing the positional information to be recognized as the same location.

Additionally, for nnUNet to recognize the CTA images, all data must be in the same size. Therefore, any CTA images that were not in 512 x 512 pixels were cropped and adjusted to 512 x 512 pixels. This crop applied to three CTA image series in the learning group.

Then nnUNet model was trained with the pre-processed data. The learning group for the nnUNet deep learning is composed of 12000 slices. The training was repeated 1000 epochs because repeated learning can improve the performance of the AI model.

2.3.1 nnUNet one

The AI model trained on perforators confirmed through surgery was referred to as nnUNet one. Perforators identified during surgery are actual, existing perforators, and the AI model was trained to learn the true perforators. In the train group, a total of 105 perforators were confirmed through surgery in 48 patients. These perforators were marked on the CTA series, which served as the training data for nnUNet one. In other words, nnUNet one is an AI model trained on "surgically confirmed perforators."

2.3.2 nnUNet adv

The AI model trained on data marked by two experts as perforators in the CTA series, regardless of whether they were confirmed through surgery, was referred to as nnUNet adv. In the train group, a total of 186 perforators were marked by the two experts across 48 CTA series, and this data was used to train nnUNet adv. Since the perforators were identified solely through interpretation of the CTA series and not confirmed via surgery, the dataset may include cases where the perforators do not actually exist. In other words, nnUNet adv is an AI model trained on "radiologically confirmed perforators."

2.4. Assessment of artificial intelligence model performance

After developing the artificial intelligence (AI) model using deep learning techniques, the model's performance was evaluated using two methods. These two methods were referred to as "perforator-specific performance" and "slice-specific performance."

2.4.1 Ground truth

To evaluate the performance of the AI model, data to be used as ground truth is required. While it would be ideal to use perforators identified through surgery as the ground truth, this approach poses ethical concerns. Performing unnecessary surgeries solely for research purposes would be unacceptable, as patients should not undergo procedures beyond what is necessary for their reconstruction. Any unnecessary surgeries to locate additional perforators beyond those required for reconstruction must be avoided.

As an alternative, two experts interpreted CTA series and marked perforators to create ground truth data. According to previous studies, it has been reported that perforators were sometimes overlooked during CTA interpretation by experts but were found to exist during surgery (Kim et al., 2023). To identify as many perforators as possible in the given CTA series, two experts independently searched for perforators in CTA series. Cross-checking of the perforator marking areas was performed using the Supervisely application. Based on this ground truth data, the performance of the AI model was evaluated by comparing whether the areas identified as perforators by the AI model matched the expert annotations.

2.4.2 Internal and external validation

To evaluate the performance of the AI model, two approaches are employed: Internal validation, which tests the model using the data it was trained on, and External validation, which tests the model on new data not used during training. Naturally, the performance values from internal validation are expected to be higher.

For internal validation, 5 CTA series were randomly selected from the 48 CTA series used during training. For external validation, 5 CTA series not used in the training process were utilized.

2.4.3 Perforator-specific performance

Perforator-specific performance is a method for evaluating the performance of an AI model by determining whether the AI model has detected the perforator along its pathway. If the AI model detects part of the perforator prior to surgery, it can assist surgeons in locating the perforator more efficiently.
When the perforator's location is interpreted based on the position identified by the AI model, it reduces interpretation time and enables the planning of various designs for flap surgery, making this evaluation clinically significant.
Therefore, Perforator-specific performance was assessed based on whether the AI model detected the perforator on at least one slide of each perforator. If the AI model detected the perforator on one or more slides along its pathway, from its origin to the end of its distribution in the tissue, it was considered to have achieved Perforator-specific performance.

To evaluate the perforator-specific performance of the AI model, two experts identified perforators in five CTA scans from the test group and compared them with the perforators detected by the AI model.

2.4.4 Slice-specific performance

Slice-specific performance is a method used to determine the accuracy of the AI model by assessing whether the manual annotations and the AI model's annotations match on each axial slide.

To evaluate slice-specific performance, the data used for the test included 5 CTA series in test group. Each CTA of the five patients had 250 axial slides, resulting in a total of 1,250 slides for evaluation. The results for each slide were represented using colors. If the area detected by the AI model matched the area marked by two experts, the region was marked

in green (Figure 6). Areas detected by the AI model but not marked by two experts were marked in red (Figure 7). Areas marked by two experts but not detected by the AI model were marked in blue (Figure 8).

Through this color-coding system, green regions indicate true positives (TP), red regions indicate false positives (FP), and blue regions indicate false negatives (FN). By scoring slides for TP, FP, and FN, the performance of the AI model was evaluated across the 1,250 slides in 5 CTA series of test group.

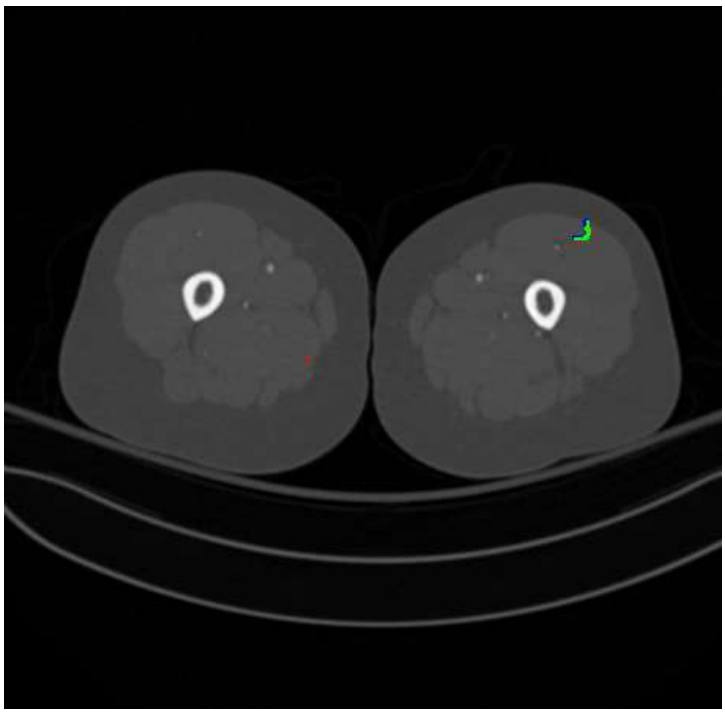

Figure 6. True positive area. The region where the location of the perforator identified by the AI model aligns with the manual brush annotation(Green mark).

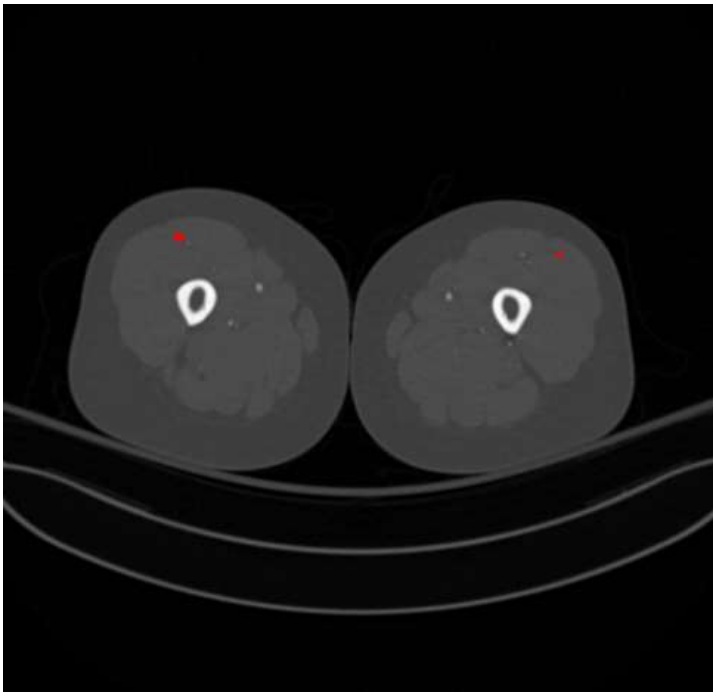

Figure 7. False positive area. The region where the AI model identified a perforator, but no manual brush annotation(Red mark) is present.

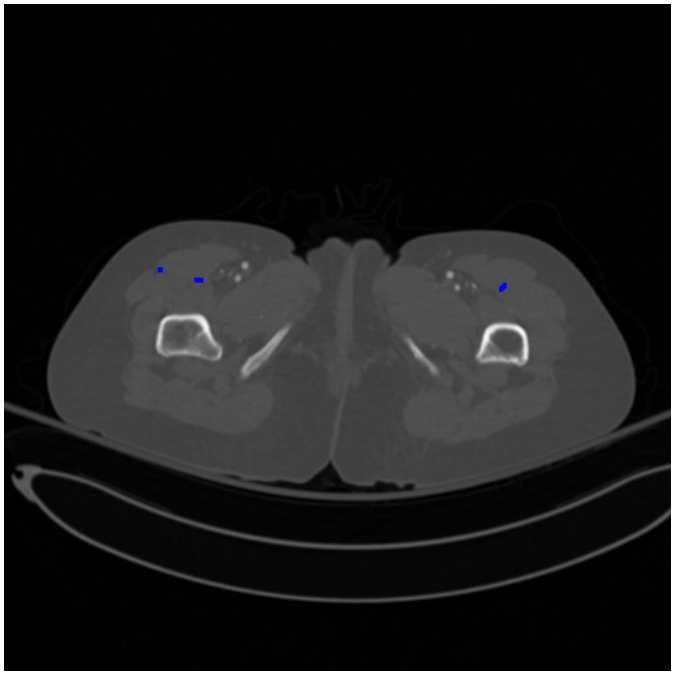

Figure 8. False negative area. The region where the AI model failed to identify a perforator, but manual brush annotation(Blue mark) is present.

2.4.5 Statistic analysis

Common metrics for evaluating the performance of AI models include Dice, IoU, Precision, and Recall (Huang et al., 2024). Dice is a performance measure typically used when labels are sparse and is one of the most widely used metrics in medical imaging and computer science. A higher Dice score indicates better performance, with the score increasing when the AI model's annotations match the ground truth and decreasing with false positives. The formula for Dice is as follows:

$$Dice = \frac{2TP}{2TP + FP + FN}$$

IoU, or Intersection over Union, is also used to evaluate AI model performance and serves as a supporting metric for Dice. Like Dice, higher values indicate better performance:

$$IoU = \frac{TP}{TP + FP + FN}$$

Precision indicates the proportion of perforators annotated by the AI model that are present:

$$Precision = \frac{TP}{TP + FP}$$

Recall, also known as sensitivity or true positive rate, represents the proportion of actual perforators present in the data that the AI model correctly annotated:

$$Recall = \frac{TP}{TP + FN}$$

3. Results

A total of 117 surgery-validated perforators were identified, with 105 observed in the 48 CTA series in the Learning Group and 12 observed in the 5 CTA series in the Test Group. Subsequently, two oral and maxillofacial surgery specialists analyzed 53 CTA series and identified a total of 207 expert-validated perforators. Among these, 186 were found in the 48 CTA series in the Learning Group, and 21 were found in the 5 CTA series in the Test Group. The Cohen's Kappa coefficient for the perforators confirmed by the two experts was 0.92, indicating a very high level of agreement (Table 2).

	Number of perforators		
	Total	Learning group (48 CTA series)	Test group (5 CTA series)
Surgically confirmed perforators	117	105	12
Radiologically confirmed perforators	207	186	21

Table 2. Numbers of perforators validated by surgery (surgically confirmed perforators) and two experts (radiologically confirmed perforators).

3.1 Perforator-specific performance

The performance of the AI model developed in this study was analyzed from a clinical perspective. The evaluation focused on the AI model's ability to identify perforators, referred to as perforator-specific performance. If a surgeon starts to interpret CTA series from the location identified by AI model, it can significantly reduce interpretation time and facilitate planning of various flap surgery designs. It can be very helpful for a surgeon. Perforator-specific performance was evaluated solely for nnUNet adv.

Expert-validated perforators in the 5 CTA series in test group were used to evaluate Perforator-specific performance of AI model. The two experts detected a total of 21 perforators in the 5 CTA series in the test group.

The AI model detected 23 perforators in the 5 CTA series in the test group. The Perforator-specific performance of AI model was evaluated by comparing the perforators annotated by AI model with the manually annotated perforators.

Out of the 21 perforators identified by the two experts, the AI model detected 15 perforators, failing to detect 6 perforators. It results in a true positive rate of 71.4%.

On the other hand, of the 23 perforators detected by the AI model, only 15 perforators were matched with the manual annotations. 8 other perforators of the 23 perforators were not identified by two experts (Table 3).

Experts \ nnUNet adv	Detection	No detection
Detection	15 (71.4%)	6 (28.6%)
No detection	8 (100%)	0 (0%)

Table 3. The numbers of perforators detected by two experts and nnUNet adv.

A statistical analysis was conducted to evaluate the perforator-specific performance. And external validation of the nnUNet adv model was performed to result in DICE value, IoU, Precision, and Recall that could be compared with other studies. The nnUNet adv showed a DICE value of 0.682, IoU of 0.517, Precision of 0.652, and Recall of 0.714.

nnUNet adv	DICE	IoU	Precision	Recall
External Validation	0.682	0.517	0.652	0.714

Table 4. Perforator-specific performance of nnUNet adv.

3.2 Slice-specific performance

The performance of the AI model developed in this study was analyzed from a computer science perspective. Axial slides from the CTA images of each patient were evaluated for slice-specific performance of AI model. For each CTA, 250 slides from the anterior superior iliac spine were evaluated. Out of a total of 53 CTA series, five CTA series, which were not used for training, were utilized as the test group. Each CTA of the five patients had 250 axial slides, resulting in a total of 1,250 slides for evaluation.

Slice-specific performance was tested for both nnUNet one and nnUNet adv. And internal validation and external validation were conducted for both nnUNet one and nnUNet adv to evaluate the AI model's performance from a computer science perspective.

First, the slice-specific performance of nnUNet one was analyzed. For internal validation, 5 CTA series were randomly selected from the 48 series used for training. For external validation, 5 CTA series that were not used during training were utilized. The DICE value was 0.0957 for internal validation and 0.0275 for external validation. The IoU value was 0.054 for internal validation and 0.014 for external validation.

nnUNet one	DICE	IoU	Precision	Recall
Internal Validation	0.0957	0.054	0.101	0.123
External Validation	0.0275	0.014	0.023	0.03

Table 5. Slice-specific performance of nnUNet one.

Using the same approach, the performance of nnUNet adv was analyzed. The DICE value was 0.284 for internal validation and 0.184 for external validation. The IoU value was 0.173 for internal validation and 0.103 for external validation.

nnUNet adv	DICE	IoU	Precision	Recall
Internal Validation	0.284	0.173	0.42	0.251
External Validation	0.184	0.103	0.284	0.144

Table 6. Slice-specific performance of nnUNet adv.

4. Discussions

The AI model's performance was assessed with two criteria: perforator-specific performance and slice-specific performance. For perforator-specific performance, nnUNet adv successfully detected 15 out of the 21 perforators identified by experts, achieving a detection rate of approximately 71%. This indicates a reasonable level of perforator-specific performance, suggesting that nnUNet adv can aid surgeons by providing hints about the location of perforators. Surgeons can locate perforators quickly and easily by starting their search from slices annotated by nnUNet adv before surgery. While experts must manually search each CTA slide to find perforators, the AI model can analyze many slides quickly and present results, potentially saving surgeons significant time in perforator localization for surgical planning.

The DICE value for nnUNet adv is 0.682. Typically, AI models analyzing medical images are considered to have good performance when a DICE value is between 0.7 and 1. According to the study by Josephine Chen et al., an AI model developed to detect deep inferior epigastric vessels using 100 CTA series achieved a DICE value of 0.703. Despite having half the number of CTA series, the DICE value in this study is similar to the DICE value reported in Josephine Chen's study. It indicates that the perforator detection performance of nnUNet adv can be considered satisfactory.

The nnUNet adv has advantages over human interpreters. Having undergone training in this study, nnUNet adv does not experience inexperience when interpreting CTA series. In contrast, all novice interpreters require a long learning curve to become proficient in interpretation. Moreover, many skilled interpreters eventually retire over time, necessitating the training of new novice interpreters to learn how to interpret CTA series

from the beginning. However, nnUNet adv does not retire, meaning it never reverts to a novice state. Instead, it continues to improve in precision and accuracy over time.

It is also worth noting the 8 perforators identified by nnUNet adv among the 23 perforators it detected, which were not identified by the two experts. If these 8 perforators detected by nnUNet adv are indeed real perforators, it would provide significant clinical benefits. Additionally, the performance of nnUNet adv could be higher than the reported DICE value of 0.682. Previous studies have reported cases where perforators not identified in CTA series by experts were later discovered during surgery. Kim et al. reported 79 perforators identified in CTA series, but 85 perforators were found during surgery in 53 patients. It has been reported that six perforators were not identified during the interpretation of the CTA series (Kim et al., 2023). If nnUNet adv successfully detected perforators missed by experts, its performance could be considered superior to that of human interpreters. Further research is required to validate these 8 false positives identified by the AI model.

Next, slice-specific performance was assessed to find the performance of nnUNet one and nnUNet adv developed in this study. The Dice scores for nnUNet one and nnUNet adv were 0.0275 and 0.184, respectively.

Using five CTA series not included in the deep learning training, external validation was performed. Dice values for nnUNet one and nnUNet adv in external validation were 0.0275 and 0.184, respectively. Additionally, internal validation was conducted using five randomly selected CTA series from a total of 48 CTA series. Dice values for nnUNet one and nnUNet adv in internal validation were 0.0957 and 0.284, respectively. nnUNet adv demonstrated superior performance compared to nnUNet one in both internal and external validation. Anouk van der Schot et al. investigated placenta vessel segmentation using U-Net and pix2pix models. After training the U-Net model with 483 images and evaluating its performance, internal validation yielded Dice and IoU values of 0.53 [0.49; 0.64] and 0.49 [0.17; 0.56], respectively (van der Schot et al., 2023). These values appear higher than the Dice (0.284) and IoU (0.173) achieved by nnUNet adv in this study. However, in van der Schot et al.'s research, when the test set was changed, the Dice score ranged from 0.49 to 0.64, and the IoU ranged from 0.17 to 0.56, indicating significant variation. Therefore, for a more accurate evaluation of nnUNet adv's performance, it is necessary to conduct internal and external validations using alternative test sets in this study.

The brush annotation tool used for marking perforators creates circular regions. If the cross-sectional shape of the perforator on a slice is oval or another shape, this can lead to pixel-level discrepancies between the perforator's actual shape and the circular brush

annotation. These discrepancies may introduce errors in the AI model's development (Figure 9). If the Supervisely app develops annotation tools with varied shapes in the future, it could help reduce these discrepancies and potentially improve the AI model's performance.

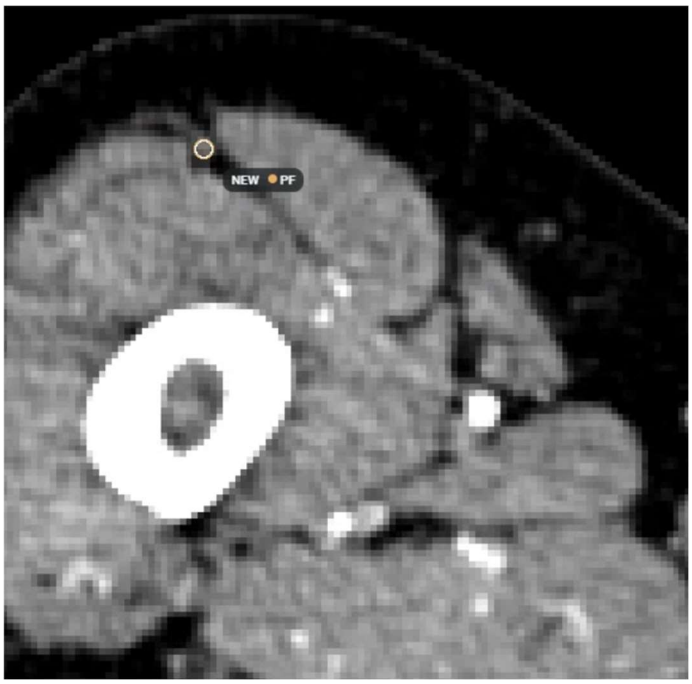

Figure 9. Inaccuracy between the perforator and brush annotation.

There are two ways to enhance the performance of nnUNet adv. The first is to increase the number of CTA series used for training. Josephine Chen et al. conducted a study using 100 CTA series to train nnUNet for detecting inferior deep epigastric vessels. In contrast, this study utilized 53 CTA series. It is anticipated that increasing the number of CTA series will result in an improved DICE value. Secondly, data augmentation allows the AI model to learn more effectively from the same dataset. Data augmentation involves creating multiple variations of a single image by applying transformations such as scaling, zooming, and rotation. This technique effectively increases the amount of data available for training. Although data augmentation was not used in this study, implementing this technique in future training of nnUNet adv is expected to significantly enhance its performance.

5. Conclusion

In this study, nnUNet adv was developed to detect perforators for ALT flaps using CTA series. nnUNet adv can assist surgeons in locating perforators quickly. With further deep learning on an expanded dataset, it could potentially surpass expert's accuracy in the future.

An AI model that undergoes more extensive deep learning is expected to reduce the time experts need to analyze CTA images preoperatively, improve the accuracy of perforator detection, and ultimately enhance surgical outcomes. Future studies are anticipated to make all of this possible.

References

Besharah BO, Ghazzawi RA, Al-Kaff HH, Abdelmonim SK, Al-Essa MA (2020). Reconstruction of facial dermatofibrosarcoma protuberans using an anterolateral thigh flap: a case report and literature review. *J Surg Case Rep* 2020(9): rjaa340.

Çiçek Ç (2023). Comparison of the Functional and Esthetic Outcomes of Free Parascapular and Anterolateral Thigh Flaps for Ankle Reconstruction. *Southern Clinics of Istanbul Eurasia* 34(3): 220.

De Beule T, Van Deun W, Vranckx J, de Dobbelaere B, Maleux G, Heye S (2016). Anatomical variations and pre-operative imaging technique concerning the anterolateral thigh flap: guiding the surgeon. *Br J Radiol* 89(1063): 20150920.

Hashimoto I, Nakanishi H, Yamano M, Abe Y (2012). Usefulness in combined free anterolateral thigh and vastus lateralis muscle flaps. *European Journal of Plastic Surgery* 35(12): 867.

Huang L, Miron A, Hone K, Li Y Segmenting Medical Images: From UNet to Res-UNet and nnUNet. 2024. p. 483.

Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18(2): 203-211.

Kalra GS, Gupta S, Kalra S (2022). Pedicle First Anterior Approach to Harvest Anterolateral Thigh Flap-Review of 304 Cases. *Indian J Plast Surg* 55(3): 272-276.

Khan MM, Chowdhury MEH, Arefin A, Podder KK, Hossain MSA, Alqahtani A, et al. (2023). A Deep Learning-Based Automatic Segmentation and 3D Visualization Technique for Intracranial Hemorrhage Detection Using Computed Tomography Images. *Diagnostics (Basel)* 13(15).

Kim H, Cha IH, Kim HJ, Nam W, Yang H, Shin G, et al. (2023). Perforators Detected in Computed Tomography Angiography for Anterolateral Thigh Free Flap: Am I the Only One Who Feels Inaccurate? *J Clin Med* 12(12).

Kim JS, Lee HH, Koh SH, Lee DC, Roh SY, Lee KJ (2024). Hand Reconstruction Using Anterolateral Thigh Free Flap by Terminal Perforator-to-Digital Artery Anastomosis: Retrospective Analysis. *Archives of Plastic Surgery* 51(1): 87.

Mavioso C, Araujo RJ, Oliveira HP, Anacleto JC, Vasconcelos MA, Pinto D, et al. (2020). Automatic detection of perforators for microsurgical reconstruction. *Breast* 50(19-24): 19.

Mavioso C, Correia Anacleto J, Vasconcelos MA, Araújo R, Oliveira H, Pinto D, et al. (2018). 274 (PB-069) - The development of an automatic tool to improve

perforators detection in Angio CT in DIEAP flap breast reconstruction. *European Journal of Cancer* 92: S74.

Mittal G, Agarwal A, Kataria G (2018). Flaps for Oral and Maxillofacial Reconstruction: Review of Literature and a Clinical Guide to the Clinicians. *Asian Journal of Oncology* 4: 37–42.

Nghija PT, Son TT (2024). The free anterolateral thigh flap substituted for a pedicled biparietal flap in reconstructing complex frontal defect. *J Surg Case Rep* 2024(3): rjae115.

Ranganath K, Jalisi SM, Naples JG, Gomez ED (2022). Comparing outcomes of radial forearm free flaps and anterolateral thigh free flaps in oral cavity reconstruction: A systematic review and meta-analysis. *Oral Oncol* 135: 106214.

Smith RK, Wykes J, Martin DT, Niles N (2017). Perforator variability in the anterolateral thigh free flap: a systematic review. *Surg Radiol Anat* 39(7): 779-789.

Song YG, Chen GZ, Song YL (1984). The free thigh flap: a new free flap concept based on the septocutaneous artery. *Br J Plast Surg* 37(2): 149-159.

van der Schot A, Sikkel E, Niekolaas M, Spaanderman M, de Jong G (2023). Placental Vessel Segmentation Using Pix2pix Compared to U-Net. *J Imaging* 9(10).

Vijayasekaran A, Gibreel W, Carlsen BT, Moran SL, Saint-Cyr M, Bakri K, et al. (2020). Maximizing the Utility of the Pedicled Anterolateral Thigh Flap for Locoregional Reconstruction: Technical Pearls and Pitfalls. *Clin Plast Surg* 47(4): 621-634.

Abstract in Korean

딥러닝으로 하지 전산화단층촬영 영상에서 악안면 재건에 필요한 대퇴피판의 천공지를 감지하는 연구

전외측 대퇴부 유리피판은 외상, 질병 또는 선천적 기형으로 인해 해부학적 결손이 있는 경우 재건을 위해 널리 이용하고 있는 유리 피판이다. 전외측 대퇴부 유리피판의 이식을 성공적으로 하기 위해 산소와 혈액을 공급하는 천공지 혈관이 반드시 필요하다. 본 연구는 전산화 단층촬영 혈관 조영술 이미지에서 전외측 대퇴부 유리피판의 천공지 혈관을 정확하게 탐지하기 위한 인공지능 모델을 개발하는 것을 목표로 하였다. 이미지 데이터로 전외측 대퇴부 유리피판 이식 수술을 받은 환자의 53개 전산화 단층촬영 혈관조영술 이미지를 사용하였다. 이 중 48개는 학습용으로 사용하였고 5개는 테스트용으로 사용하였다. 두 명의 구강악안면외과 전문의가 수동으로 주석 처리한 정보를 이용하여 nnUNet 모델이 딥러닝을 수행하였다. 본 연구에서는 수술로 확인된 천공지 혈관을 학습한 nnUNet one과 방사선학적으로 추가로 확인된 천공지 혈관을 학습한 nnUNet adv라는 두 가지 모델을 개발하였다. nnUNet one보다 nnUNet adv가 내부 검증, 외부 검증에서 Dice, IoU, Precision, Recall 모두 높은 수치를 보였다. nnUNet adv는 천공지 혈관 식별에 대한 성능 평가를 위한 외부 검증에서 Dice 점수 0.682와 71.4% 탐지율을 보였다. 본 연구에서 개발한 nnUNet adv는 외과의사의 전산화 단층촬영 혈관조영술의 판독 시간을 줄이고 수술 전 다양한 유리피판의 모양을 술자가 설계하는데 도움을 줄 수 있는 성능으로 평가한다. 또한, nnUNet adv는 전문의가 판독에 실패한 천공지 혈관을 탐지할 수 있는 가능성을 보였다. 향후 학습 데이터 세트를 추가하면 모델의 정확성과 신뢰성이 더욱 향상될 것으로 기대된다.

핵심되는 말 : 전외측 대퇴부 유리 피판, 천공지 혈관, 인공지능, 딥러닝, 컴퓨터 단층촬영 혈관조영술, 악안면부위, 재건술