

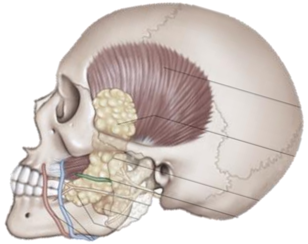

Can AI Models See What Anatomists See?

*Identifying Deep Learning Architectures Optimized for **Oral Ultrasonography**: A Comparative Study in **Buccal Mucosa Classification***

Goun Jung, Yu-Ran Heo, Hee-Jin Kim

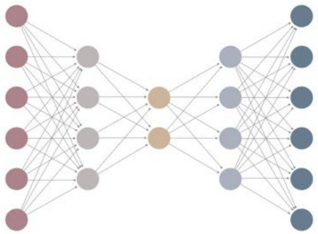
Division in Anatomy and Developmental Biology, Department of Oral Biology
Human Identification Research Institute, BK21 FOUR Project
Yonsei University College of Dentistry

Opening 4 Keywords

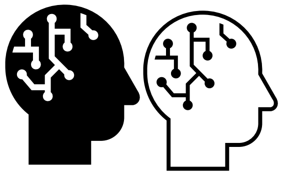


Infraoral Anatomy

Ultrasonography



Artificial Intelligence



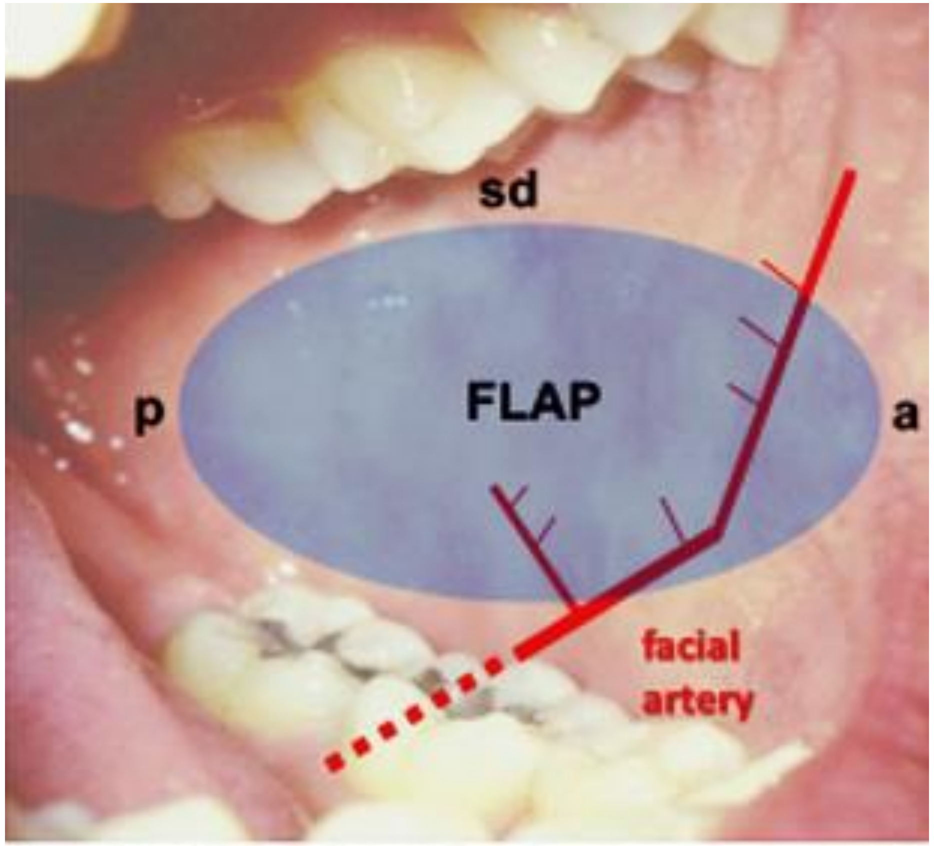
Explainability

"Can deep learning models classify upper and lower buccal mucosa regions from ultrasound images in a way that aligns with anatomical understanding?

Background

Anatomical Significance of the Buccal Mucosa

- Forms the **Buccinator Musculo-Mucosal Flap (BMMF, Bozola Flap)** with the buccinator muscle
- First choice for the reconstruction of defects in oral cavity, oropharynx, and nasal septum
- The **BMMF** is an extremely versatile 'like for like' local flap option due to its long arc of rotation
- Receives blood supply from fascial artery and buccal artery: Classified as an axial flap: classified as an axial flap

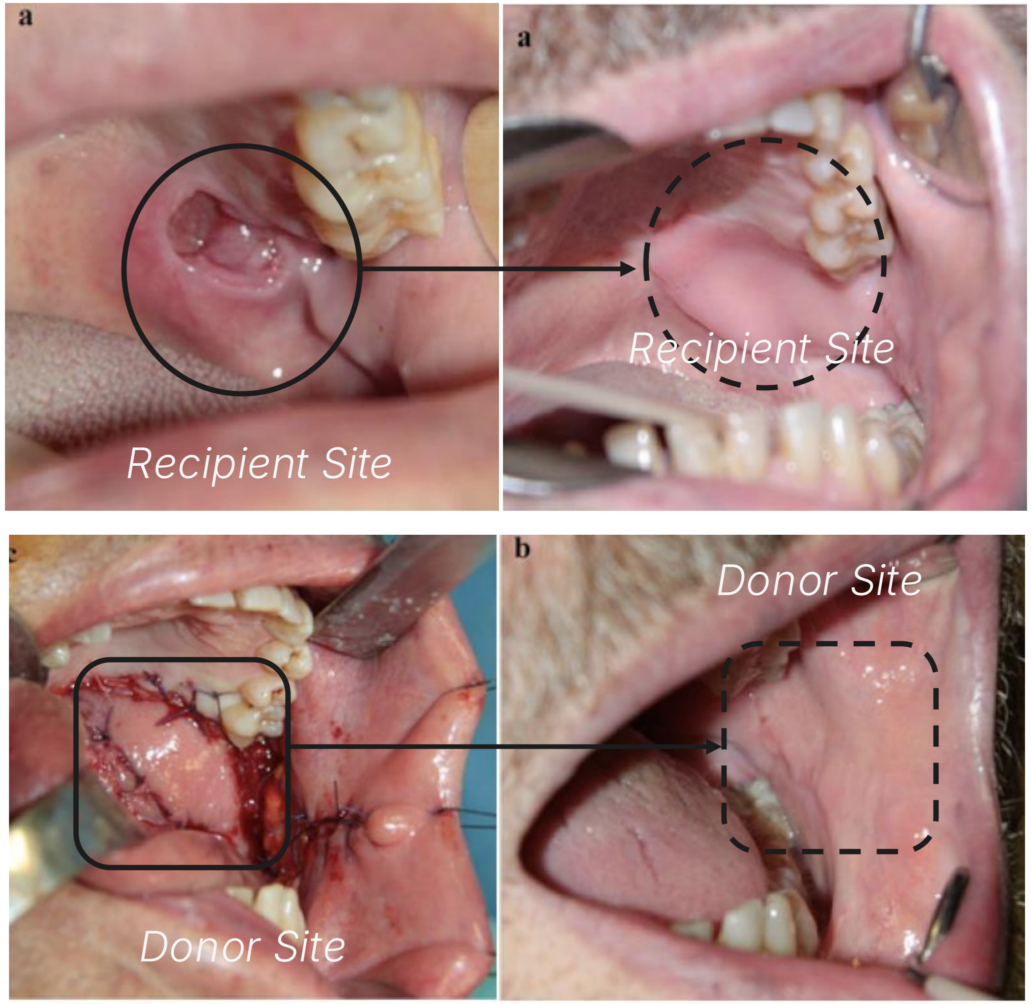


a long arc of rotation of BMMF: 2013, K.Khan et al.

Background

Clinical Advantages of the Buccal Mucosa

- Provides mucosal coverage, not cutaneous, and also maintains sensation
- A valuable reconstructive option that can cover extensive areas while maintaining vascular supply
- Donor site can be closed primarily with minimal deformty or scarring



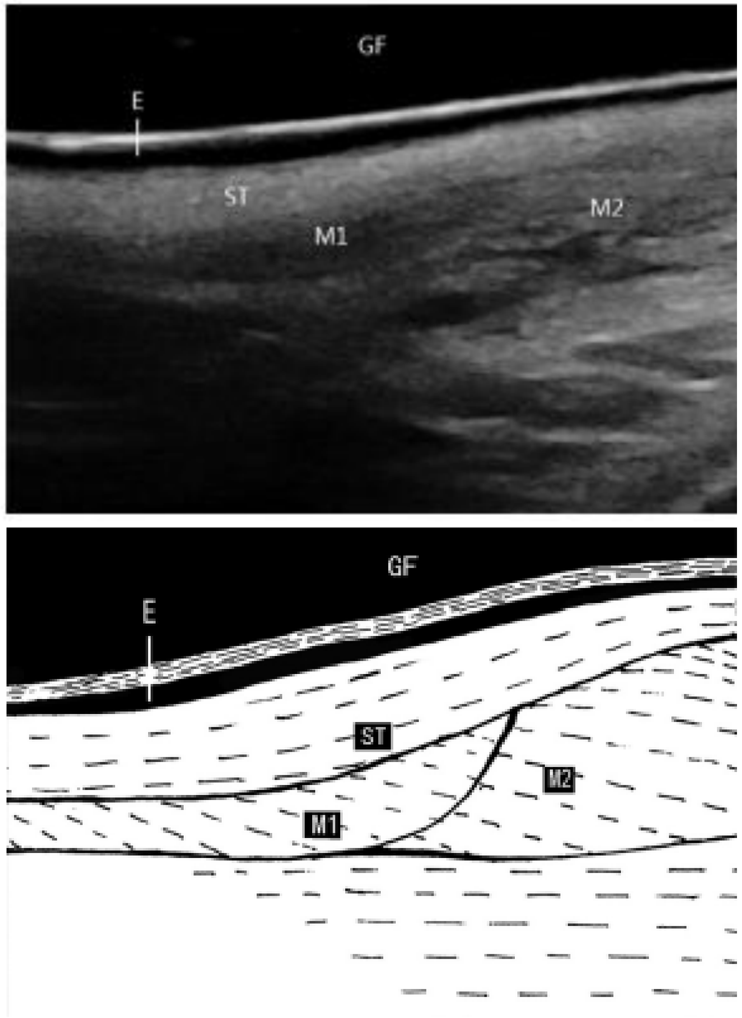
Nine-Month Post-surgery: Palatal Mucoepidermoid Carcinoma Site Reconstruction with BMMF - Surgical and Donor Site Healing: 2021, Sesenna et al.

Background

Preoperative Challenge

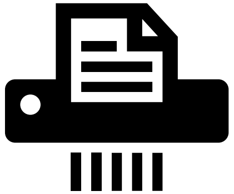
- Precise identification of **anatomical structures like arteries or duct** is critical
- **Ultrasound is a non-invasive tool**, but interpretation is difficult
    - Anatomically ambiguous region
    - Few previous studies
    - High subjectivity in image reading

*An **objective, AI-assisted approach** is needed to assess vascular anatomy and flap viability.*



presents unique challenges due to complex tissue structures and low contrast in the oral region: 2018, Yiqun Liu et al.

Research Objective

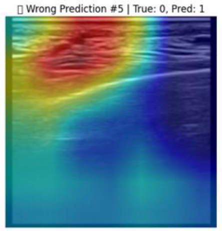


Automated Classification

To classify buccal mucosa ultrasound images into superior and inferior regions using deep learning

Model Optimization

To improve model robustness and generalization via hyperparameter tuning and patient-level data separation



Visual Explainability

To visualize model attention using Grad-CAM and evaluate anatomical focus

Anatomical Validation

To validate model predictions against known anatomical structures, ensuring clinical interpretability

Methods Workflow Overview

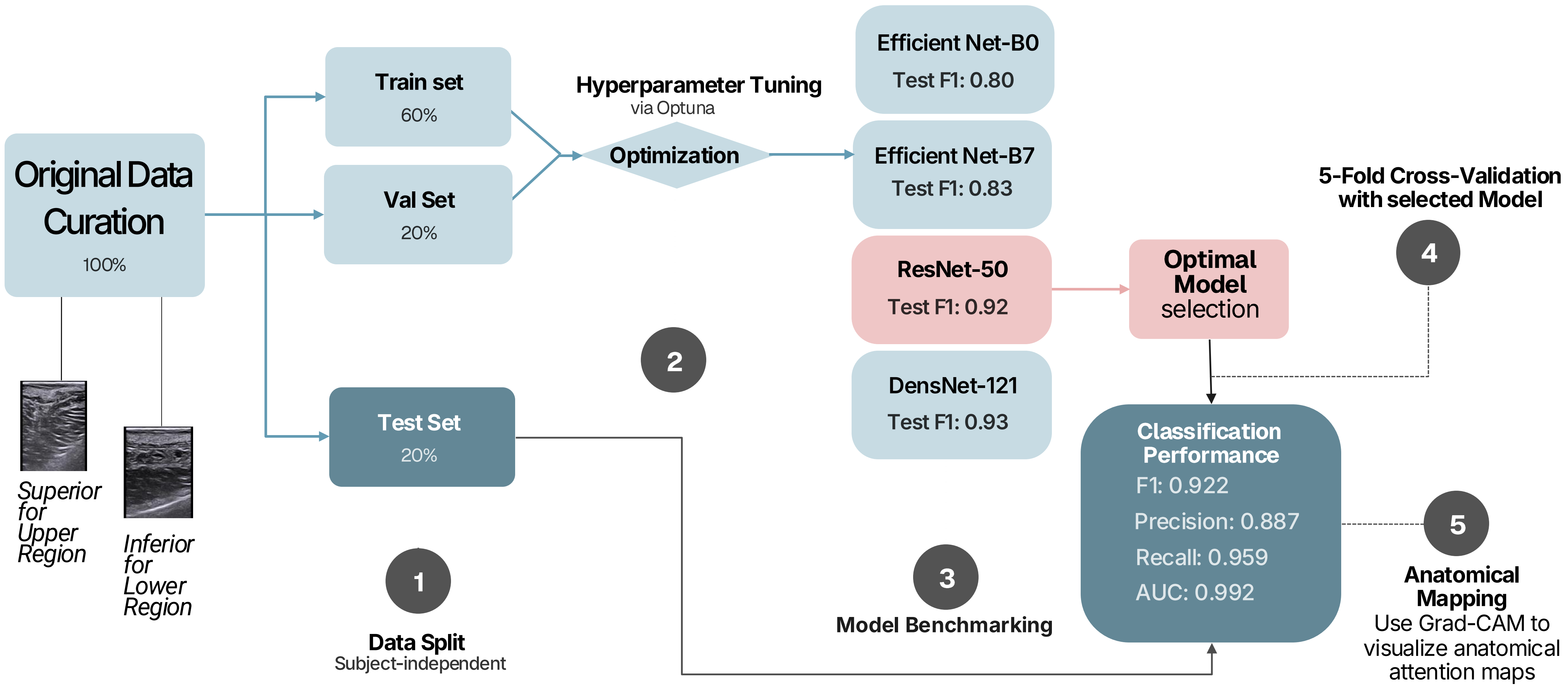
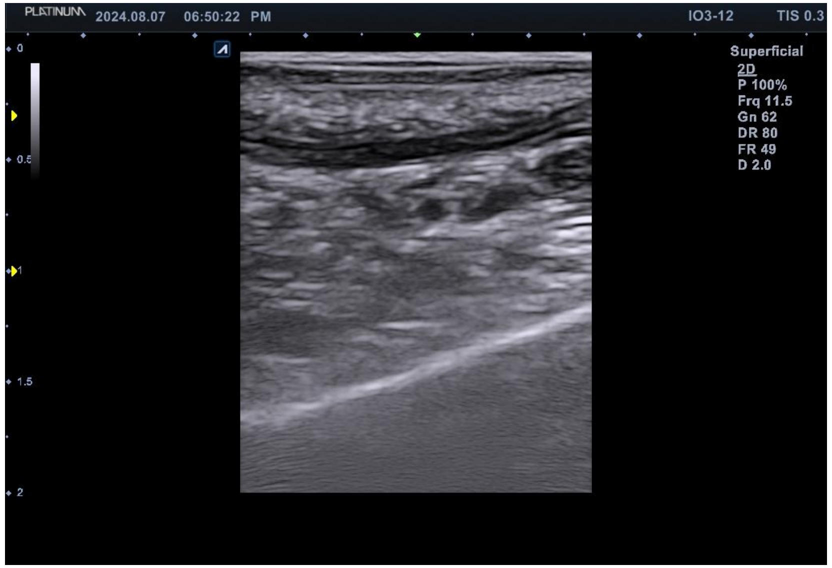
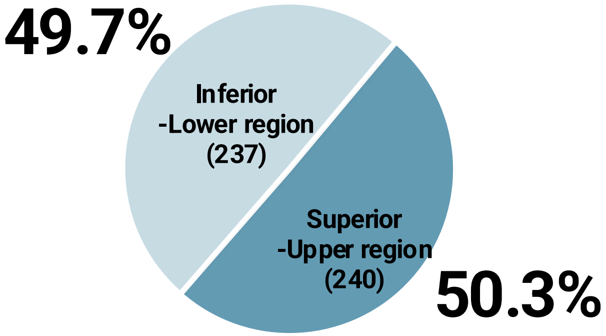


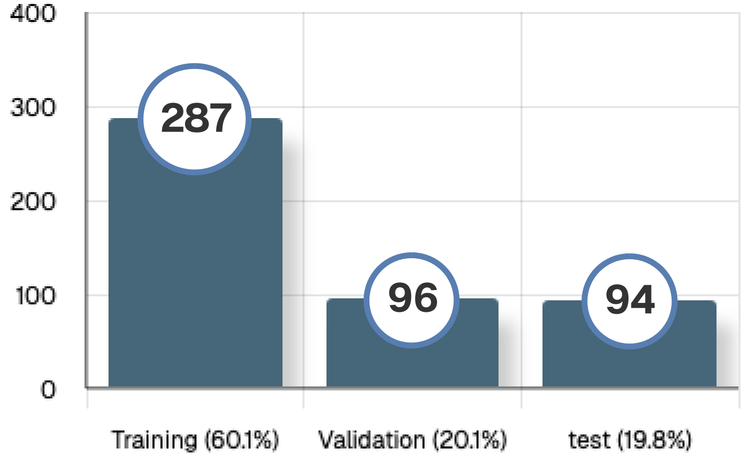
Image Set Description

477 intraoral ultrasound images
collected from 40 volunteers





✓ Less than 1% difference between classes



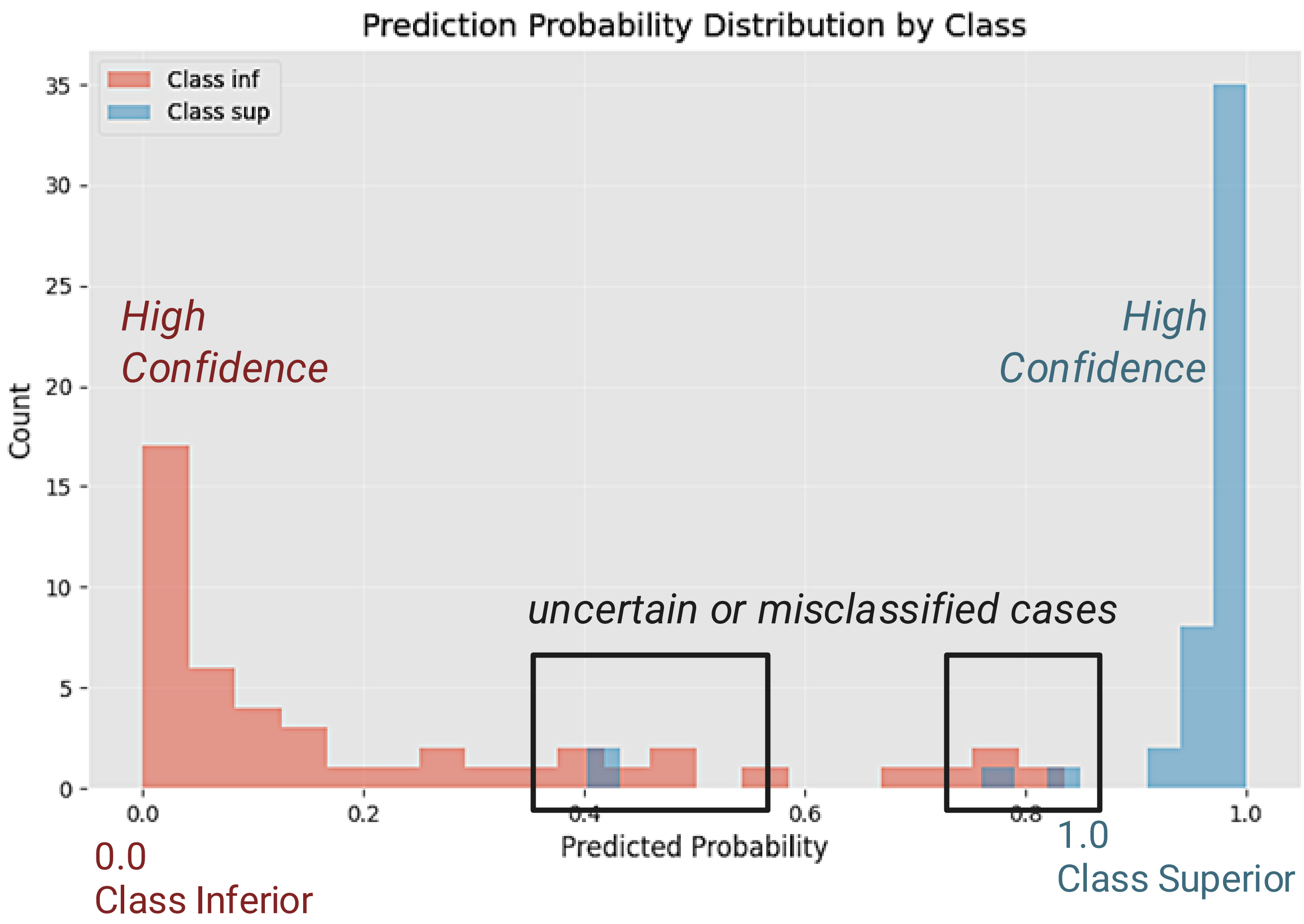
- ***Subject-independent evaluation***

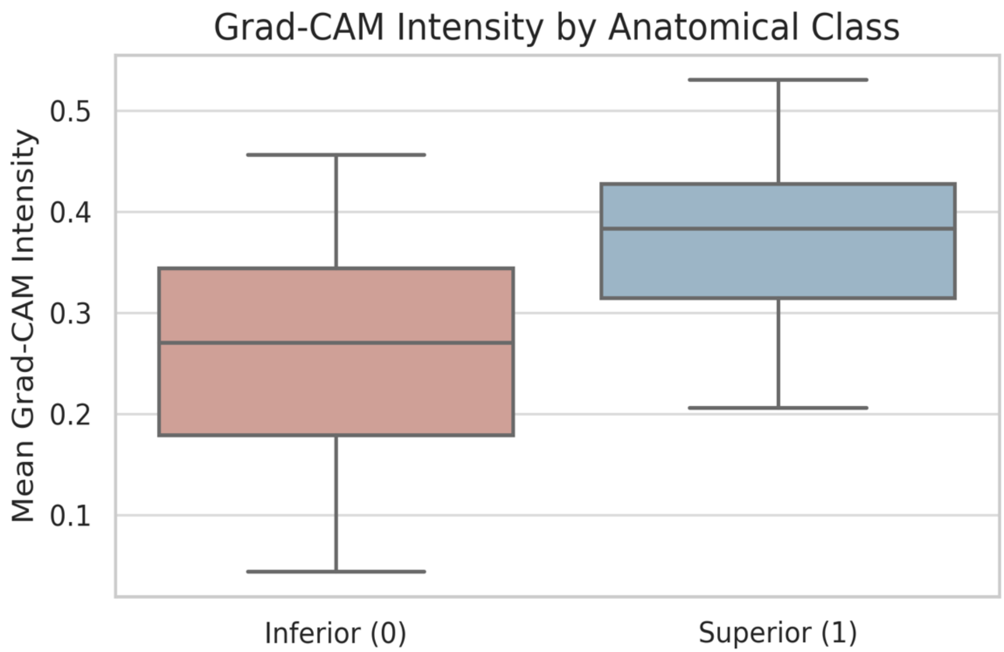
  No patient overlap across training, validation, and test sets.

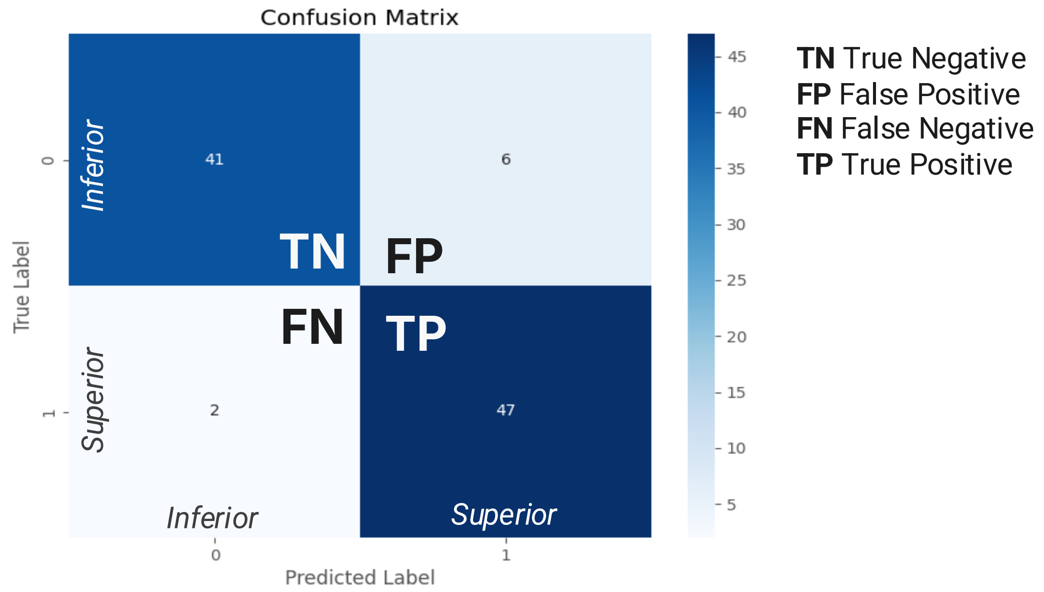
  To ensure generalizability and prevent data leakage.

*Obtained IRB Approval (2024-2-0006)

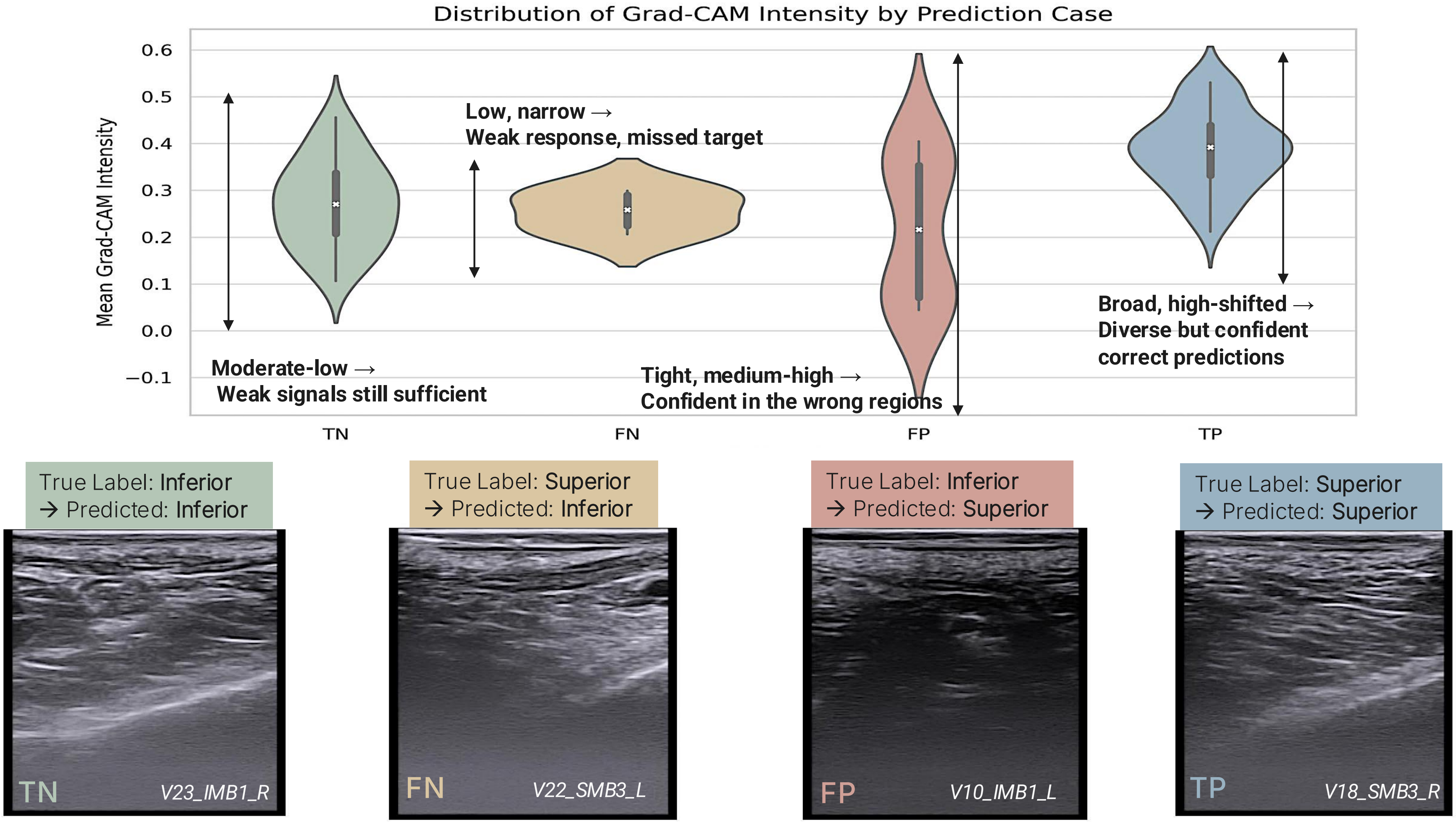
Classification **Results**





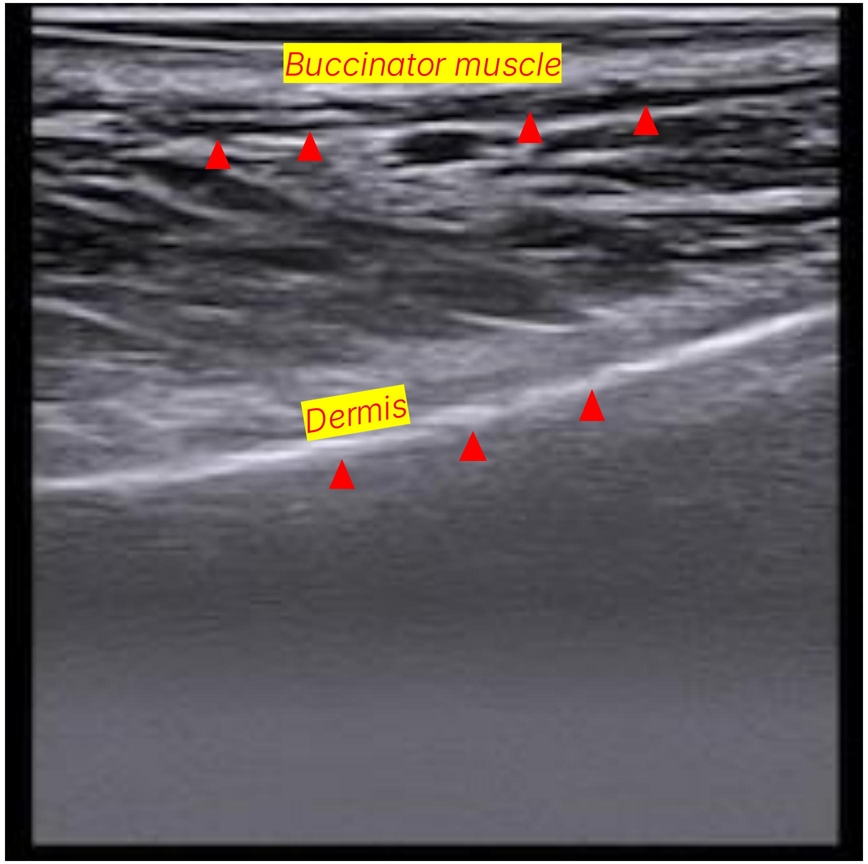


Classification **Results**

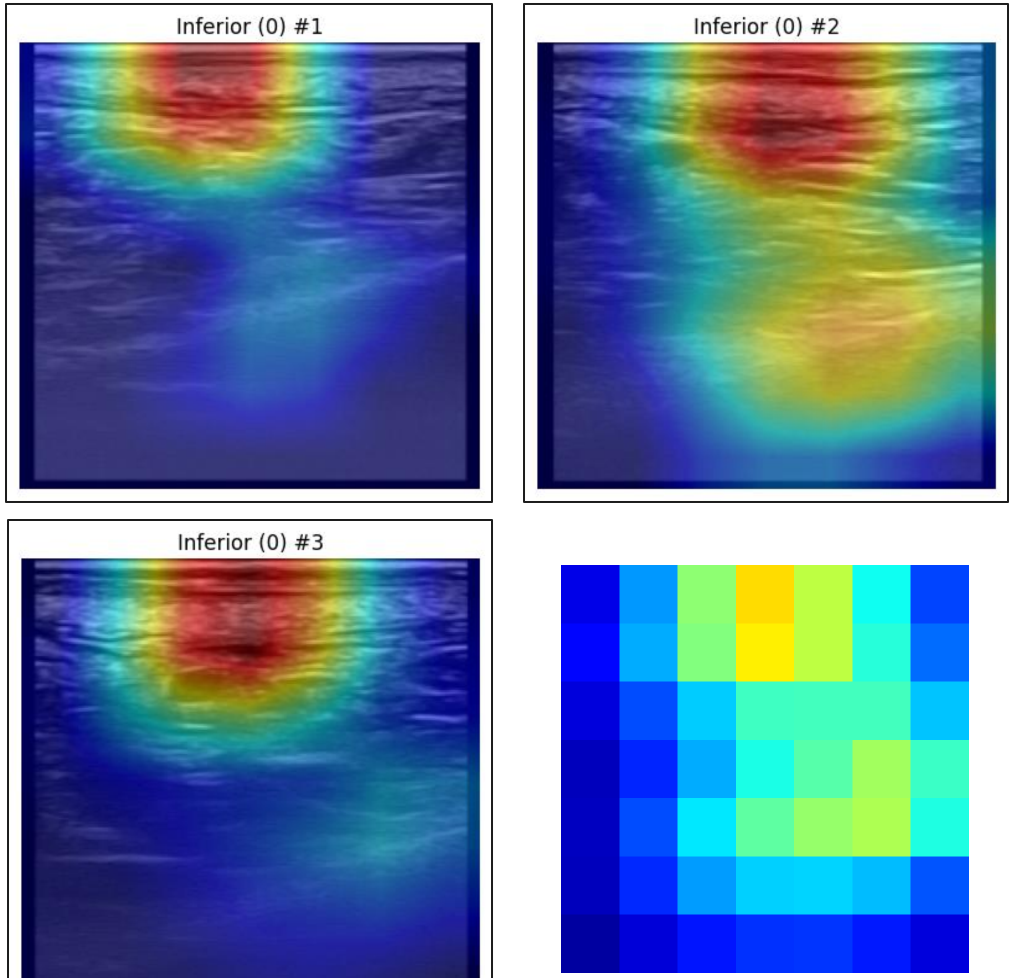


Grad-CAM **Results**

Representative Attention Maps



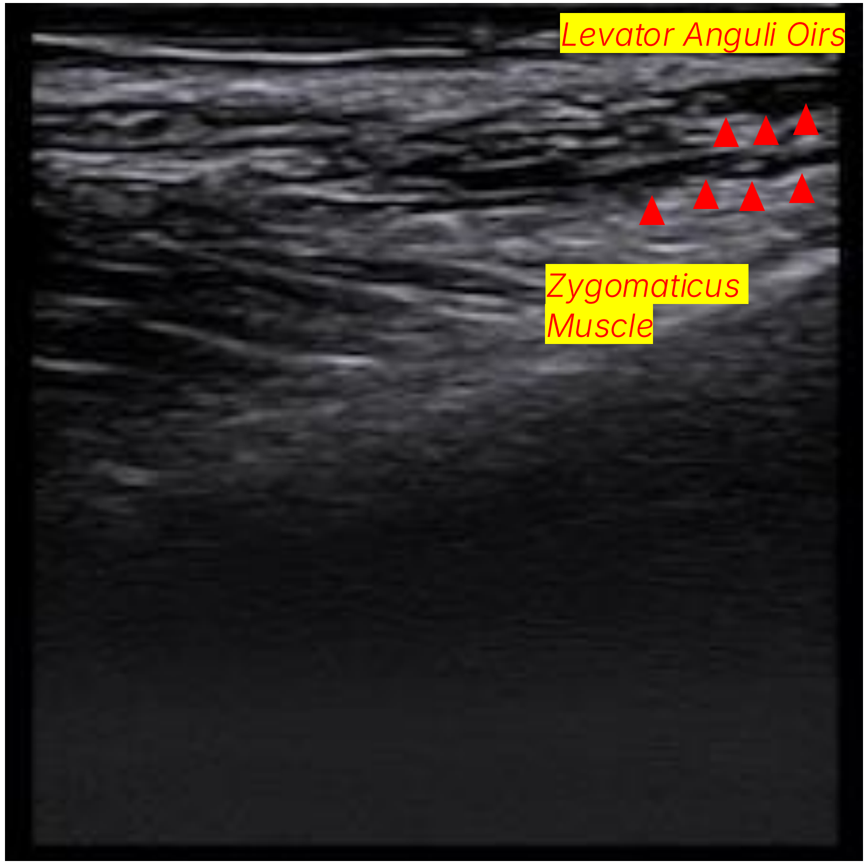
IO3_IBM3_R



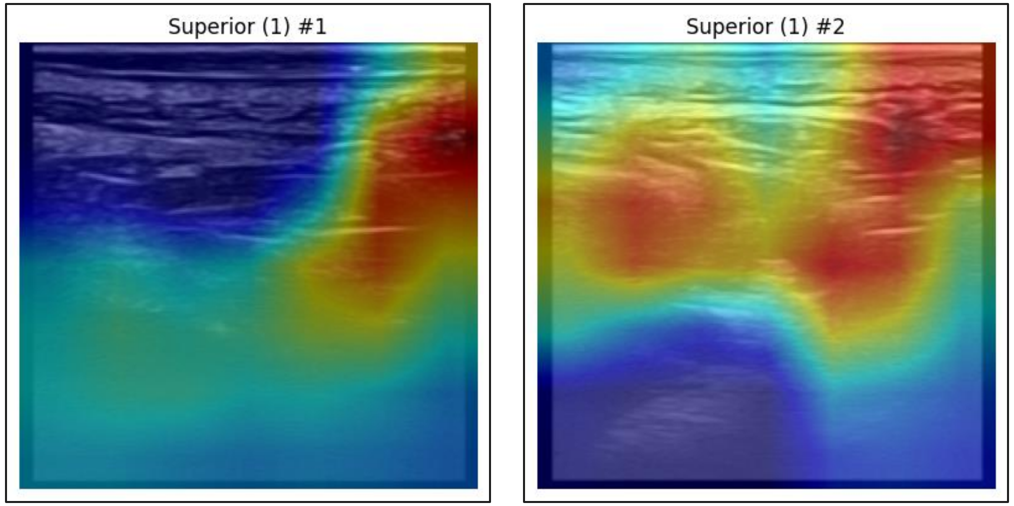
AvgCAM_inferior

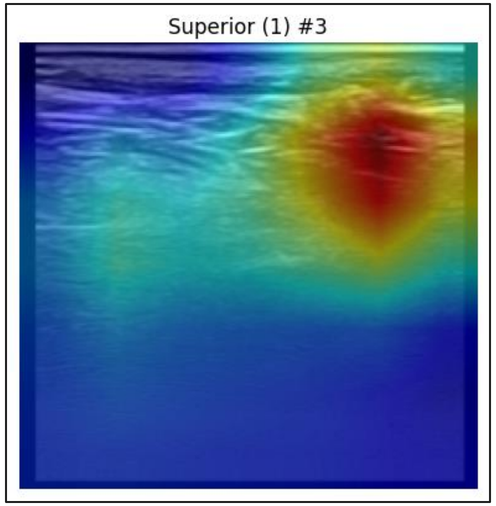
Grad-CAM **Results**

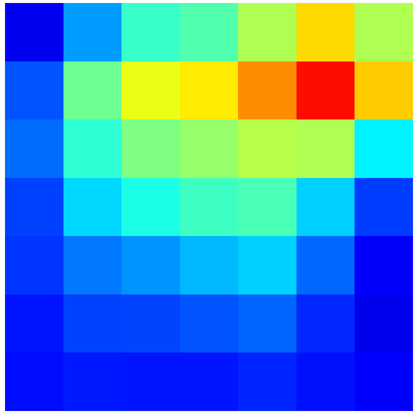
Representative Attention Maps



V6_SBM3_L



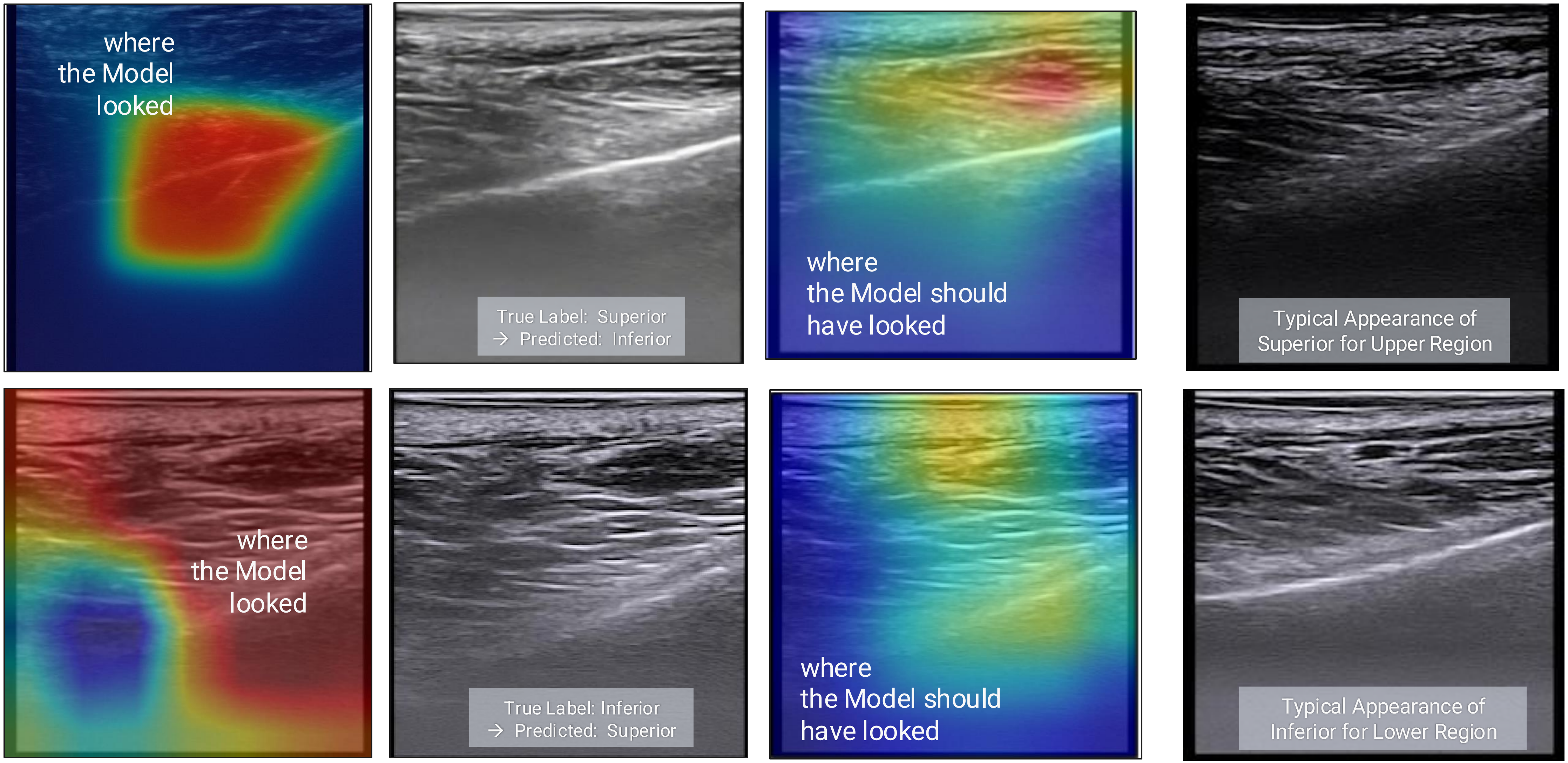




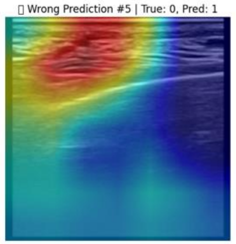
AvgCAM_superior

Grad-CAM **Results**

Misguided Model Attention



Conclusion



Deep learning model successfully classified buccal mucosa ultrasound images

Grad-CAM intensity patterns offer insight into model behavior and error characteristics.

Potential applications in various clinical domains with Detection of key anatomical structures

" By aligning AI attention with anatomical expertise, we demonstrate how explainable deep learning can advance ultrasound interpretation and clinical decision-making."

You might ask...

❓*Why CNN Models? Why ResNet-50 over DenseNet-121?*

❓*No augmentation. Why?*

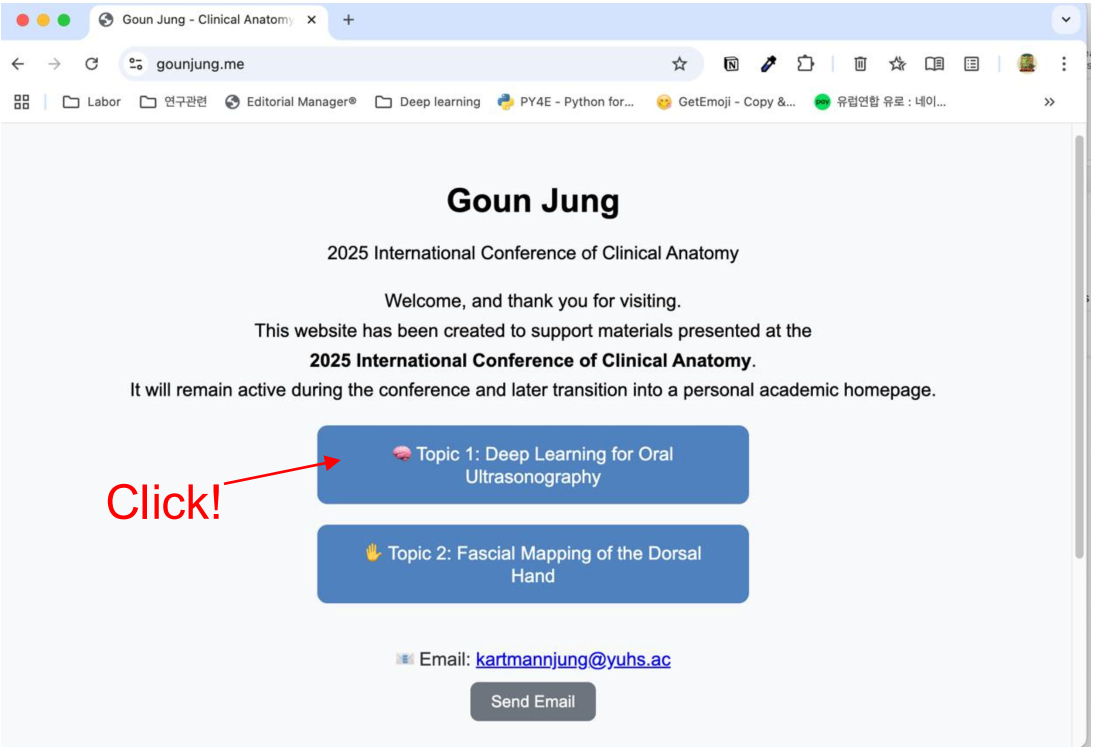
❓*Why Grad-CAM?*

❓*Ultrasound-Specific Layer?*

❓*Overfitting Problem?*

❓*What is Optuna?*

❓*Why F1?*



gounjung.me

• *In the age of AI, curiosity is our most powerful tool.*

• *Data is fuel, AI is the engine— but direction comes from us.*

You might ask…

- ❓ *Why CNN Models? Why ResNet-50 over DenseNet-121?*
- ❓ *No augmentation. Why?*
- ❓ *Why Grad-CAM?*
- ❓ *Ultrasound-Specific Layer?*
- ❓ *Overfitting Problem?*
- ❓ *What is Optuna?*
- ❓ *Why F1?*

QnA

❓ Why CNN Models? Why ResNet-50 over DenseNet-121?

We chose CNN models as they have proven effective in medical image classification, especially where anatomical structures need to be spatially interpreted.

Among the tested architectures, **ResNet-50** provided a good balance between **model complexity, training time, and interpretability**, which was critical for our Grad-CAM–based explainability.

Although **DenseNet-121** showed slightly higher accuracy, ResNet-50 was chosen for final Grad-CAM visualization because:

- Its skip connections make attention maps more localized and interpretable.
- It had more **stable convergence** during cross-validation.
- And it is **widely adopted in clinical imaging research**, supporting reproducibility.

❓ No augmentation. Why?

We deliberately **did not apply data augmentation** in this experiment because:

- The dataset is **anatomically sensitive**. Flipping, rotation, or shifting could disrupt spatial orientation (e.g., left vs. right cheek, superior vs. inferior).
- Our **primary goal was model interpretability**, not just accuracy.
- We focused on **inherent spatial patterns**, as learned from unaltered, real patient data.

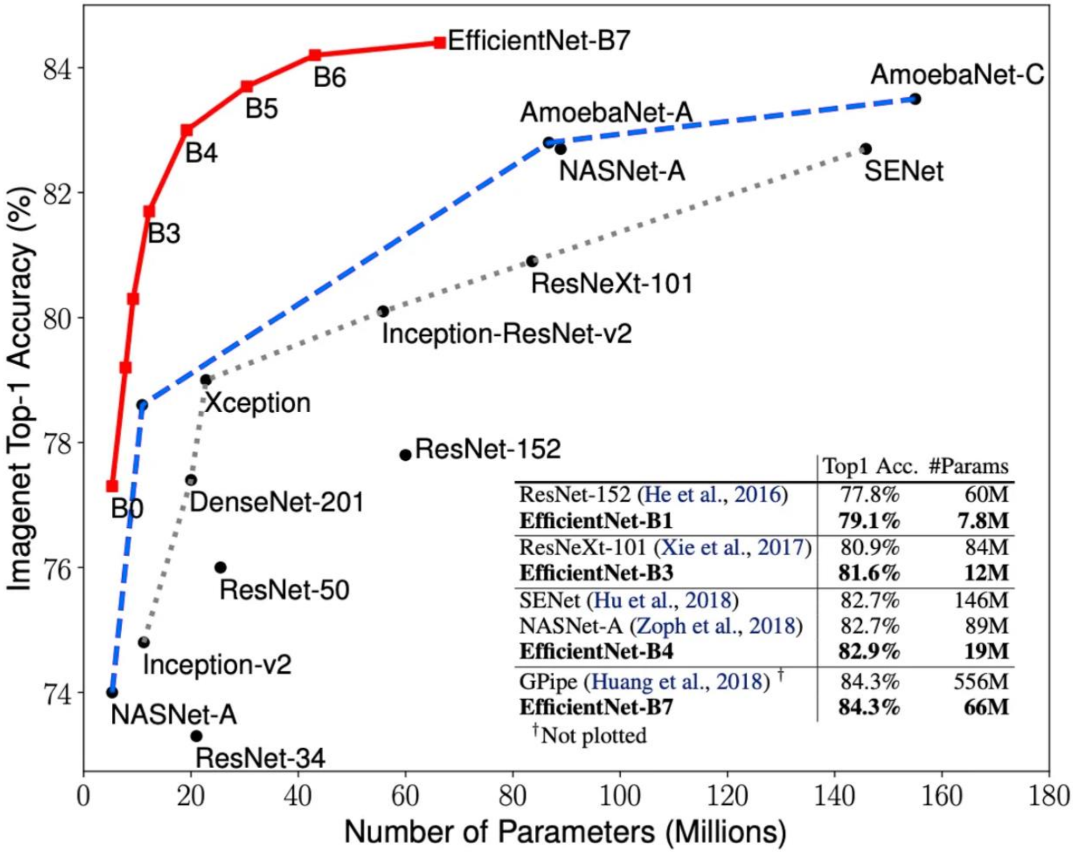
In future work, domain-specific augmentation strategies could be explored—such as **intensity variation or speckle simulation**.

gounjung.me

- *In the age of AI, curiosity is our most powerful tool.*
- *Data is fuel, AI is the engine— but direction comes from us.*

Why these 4 Models?

https://medium.com/@enrico.randellini/image-classification-resnet-vs-efficientnet-vs-efficientnet-v2-vs-compact-convolutional-c205838bbf49



◆ **ResNet-50**
is a well-established baseline model: moderately sized and widely used in medical imaging.

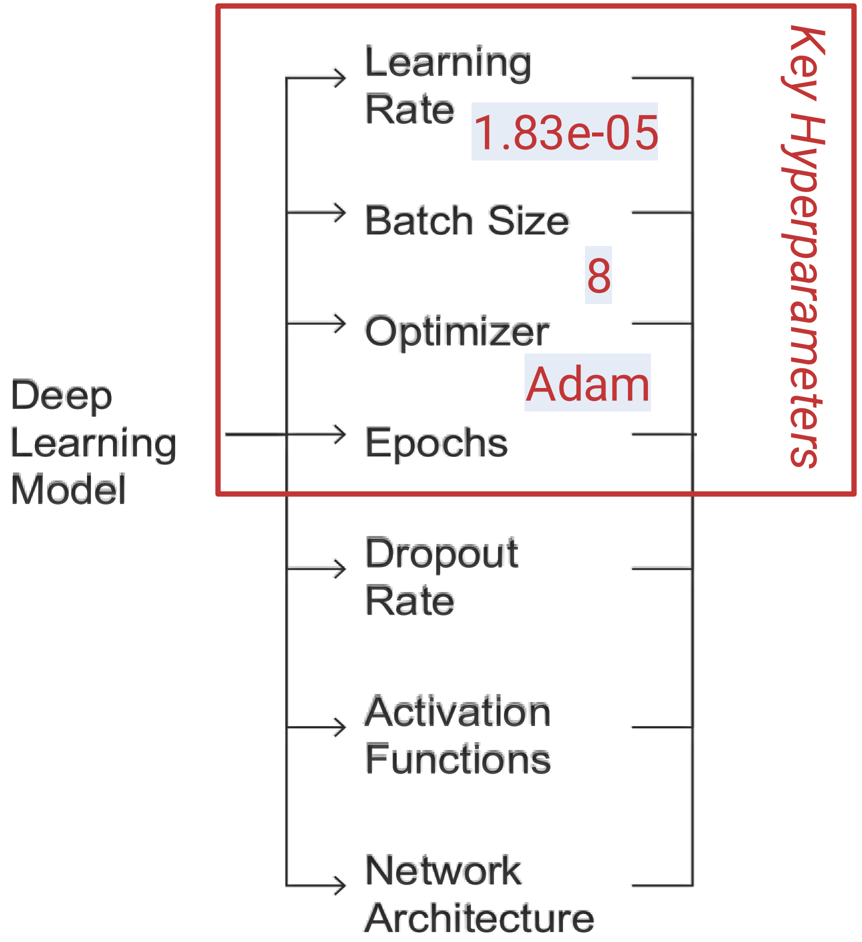
◆ **DenseNet-121**
offers improved accuracy with better feature reuse, though at higher computational cost.

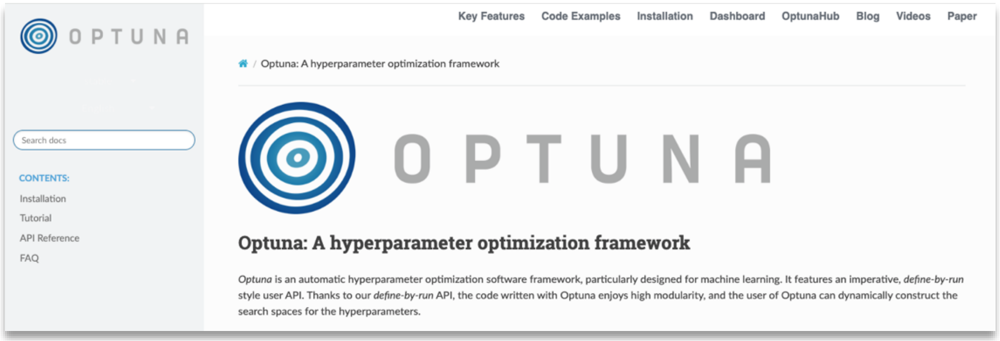
◆ **EfficientNet-B0**
was selected for its **small size and lightweight architecture** : useful in clinical deployment scenarios.

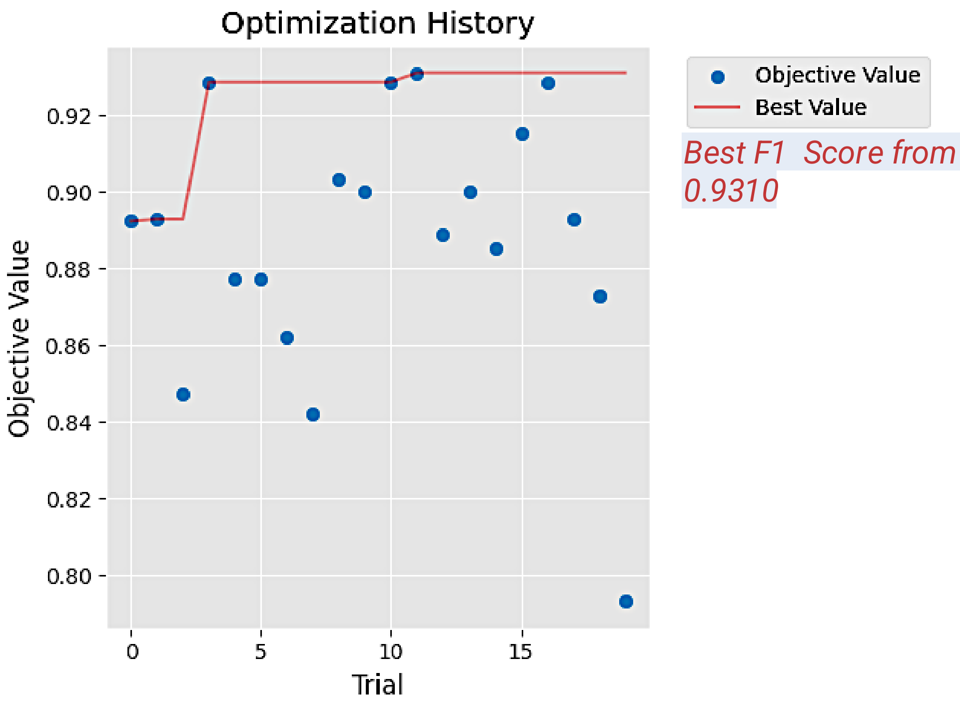
◆ **EfficientNet-B7**
on the other hand, represents a **high-end accuracy-focused model**, allowing us to test the upper bound of performance.

Hyperparameter Optimization **Results**

Deep Learning Hyperparameters







Best F1 Score from Optuna Trial 0.9310