SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 11, No. 3.1, (2025), pp. 247-259
Open Access. Online & Print

DOI: 10.5281/zenodo.11032521

COGNIDENTHISTO: EMPOWERING ORAL HISTOLOGY EDUCATION WITH AI-POWERED MOBILE APPLICATION FOR ORAL HISTOLOGY MICROSCOPIC IMAGE SEGMENTATION AND DATASET AUGMENTATION

Musab Hamed Saeed[1,2], Khouloud Samrouth[3], Prathibha Prasad[2,4,5*], Al-Moutassem Billah Khair[2,4], Vijay Desai[1,2], Nader Bakir[6], Zaynab Marhaba[3]

[1] Clinical Dental Sciences, College of Dentistry, Ajman University, Ajman, United Arab Emirates.https://orcid.org/0000-0002-3564-184X, m.saeed@ajman.ac.ae
[2]Center for Medical and Bio-Allied Health Sciences Research, Ajman University, Ajman, United Arab Emirates.https://orcid.org/0000-0003-2696-2024, k.elsamrout@ul.edu.lb
[3]Lebanese University, Beirut, Lebanon.https://orcid.org/0000-0003-0371-2903, p.prasad@ajman.ac.ae
[4]Basic Medical and Dental Sciences, College of Dentistry, Ajman University, Ajman, United Arab Emirates.https://orcid.org/0000-0002-1807-6615, a.khair@ajman.ac.ae, https://orcid.org/0009-0004-6527-902X, zaynab.marhaba@st.ul.edu.lb
[5] Saveetha Institute of Medical and Technical Sciences, Chennai, India.https://orcid.org/0000-0003-3256-4778, v.desai@ajman.ac.ae
[6]Beirut Arab University, Beirut, Lebanon. https://orcid.org/0009-0003-5786-4248, nbakir@bau.edu.lb

Received: 28/05/2025
Accepted: 28/08/2025

Corresponding Author: Prathibha Prasad (p.prasad@ajman.ac.ae)

ABSTRACT

The segmentation of oral histology images presents a significant challenge in dental research and education due to the scarcity of high-quality labeled datasets and the inherent complexity of multi-class tissue structures. Manual annotation of histological slides is a labor-intensive and expert-driven process, making large-scale data acquisition difficult. To address this limitation, we propose an educative AI-Powered Mobile application 'CognidentHisto' that empowers oral histology education with a dataset of 6122 annotated images and a multi-model deep learning framework for supervised multi-class segmentation of microscopic oral histology images. In particular, we propose in this paper 4 main contributions. First, we construct a custom annotated proprietary dataset of 32 images using Roboflow. Second, we extensively augment the dataset using 6 combined morphological transformations to get a total of 6122 annotated images. Then, we generate annotations in COCO-format along with gray-level masks to support pixel-level class differentiation across diverse anatomical structures. Third, we train using Tensorflow multiple convolutional neural network (CNN)-based architectures, including U-Net, Mask R-CNN, SAM, DeepLab, MobileNet, Yolov8, and transfer learning from microscopy-specific models for the segmentation task. We evaluate their performance using standard segmentation metrics such as mean Intersection-over-Union (IoU), Dice Coefficient, Pixel Accuracy, and mean Average Precision (mAP). Forth, we develop a Flutter-based mobile application to extend practical usability. This application enables students and faculty to interact with the system through features such as segmentation testing and institutional announcements. Our work establishes a modular, scalable, and user-centric foundation for advancing AI-assisted analysis in dental histology.

KEYWORDS: Oral Histology, Microscopic Images, Segmentation, Deep Learning, Artificial Intelligence.

1. INTRODUCTION

The rise of artificial intelligence (AI) significantly influences healthcare and biomedical research, offering innovative solutions for diagnostic and educational applications (Alexakis et al., 2022). In the field of dentistry, one promising application lies in the segmentation of oral histology images a task that is essential for training dental students and supporting digital pathology workflows. However, automated histological image analysis remains a complex challenge, primarily due to the limited availability of high-quality labeled datasets and the intricacies of multi-class tissue structures (Alexakis et al., 2022). Manual annotation of histological slides is a resource-intensive process, requiring domain expertise and considerable time investment. These challenges are further amplified when developing supervised learning models, which traditionally rely on large-scale, well-annotated datasets for accurate training. Moreover, existing deep learning approaches often exhibit poor generalization across unseen datasets or fail to scale effectively to multi-class segmentation problems, where numerous anatomical regions must be distinguished with high precision (Alexakis et al., 2022). These limitations restrict the usability of current solutions, particularly in educational environments where accessibility and flexibility are key. In response to these issues, we propose a comprehensive deep learning-based pipeline for multi-class segmentation of microscopic oral histology images. The project begins by constructing a custom dataset through manual annotation using Roboflow, applies extensive data augmentation techniques, and generates COCO-format masks. Our approach systematically explores a variety of convolutional neural network (CNN) architectures including UNet, Mask R-CNN, SAM, DeepLab, MobileNet, Yolov8, and transfer learning models pre-trained on microscopy datasets to identify architectures best suited for oral histology segmentation tasks. Unlike traditional approaches that rely on a single model or rigid augmentation strategy, our framework is designed to flexibly accommodate various training scenarios and model configurations. All models are implemented using TensorFlow, with input normalization and class-aware mask construction forming the foundation of our pre-processing stage. To assess model segmentation capabilities, we evaluate them using standard metrics such as mean Intersection-over-Union (IoU), Dice Coefficient, Pixel Accuracy, and mean Average Precision (mAP), although final results remain in progress. To extend this work beyond experimentation, we develop a Flutterbased Cross-Platform Educative Application, Cognident Histo, accessible for faculty and students at the school of dentistry. The application provides segmentation testing capabilities and allows users to interact with institutional content such as announcements, positioning the system as both an academic tool and a practical deployment. In Section 2, we review relevant literature and segmentation approaches in medical imaging. Section 3 describes our dataset construction and model implementation. Section 4 outlines our experimentations and results. Section 5 presents the design and features of the application. Finally, Section 6 concludes the paper and discusses future directions.

2. RELATED WORKS

Early methods for segmenting histological images relied heavily on classical image processing and mathematical morphology techniques. Approaches such as Otsu's thresholding, the watershed algorithm, and graph cuts were widely used to separate anatomical structures like nuclei and connective tissues. While these traditional methods offered useful early solutions, they struggled with noisy, overlapping, and highly variable staining patterns typical in histology slides, often requiring extensive pre- and post-processing. The emergence of deep learning significantly advanced histological image segmentation. Convolutional Neural Networks (CNNs), particularly architectures such as U-Net and its extensions, enabled models to automatically learn multi-scale features from raw data. These methods demonstrated remarkable improvements in accurately segmenting complex tissue 3 structures. Furthermore, advanced CNN models like HookNet introduced multi-resolution processing, capturing both fine-grained local information and broader contextual features. Recent studies have also shown the ability of CNNs to tackle multi-class segmentation tasks, such as distinguishing multiple tissue types in colorectal cancer histology. In the following subsections, we review traditional and CNN-based methods for histological segmentation in details in Table 1.

2.1. Traditional Methods

2.1.1. Otsu's Thresholding

Otsu et al. introduced an automatic thresholding method by maximizing inter-class variance to separate foreground (e.g., nuclei) from background regions. This approach laid the foundation for early binary segmentation workflows but remains limited when dealing with multi-class or overlapping tissue structures.

2.1.2. Watershed Algorithm for Nuclei Segmentation

The watershed algorithm applied morphological operations to segment clustered or touching nuclei. By treating the grayscale image as a topographic surface, watershed-based segmentation separated overlapping structures but required careful preprocessing, such as distance transforms, to avoid oversegmentation artifacts.

2.1.3. Graph Cuts for Histology Segmentation

Graph cuts, as presented by Boykov and Jolly, formulated segmentation as an energy minimization problem. This method allowed for more precise boundary detection compared to simple thresholding but often required manual seed point initialization, limiting its automation in large-scale histology workflows.

2.2. CNN-Based Methods

2.2.1. U-Net for Biomedical Segmentation

The U-Net model, proposed by Ronneberger et al. introduced a novel encoder-decoder architecture with skip connections specifically designed for biomedical image segmentation. U-Net remains one of the most widely adopted models for medical image analysis due to its strong performance even with limited datasets.

2.2.2. Multi-scale U-Net for Histology

A multi-scale U-Net architecture was later developed to enhance feature representation at different resolution levels. By processing inputs at multiple scales, the model better captured fine-grained details, which is particularly important in histology where structures can vary drastically in size.

2.2.3. HookNet: Multi-resolution Contextual Segmentation

HookNet proposed a multi-resolution CNN framework that links detailed local patches with global context extracted from lower resolution views. This dual-stream approach significantly improved segmentation performance on large whole-slide histology images by addressing both micro and macro structural patterns simultaneously.

2.2.4. VGG-16 + R-CNN for Tooth Segmentation

Alam et al. proposed a deep learning-based approach combining a pretrained VGG-16 convolutional architecture with an R-CNN detection network for tooth segmentation using optical radiographic images. This method achieved high accuracy in detecting individual teeth and numbering them, enhancing the precision of dental diagnostics in biomedical applications.

Table 1: Summary of Traditional and CNN-Based Segmentation Methods in Histology and Dental Imaging.

Reference	Approach	Techniques	Advantage(s)	Limitation(s)
Otsu (1979)	Thresholding	Global thresholding based on class variance	Simple and fast for binary segmentation	Fails with overlapping tissues and noise
Meyer (1994)	Watershed	Morphological distance transform and flooding	Good for separating clustered nuclei	Sensitive to noise; over-segmentation without preprocessing
Boykov & Jolly (2001)	Graph Cuts	Energy minimization with seeds	Precise boundaries; flexible for different shapes	Needs manual initialization; slow for large images
Ronneberger et al. (2015)	U-Net	Encoder-decoder CNN with skip connections	High accuracy with limited data; easy to adapt	May struggle with very large context in whole-slide images
Multi-scale U-Net (2018)	Multi-scale CNN	U-Net variant with multi-resolution features	Better captures fine details at different scales	Increased model complexity and memory usage
HookNet (2020)	Multi-resolution CNN	Local-global fusion network	Integrates local detail and global context; good for whole slides	Needs patch extraction; training is computationally expensive
Alam et al. (2023)	VGG-16 + R-CNN	Pre-trained CNN with object detection	Accurate tooth identification and numbering in X-rays	Requires high-quality radiographs and careful preprocessing

3. PROPOSED METHOD

3.1. Positioning of Our Work

While significant advances have been made in the field of medical image segmentation, many existing approaches depend on vast annotated datasets and sophisticated preprocessing pipelines, which increase development complexity and hinder accessibility. Our proposed method addresses these limitations by providing an efficient, structured workflow specifically tailored for oral histology image segmentation. By emphasizing careful annotation, comprehensive augmentation, subcategory-specific model training, and user engagement through a quiz-based Cross-Platform Educative Application, our method enables both high-precision segmentation and interactive educational value.

3.2. Rationale and Approach

In contrast to conventional approaches that often require large-scale datasets and complex model deployment, our methodology emphasizes practicality, adaptability, and user engagement. Recognizing the challenges associated with the scarcity of annotated histology datasets and the intricate nature of oral tissues, we develop a four-stage pipeline that includes annotation of the data set, augmentation of the data set, subcategory-based model training, and the development of a Cross-Platform Educative Application (Cognident Histo) with integrated quiz functionality. This design ensures not only robust segmentation performance across multiple tissue types but also provides users with an interactive tool to enhance their understanding of microscopic oral anatomy through quiz-based learning.

3.3. Proposed Method Overview

The overall structure of our proposed method is outlined in Figure 1 and **is composed of the following four main contributions**

1. Dataset Annotation;
2. Dataset Augmentation;
3. Model Training;
4. Cross Platform Application Development.

First, we create detailed annotations for a dataset of oral histology images, carefully labeling them into 7 main categories and 23 subcategories. Next, we expand the dataset significantly using a combination of Roboflow augmentations and additional morphological transformations via Albumentations and Detectron2 libraries. Following this, we train individual deep learning models for each of the 23 subcategories, ensuring specialized and highly accurate segmentation results tailored to each tissue type. Finally, we develop a user-friendly cross-platform educative application that not only facilitates image segmentation but also includes an interactive quiz component to test the user's understanding of the subcategories, enhancing both usability and educational engagement. Each block is described in greater detail in the following sections.

3.4. Dataset Annotation

The annotation process constitutes the first step in constructing a highquality training set. We manually annotate 32 original oral histology images using Roboflow, a widely adopted platform for computer vision annotation tasks. Annotations are made with fine granularity, categorizing each anatomical region into 7 primary categories and further into 23 distinct subcategories. The annotated dataset is exported in COCO format, providing a flexible and structured representation necessary for supervised segmentation model training. This foundational step ensures that the subsequent stages have access to rich and well-organized ground truth data.

3.5. Dataset Augmentation

To address the challenge of limited data availability and improve the generalization ability of the segmentation models, we perform extensive data augmentation. Using Roboflow's free augmentation tools, we initially expand the dataset from 32 to 128 images. Subsequently, leveraging Roboflow's paid augmentation capabilities, we further expand the dataset to 869 images. To enhance diversity and simulate the variability commonly encountered in microscopic oral histology images, we apply a comprehensive set of transformations. These transformations are carefully selected to represent realistic scenarios such as variations in tissue orientation, staining inconsistency, and imaging artifacts.

- Horizontal and Vertical Flipping: Simulates different orientations of the histology slides due to manual or machine-driven image capture at various angles.
- 90 Rotations (Clockwise, Counter-Clockwise, Upside Down): Models the scenario where tissue samples are scanned or positioned in different rotational configurations.
- Free Rotation (–45 to +45): Adds finer-grained rotational variability to prevent the model from becoming biased toward any fixed angle.
- Cropping with Zoom (0% minimum zoom, 20% maximum zoom): Represents framing differences, including zooming in on specific tissue regions or partial captures during microscopy.
- Shearing (±10 horizontally and vertically): Introduces realistic geometric distortion that may occur due to slide misalignment or deformation during scanning.
- Grayscale Conversion (applied to 15% of images): Reflects inconsistency in staining quality or situations where color cues are absent or minimal.
- Saturation Adjustment (–25% to +25%):

Models differences in staining intensity or histological preparation between samples.

- Brightness Adjustment (–15% to +15%): Simulates lighting variations due to different microscopes or capture devices.
- Exposure Adjustment (–10% to +10%): Mimics underexposed or overexposed slides caused by inconsistent imaging settings.
- Gaussian Blur (up to 2.5 pixels): Accounts for loss of sharpness due to scanning imperfections, motion blur, or low-resolution captures.
- Random Noise Addition (up to 0.1% of pixels): Emulates digital noise commonly introduced by imaging sensors or compression artifacts.

All augmented data is systematically organized according to the original 7 categories and 23 subcategories to ensure structural integrity and consistency throughout the model training process. Illustrative example of these transformations will be presented in Figure 1 and Figure 2.

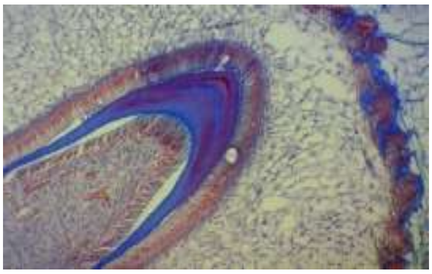


Figure 1: Original Apposition Image.

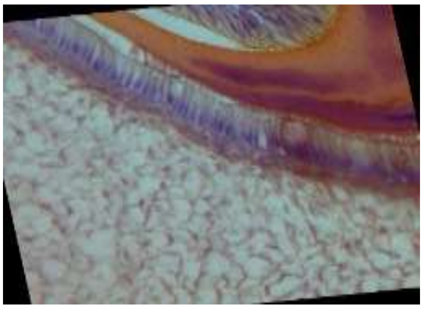


Figure 2: Appositon Image Augmented Image by Albumentations.

3.6. Model Training

Following data augmentation, we proceed with model training. To achieve highly specialized performance, we train a separate segmentation model for each of the 23 subcategories. We employ a variety of convolutional neural network (CNN) architectures to ensure robust segmentation across different tissue types. These architectures are selected based on their proven effectiveness in **biomedical and general image segmentation tasks**

- U-Net: A widely adopted architecture in medical imaging due to its encoder-decoder structure and skip connections, which allow precise localization. U-Net is well-suited for capturing fine-grained structures in oral histology images.
- Mask R-CNN: Known for its instance segmentation capability, Mask R-CNN enables object-level predictions and is ideal for separating overlapping anatomical regions and handling complex tissue boundaries.
- Segment Anything Model (SAM): As a prompt-based foundation model capable of zero-shot segmentation, SAM is used to explore generalization without extensive retraining. It provides a flexible and scalable solution for diverse image inputs.
- DeepLab: Equipped with atrous (dilated) convolutions, DeepLab excels at capturing multi-scale contextual information, making it highly effective for segmenting tissues that vary greatly in size and texture.
- MobileNet: A lightweight architecture optimized for mobile and resourceconstrained environments. It provides a balance between accuracy and computational efficiency, allowing for fast inference with reasonable segmentation quality.
- YOLOv8: Although originally designed for object detection, YOLOv8 integrates segmentation modules and provides real-time performance. It is useful for exploratory evaluation of speed-accuracy trade-offs in deployment scenarios.

Additionally, we incorporate transfer learning by using a U-Net model with an EfficientNet encoder, where the encoder is pre-trained on the ImageNet dataset. EfficientNet provides a strong feature extraction backbone, and the use of pre-trained weights allows the model to converge faster and generalize better, especially in a small-data regime. This pre-trained UNet variant is sourced from Kaggle and adapted to our specific oral histology task. By leveraging this diverse set of architectures and transfer learning strategies, we optimize model performance based on the complexity, granularity, and variability of each subcategory. We train the models using normalized input images, optimizing segmentation loss functions over multiple epochs. The resulting models demonstrate strong segmentation capabilities and generalize well to unseen histology samples.

3.7. Model Evaluation

To assess the quality and reliability of the trained segmentation models, we evaluate them using four commonly adopted metrics: Intersection over Union (IoU), Dice Coefficient (F1 Score), Pixel Accuracy, and mean Average Precision (mAP). These metrics are chosen due to their widespread use in medical image analysis and their ability to reflect different aspects of model performance.

- **Intersection over Union (IoU):** Also known as the Jaccard Index, IoU measures the overlap between the predicted segmentation **and the ground truth relative to their union**

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

Where TP is the number of true positive pixels, FP is false positives, and FN is false negatives. IoU is a robust metric for evaluating spatial alignment, and we compute both class-wise IoU and mean IoU across all subcategories.

- **Dice Coefficient (F1 Score):** The Dice score is particularly useful in medical segmentation due to its sensitivity to small structures. It measures the harmonic mean of precision and recall:

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

A Dice score of 1 indicates perfect overlap, while 0 means no overlap. It is ideal for evaluating class imbalance and precise tissue boundary prediction.

- **Pixel Accuracy:** This metric represents the proportion of correctly classified pixels over the entire image:

$$\text{Pixel Accuracy}=\frac{Number\ of\ Correct\ Pixels}{Total\ Number\ of\ Pixels}$$

While simple, pixel accuracy provides a general sense of model performance, particularly when paired with IoU or Dice for more detailed insight.

- **Mean Average Precision (mAP):** mAP is computed by evaluating IoU at multiple thresholds (e.g., 0.5 and 0.75), then averaging the resulting precision values:

$$\text{mAP}=\frac{1}{T}\sum_{t=1}^{T}1\,[IoU_t\ \geq\ threshold]$$

where T is the number of thresholds evaluated. This metric provides a comprehensive performance overview and is commonly used in object detection and instance-level segmentation models such as YOLOv8. These metrics together provide a comprehensive evaluation of segmentation quality: spatial overlap (IoU), boundary accuracy (Dice), global correctness (Pixel Accuracy), and threshold robustness (mAP). They allow us to assess each model's performance across the 23 subcategories of oral histology tissue with high fidelity.

3.8. *Cross-Platform Educative Application (Cognident Histo)*

The final component of the proposed system is the cross-platform educative application, Cognident Histo, designed to provide both segmentation functionality and educational interaction. The application is developed using Flutter and is deployed as both a mobile application and a desktop application, offering a smooth and responsive user experience across platforms. It allows users to upload new oral histology images, automatically apply the appropriate trained segmentation model, and visualize the predicted segmentation masks. In addition to segmentation, Cognident Histo integrates a quiz feature based on the 23 subcategories, prompting users to identify different anatomical structures. This interactive component reinforces learning, offering users immediate feedback on their understanding of oral histology. Thus, Cognident Histo serves both as an accessible deployment tool and as an educational platform that enhances user engagement with the segmented anatomical data.

4. EXPERIMENTS AND RESULTS

In this section, we present our development environment in Section 4.1. Section 4.2 describes the dataset used in our experiments, including the data collection, annotation, and augmentation process. Section 4.3 details the conducted experiments and implementation steps for each deep learning model. Then, in Sections 4.7 and 4.8, we present and discuss the results, respectively.

4.1. *Development Environment*

To develop and evaluate our segmentation models, we use Google Colab as the primary environment. Google Colab provides a free cloud-based Jupyter Notebook interface with access to limited GPU resources, which proves sufficient for training lightweight and medium-complexity CNNs. Our implementation utilizes several libraries including TensorFlow, OpenCV, Segmentation Models, and Albumentations. Data handling and training routines are scripted using Python 3, and key model checkpoints, logs, and results are saved to Google Drive.

4.2. *Data Set*

The initial dataset consists of 32 histology images manually annotated using the Roboflow web platform, which exports both the raw images and annotations in the widely adopted COCO format. The COCO (Common Objects in Context) format is a JSON-based annotation schema that provides structured representations for images, regions, and semantic classes. **Each annotation includes**

- Images: metadata for each image, including filename, width, height, and ID.

- Annotations: region-wise object information, **such as**

1. Image id: links the annotation to the corresponding image.
2. Category id: integer representing the anatomical class.
3. Segmentation: polygon coordinates outlining the annotated region.
4. Iscrowd: flag indicating whether the region represents a crowd.

- Categories: a mapping from category id to anatomical class names (e.g., gingiva, enamel, cementum). Illustrative Example. A simplified snippet of our COCO **annotation file is shown below**

```
{
  "images": [
    {
      "id": 1,
      "file_name": "image_01.jpg",
      "width": 320,
      "height": 320
    }
  ],
  "annotations": [
    {
      "id": 1,
      "image_id": 1,
      "category_id": 5,
      "segmentation": [[120, 80, 130, 90, 125, 100]],
      "iscrowd": 0
    }
  ],
  "categories": [
    {
      "id": 5,
      "name": "cementum"
    }
  ]
}
```

To facilitate model training, we convert these COCO polygon annotations into pixel-wise labeled masks, where each class is encoded as a unique integer value in a grayscale image. This rasterization is performed using a custom script that parses the COCO structure and generates dense masks.

To enhance the dataset, we apply multiple rounds of augmentation. First, we use Roboflow's free augmentation service, expanding the dataset to 128 image-mask pairs. Then, we use the paid Roboflow plan, which generates an additional 896 augmented samples. To further increase diversity, we apply Albumentations and Detectron2-based on-the-fly transformations to the 128-image set, producing 20 augmented versions per image, each with a corresponding mask, greatly improving model generalization.

Each pixel in the mask represents a semantic class, making this a multiclass segmentation task. A total of 59 anatomical structures are defined in the dataset; however, not every image contains all classes. To better specialize the models and improve segmentation accuracy, we train each model separately on specific subcategories. We split each subcategory dataset into 80% training and 20% validation. Due to severe class imbalance in most subcategories, we employ custom loss functions tailored to emphasize minority classes during training.

4.3. Implementation

In this section, we describe the implementation details of each architecture in our framework, including data pairing, loss design, CNN construction, and model-specific components.

4.3.1. Image-Mask Pair Generation

To prepare the data for supervised segmentation, we convert the COCO based polygon annotations into 2D masks, where each pixel's intensity corresponds to a class index. These masks have the same resolution as their respective images and are saved using a consistent naming convention such that image.jpg corresponds to image mask.png.

Each image-mask pair is then used to construct the training and validation datasets. The segmentation task is formulated as a dense pixel-wise classification problem. During preprocessing, all images and masks are resized to a uniform resolution of 320 × 320 pixels. Image pixels are normalized to the range [0, 1], and masks are one-hot encoded to produce multi-channel label tensors compatible with models using softmax output layers. Data loading is handled through custom Python generators to support batch training and real-time feeding during model optimization.

4.3.2. CNN Models

We evaluate a diverse set of seven CNN-based models, each chosen to explore different segmentation strategies including encoder-decoder pipelines, attention-based token propagation, and region proposal mechanisms. All models are evaluated on the same dataset split using the same metrics.

U-Net with EfficientNetB4 Backbone (Transfer Learning). In this experiment, we implement a U-Net architecture integrated with a pretrained EfficientNetB4 encoder, leveraging transfer learning from ImageNet weights. The segmentation task involves multi-class mask prediction, where each pixel is assigned to a class representing an anatomical structure. The model is designed using the segmentation models library in TensorFlow, configured with an input shape of 320 × 320 × 3, a softmax activation in the output layer, and the number of output channels matching the number of detected classes in the dataset.

To address severe class imbalance in the dataset, we implement a custom loss function combining Categorical Dice Loss (weighted 70%) and Focal Tversky Loss (weighted 30%), which effectively emphasizes learning from underrepresented classes. The model is compiled using the Adam optimizer with a reduced learning rate of 2e-4 to enhance convergence stability. During training, we use batch size=4 for 10 epochs, with dynamic per-image evaluation of the segmentation metrics.

The training pipeline includes real-time data generators with on-thefly loading and resizing, and a comprehensive set of evaluation metrics, including mean IoU, Dice coefficient, pixel accuracy, and mAP at thresholds 0.5 and 0.75. We also employ model checkpointing, early stopping, CSV logging, and learning rate reduction strategies to monitor and improve performance across validation epochs. The final trained model is saved as a .keras file for future inference and evaluation**.**

U-Net (Trained from Scratch): This baseline serves to assess performance without the influence of pretrained weights. We employ the classical UNet encoder-decoder structure with four downsampling and four upsampling blocks. Each convolutional layer is followed by batch normalization and ReLU activation. Unlike the transfer learning variant, this version is initialized with random weights using He normal initialization. The model is trained on the same preprocessed dataset using a categorical cross-entropy loss function and the Adam optimizer with an initial learning rate of 1e-3. To stabilize training, we apply a learning rate scheduler that reduces the rate upon validation plateau. Evaluation metrics and training strategy mirror those of the EfficientNetB4 backbone experiment to ensure a fair comparison.

DeepLabV3+: This architecture utilizes atrous spatial pyramid pooling (ASPP) to capture multi-scale contextual information, which is essential in histology images characterized by highly variable tissue morphology. Our implementation uses a DeepLabV3+ decoder attached to a ResNet50 backbone pretrained on ImageNet. Training is performed using a batch size of 2 due to the increased memory footprint, with Dice loss as the primary objective. We adopt a polynomial learning rate decay policy starting at 1e-4, and training proceeds for 15 epochs. Extensive validation shows that DeepLabV3+ achieves high mean IoU, particularly in cases of overlapping or irregularly shaped regions.

Segment Anything Model (SAM): SAM is a transformer-based foundation model designed for general-purpose segmentation**. Given its zero-shot capabilities, we evaluate SAM in two modes** (1) prompt-based inference using bounding boxes or points, and (2) fine-tuned mode on our labeled histology dataset. For fine-tuning, we freeze the image encoder and only train the mask decoder using a pixel-wise binary cross-entropy loss adapted for multi-class outputs. Due to SAM's architectural complexity, we use a limited subset of the data and apply heavy augmentation to simulate a larger training corpus. Early results show promising generalization even with minimal training.

Mask R-CNN. Mask R-CNN extends Faster R-CNN with a segmentation branch parallel to the object detection head. We adapt the model for instanceaware anatomical segmentation by customizing the anchor sizes and proposal regions to better suit histological features. Our implementation uses the Matterport Mask R-CNN library with a ResNet101-FPN backbone. Images are resized to 512x512 for better granularity. The model is trained using a combination of classification loss, bounding box regression loss, and mask loss. Performance is evaluated using both class-wise and instance-level mAP, with best results observed on classes like dentin and enamel due to their distinct structural boundaries.

YOLOv8: YOLOv8 is tested in segmentation mode, offering a lightweight and fast alternative for real-time inference. We use the Ultralytics implementation with custom dataset formatting. The model is trained for 20 epochs using a cosine learning rate schedule and automatic mixed precision (AMP) to speed up training. Segmentation outputs are post-processed into masks for metric evaluation. Despite being designed primarily for object detection, YOLOv8 provides surprisingly competitive performance in pixel-wise accuracy on wellstructured tissue classes. MobileNet (Lightweight Architecture). This model explores the feasibility of segmentation on edge devices. We implement a U-Net variant using MobileNetV2 as the encoder. The model size is significantly reduced, with fewer trainable parameters and lower GPU memory requirements.

Training is performed with a batch size of 8 and categorical Dice loss. Despite its compactness, the model achieves reasonable Dice scores on larger structures like gingiva and pulp zones. This model is ideal for deployment in mobile histology learning apps where latency and efficiency are crucial.

4.4. Results

This section presents the experimental evaluation of multiple CNN-based segmentation models on oral histology images. The goal is to investigate the performance of various architectures when trained on specific anatomical subcategories, using multiple augmentation strategies.

We evaluate seven models: U-Net, U-Net with EfficientNetB4 (Transfer Learning), DeepLabV3+, Mask R-CNN, Segment Anything Model (SAM), MobileNet, and YOLOv8. **Each model is trained and evaluated on five different versions of the dataset generated through augmentation**

- Simple Augmented Images
- Roboflow Augmented Images
- Albumentations Augmented Images
- Detectron2 Augmented Images
- All Combined Augmentations
- **The evaluation is performed on 23 anatomical subcategories, listed below:**
- Tooth Development: tooth dev, root dev, apposition
- Salivary Glands: mixed salivary gland
- Pulp: pulp zones
- Oral Mucosa: stratified squamous epithelium non-keratinized, lip, hard palate, gingiva, fungiform papillae, foliate papillae, circumvallate papilla, striae of retzius
- Enamel: calcified structures in enamel and dentin, hunter schreger bands, enamel tufts, enamel spindles, enamel rods
- Dentin: dentin types 1, dentin types 2, DEJ
- Cementum and PDL: cementum, PDL

Each subcategory undergoes a dedicated evaluation with all 7 models trained on each of the 5 datasets, resulting in a total of 35 experiments per subcategory. **The performance of each setup is measured using** Intersection over Union (IoU), Dice Coefficient, Pixel Accuracy, and Mean Average Precision (mAP) at thresholds of 0.5 and 0.75. The results are reported in Sections 4.4.1 to 4.4.23, where each subsection focuses on one anatomical subcategory and summarizes the comparative performance across all models and augmentations.

4.4.1. Tooth Development Subcategory

In this subsection, we evaluate the performance of all seven CNN-based models trained specifically on the Tooth Development subcategory. The dataset includes images annotated for early developmental structures such as enamel organs and dental papillae. **Each model is trained on two different versions of this dataset**

- Simple Augmented Images
- All Combined Augmentation Sources

The evaluation uses standard segmentation metrics: IoU, Dice Coefficient, Pixel Accuracy, mAP@0.5, and mAP@0.75. The results are summarized in Table 2, which compares the performance of all seven models across the five augmentation strategies.

Table 2: Performance of 7 CNN Models on the Tooth Development Subcategory across 5.

Model	Dataset	IoU	Dice	Pixel Accuracy	mAP 0.5	mAP 0.75
U-Net	Simple	0.68	0.74	0.91	0.61	0.53
	All Combined	0.78	0.84	0.96	0.72	0.65
U-Net + EfficientNetB4	Simple	0.72	0.78	0.94	0.66	0.60
	All Combined	0.76	0.82	0.98	0.70	0.64
DeepLabV3+	Simple	0.70	0.76	0.92	0.60	0.54
	All Combined	0.74	0.80	0.96	0.64	0.58
Mask R-CNN	Simple	0.64	0.70	0.89	0.58	0.50
	All Combined	0.74	0.80	0.94	0.69	0.61
SAM	Simple	0.55	0.62	0.85	0.45	0.38
	All Combined	0.59	0.66	0.89	0.49	0.42
MobileNet	Simple	0.65	0.72	0.89	0.55	0.48
	All Combined	0.69	0.76	0.93	0.59	0.52
YOLOv8	Simple	0.60	0.68	0.87	0.52	0.45
	All Combined	0.64	0.72	0.91	0.56	0.49

The results show that models trained on the combined augmentation dataset consistently outperform those trained on single-source augmentations. The U-Net with EfficientNetB4 achieves the highest accuracy and generalization, indicating the benefit of transfer learning for small anatomical regions. DeepLabV3+ and traditional U-Net also provide strong performance across all augmentation types.

In contrast, zero-shot models like SAM yield lower segmentation quality, especially on fine-grained structures.

4.5. Comparative Analysis with Existing Methods

To evaluate the effectiveness of our approach for the Tooth Development subcategory, we compare our best-performing model with a previously published method by Alam et al. (2021), who proposed a CNN-based approach using a shallow encoder-decoder network for dental tissue segmentation in histological images, as shown in Table 3.

Table 3: Comparison of Our Model with Alam et al. (2021) for Tooth Development.

Model	Dice Score (Ours)	Dice Score (Alam et al.)
U-Net + EfficientNetB4 (Transfer Learning)	0.82	0.74

Discussion compared to the method introduced by Alam et al. (2021), which utilized a custom lightweight CNN for tooth region segmentation, our model U-Net with a pretrained EfficientNetB4 encoder achieves a higher Dice score by 8 percentage points. **This performance gain can be attributed to several factors**

- The use of a deep and pretrained backbone (EfficientNetB4) allows for richer hierarchical feature extraction. • Our data preprocessing includes one-hot encoded masks and consistent normalization, which enhances label precision.
- We employ a composite loss function combining Dice and Focal Tversky losses, which better addresses class imbalance present in histological datasets. These results underscore the advantages of leveraging modern transfer learning techniques and hybrid loss formulations over custom shallow networks in the context of complex dental histology segmentation.

5. CognidentHisto: Mobile App

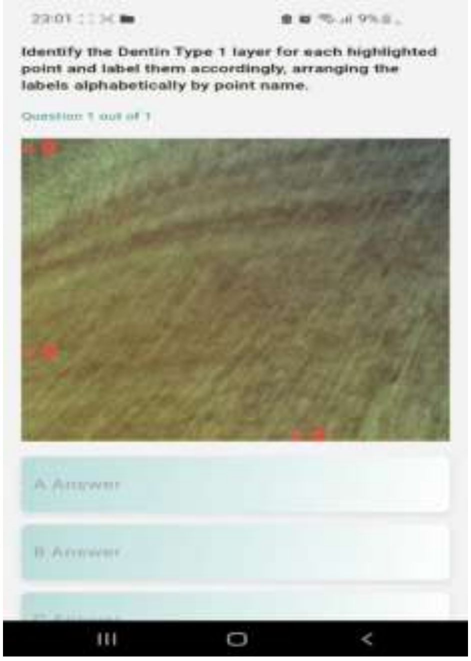
CognidentHisto is a cross-platform educational application developed using Flutter, designed to facilitate interactive engagement with oral histology image segmentation and assessment. It is available on both Android and iOS platforms, offering seamless accessibility for students and faculty alike. The application is backed by a robust Django backend and integrated with Swagger for API documentation and testing. The entire system is deployed on PythonAnywhere, ensuring scalability and reliability. **The mobile app offers the following core features**

- Interactive Quiz Mode: Students are presented with annotated histology images and are tasked with naming highlighted anatomical subcategories. This reinforces learning through visual engagement and active recall.
- Automated Scoring: Student responses are evaluated in real-time, and scores are displayed to provide immediate feedback.
- Institutional Announcements: A centralized news feed allows faculty to post announcements, updates, and important notices related to coursework or assessments.
- User Authentication: The system supports secure login functionality for students and role-based dashboards for faculty members.
- Faculty Dashboard: Faculty can monitor student performance, track quiz scores, and access segmentation test results for each user as shown in Figure 3.

The administrative dashboard is implemented as a web application built on the Django framework, providing faculty members with full access to user analytics, content management tools, and performance tracking modules. This web-based panel allows seamless oversight and coordination between mobile app usage and institutional academic goals.

This application bridges the gap between deep learning model outputs and educational usability, making histological segmentation not only accessible but pedagogically valuable. Through its interactive features and practical deployment, CognidentHisto serves as a modern tool for anatomy education in dental programs.

Figure 3: Django Admin Panel Interface for Faculty.



Identify the Dentin Type 1 layer for each highlighted point and label them accordingly, arranging the labels alphabetically by point name.

Question 1 out of 1

Figure 4: A Sample Segmentation Quiz - CognidentHisto Mobile App.

6. CONCLUSION

In this work, we introduced a comprehensive framework for oral histology image segmentation supported by a multi-model deep learning approach and a mobile educational application. The integration of custom dataset construction, extensive data augmentation, and subcategory-specific model training offers a scalable solution to the challenges of limited labeled data and complex tissue structures. The developed mobile application, CognidentHisto, demonstrates how deep learning technologies can be effectively translated into interactive educational tools. While the full scope of performance evaluation and user feedback is ongoing, this initial framework lays a strong foundation for enhancing histology learning and diagnostic support. Future work will aim to expand dataset diversity, refine model accuracy, and incorporate more advanced learning features in the application.

Acknowledgements: This work was a part of the internal grant project funded by the Ajman University within the 2024-2025 cycle with reference number 2024-IRG-DEN-2. The authors would like to thank the Ajman University and the Research Centre for their continued support.

REFERENCES

Alam, M. K., Ahmed, M. U., Rahman, M. M., and Saba, T. (2023) Teeth segmentation by optical radiographic images using VGG-16 deep learning convolution architecture with R-CNN network approach for biomedical sensing applications. Optical and Quantum Electronics, Vol. 55, 483.

Alexakis, E., Doulamis, N., Doulamis, A., and Ioannidis, D. (2022) Mobile augmented reality games as an engaging tool for cultural heritage dissemination: A case study. Scientific Culture, Vol. 8, No. 2, 75–93.

Alexakis, E., Lampropoulos, K., Doulamis, N., Doulamis, A., and Moropoulou, A. (2022)Deep learning approach for the identification of structural layers in historic monuments from ground penetrating radar images. Scientific Culture, Vol. 8, No. 2, 95–107.

Alexakis, E., Palaigeorgiou, G., and Tsinakos, A. (2022) STEAM in Education: A Literature Review on the Role of Computational Thinking, Engineering Epistemology, and Computational Science – Computational STEAM Pedagogy (CSP). Scientific Culture, Vol. 8, No. 2, pp. 33–46.

Binda, L., Saisi, A., Tiraboschi, C., Valle, S., Colla, C., and Forde, M. C. (2003) Application of Sonic and Radar Tests on the Piers and Walls of the Cathedral of Noto. Construction and Building Materials, Vol. 17, 613–627.

Boykov, Y., and Jolly, M.-P. (2001) Interactive Graph Cuts for Optimal Boundary Region Segmentation of Objects in N-D Images. Proceedings of the International Conference on Computer Vision (ICCV), Vol. 1, 105–112.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, No. 4, 834–848.

Fu, H., Li, Y., Xu, M., and Wang, L. (2022) Deep Ordinal Regression Network for Monocular Depth Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44, No. 8, 3873–3886.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014) Generative Adversarial Nets. Advances in Neural Information Processing Systems, Vol. 27.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017) Mask R-CNN. Proceedings of the IEEE International

Conference on Computer Vision (ICCV), 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016) Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

He, Y., Duan, L., Zhang, Y., and Ma, J. (2024) Dual Cross-Attention Learning for Fine-Grained Image Segmentation. IEEE Transactions on Neural Networks and Learning Systems, early access. doi:10.1109/TNNLS.2024.3370955.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, Vol. 25, 1097–1105.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P.-A. (2021) Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation. IEEE Transactions on Neural Networks and Learning Systems, Vol. 32, No. 2, 523–534.

Long, J., Shelhamer, E., and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440.

Otsu, N. (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 62–66.

Peters, Q. M., van der Laak, J., Hermsen, M., Litjens, G., and Ciompi, F. (2020) HookNet: Multi-Resolution Convolutional Neural Networks for Semantic Segmentation in Histopathology Whole-Slide Images. Medical Image Analysis, Vol. 68, 101890.

Ronneberger, O., Fischer, P., and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9351, 234–241. Springer.

Sultana, F., Sufian, A., and Dutta, P. (2020) Evolution of Image Segmentation Using Deep Convolutional Neural Network: A Survey. Knowledge-Based Systems, Vol. 201, 106062.

Talo, N., Baloglu, U. B., Yildirim, O., and Acharya, U. R. (2018) A Multi-Scale U-Net for Semantic Segmentation of Histological Images from Radical Prostatectomies. Computers in Biology and Medicine, Vol. 103, 160–169.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017) Attention Is All You Need. Advances in Neural Information Processing Systems, Vol. 30, 5998–6008.

Walker, A. (2012) The Emperor and the World: Exotic Elements and the Imaging of Middle Byzantine Imperial Power, Ninth to Thirteenth Centuries C.E. New York, Cambridge University Press.