AAO-HNSF 2025 Annual Meeting & OTO EXPO™
INDIANAPOLIS, INDIANA
OCTOBER 11-14

(A531)

Deep Learning-Based Web Application for the Detection and Classification of Salivary Gland Tumors

Ping-Chia Cheng; Po-Wen Cheng; Li-Jen Liao
Department of Otolaryngology Head and Neck Surgery, Far Eastern Memorial Hospital, New Taipei City 22060, Taiwan

Introduction

Salivary gland tumors (SGTs) are one type of the head and neck tumors, with an annual incidence ranging from 2.5 to 8.6 cases per 100,000 individuals[1,2]. Although malignant SGTs constitute less than 20 % of these cases, they require more intensive treatment than benign SGTs, underscoring the need for comprehensive treatment planning. Current pre-operative assessments include imaging and cytopathological examinations, with ultrasound serving as the primary imaging tool for evaluating SGTs. Ultrasound offers high-resolution, radiation-free, and rapid imaging for superficial tissues such as the salivary glands, and it facilitates the simultaneous acquisition of fine-needle aspiration cytology (FNAC). However, current diagnosis of SGTs relies on subjective features observed in ultrasound imaging. In our previous study, we constructed a subjective ultrasound score for evaluating SGTs[3]. We further compared this score with ultrasound elastograpy and FNAC. The results indicated that in differentiating between malignant and benign SGTs, the subjective ultrasound score had a sensitivity of 58%, specificity of 89%, and accuracy of 85%, while ultrasound elastography had a sensitivity of 69%, specificity of 70%, and accuracy of 70%, and FNAC had a sensitivity of 74%, specificity of 93%, and accuracy of 91%[4]. Despite the high accuracy, all methods suffered from low sensitivity, even with cytology. Furthermore, diagnoses based on subjective ultrasound imaging may vary among different specialists. Consequently, we have shifted our research focus to the application of deep learning (DL) to establish an objective and automated diagnostic method.

DL techniques are increasingly being applied to medical image analysis, including tasks such as classification, detection, and segmentation[5,6]. In our previous study, we developed a DL model using a modified ResNet50V2 architecture for classifying SGTs under ultrasound image, and this model achieved high diagnostic performance on the testing set[7]. However, it relied on cropped ultrasound images restricted to the tumor region alone, necessitating an additional step for manual cropping and potentially introducing bias during clinical application. To address this limitation, we explored object detection models for automated tumor region cropping. Object detection not only identifies objects but also precisely localizes them by drawing bounding boxes around their boundaries[8,9]. Our approach combines object detection with our existing classification model, allowing us to achieve both automated tumor detection and subsequent classification. We considered both one-stage and two-stage object detection models. For one-stage model, it directly predicts bounding boxes and class probabilities, making them faster than two-stage models. For two-stage model, it involves a region proposal step followed by classification and refinement. While it offers higher accuracy, it is more complex and time-consuming. For our implementation, we chose YOLOv8, the latest version of the YOLO (You Only Look Once) family, due to its smaller size and faster inference speed[10,11]. Additionally, we developed a user-friendly web application using Streamlit, which allows clinicians and researchers to easily interact with the model.

In summary, our study aims to seamlessly integrate object detection and classification to automate SGT analysis in ultrasound images. By developing a web application, we hope to facilitate clinical and research use.

Methods

Ethical Considerations

The study was carried out at a tertiary medical center. We reviewed the records of patients who visited our outpatient department between January 2007 and December 2022 and underwent ultrasound examinations for SGT. The ultrasounds were performed by seasoned otolaryngologists using a Toshiba Aplio 500 (Canon Medical Systems, Tochigi-ken, Japan) equipped with a 5-14 MHz linear-array transducer. We included B-mode ultrasound images from adult patients who subsequently underwent excision or core needle biopsy (CNB), with pathological reports available. CNB was the chosen procedure for patients for whom open surgery was not suitable. Pathological diagnoses from these reports served as the definitive standard for classifying tumors as malignant or benign. Ultrasound images with poor quality were excluded. The process for inclusion and exclusion is depicted in Figure 1.

Data Collection and Preparation

The study protocol is illustrated in Figure 2. Ultrasound images were retrieved from the picture archiving and communication system (PACS), covering various views of SGTs. We preserved ultrasound images while ensuring all identifiable information was removed to protect patient confidentiality. This process involved the elimination of names, medical record numbers, birth dates, and dates of ultrasound execution. We achieved this by utilizing the Snipping Tool provided by Microsoft. For effective model establishment and evaluation, we categorized these ultrasound images into three sets. The training set, which includes 684 ultrasound images (475 benign and 212 malignant) from patients diagnosed between January 2007 and December 2020, was utilized for model establishment. The validation set, comprising 78 ultrasound images (54 benign and 24 malignant) from patients diagnosed between January 2021 and December 2021, was employed for model validation. The testing set, containing 100 ultrasound images (82 benign and 18 malignant) from patients diagnosed between January 2022 and December 2022, was used to assess the model's predictive capabilities. During the object detection training process, we employed labelImg, a Python-based open-source image annotation tool, to delineate bounding boxes (Figure 3). Each bounding box was labeled as 'tumor'. To facilitate integration with our classification model, we further assigned class labels: benignity as class 0 and malignancy as class 1.

Model Establishment

We developed our model utilizing the Python programming environment on Google Colaboratory (Colab) with an NVIDIA T4 GPU (NVIDIA Corp., Santa Clara, CA, USA). Colab provides complimentary GPU resources and functions as a web-based Jupyter Notebook. We use the YOLOv8 for the object detection training[12]. The input image undergoes resizing to a 128×128 dimension to facilitate the training process before being inputted into the CNN model. This model divides the input image into a grid structure, where each grid cell is tasked with detecting objects whose centers are located within its boundaries. The fully connected layers bridge the convolutional layers to YOLO's output layer, enabling the identification of target objects. In terms of bounding box prediction, YOLOv8 employs two loss functions: Distribution Focal Loss (DFL) and Complete Intersection over Union (CIoU) Loss. DFL, an augmented variant of Focal Loss, is designed to address class imbalance by adjusting sample weights based on class information. CIoU loss, an evolved form of Intersection over Union (IoU) loss, measuring the precision of object detectors by assessing the congruence between predicted bounding boxes and ground truth. We trained the model for 100 epochs using a training set to construct the detection model. The validation set was used to validate the precision of bounding box predictions

Web Application Integration

We connected the detection model with our previously trained classification model, and developed a web application using Streamlit. The operational code is stored in a GitHub repository, and the application is deployed under the Streamlit domain. When a user uploads an ultrasound image of a SGT, the application automatically detects the tumor, crops the relevant region, and classifies the benignity or malignancy of the SGT (Figure 3). This entire process can also be executed locally.

Statistical Analysis

We evaluated the diagnostic performance of our web application model using the testing set. A confusion matrix was generated, including metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Internal and External Validation

Following the establishment and validation of our web application model, we proceeded to validate its diagnostic consistency using distinct patient groups. The internal validation set included 24 benign and 3 malignant ultrasound images. Each image was sourced from a patient diagnosed between January 2023 and June 2023 in our hospital. The external validation set comprised 36 benign and 21 malignant ultrasound images. Each image was sourced from a patient whose case was reported on an online case report website (https://www.ultrasoundcases.info). We obtained their permission to use these images for research purposes

Results

The flow chart of inclusion and exclusion criteria is presented in Figure 1. The study protocol is illustrated in Figure 2. There were total 862 ultrasound images of SGTs in our study, including 611 benign and 251 malignant ultrasound images. We split the dataset as 684, 78, and 100 ultrasound images in the training, validation, and testing set. Figure 4 showed the predicted result of validation set using YOLOv8 after fine tuning on our training set. During the validation process of objection detection, the precision and recall of bounding box was 0.939 and 0.945, and mAP50 (mean average precision at IoU threshold of 0.5) was 0.958.

After the objected detection model establish, we connected with the pre-trained classification model, and developed a web application. This web application model obtained an accuracy of 85%, a sensitivity of 78%, a specificity of 87%, a PPV of 56%, and a NPV of 95% in the testing set.

Further internal and external validation was performed (Figure 5), and the web application model achieved an accuracy of 85%, a sensitivity of 100%, a specificity of 83%, a PPV of 60%, and a NPV of 93% in the internal validation set, and showed an accuracy of 79%, a sensitivity of 76%, a specificity of 81%, a PPV of 70%, and a NPV of 85% in the external validation set.

Discussion

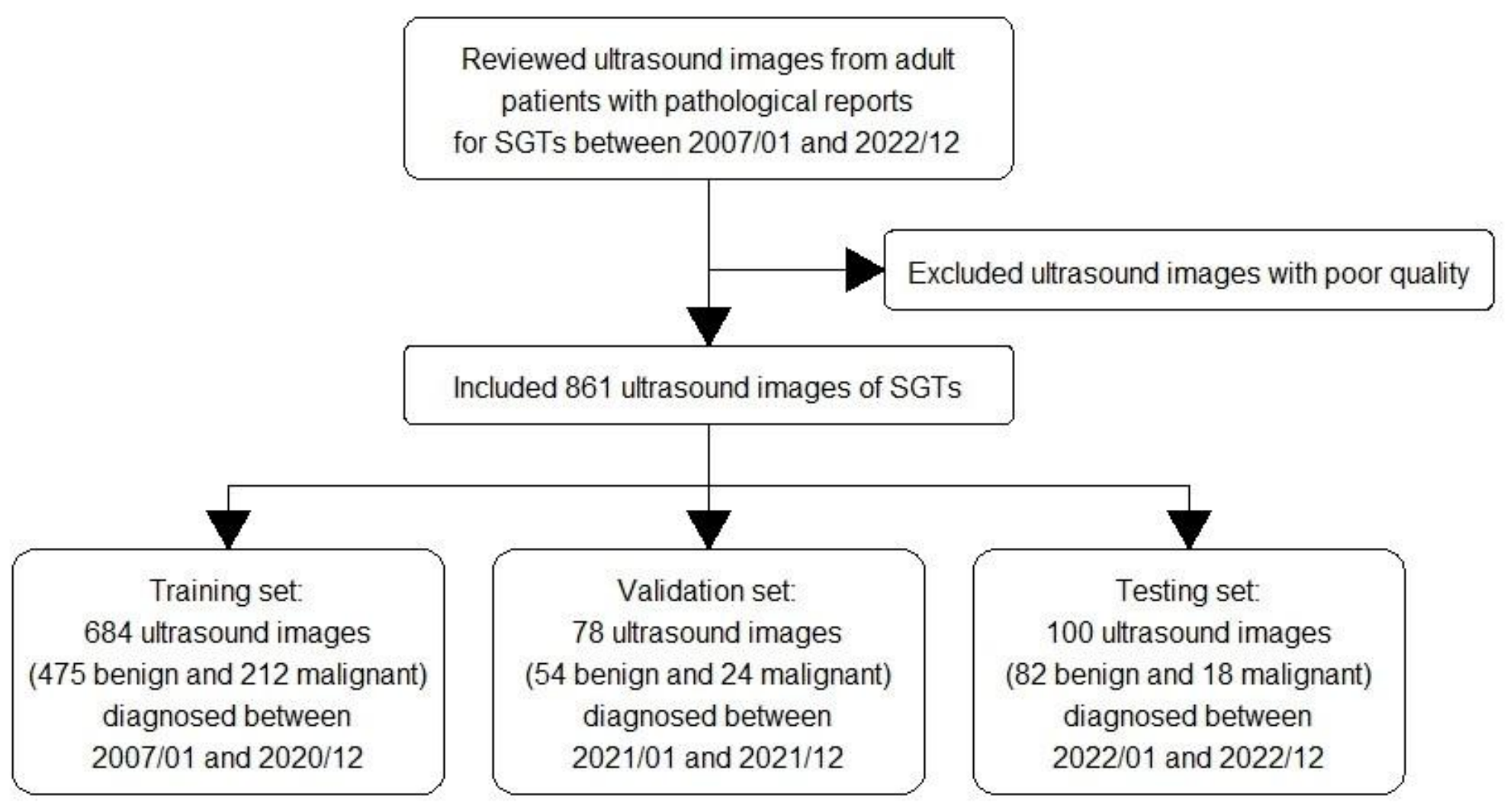
This study presents a compelling study that leverages the use of DL for the automated detection and classification of SGTs using ultrasound images. We had seamlessly integrated these DL models into a user-friendly web application, which simplifies the process for users by requiring the upload of a single ultrasound image of the SGT. For optimal results, we recommend using images of high resolution or those that appear more suspicious for malignancy. This application is designed for both local and online utilization. To safeguard patient privacy, we recommend removing all identifiable information from the ultrasound image before uploading. Our web application model was trained on a dataset of ultrasound images from patients with pathology-confirmed SGTs. The distinct cohorts for training, validation, and testing were employed to ensure a comprehensive evaluation of the model's performance. Our model demonstrated an accuracy of 85%, a sensitivity of 78%, a specificity of 87%, a Positive Predictive Value (PPV) of 56%, and a Negative Predictive Value (NPV) of 95% in the testing set. Furthermore, we compiled two separate datasets for internal and external validation, where the model achieved an accuracy of 79%, a sensitivity of 76%, a specificity of 81%, a PPV of 70%, and an NPV of 85% in the external validation set. The model's high diagnostic performance in both internal and external validations underscores the potential of this method. To further enhance diagnostic consistency and accuracy, histogram equalization was utilized during the classification process to adjust for operator-controlled gray level variations. When compared to the subjective ultrasound score[4], which exhibited a sensitivity of 58%, specificity of 89%, and accuracy of 85%, our web application model showed a notably higher sensitivity. It proved to be an effective diagnostic method for the automatic classification of benign and malignant SGTs, offering a balanced sensitivity and specificity, and mitigating the variability of interpretation when we evaluated under subjective ultrasound features.

The user-friendly interface of our model promotes its application in both clinical and research settings. For deployment purposes, we have chosen to develop these DL models as a web application, facilitating ease of access and use. Currently, there are two python software development kits (SDKs), Gradio[13] and Streamlit[14], that can use to streamline the creation of web components directly through Python code. Both SDKs can be easily developed and executed on a local host. However, their approaches to online deployment diverge. The Python code crafted with Gradio is stored within Gradio's infrastructure and is inherently public, posing a challenge for medical image analysis due to privacy considerations. Conversely, code developed with Streamlit is hosted on GitHub, allowing for private repositories, and the web application is deployed under the Streamlit domain. While deploying on the Streamlit domain may still raise concerns regarding patient confidentiality, clinicians and researchers have the option to anonymize the ultrasound images before uploading or to opt for local deployment of the model.

In the realm of object detection, there are one-stage and two-stage object detection models. Although the two-stage object detection model are known for their superior accuracy, their slower processing speed limits its use in real-time object detection[15]. In our study, we employed the one-stage detection model, YOLOv8, for our detection training of SGT. It yielded impressive results in bounding box prediction for SGTs, achieving a precision of 0.939, recall of 0.945, and mAP50 of 0.958. The YOLOv8 algorithm has gained traction in medical image detection tasks, such as identifying brain tumor, breast cancer, and lung disease[16]. Its application also extends to cytology and pathology image detection, where it has demonstrated high precision and recall[17]. Notably, YOLOv8 supports multi-class prediction, making it suitable for both detection and classification tasks. However, our initial attempts to use YOLOv8 for both detection and classification yielded suboptimal results. Consequently, we adapted our approach, utilizing YOLOv8 exclusively for tumor detection and employed our previously trained DL model for classification. This strategic adjustment led to high precision and recall in object detection, and the coupled classification model demonstrated promising results, with commendable accuracies, sensitivities, and specificities in both internal and external validation sets. Although this modification may augment the model's size, it significantly enhanced both detection and classification performance compared to using YOLOv8 alone.

Limitations

This study, while insightful, still presents several limitations that warrant discussion. Firstly, the retrospective design introduces potential selection bias, which may constrain the broader applicability of the findings. Secondly, the dataset utilized was relatively small in size, consisting of 862 ultrasound images of SGTs used for training, validation, and testing phases. Despite the implementation of distinct internal and external validation cohorts to assess the model's diagnostic accuracy, the results derived from this specific dataset might not be generalizable to other demographic groups. Thirdly, the model's detection and classification capabilities hinge on the gray-level intensity of ultrasound image. Although histogram equalization was employed to mitigate inconsistencies, inherent variations in ultrasound equipment and their respective settings could potentially alter the gray-level distribution, thereby influencing the predictive outcomes. Given these considerations, it is imperative to pursue additional research across multiple institutions to substantiate the model's validity. Such studies would provide valuable insights into the model's performance in actual clinical environments and its potential integration within existing diagnostic workflows.



Reviewed ultrasound images from adult patients with pathological reports for SGTs between 2007/01 and 2022/12

Excluded ultrasound images with poor quality

Included 861 ultrasound images of SGTs

Training set: 684 ultrasound images (475 benign and 212 malignant) diagnosed between 2007/01 and 2020/12

Validation set: 78 ultrasound images (54 benign and 24 malignant) diagnosed between 2021/01 and 2021/12

Testing set: 100 ultrasound images (82 benign and 18 malignant) diagnosed between 2022/01 and 2022/12

Figure 1. Flow chart to illustrate the study's inclusion and exclusion criteria

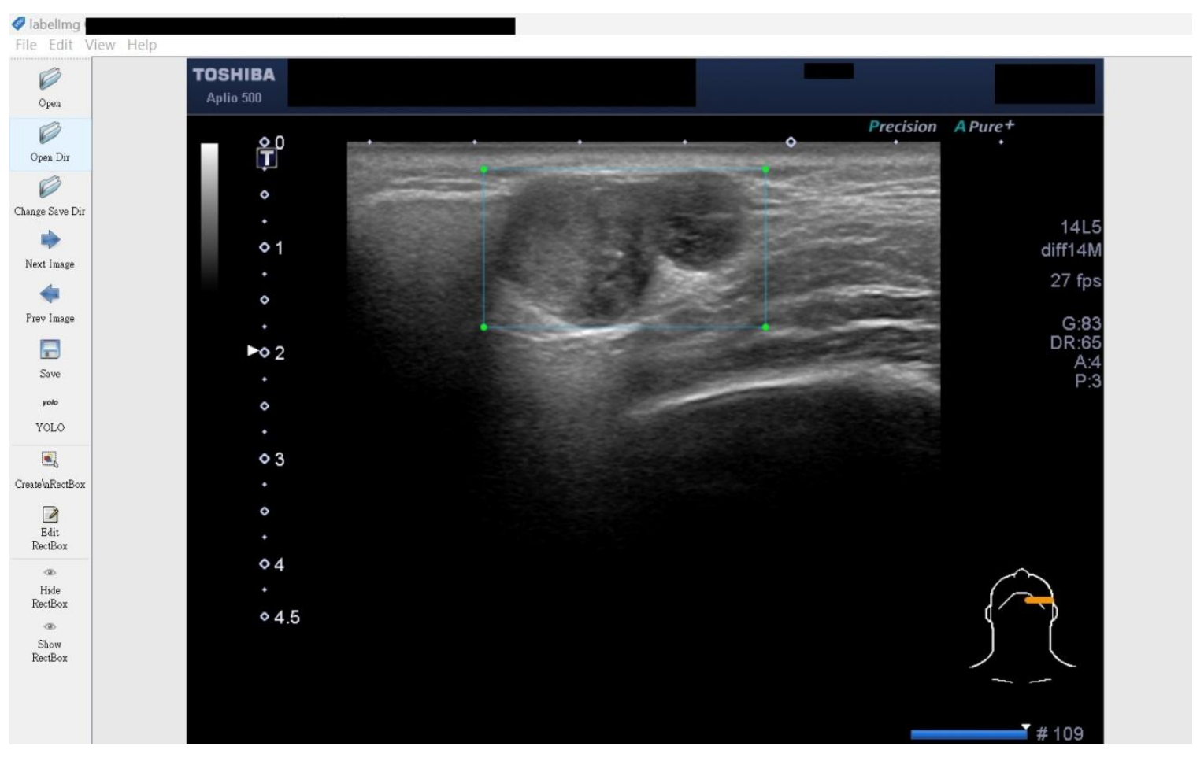
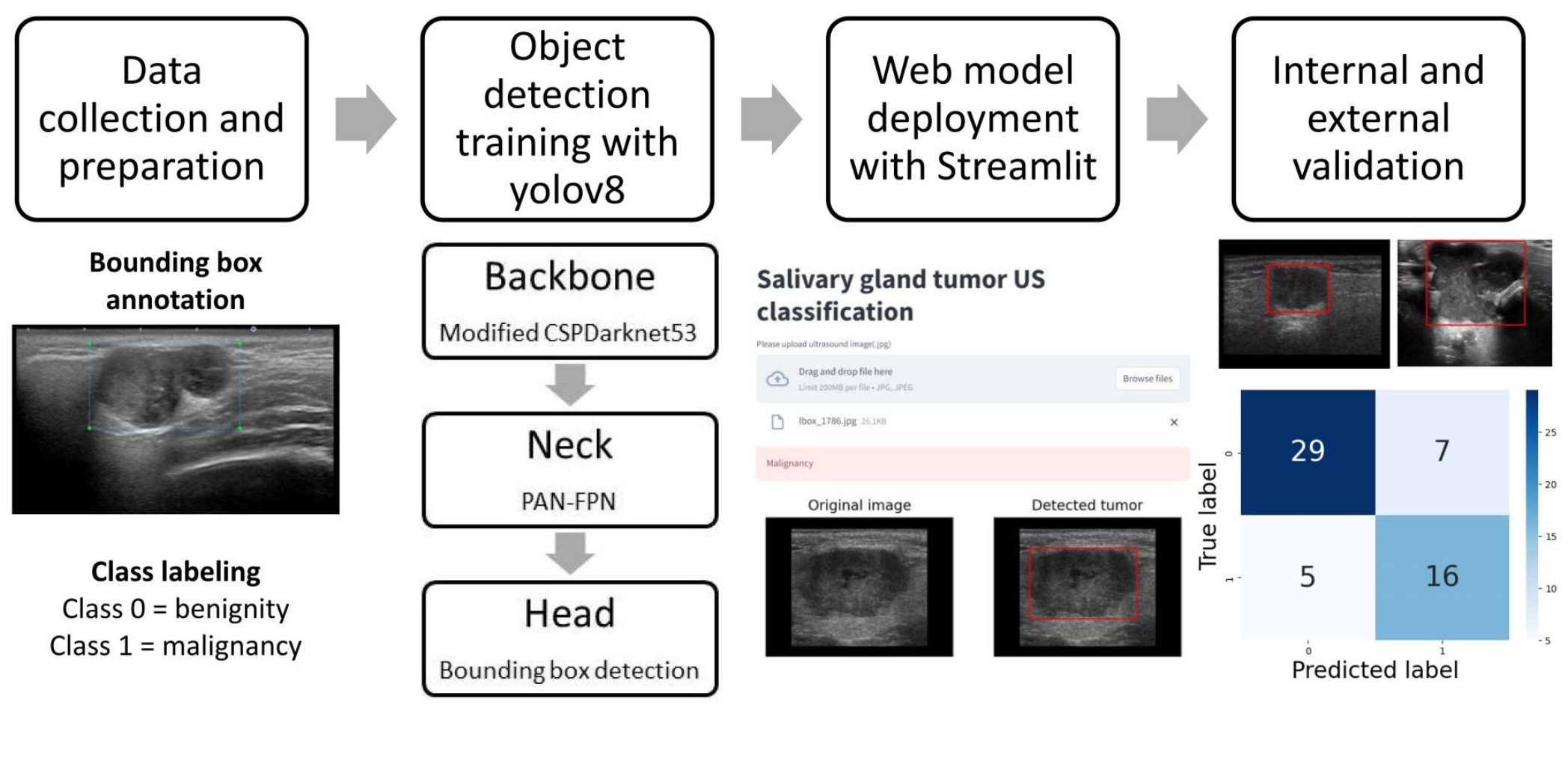


Figure 3. Bounding box annotation using labelImg



Data collection and preparation → Object detection training with yolov8 → Web model deployment with Streamlit → Internal and external validation

Bounding box annotation

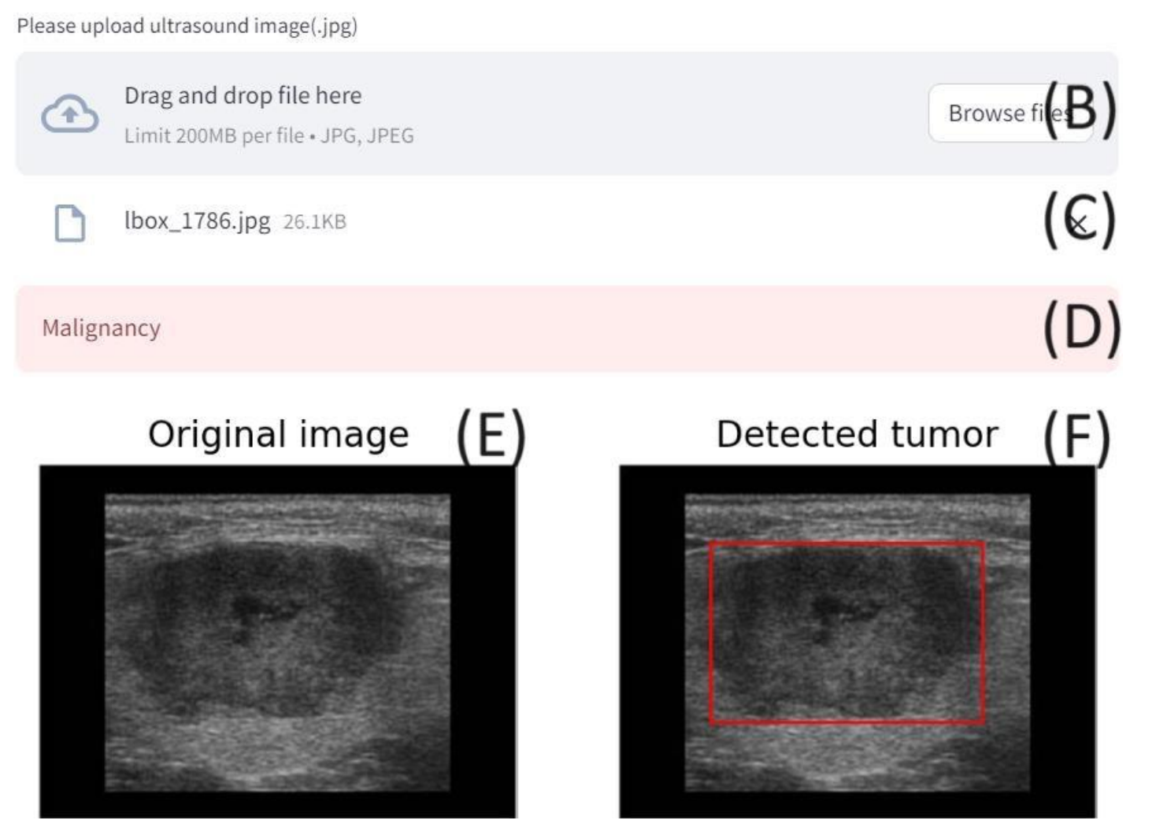
Class labeling
Class 0 = benignity
Class 1 = malignancy

Backbone
Modified CSPDarknet53

Neck
PAN-FPN

Head
Bounding box detection

Figure 2. The study protocol



Salivary gland tumor US classification (A)

Please upload ultrasound image(.jpg)

Drag and drop file here
Limit 200MB per file • JPG, JPEG
Browse files (B)

lbox_1786.jpg 26.1KB (C)

Malignancy (D)

Original image (E)

Detected tumor (F)

Figure 4. Web application interface

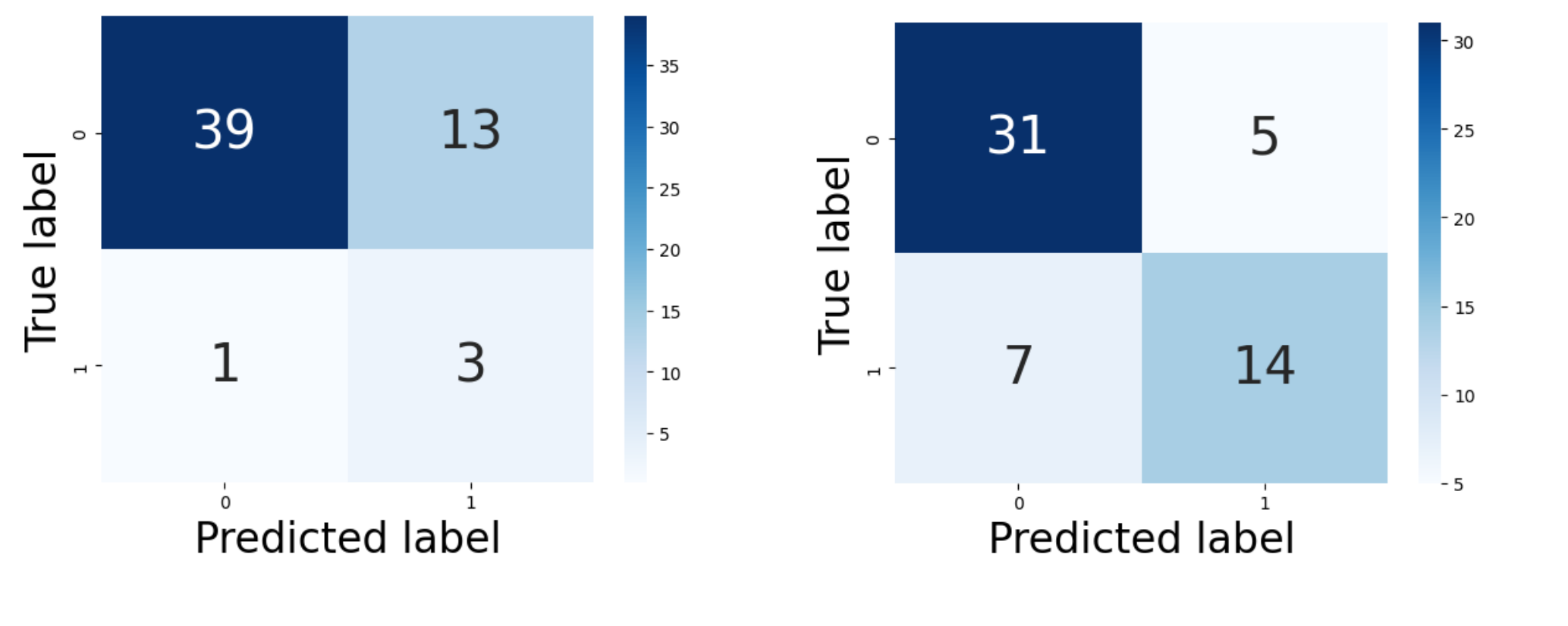


Figure 5. The confusion matrix for the internal validation set (A) and the external validation set (B), utilizing the web application model.