International Journal on Robotics, Automation and Sciences

AI-Assisted Analysis for Breast Cancer Imaging and Diagnostics

Kai Liang Lew, Chean Khim Toa*, Pengfei Zhou, Chia Shyan Lee, Tetuko Kurniawan, Suleiman Aliyu Babale and Cheng Zheng

***Abstract* – Breast cancer cases have increased by 0.5% each year. X-ray, CT-Scan, and magnetic resonance imaging have been used to detect cancer without harming the patient. However, these methods usually used manual screening to process medical images, which leads to longer processing time and increases the burden on the expert. With the help of deep learning, automation-driven breast cancer detection, segmentation, and explanation can be performed in the process, which can greatly reduce the processing time and the burden on experts. This paper proposed a deep learning model, S-YOLOv11 by combining YOLOv11 with SimAM attention mechanism and a GUI with integration of a large language model. The model is trained with 624 images and tested with 156 images. Several YOLO architectures were compared, including YOLOv8, YOLOv9, YOLOv10, and YOLOv11. The proposed model has outperformed the other models. In the detection task, 0.806 precision, 0.635 recall, and 0.724 mAP were achieved. In the segmentation task, 0.833 precision, 0.65 recall, and 0.739 mAP were achieved. In addition, the study also improved the functionality of the GUI by accessing the ChatGPT API. It is possible to generate medical analysis for breast cancer tumors, with the use of GUI for visualization. However, current research is still in the development stage and it needs to be put into clinical trials before it can be used.**

Keywords*— *Breast Cancer, Image Segmentation, Deep Learning, Supervised Learning, Image Detection.

I. INTRODUCTION

In 2023, according to the American Cancer Society, breast cancer accounts for 31% of all cancers in women [1]. Its incidence rate and mortality rate rank first among women. The incidence rate of female breast cancer has increased by approximately 0.5% each year since the mid-2000s. To improve this situation, cancer examinations and treatment should be started as soon as possible. For cancer detection, it is best to use imaging techniques that are at lower risk to the patient. X-ray, CT-Scan, ultrasound, and magnetic resonance imaging (MRI) are the most

*Corresponding Author email: cheankhim.toa@xmu.edu.my, ORCID: 0000-0003-0879-4848
Kai Liang Lew is with Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia (e-mail: 1132703002@mmu.edu.my).
Chean Kim Toa is with School of Electrical Engineering and Artificial Intelligence, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor. (e-mail: cheankhim.toa@xmu.edu.my).
Pengfei Zhou is with School of Electrical Engineering and Artificial Intelligence, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor. (e-mail: AIT2109109@xmu.edu.my)
Chia Shyan Lee is with Curtin University, Perth, Australia (email: cat_lee97@hotmail.com).
Tetuko Kurniawan is with Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland (email: tkurniaw@ippt.pan.pl)
Suleiman Aliyu Babale is with Department of Electrical Engineering, Bayero University Kano, Kano, Nigeria (email: sababale.ele@buk.edu.ng)
Cheng Zheng is with Wireless Signal Processing Technology Inc, Canada. (e-mail: zheng.cheng@wirelessignal.com)

International Journal on Robotics, Automation and Sciences (2025) 7, 1:111-119
https://doi.org/10.33093/ijoras.2025.7.1.13
Manuscript received: 3 Jan 2025 | Revised: 20 Feb 2025 | Accepted: 27 Feb 2025 | Published: 31 Mar 2025
Published by MMU PRESS. URL: http://journals.mmupress.com/ijoras
This article is licensed under the Creative Commons BY-NC-ND 4.0 International License

common methods because they help in detection without putting the patient in harm's way.

Currently, many radiologists are using manual screening to process medical images. This approach may lead to inefficiencies and increase the burden on radiologists. However, fortunately, radiologists' workload can be reduced by AI-based computer technology, which is a crucial tool that helps to increase screening efficiency while minimizing missed or incorrect diagnoses that would have resulted from the subjectivity of human judgment. Among these technologies, object detection and instance segmentation stand out. Despite advancements in AI-driven breast cancer detection, two key challenges remain. First, most AI models primarily rely on object detection, which only provides bounding boxes around tumors rather than precise segmentation of their boundaries. This limitation reduces their clinical utility, as segmentation is essential for surgical planning and treatment decisions. Second, existing AI-based imaging models lack explainability, making it difficult for radiologists and patients to interpret the results. Without clear medical explanations, clinicians may hesitate to rely on AI-based diagnoses

Object detection can locate and even classify areas suspected of being pathological in cancer images. On this basis, instance segmentation can segment the area of the tumour in the bounding box located by object detection.

Thus, this paper proposes a deep learning model, S-YOLOv11, by integrating YOLOv11 with the SimAM attention mechanism. Moreover, the developed GUI is enhancing with more functionality and combining with a ChatGPT Application Programming Interface (API) to provide reasoning on cancer related information. The proposed framework integrates AI-guided mechanisms, thus improving the interpretability, robustness, and clinical utility of automated medical image analysis systems.

The paper is organized as follows. Literature review section reviews breast cancer, LLM and the work that is related to breast cancer and YOLO methods. Methodology section explains the proposed methodology for breast cancer using the proposed architecture. This section provides a detailed explanation of the architecture and training procedures used to develop the classification model. Experiment, result and discussion section shows the results of the assessment and experiments, highlighting key findings and areas of strength and weakness in the model, discusses the implications of these results, and compares them with deep learning models. Lastly, Conclusion section concludes the effectiveness of the work and suggests recommendations for further work.

II. Literature Review

This section explained breast cancer, LLM and the work that is related to breast cancer and YOLO methods.

A. Breast Cancer

Cancer is usually named for the part of the body. For example, if cancer develops in the brain, it will be called brain cancer. Due to the unstable growth and proliferation of cells in breast tissue, it will cause breast cancer. The breast mainly consists of glandular tissue and stromal tissue [2]. The glandular tissue consists of glands and milk producing ducts. In addition, stroma tissue consists of fat and fibrous connective tissue. When benign changes are in the breast, it will cause cancer. Mostly, when people get ductal carcinoma, it will cause breast cancer. Sometimes, breast cancer can be caused by lobular carcinoma. Some are caused by other tissue lesions. Breast cancer is commonly detected by dedicated breast magnetic resonance imaging (DBMRI), ultrasound (US) and mammography (MG) [3]. All methods have their own pros and cons. When using DBMRI for the detection of breast cancer, it has high sensitivity. But the specificity is low. When detecting dense breast cancer, both DBMRI and US have high sensitivity. But compared to the DBMRI, the US is cheaper. There are studies that use deep learning to classify the breast cancer. Wang et al. [4] developed a deep learning model, Densely Connected Convolutional Networks to classify the breast cancer. The experiment demonstrated 84% accuracy in their test set which showing improvement in classification. Shah et al. [5] presented an ensemble deep learning model by using EfcientNet, AlexNet, ResNet, and DenseNet. They were able to achieve a precision of up to 94.6%, sensitivity of 92.4%, specifcity of 96.1%, and area under the curve (AUC) of 98.0% in classifying the breast cancer. Kebede et al. [6] developed an ensemble model by combing EfficientNet-based classifiers withYOLOv5 to identify abnormalities. Their results demonstrated an F1-score of 0.87 and a sensitivity of 0.82. Toma et al. [7] had conducted experiments with ResNet, ResNeXt, SENet, Dual Path Net, DenseNet, NASNet, and Wide ResNet by applying transfer learning to detect the breast cancer. They used a BreaKHis database, which has 7909 histopathological pictures. Their results showed that ResNeXt-50 has the highest accuracy, 99.8%. Chan [8] had reviewed both non-invasive and invasive breast cancer detection methods, comparing traditional approaches like self-exams, mammography, and ultrasound with emerging techniques including deep learning-based diagnostics and liquid biopsy innovations.

Object detection technology is used most to locate and classify a variety of objects in an image [9]. There are three main com`ponents: selecting target regions, extracting object features, and classification. Most of the attributes of the targets are unknown when we select target regions and locate the objects. So, we need to add many multi-scale sliding windows to cover all target regions. Feature extraction usually extracts features for classification by improving the ability to resist deformation and improving expression capabilities. After that, one classifiers is used for traditional object detection to perform classification [10], such as VJ (Viola Jones) [11], HOG (histogram of oriented gradients) [12]. Due to the low precision of traditional object detection, deep learning-based object detection techniques have been rapidly developed in recent years. They have higher precision and greater robustness in complex environments. So, they may be more suitable for detecting tumours. In recent years, the YOLO (You Only Look Once) algorithm has

developed rapidly. YOLO variants are underpinned by the principle of real-time and high-classification performance, based on limited but efficient computational parameters to complete designated tasks more efficiently [13]. In the first few versions of the YOLO series, they can only complete the object detection task. But in the versions after YOLOv7, YOLOv8, YOLOv9, and YOLOv11 can complete multiple tasks such as object detection and instance segmentation. Currently, some studies have shown that the YOLO model performs well in breast cancer detection. However, these studies focus on object detection. The proposed methods were only used to find the bounding box of the tumour, but could not segment its area. But segmenting the tumour area is crucial. Without segmenting the tumour area, the surgeon cannot accurately diagnose the tumour. Furthermore, during subsequent tumour removal surgery, the surgeon was unable to develop a surgical plan in advance to remove the tumour. This may lead to surgical failure and threaten the patient's life.

In the paper [14], a YOLO-based Computer-Aided Diagnosis (CAD) was proposed. It was used to perform mass detection and classification. It can show that the system can work in areas with masses in the pectoral muscles or in dense regions. Sun et al. [15] introduced the principle of classic YOLO algorithms such as YOLOv1, YOLOv2, YOLOv3 to YOLOv5 and sorted out the application of those algorithms in the detection of breast cancer. Prinzi et al. [16] compared several YOLO architectures, including the YOLOv3, YOLO-v5 and YOLOv5 transformers in breast cancer detection. They found that the small YOLO-v5 model resulted in the best developed solution, obtaining an mAP of 0.621 in the proprietary data set. Samanta et al. [17] used the YOLOv8 model to detect ultrasound breast cancer images and achieved mAP(mean average precision) 99.5%, recall 98.4% and accuracy 96.5%. Gui et al. [18] proposed a model called FS-YOLOv9 for breast cancer detection. This model added an additional max pooling layer before Conv1 of Adown and replaced Adown of P3 in the backbone with a high-frequency Haar wavelet 18 convolution kernel. Compared to the official YOLOv9 model, FS-YOLOv9 showed a higher average F1 score (0.700 vs 0.669), FAUC (0.695 vs. 0.662), and mAP50 (0.713 vs. 0.679) in the internal dataset; in the external testing dataset, the FS-YOLOv9 improved the average F1 score, FAUC and mAP50 by 4.58%, 5.78%, and 4.41% respectively.

B. Large Language Model

Nowadays, large language models (LLMs) have rapidly developed. This changed chatbot systems and results in unprecedented levels of intelligence. LLM is trained on massive amounts of text data to gain a deeper understanding of the language input by users.

ChatGPT is an OpenAI-developed artificial intelligence program. It can interact with humans in natural language. GPT stands for Generative Pre-training, which is a natural language processing technology based on deep learning. OpenAI trained ChatGPT with a large amount of data to enable it to perform well in natural language tasks. The basic idea of training ChatGPT is to let AI learn text chaining on general data and master the ability to generate text later. This method does not require humans to annotate data, but only needs to input data into ChatGPT. In this way, 21 can train the data by itself, and then people can evaluate the responses it generates. For ChatGPT, a question usually has multiple answers. When GPT generates text, people need to provide some guidance. For example, when the user needs GPT to generate a 2D graphic, GPT does not know what kind of 2D graphic the user needs. At this time, the user can tell GPT that a 2D graphic similar to a dog is needed. This method is called reinforcement learning based on human feedback (RLHF).

ChatGPT as an advanced chatbot is well known by most people. However, the training and architecture of ChatGPT are still unpublished. Therefore, the Vicuna team introduced an open source chatbot called Vicuna-13B. Vicuna-13B is backed by enhanced datasets and an easy-to-use, scalable infrastructure [19]. First, they collected around 70K conversations from ShareGPT.com. Then, they enhanced the training scripts provided by Alpaca. The training was done with PyTorch FSDP. For serving the demo, they implemented a lightweight distributed serving system. At the same time, they implemented the preliminary evaluation of the model quality. And they used the GPT-4 to judge the output. In addition, they compared two different models sending the prompts to GPT-4 and accessed which model provides better responses by GPT-4.

There is a GUI developed that uses the YOLOv8 model to detect tumours. The GUI presents the data, but does not have any interpretation of the data.

C. Research Gap

In breast cancer diagnosis, it is very important to detect the location of the tumor. In the related work, many researchers have shown that the YOLO model has a high accuracy in detecting the location of the tumor. By using various versions of the YOLO model, breast cancer images are classified, located and detected. However, most research can only generate a square area of the breast tumor, rather than the area and specific location of the breast tumor. For oncologists, this will undoubtedly increase the risk of misjudgment. Therefore, segmenting the specific location of the tumor is crucial to improving the accuracy of diagnosing breast cancer tumors. This can not only provide oncologists with more data for diagnosis to improve the accuracy of diagnosis, but also allow patients and doctors to see the tumor area more intuitively, so as to better understand breast cancer tumors. However, at present, lack of related research on the segmentation of breast cancer tumors using the YOLO model. Therefore, it is a good research direction to use the YOLO model to segment breast cancer tumor images. On the other hand, with the increase in the amount of medical image data, the workload of doctors has increased significantly [20]. Oncologists not only have to look at a large number of breast cancer tumor images, but also have to interpret the content of the images with patients. This greatly

reduces the time of oncologists. Currently, large language models such as ChatGPT, Gemini and Vicuna are very popular. These large language models can interpret data about breast cancer tumors to provide medical diagnosis and treatment plans, which will undoubtedly reduce the workload of oncologists and assist them in diagnosis and treatment. Therefore, whether it is feasible to use ChatGPT to interpret images and generate medical diagnosis reports requires further investigation and research.

III. METHODOLOGY

This section introduces the steps of developing an integrated AI framework combining the S-YOLOv11model with reasoning mechanisms to classify, detect, and segment breast cancer images. Figure 1 shows the flowchart of the entire process in this paper.

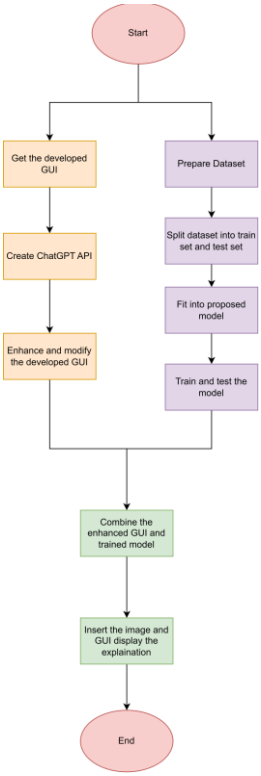

FIGURE 1. Flowchart of the entire process.

A. You Only Look Once v11 (YOLO v11)

YOLOv11 is the latest generation of computer vision model launched by Ultralytics. It represents the latest progress in the field of target detection. It adopts a new Transformer architecture, supports multi-modal input, and provides flexible deployment solutions. It has significant advantages in object detection, instance segmentation, and image classification, posture estimation, oriented object detection, object tracking, and other computer vision tasks, showing excellent performance and accuracy, which has been significantly improved compared to YOLOv8. Figure 2 shows the architecture of YOLOv11.

B. Simple, Parameter-Free Attention Module (SimAM)

SimAM is a simple parameter-free attention module. It can improve the detection of small targets and has the advantages of being lightweight and efficient. It defines an energy function based on neuroscience theory to calculate 3-D attention weights, which can effectively improve the representation ability of the network. SimAM can provide weights for each neuron. Researchers better achieve attention by evaluating the importance of each neuron. Information-rich neurons often exhibit different firing patterns from surrounding neurons. Moreover, activated neurons inhibit surrounding neurons. Researchers need to find activated neurons. The simplest way to find activated neurons is to measure the linear separability between the desired neuron and other neurons. Equation 1 shows the equation of an energy function for neurons.

$$e_t(w_t, b_t, \boldsymbol{y}, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}(y_o - \hat{x}_i)^2, \qquad (1)$$

where e_t is the overall loss (or error), w_t represents weights (parameters) of a model, $b_t$is the bias term, $\boldsymbol{y}$ denotes the set of target values and x_i is input sample from the dataset. y_t is the actual (ground truth) target value at time step *t*. $\hat{t}$ is the predicted value at time step *t*. *M* is the total number of samples in the set. y_o is a

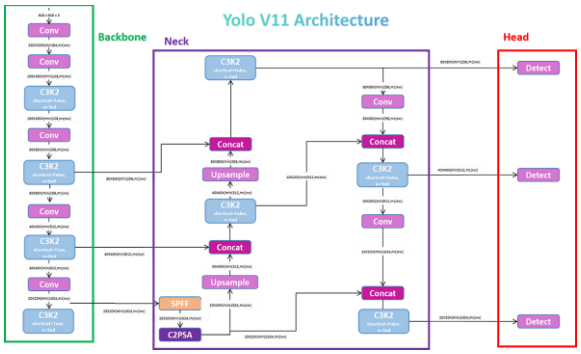

FIGURE 2. The architecture of YOLOv11.

shared reference value. $\hat{x}_i$ is the prediction for the *i*-th sample in the set. Equation 2 shows the equation of final energy function.

$$e_t(w_t, b_t, \boldsymbol{y}, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}\left[(-1-(w_t x_i + b_t))^2 + (1-(w_t + b_t))^2\right] + \lambda w_t^2 \qquad (2)$$

Each channel has an energy function. Equation 3 shows the equation of minimum energy function. This equation means that the greater the difference between a neuron and its surrounding neurons, the more important the neuron is.

$$e_t^* = \frac{4(\widehat{\sigma^2}+\lambda)}{(t-\hat{t})^2 + 2\widehat{\sigma^2} + 2\lambda} \qquad (3)$$

where e_t^* is some optimal or adjusted error/loss value and $\widehat{\sigma^2}$ is the estimated variance or a similar statistical

measure. λ is a regularisation parameter. t is the true value. $\hat{t}$ is the predicted value.

C. YOLOv11-SimAM

SimAM improves feature discriminable by explicitly modeling the importance between channels and in space, suppressing redundant features, and highlighting key information. By introducing SimAM into the backbone, this paper can more effectively focus on the prominent features of the target area. By enhancing feature representation, the target detection rate is improved, especially for small targets and targets in complex backgrounds. Figure 3 shows the architecture of YOLOv11-SimAM.

The proposed method is to add the SimAM attention mechanism to the eleventh layer of the backbone. SimAM models the saliency of each channel by display, highlighting key features and helping to eliminate background noise or irrelevant information. In the eleventh layer of the backbone, SimAM can further optimize semantic features, making the YOLOv11 model more accurate in identifying complex targets. In addition, since SimAM is a parameter-free attention module, no additional parameters are added. This will not increase the computational complexity of the model, and is more suitable for the fast characteristics of the YOLO model.

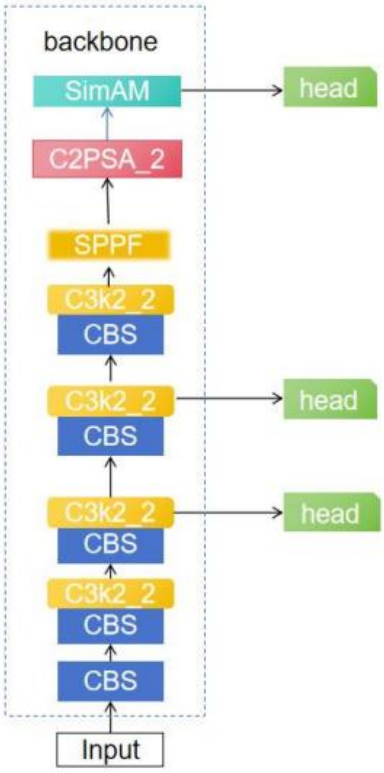

FIGURE 3. The architecture of YOLOv11-SimAM.

The structure of YOLO consists of three parts: backbone, neck, and head. The backbone is mainly used to extract and aggregate image features. The main function of the neck is to optimize the extraction of features. The head takes the optimized feature map as input and then makes predictions. YOLO divides the image into a grid and then makes predictions for each grid. At the same time, it can also allow multiple objects to be predicted in the same grid.

C. Graphical User Interface (GUI) Integration

The GUI is developed by using PyQT5. The GUI allows the user to use it easily without any coding experience. The enhancement for the GUI is to integrate with the ChatGPT API so that it can upload images and data of breast cancer tumors to obtain treatment plans. If wrong information is entered, abnormal information will not be sent to the GUI. Figure 4 shows the flowchart of the GUI and figure 5 shows the developed GUI before integrate with ChatGPT.

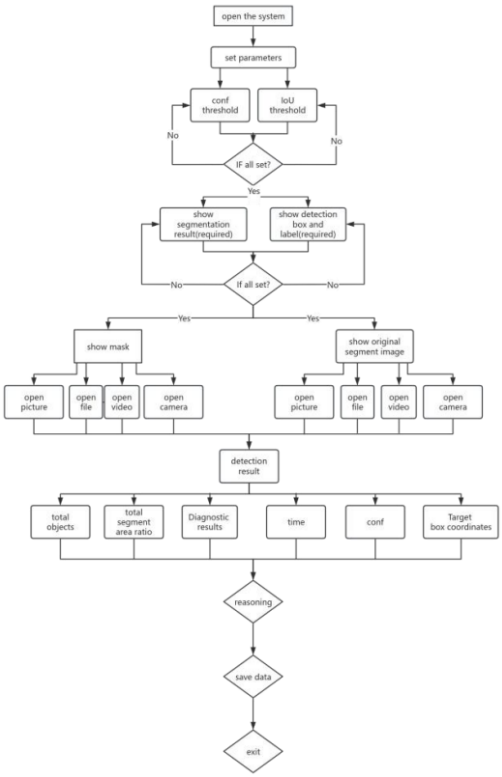

FIGURE 4. The flowchart of GUI.

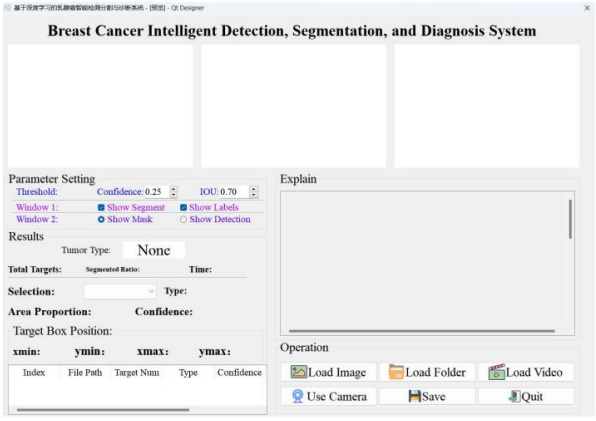

FIGURE 5. The developed GUI before integrating with ChatGPT.

IV. EXPERIMENT, RESULT & DISCUSSION

A. Dataset

Al-Dhabyani et al. [21] completed the collection of this dataset in 2018. These ultrasound images are from women aged 25 to 75 years old. The number of patients is 600 female patients. In this dataset, a total of 780 images are included, and the format of the images is Portable Network Graphic (PNG). These data are divided into three categories, namely benign, malignant, and normal. There are 437 benign breast cancer tumours, 210 malignant tumours, and 133 normal ones.

Ultrasound images were classified as benign, malignant, and normal. There are a total of 780 images in the datasets, where the training and testing set are contained in different folders. The training dataset includes 624 images and the testing dataset includes 156 images.

B. Evaluation Metrics

Evaluation metrics are important because they can be used to evaluate the performance of the model. The evaluation metrics used in this paper are classification evaluation, localization evaluation, and segmentation evaluation. Classification evaluation metrics consist of accuracy, precision, recall, localization evaluation metric is using mAP, and lastly, segmentation evaluation metric is Intersection over Union (IoU).

The accuracy is the ratio of correctly predicted instances out of the total number of instances in the dataset. Equation 4 shows the equation of the accuracy.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}, \tag{4}$$

where True Positives (TP) is the correctly predicted positive instances. True Negatives (TF) is correctly predicted negative instances. False Positives (FP) is incorrectly predicted positive instances. False Negatives (FN) is incorrectly predicted negative instances. Precision is the ratio of TP out of all positive predictions. Equation 5 shows the equation of precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Recall is the ratio of TP out of all positive predictions. Equation 6 shows the equation of recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

Average Precision (AP) is the area under the Precision-Recall curve. It is used to summarise the model performance across confidence thresholds. Equation 7 shows the equation of AP.

$$\text{AP} = \int_0^1 \text{Precision}(r)\ dr \tag{7}$$

mAP(mean Average Precision) the average value of all AP values. It is used to evaluate the performance of object detection. Equation 8 shows the equation of mAP.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \tag{8}$$

Intersection over Union (IoU) is the ratio of the overlapping area between the predicted region and the ground truth region to the combined area of both region. Equation 9 shows the equation of recall.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{9}$$

C. Model Setting

The image size is resized to 640 x 640 pixels. The epoch for model training is 150. The batch size used in this paper is 4 and the learning rate is 0.001. The optimiser is Stochastic Gradient Descent (SGD). There are three loss functions used in this model, including Bounding Box Loss, Objectness Loss and Classification Loss. Bounding Box Loss is to ensure accurate localisation of objects. Objectness Loss focuses on detecting objects against background. Classification Loss ensures that the object is classified correctly.

D. Results & Discussion

Table 1 shows the results of YOLOv8 model to S-YOLOv11model in detection with all classes. The precision of the S-YOLOv11model is 0.102 higher than that of the YOLOv8 model. The mAP50 is 0.04 higher than that of YOLOv8. The mAP50-95 is 0.022 higher than YOLOv8. This shows that the S-YOLOv11model performs better than the YOLOv8 model in breast cancer detection tasks.

TABLE 1. Comparison of YOLOv8 model to S-YOLOv11model in detection (class all)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11
Precision	0.704	0.666	0.561	0.638	0.806
Recall	0.653	0.682	0.613	0.662	0.635
mAP50	0.684	0.677	0.620	0.650	0.724
mAP50-95	0.421	0.410	0.398	0.382	0.443

Table 2 shows the results of YOLOv8 model to S-YOLOv11model in detection with benign class. The performance of the S-YOLOv11 model is similar to that of the YOLOv8 model in the benign class. The precision of the S-YOLOv11 model is higher than that of the YOLOv8 model, but other indicators are slightly lower than those of the YOLOv8 model.

TABLE 2. Comparison of YOLOv8 model to S-YOLOv11model in detection (class benign)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11

Precision	0.722	0.791	0.594	0.781	0.806
Recall	0.750	0.761	0.715	0.716	0.693
mAP50	0.783	0.804	0.698	0.772	0.764
mAP50-95	0.577	0.567	0.521	0.529	0.565

Table 3 shows the results of YOLOv8 model to S-YOLOv11model in detection with malignant class. Although the S-YOLOv11 model performs worse than YOLOv8 in the benign class, it performs much better than YOLOv8 in the malignant class. The precision of the S-YOLOv11 model is 0.12 higher than YOLOv8. The recall is 0.022 higher than that of YOLOv8. The mAP50 is 0.1 higher than that of YOLOv8. The mAP50-95 is 0.056 higher than YOLOv8.

TABLE 3. Comparison of YOLOv8 model to S-YOLOv11model in detection (class malignant)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11
Precision	0.686	0.541	0.529	0.494	0.806
Recall	0.556	0.603	0.511	0.609	0.578
mAP50	0.584	0.551	0.543	0.529	0.684
mAP50-95	0.265	0.252	0.275	0.235	0.321

Table 4 shows the results of YOLOv8 model to S-YOLOv11model in segmentation with all classes. The proposed model performed the best among the other YOLO, followed by YOLOv8. Since YOLOv10 does not have a segment-related model, there are no data related to the YOLOv10 model in the segmentation task. Therefore, the model that performs the worst in the segmentation task is YOLOv11. After adding the SimAM attention mechanism to the backbone of the YOLOv11 model, the performance of the model has been greatly improved. The precision is 0.157 higher than that of the original YOLOv11 model. However, therecall is 0.06 lower than that of YOLOv11. The mAP50 is 0.052 higher than the YOLOv11 model. The mAP50-95 is 0.028 higher than that of YOLOv11

TABLE 4. Comparison of YOLOv8 model to S-YOLOv11model in segmentation (class all)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11
Precision	0.746	0.711	N.A.	0.676	0.833
Recall	0.692	0.639	N.A.	0.710	0.650
mAP50	0.732	0.683	N.A.	0.687	0.739
mAP50-95	0.461	0.450	N.A.	0.439	0.467

Table 5 shows the results of YOLOv8 model to S-YOLOv11model in segmentation with benign class. The performance of the S-YOLOv11 model is similar to that of the YOLOv8 model. However, the precision is almost 0.1 higher than that of the YOLOv8 model. Other S-YOLOv11 evaluation indicators are similar to YOLOv8.

TABLE 5. Comparison of YOLOv8 model to S-YOLOv11model in segmentation (class benign)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11
Precision	0.754	0.849	N.A.	0.816	0.852
Recall	0.784	0.767	N.A.	0.750	0.721
mAP50	0.807	0.830	N.A.	0.806	0.795
mAP50-95	0.591	0.610	N.A.	0.595	0.590

Table 6 shows the results of YOLOv8 model to S-YOLOv11model in segmentation with malignant class. The performance of the S-YOLOv11 model is better than those of the other four models. The precision of the S-YOLOv11 model is 0.284 higher than YOLOv11, indicating a great improvement. The recall is also 0.067 higher than that of YOLOv11. The mAP50 is 0.14 higher than that of 51 YOLOv11. The mAP50-95 is 0.069 higher than that of YOLOv11. This shows that the performance of the S-YOLOv11 model is greatly improved compared to the original YOLOv11 model.

TABLE 6. Comparison of YOLOv8 model to S-YOLOv11model in segmentation (class malignant)

Metric	YOLOv8	YOLOv9	YOLOv10	YOLOv11	S-YOLOv11
Precision	0.738	0.574	N.A.	0.529	0.813
Recall	0.657	0.511	N.A.	0.511	0.578
mAP50	0.600	0.537	N.A.	0.543	0.683
mAP50-95	0.330	0.291	N.A.	0.275	0.344

The enhanced GUI allows the usage of ChatGPT. After the user done with the image segmentation, the GUI explains the segmented image. Figure 6 shows

the enhanced GUI with an explanation for the segmented tumour.

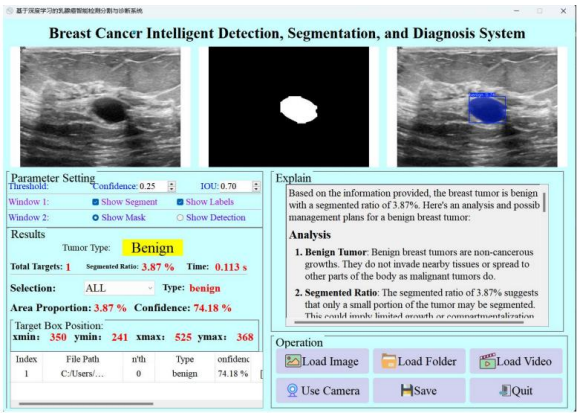

FIGURE 6. The enhanced GUI displaying reasoning.

The S-YOLOv11 model proposed in this study has improved performance in object detection and segmentation tasks compared to the YOLOv8, YOLOv9, YOLOv10, and YOLOv11 models. In the detection task, the mAPis 0.724, the recall is 0.635, and the precision is 0.806. In the segmentation task, the mAP is 0.739, the recall is 0.65, and the precision is 0.833. The S-YOLOv11 model has made significant progress compared to the YOLOv11 model. This study also realized the interaction between ChatGPT and GUI. Inference was achieved by accessing ChatGPT's API. This allowed oncologists to have more time to handle the cases, thus improving their efficiency.

The integration of ChatGPT into the GUI to provide natural language explanations for breast cancer tumor segmentation is part of the innovation. However, it has not been clinically validated by medical experts. This is an important limitation that must be addressed before deploying the system in real-world medical practice.

V. CONCLUSION

In this paper, four models from YOLOv8 to YOLOv11 were compared. The study found that YOLOv8 performed the best. Although the performance of YOLOv11 is not as good as that of YOLOv8, this study achieved better results than YOLOv8 by adding the SimAM attention mechanism to the backbone of YOLOv11 to improve the feature extraction ability of the model. For comparison, in the detection task, the mAP of S-YOLOv11is 0.724, the recall rate is 0.635, and the precision is 0.806. In the segmentation task, the mAP is 0.739, the recall rate is 0.65, and the precision is 0.833. The performance is approximately 10% higher than that of YOLOv8. Additionally, by utilizing the ChatGPT API, breast cancer tumor images can be effectively interpreted. This function was integrated into GUI to enhance it's capabilities. Finally, the user can use the GUI to open the breast cancer tumor image and analyze the breast cancer tumor through the S-YOLOv11model to generate relevant data. After that, the GUI will generate an explanation, including the interpretation of the breast cancer tumor data and the treatment plan.

A. Future Work

There is still some work to be done in the future. This study only uses ChatGPT's API, which takes a long time to access. And it is very difficult to obtain ChatGPT's API, which requires a lot of money to maintain. In addition, the data set used in this study is small and the amount of data that can be collected is limited. In order to conduct further research, more relevant data need to be obtained from hospitals for model training. In the future, the research will independently develop large language models, which can save costs and make the GUI respond faster.

Currently, the integration of ChatGPT provides natural language explanations of tumor characteristics. However, there is no clinical validation of the AI-generated medical reports. Future work should involve expert evaluation from radiologists or oncologists to assess the reliability and accuracy of these explanations. A comparison study between ChatGPT-generated reports and expert-written reports could provide insights into the credibility and usefulness of AI-driven explanations in clinical practice.

This study is currently limited to ultrasound imaging. Future research should explore the integration of mammography, MRI, and CT scans to enhance model generalizability. A multi-modal dataset would enable the model to learn cross-domain representations, improving detection and segmentation accuracy across different imaging techniques.

ACKNOWLEDGMENT

There is no funding agencies supporting the research work

AUTHOR CONTRIBUTIONS

Kai Liang Lew: Writing – Review, Edit, Original Draft Preparation;

Chean Khim Toa: Project Administration, Writing – Review & Editing, Supervision;

Pengfei Zhou: Conceptualization, Data Curation, Methodology, Validation,

Chia Shyan Lee: Writing – Review & Editing;

Tetuko Kurniawan: Writing – Review & Editing;

Suleiman Aliyu Babale: Writing – Review & Editing;

Cheng Zheng: Writing – Review & Editing;

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

REFERENCES

[1] R.L. Siegel, K.D. Miller, N.S. Wagle and A. Jemal, "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17-48, 2023.
DOI: https://doi.org/10.3322/caac.21763

[2] H. Gandhi and K. Kumar, “Artificial Intelligence for the Management of Breast Cancer: An Overview,” *Current Drug Discovery Technologies*, vol. 21, no. 4, Art. no. e031123223115, 2024. DOI: https://doi.org/10.2174/0115701638262066231030052520

[3] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei and X. Wei, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," *arXiv*, 2022. DOI: https://doi.org/10.48550/arXiv.2209.02976

[4] Y. Zhang, P. Li, Y. Lan, X. Jia, and Y. Lv, “Research on Breast Cancer Detection Methods Based on ODMV-MulDyHead-YOLO,” *IEEE Access*, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3508780

[5] D. Shah, M.A.U. Khan, M. Abrar and M. Tahir, "Optimizing Breast Cancer Detection With an Ensemble Deep Learning Approach," *International Journal of Intelligent Systems*, vol. 2024, no. 1, 2024. DOI: https://doi.org/10.1155/2024/5564649

[6] S.R. Kebede, F.G. Waldamichael, T.G. Debelee, M. Aleme, W. Bedane, B. Mezgebu and Z.C. Merga, "Dual view deep learning for enhanced breast cancer screening using mammography," *Scientific Reports*, vol. 14, no. 1, 2024. DOI: https://doi.org/10.1038/s41598-023-50797-8

[7] T.A. Toma, S. Biswas, M.S. Miah, M. Alibakhshikenari, B.S. Virdee, S. Fernando, M.H. Rahman, S.M. Ali, F. Arpanaei, M.A. Hossain, M.M. Rahman, M. Niu, N.O. Parchin and P. Livreri, "Breast Cancer Detection Based on Simplified Deep Learning Technique With Histopathological Image Using BreaKHis Database," *Radio Science*, vol. 58, no. 11, 2023. DOI: https://doi.org/10.1029/2023RS007761

[8] W. T. Chan, "Review on Present-day Breast Cancer Detection Techniques,"*International Journal on Robotics, Automation and Sciences*, vol. 6, no. 1, pp. 94–101, 2024. DOI: https://doi.org/10.33093/ijoras.2024.6.1.13

[9] K. V. Sriram, R. H. Havaldar, "Analytical review and study on object detection techniques in the image,"*International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, no. 05, pp. 2150031, 2021. DOI: https://doi.org/10.1142/S1793962321500318

[10] R. Ye, J. Shao, J. Shao, and J. Ji, “Review of deep learning based target detection algorithm research,” in *Proceedings of SPIE, 6th International Conference on Mechatronics and Intelligent Robotics (ICMIR2022)*, vol. 12301, Art. no. 123011T, 2022 DOI: https://doi.org/10.1117/12.2644526

[11] P. Viola, M. Jones, "Robust real-time face detection,"*Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, pp. 747–747, 2001. DOI: https://doi.org/10.1109/ICCV.2001.937709

[12] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection,"*2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, 2005. DOI: https://doi.org/10.1109/CVPR.2005.177

[13] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, pp. 677, 2023. DOI: https://doi.org/10.3390/machines11070677

[14] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain and A. Elmaghraby, "Tissue classification and segmentation of pressure injuries using convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 51-58, 2018. DOI: https://doi.org/10.1016/j.cmpb.2018.02.018

[15] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio, and S. Vitabile, “A Yolo-Based Model for Breast Cancer Detection in Mammograms,” *Cognitive Computation*, vol. 16, pp. 107–120, 2024. DOI: https://doi.org/10.1007/s12559-023-10189-6

[16] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio and S. Vitabile, "A Yolo-Based Model for Breast Cancer Detection in Mammograms," *Cognitive Computation*, vol. 16, no. 1, pp. 107-120, 2024. DOI: https://doi.org/10.1007/s12559-023-10189-6

[17] P.K. Samanta, A. Basuli, N.K. Rout and G. Panda, "Improved Breast Cancer Detection from Ultrasound Images Using YOLOv8 Model," *2023 IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, pp. 1-6, 2023. DOI: https://doi.org/10.1109/AESPC59761.2023.10390341

[18] H. Gui, T. Su, X. Jiang, L. Li, L. Xiong, J. Zhou and Z. Pang, "FS-YOLOv9: A Frequency and Spatial Feature-Based YOLOv9 for Real-time Breast Cancer Detection," *Academic Radiology*, vol. 32, no. 3, pp. 1228-1240, 2025. DOI: https://doi.org/10.1016/j.acra.2024.09.048

[19] L. Zheng, W. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E.P. Xing, H. Zhang, J.E. Gonzalez and I. Stoica, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," *arXiv*, 2023. DOI: https://doi.org/10.48550/arXiv.2306.05685

[20] A.Y. Yuan, Y. Gao, L. Peng, L. Zhou, J. Liu, S. Zhu and W. Song, "Hybrid deep learning network for vascular segmentation in photoacoustic imaging," *Biomedical Optics Express*, vol. 11, no. 11, pp. 6445, 2020. DOI: https://doi.org/10.1364/BOE.409246

[21] W. Al-Dhabyani, M. Gomaa, H. Khaled and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, pp. 104863, 2020. DOI: https://doi.org/10.1016/j.dib.2019.104863

[22] F.F. Ting, Y.J. Tan and K.S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, pp. 103-115, 2019. DOI: https://doi.org/10.1016/j.eswa.2018.11.008

[23] Y. J. Tan, K. S. Sim and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," *2017 International Conference on Robotics, Automation and Sciences (ICORAS)*, pp. 1-5, 2017. DOI: https://doi.org/10.1109/ICORAS.2017.8308076

[24] J.J. Toh and M.S. Sayeed, "Review on Detecting Pneumonia in Deep Learning," *International Journal on Robotics, Automation and Sciences*, vol. 6, no. 1, pp. 70-77, 2024. DOI: https://doi.org/10.33093/ijoras.2024.6.1.10

[25] J. Cheng and S. Haw, "Mental Health Problems Prediction Using Machine Learning Techniques," *International Journal on Robotics, Automation and Sciences*, vol. 5, no. 2, pp. 59-72, 2023. DOI: https://doi.org/10.33093/ijoras.2023.5.2.7

[26] N.M. Loorutu, H. Yazid and K.S.A. Rahman, "Prostate Cancer Classification Based on Histopathological Images," *International Journal on Robotics, Automation and Sciences*, vol. 5, no. 2, pp. 43-53, 2023. DOI: https://doi.org/10.33093/ijoras.2023.5.2.5