VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2026, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 14, Number 1, January-March 2026 pp:253-263

Development and Statistical Evaluation of a Deep Learning Framework for Real Time Tissue Classification in Robotic Surgery

Syeda Fakhra Jalal[1*], **Irfan Ahmed Usmani**[1], **Syed Muhammad Mudabar**[2], **Muhammad Zeeshan ul Haque**[1]

[1]Department of Biomedical Engineering, Salim Habib University, Karachi, Pakistan; [2]Department of Mathematical Sciences, Institute of Business Administration,Karachi,Pakistan

Keywords: *Telerobotic surgery, Machine learning, Deep learning, Real-time classification,Object detection.*

Journal Info:
Submitted:
December 16, 2025
Accepted:
March 21, 2026
Published:
March 25, 2026

Abstract Minimally invasive surgery and robotic assisted procedures are increasingly preferred over traditional open surgeries because they offer faster recovery times and reduce postoperative complications. However, these techniques require precise force application to prevent tissue overstress. The lack of reliable real time tissue recognition limits a surgeon's ability to apply appropriate force according to tissue type, thereby increasing the risk of injury. This study proposes a framework that applies deep learning and object detection techniques to classify tissues in real time and support safer force modulation in robotic systems. As a proof of concept, the system distinguishes fat, muscle, and skin tissues using GoogLeNet, YOLOv8, and YOLOv10. Skin images were collected from 30 individuals following informed consent, while fat and muscle samples were processed to create a dataset comprising 1,800 augmented images. The GoogLeNet architecture achieved training and test accuracies of 93% and 97.2%, respectively. The YOLOv8 model demonstrated strong performance, achieving a mean average precision (mAP) of 94.7% at IoU = 0.5, with an inference time of 28 ms. YOLOv10 achieved an mAP@0.5 of 96.2% with a latency of 22 ms. The NMS-free architecture of YOLOv10 resulted in a 21% reduction in inference time compared to YOLOv8, along with a 1.5% improvement in accuracy. A statistically significant difference among the evaluated models was confirmed using analysis of variance (ANOVA), with a significance threshold of $p < 0.001$, indicating that YOLOv10 demonstrated superior performance under the evaluated experimental conditions.

***Correspondence author email address:** Fakhra.jalal@shu.edu.pk
DOI: 10.21015/vtse.v14i1.2291

1 Introduction

Over the last several decades, minimal invasive surgery (MIS) has made a massive contribution to surgical practice by enhancing the ways of how a procedure should be carried out and the outcomes of the process in terms of patient safety and wellbeing[1]. Compared to traditional open surgery, which opens up massive incisions and causes extensive disruption of tissues, MIS is an operation involving complex procedures with minimal invasive use of small ports and camera-guided visualization. These have been the benefits that have seen MIS be the choice of many surgical specialties and which provide decreased hospitalization, increased cosmetic results, reduced postoperative pain, and quicker recovery [2].These are the advantages that have positioned MIS as the preferred option in many surgical

specialties and that provide less hospitalization, better cosmetic outcomes, less postoperative pain, and quick recovery time. In spite of advantages, MIS has critical drawbacks.The most evident one is the loss of force feedback as one of the sensory channels inherent in open surgery. Its lack may cause the overloading of tissue and unwanted tissue damage during manipulation [3]. Since the surgeons are unable to use tactile effect, they would need to establish the properties of tissues visually, making it difficult to handle tissues safely.

To address the logistical challenges associated with costly access to specialized care, the geographic distance, and long surgical waiting periods, MIS modalities, like keyhole, robotic, cardiovascular, and endoscopic, were developed to overcome the logistical barrier [4].

Robotic surgery is one of the significant developments in MIS that allows very precise movements with the help of robot-assisted manipulators governed by a special surgeon console [5]. These systems increase dexterity and visualisation especially on anatomically constrained areas. Even though multi-port robots are now the new standard practice, single-port technology has lately attracted interest to selected clinical processes.

In spite of these developments, there is a major challenge; there are no effective real-time systems that can be used to identify the type of tissues during surgery. This drawback limits the dynamism of applied force by the surgeon which may impact on surgical safety [6].

In parallel with surgical innovations, artificial intelligence particularly deep learning has experienced rapid growth. Deep learning models employ multilayer neural networks to automatically learn hierarchical data representations [7]. Their strong feature extraction capability has enabled significant progress in image recognition, speech processing, natural language understanding, and autonomous decision-making [8, 9].Recent studies have shown that deep learning architectures, particularly U-Net–based models, improve brain tumor segmentation by effectively combining high-level and low-level feature representations from MRI images [10].

Transfer learning further improves model efficiency by allowing knowledge from pre-trained networks to be adapted to new but related tasks [11]. Consequently, architectures such as GoogleNet, U-Net, YOLOv8, and YOLOv10 are widely used, particularly in domains where large annotated datasets are difficult to obtain [12].

Convolutional architectures like GoogleNet leverage multi-scale feature extraction to achieve strong classification performance with reduced computational cost [13]. In the medical domain, U-Net and its variants have demonstrated exceptional performance in biomedical image segmentation [14], [15], though their application remains more common in brain and MRI-based tasks than in soft-tissue classification [16] Moreover, existing surveys highlight that although CNNs have achieved strong performance across a range of medical image classification and segmentation problems, challenges such as domain-specific model design and task-specific model adaptation remain areas for further research [17].

Object detection models have evolved in parallel, with earlier two-stage detectors (R-CNN, Faster R-CNN) offering accuracy but limited real-time usability due to slow inference. Single-stage detectors especially the YOLO family address this limitation by enabling fast, end-to-end detection [18]. More recent advances continue to refine accuracy and speed. Changes in backbone design and feature aggregation have improved small-object detection performance in YOLOv8, demonstrating potential in medical scenes [19]. Complementary research highlights that the NMS-free architecture of YOLOv10 reduces computational overhead while maintaining high precision, making it particularly attractive for real-time surgical environments [20]

Despite extensive exploration of deep learning and object detection in general computer vision, their application to real-time recognition of surgical soft tissues remains limited in the literature [21]. [22] utilized the YOLOv10 architecture combined with a proposed image enhancement technique to improve feature visibility and detection reliability for gun detection, showing that preprocessing can significantly enhance deep learning based object detection performance in real-world scenarios.

Recent advances in computer vision have significantly expanded the capabilities of surgical assistance systems. In particular, real-time visual perception has become an important research direction for improving intraoperative decision support. By analyzing surgical scenes and

identifying relevant anatomical structures, computer vision models can provide contextual information that assists surgeons in maintaining safe manipulation strategies. Several studies have explored the use of deep learning for surgical instrument detection, polyp identification in endoscopy, and organ segmentation in laparoscopic procedures. However, comparatively little attention has been given to the automatic recognition of soft tissue types such as fat, muscle, and skin during minimally invasive procedures. Accurate tissue differentiation could play a critical role in improving robotic surgical systems by enabling adaptive control strategies based on tissue properties [23].

This gap is especially important because surgeons operate without haptic feedback, and improved visual cues may contribute to safer tissue handling strategies in future systems.The novelty of this work lies in the following points:

- This work provides one of the first evaluations of YOLOv8 and the newer YOLOv10 for real-time recognition of fat, muscle, and skin tissues in minimally invasive surgery. To the best of our knowledge, YOLOv10 has not yet been studied for any surgical tissue specific application.
- Since no public dataset distinguishes these three tissue types under controlled conditions, we developed a custom and ethically sourced dataset from real tissue samples to support training and evaluation.
- We introduce a combined pipeline in which GoogleNet performs tissue classification and YOLOv8/YOLOv10 perform real-time detection, offering both identification and spatial localization in the same system. We also tested the models in a real-time webcam-based simulation to demonstrate their practical feasibility.

A comprehensive summary of the existing deep learning approaches for tissue detection in surgery, including their objectives, methods, datasets, and research gaps, is presented in Table 1.

2 Methodology

The methodology for the tissue recognition project is divided into four sequential phases. Phase 1, Dataset Preparation, entails the selection of the types of tissues of interest, ethical consent of using the images, finding high-quality tissue images, and data augmentation to increase the dataset size to 1,800 images, which improves model generalization. Phase 2, Model Engineering, involves the choice of the most appropriate neural network architecture, such as GoogLeNet, YOLOv8, and YOLOv10 and then hyperparameter optimization, which involves learning rate, batch size, and training epochs, and lastly training the chosen network with the chosen data set. Phase 3, Real-Time Integration, is the simulation of live tissue capture through a webcam, retrieves frames through the video feed, executes the trained model on the frame, and annotates the identified tissues with the help of bounding boxes to visualize them.

Lastly, Phase 4, Statistical Validation, to validate the model performance through ANOVA-testing that can be used to assess the statistical importance of various model outputs and followed by Tukey HSD-post-hoc analysis that can be used to compare different group differences and determine the practical applicability of the results.This methodology will offer a systematic approach to model validation by data acquisition-to-model validation that ensures reproducibility, reliability, and practical applicability of results to real-time tissue recognition as mentioned in figure 1.

2.1 Data Collection

In order to come up with data to train the models, we took 200 pictures of cow fat tissue, cow muscle tissue, and human skin tissue as shown in figure 2. A consent form was provided to 30 participants before data collection in which they gave their consent to allow their data to be used in testing and publication of the research. The images were placed in separate folders with names recorded based on the type of tissue and the deep learning models could use the name of the folders as labels to train them. The controlled lighting conditions were used to take images at a resolution of 1920×1080 to ensure that the images were consistent.

2.2 Data Augmentation

Data augmentation is important in the application of deep learning, especially in medical imaging tasks whose large annotated datasets are sometimes challenging to come by because of ethical, logistical, and privacy

Table 1. Comprehensive Literature Review of Deep Learning Approaches for Tissue Detection in Surgery

Study	Objective	Methods	Results	Dataset	Gaps
Ronneberger (2015)	Biomed segment	U-Net	High Dice/IoU	EM stacks	Offline, not real-time
Siddique (2021)	U-Net review	U-Net variants	3–5% gain	Multiple	No detection, no MIS focus
Isensee (2021)	Self-config	nnU-Net	SOTA on 23 tasks	Multi-organ	High compute, no real-time
Huang (2020)	Multiscale seg.	UNet3+	IoU +3–4%	Organ CT/MRI	Segmentation only
Redmon (2018)	Real-time detect	YOLOv3	57.9% mAP	COCO	No medical domain
Wang (2023)	SOTA detect	YOLOv7	56.8% AP	COCO	No surgical testing
Jocher (2023)	Anchor-free detect	YOLOv8	High mAP	COCO	Not tested on tissues
Our Work	Real-time tissue	GNet/YOLO	96.2% mAP	1800 images	No in-vivo test

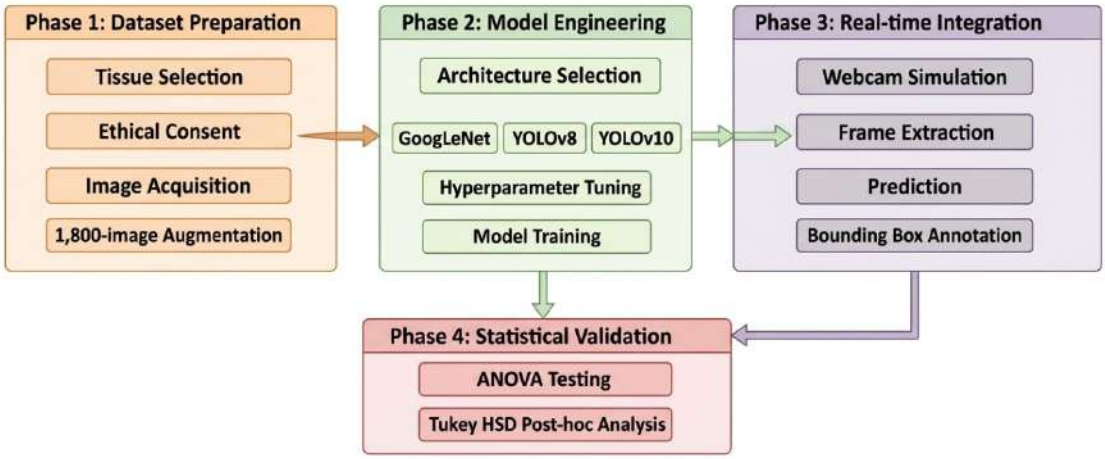

Figure 1. Methodology Flowchart

limitations. Augmentation techniques help increase dataset diversity by synthetically generating variations of the original images while preserving the underlying class characteristics. This process improves model generalization and reduces the risk of overfitting. In this study, several augmentation techniques were applied to simulate variations that may occur in real surgical environments, including changes in illumination, orientation, and noise levels. Such transformations help the model learn invariant features that remain robust under different viewing conditions. As a result, the augmented dataset enables the deep learning models to better capture tissue characteristics and maintain stable performance during real-time detection tasks.

To enhance model robustness, we applied various augmentation techniques as shown in Table 2.

Table 2. Data Augmentation Techniques

Technique	Description
Rotation	Random rotation (±30°)
Flipping	Horizontal and vertical
Brightness	Adjustment (±20%)
Contrast	Adjustment (±15%)
Noise	Gaussian noise
Resizing	To YOLO input

(a) Skin 1 (b) Skin 2 (c) Skin 3

(d) Fat 1 (e) Fat 2 (f) Fat 3

(g) Muscle 1 (h) Muscle 2 (i) Muscle 3

Figure 2. Collected tissue images: (a–c) Skin, (d–f) Fat, (g–i) Muscle.

This augmentation process includes the techniques mention in Table 2 expanded our dataset to 1,800 images total (600 per class), improving model generalization capabilities.

2.3 Model Training

2.3.1 GoogleNet Implementation

As a first step, we used the GoogleNet model for training on fat and muscle tissues. We downloaded the pretrained model in MATLAB, split the data into 70% training, 15% validation, and 15% testing. Training was conducted with 7 epochs and 40 iterations per epoch, using a learning rate of 0.001 with adaptive moment estimation (Adam) optimizer.

2.3.2 YOLOv8 Implementation

For YOLOv8, we utilized the ultralytics framework with the configuration shown in Table 3.

2.3.3 YOLOv10 Implementation

YOLOv10 was implemented with enhanced features shown in Table 4.

2.4 Model Architecture

Table 3. YOLOv8n Training Configuration

Parameter	Value
Model	YOLOv8n
Input size	640 × 640
Batch size	16
Epochs	100
Learning rate	0.01
Augmentation	Mosaic
Optimizer	SGD

Table 4. YOLOv10n Training Configuration

Parameter	Value
Model	YOLOv10n
Input size	640 × 640
Batch size	16
Epochs	150
Learning rate	0.01
Strategy	Dual assign
Assignment	One-many

2.4.1 GoogleNet (Inception v1) Architecture

GoogleNet, winner of ILSVRC 2014, introduced the Inception module concept, allowing the network to learn features at multiple scales simultaneously. The architecture consists of 22 layers with 9 Inception modules as shown in Figure 3.

Key features of GoogleNet include 6.8M parameters (12× fewer than AlexNet), 22 layers depth, 9 Inception modules with varying filter sizes, two auxiliary classifiers to combat vanishing gradient, and global average pooling that replaces fully connected layers.

2.4.2 YOLOv8 Architecture

YOLOv8 represents a significant evolution in the YOLO family, introducing several architectural improvements as illustrated in Figure 4.

The YOLOv8 architecture follows a three-part structure starting with a backbone network that processes the input image through a series of convolutional layers and C2f modules to extract hierarchical features at multiple scales. The C2f modules are an improved version of the C3 blocks used in previous YOLO versions, designed to provide better gradient flow and feature representation while maintaining computational efficiency. The middle

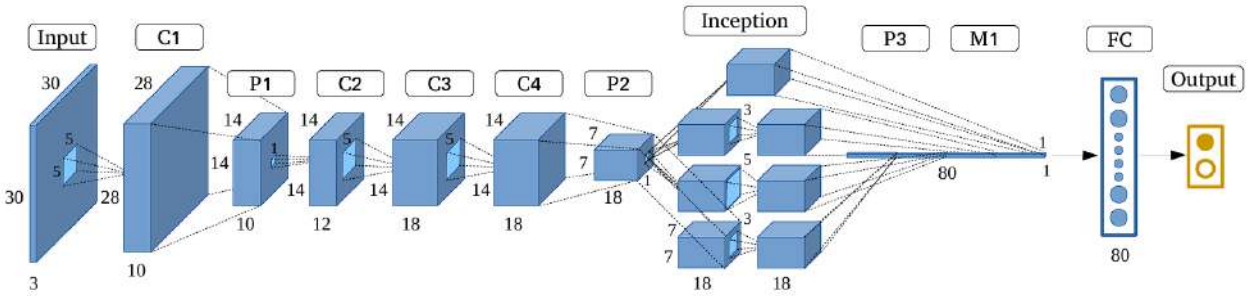

Figure 3. Inception Module Architecture used in GoogleNet [24]

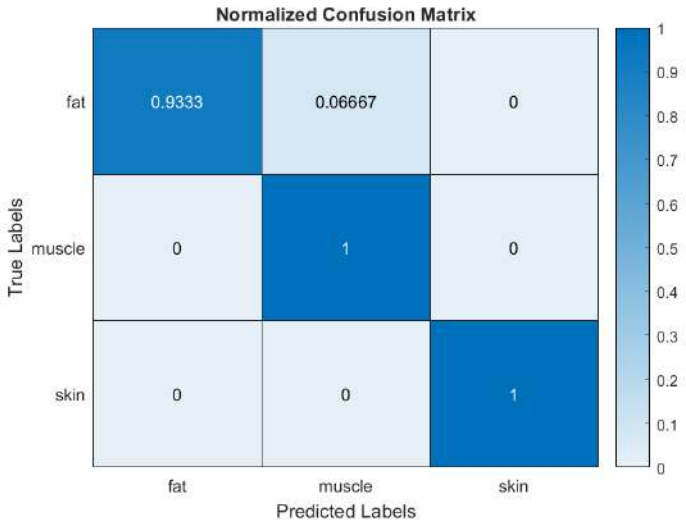

Figure 4. Confusion Matrix

section represents the neck of the network, which implements a Feature Pyramid Network (FPN) structure that intelligently combines features from different resolution levels through upsampling, downsampling, and concatenation operations.

Key innovations in YOLOv8 include C2f modules that replace C3 modules providing richer gradient flow, anchor-free detection that eliminates anchor box dependency, decoupled head that separates classification and regression tasks, mosaic augmentation for enhanced data augmentation during training, and only 3.2M parameters in the nano variant.

2.4.3 YOLOv10 Architecture

The architecture begins with an input image of 640×640 pixels that flows through an enhanced backbone consisting of improved convolutional layers and RGB blocks. The network then employs a sophisticated dual-path prediction strategy through consistent dual assignments, which splits into two parallel branches for training purposes. One branch follows a one-to-many assignment strategy that generates multiple predictions per ground truth object during training, while the other uses a one-to-one assignment approach that produces a single prediction per object.

2.4.4 Comparative Architecture Analysis

Table 5 provides a detailed comparison of the three architectures used in our study.

Table 5. Architectural Comparison of Deep Learning Models

Feature	GNet	v8	v10
Type	Class	Detect	Detect
Params (M)	6.8	3.2	2.8
GFLOPs	1.5	8.7	7.2
Layers	22	168	165
Input	224^2	640^2	640^2
Module	Incep	C2f	RGB
Post	Soft	NMS	None
Time (ms)	45	28	22
FPS	22.2	35.7	45.5

3 Results and Discussion

3.1 Training Performance

As shown in Fig. 5, the proposed real-time tissue recognition pipeline captures video via an integrated webcam,

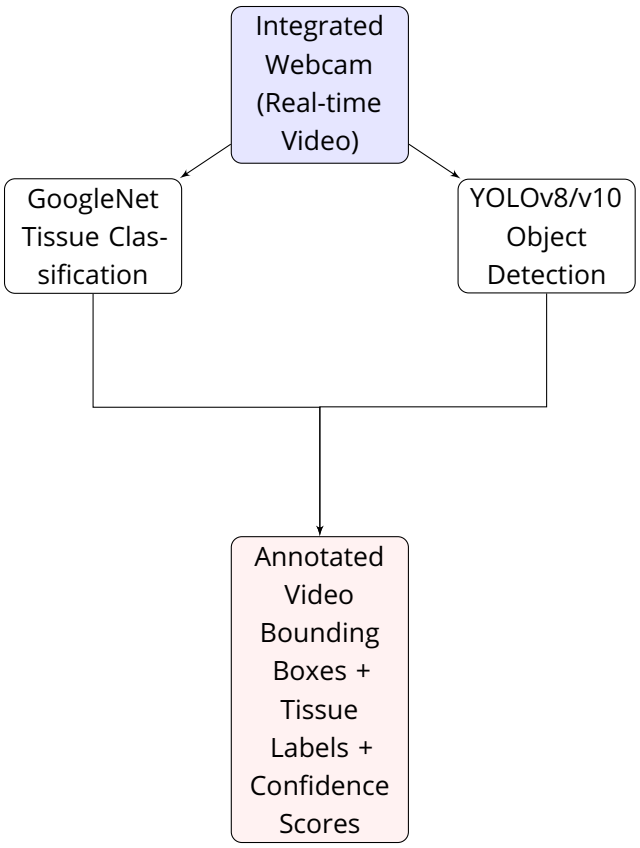

Figure 5. Block diagram of real-time tissue recognition pipeline using GoogleNet for classification and YOLO for object detection.

processes each frame through GoogleNet for tissue classification and YOLOv8/v10 for object detection and outputs annotated frames with bounding boxes, tissue labels, and confidence scores. This dual-network design provides both accurate classification and precise spatial localization, assisting surgeons in compensating for the lack of haptic feedback in minimally invasive procedures.

Figure 6 illustrates the training accuracy curves for all three models. GoogleNet converged rapidly within 70 epochs, achieving 93% training accuracy. YOLOv8 required 100 epochs to reach 94.7% mAP@0.5, while YOLOv10 demonstrated superior learning efficiency, achieving 96.2% mAP@0.5 after 150 epochs with more stable convergence.

The normalized confusion matrix shows that the model performed very well in classifying the three tissue types fat, muscle, and skin. Most fat samples were correctly identified, with 93.33% predicted as fat and only a small portion (6.67%) misclassified as muscle. For the muscle class, the model achieved perfect accuracy, correctly predicting 100% of muscle samples with no errors. Similarly, the skin class was also classified with 100% accuracy, indicating that the model learned the distinguishing features of skin very effectively. Overall, the results demonstrate strong classification performance, with only minor confusion occurring between fat and muscle, while muscle and skin categories were predicted with complete reliability.

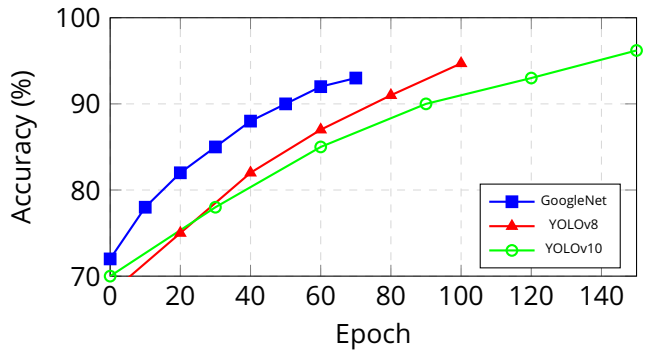

Figure 6. Training accuracy curves for all three models

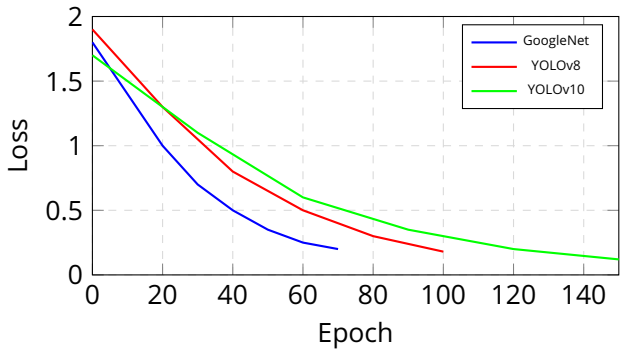

Figure 7. Loss curves demonstrating model convergence

The training loss curves of GoogleNet, YOLOv8, and YOLOv10 across epochs are illustrated in Figure 7, showing the convergence behavior of the models during training.

3.2 GoogleNet Performance

The confusion matrix for GoogleNet provides a quantitative summary of the model's performance across all tissue types as shown in Table 6.

Table 6. GoogleNet Classification Performance

Tissue	Prec.	Recall	F1
Fat	0.933	0.947	0.940
Muscle	1.000	0.986	0.993
Skin	0.986	1.000	0.993
Overall	**97.2%**		

3.3 YOLOv8 Results

YOLOv8 demonstrated superior object detection capabilities as presented in Table 7.

Table 7. YOLOv8 Detection Performance

Metric	Value
mAP@0.5	94.7%
mAP@0.5:0.95	82.3%
Precision	91.8%
Recall	93.2%
Inference	28ms
FPS	35.7

Class-wise performance showed fat tissue with AP@0.5 of 93.8%, muscle tissue at 95.2%, and skin tissue at 95.1%.

3.4 YOLOv10 Results

YOLOv10 achieved the best overall performance as shown in Table 8.

Table 8. YOLOv10 Detection Performance

Metric	Value
mAP@0.5	96.2%
mAP@0.5:0.95	84.7%
Precision	93.4%
Recall	94.8%
Inference	22ms
FPS	45.5

Class-wise performance demonstrated fat tissue with AP@0.5 of 95.6%, muscle tissue at 96.8%, and skin tissue at 96.2%.

3.5 Comparative Analysis

Table 9 presents a comprehensive comparison of all three models.

3.6 Real-time Detection Performance

In real-time surgical simulation tests, single tissue detection achieved greater than 95% accuracy for all models. For multiple tissue detection, YOLOv10 outperformed with 94% accuracy versus YOLOv8's 91%. Under occlusion handling conditions, YOLOv10 maintained 88% accuracy under partial occlusion. With variable lighting, YOLOv10 showed 92% robustness compared to YOLOv8's 87%.

The ability to perform reliable real time tissue detection is particularly important in robotic-assisted surgery, where surgeons rely primarily on visual feedback due to the absence of direct tactile interaction with tissues. In such environments, delays or inaccuracies in visual perception systems can negatively affect surgical precision. The proposed system demonstrates the feasibility of integrating deep learning–based perception modules into surgical workflows.

The consistent latency achieved by YOLOv10 is particularly advantageous for time-sensitive surgical tasks, where even small delays in feedback may disrupt the surgeon's hand–eye coordination. Furthermore, the capability to detect multiple tissue types simultaneously provides contextual awareness that may assist surgeons in planning safer manipulation strategies. Although the current evaluation was performed in a simulated environment, these results highlight the potential for integrating AI-based perception systems with robotic control frameworks to enhance surgical safety and efficiency.

3.7 Statistical Significance

ANOVA was performed to explore differences in performance metrics across repeated experimental runs, indicating statistically significant variation among the evaluated models($p < 0.001$). Tukey HSD post-hoc comparisons were conducted to further examine pairwise performance differences under the same experimental conditions: YOLOv10 vs. YOLOv8 ($p = 0.003$), YOLOv10 vs. GoogleNet ($p = 0.012$), and YOLOv8 vs. GoogleNet ($p = 0.041$). Although GoogleNet achieved the highest classification accuracy on individual tissue samples, the YOLO models provided superior real-time detection

Table 9. Comparative Analysis of All Models

Model	Accuracy	mAP@0.5	Time	Params	GFLOPs
GoogleNet	97.2%	-	45ms	6.8M	1.5
YOLOv8n	-	94.7%	28ms	3.2M	8.7
YOLOv10n	-	96.2%	22ms	2.8M	7.2

performance.YOLOv8 uses Non-Maximum Suppression (NMS) to remove overlapping "duplicate" bounding boxes. NMS is often a computational bottleneck because it typically runs on the CPU or requires heavy GPU post-processing, which adds jitter to the video feed.

In particular, YOLOv10's uses "consistent dual assignments" during training. This allows the model to learn to predict only one box per object during inference, completely removing the need for NMS, which reduces latency while maintaining high accuracy, making it well suited for time critical surgical applications. These architectural improvements resulted in a 21% reduction in inference time and a 1.5% increase in mAP@0.5 compared with YOLOv8.

The superiority of YOLOv10 in this surgical context is primarily attributed to its NMS-free inference. In traditional surgical vision tasks, NMS creates a variable delay depending on the number of objects detected, which can cause 'lag' in the surgeon's display,eliminating this post-processing step, YOLOv10 provides a deterministic and consistent 22ms latency. This consistency can also be vital for compensating for the lack of haptic feedback; a surgeon can rely on a perfectly synchronized visual cue to adjust their grip strength on delicate fat vs. dense muscle tissue, thereby reducing the risk of unintended tissue overstress.

In preparing this manuscript, AI-based tools were employed to assist in text refinement, including improving clarity, grammar, and overall readability. These tools were utilized strictly as linguistic aids; all scientific content, analyses, and interpretations presented in this work were conceived and validated by the authors. The use of AI ensured consistent language quality and helped convey complex ideas more effectively, facilitating comprehension for a broader audience.Ethical considerations were observed throughout, with full transparency regarding the role of AI assistance, as acknowledged in the manuscript.

4 Conclusion

This paper will prove that the deep learning models can facilitate the achievement of highly accurate and reliable tissue detection in the simulated surgical setting. Of the assessed models, GoogleNet model had a classification accuracy of 97.2%, and the object detection models YOLOv8 and YOLOv10 have good real-time performance with mAP scores of 94.7% and 96.2%, at 0.5 (respectively). It is interesting to note that the NMS-free YOLOv10 recorded very low inference times of only 22 ms yet records high accuracies, which also suggests that it will be suitable in time-sensitive computational surgical assistance systems. These findings should emphasize the computational viability of AI-assisted tissue detection in a controlled experimental context.

The suggested system will offer a potential starting point of an intelligent surgical assist system in the future since it would allow real-time detection of tissues and allow regulating the amount of force applied on them according to tissue-specific thresholds.The proposed system would help to reduce the reliance on haptic sensations and prevent extensive tissue damage caused by the surgeon in the delicate context of the interventions. The system seeks to fill the gap between the computational vision method, and the realistic use of surgical methods using intelligent force control by incorporating real-time computer vision and intelligent force control.

The weakness of the given study is the relatively low number of data used in comparison to large scale computer vision benchmarks. Even though the augmentation methods were used to enhance the diversity of the data, it is necessary that in future research more diverse datasets that are subjected to different lighting conditions, camera angles, and surgical conditions should be collected. Also, the current structure is aimed at the static tissue classification and detection, yet

the real conditions of surgery are the moving tissue deformation, bleeding, and blocking with surgical devices. The need to tackle these complexities will add more complicated models and comprehensive clinic validation.

The dataset that was used in this study was gathered under controlled conditions and consists of mixed biological sources, which might not be able to represent the variability that exists in the real clinical surgical setting. Moreover, cross-patient and cross-device validation, as well as real-time assessment were not conducted and foundation was a simulated webcam, as opposed to an endoscopic or robotic surgical imaging system.

So the results are to be taken as evidence-of-concept computational framework as opposed to a surgical system that has been shown to be clinically valid. Future studies will involve the addition of more types of tissue and imaging conditions to the data, the inclusion of endoscopic and robotic surgical imaging data, and the assessment of cross-subject generalization. The next step in development can also be aimed at the assimilation with the force-sensing and haptic-feedback systems, and finally, the structured clinical validation research can be conducted to determine the translational feasibility of the suggested method in the real surgical processes.

Author Contributions

Syeda Fakhra Jalal & Syed Muhammad Mudabbir: Conceptualization, Methodology, Software, Dataset creation and curation, Writing-Original draft preparation.**Irfan Ahmed Usmani**: Visualization, Investigation, Validation. **Muhammad Zeeshan ul haque**: Supervision, Writing-Reviewing and Editing.

Compliance with Ethical Standards

It is declared that all authors do not have any conflict of interest. It is also declared that this article does not contain any studies with human participants or animals performed by any of the authors without proper ethical approval. Furthermore, informed consent was obtained from all individual participants included in the study for skin image collection.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The dataset used in this study was created by the authors and is not publicly available, as the research is ongoing and the data remain proprietary. However, the dataset may be shared upon reasonable request. Interested researchers can contact the corresponding author via email to request access, subject to approval.

Acknowledgment

We thank all participants who provided informed consent for skin tissue data collection.The authors also acknowledge the use of AI for enhancing the clarity and correctness of the text through grammar refinement and rephrasing. The authors declare that the artificial intelligence (AI) tool Chatgpt was used only for language editing, formatting, or technical refinement. No AI tool was used for the generation of research data, analysis, results, interpretations, or cited scholarly content.All AI-assisted content was reviewed and validated by the authors, who take full responsibility for the final manuscript.

References

[1] A. F. Khan, M. K. MacDonald, C. Streutker, C. Rowsell, J. Drake, and T. Grantcharov, "Tissue stress from laparoscopic grasper use and bowel injury in humans: Establishing intraoperative force boundaries," *BMJ Surgery, Interventions, & Health Technologies*, vol. 3, no. 1, pp. 265–289, 2021.

[2] M. Gómez Ruiz, M. Lainez Escribano, C. Cagigas Fernandez, L. Cristobal Poch, and S. Santarrufina Martinez, "Robotic surgery for colorectal cancer," *Annals of Gastroenterological Surgery*, vol. 4, no. 6, pp. 646–651, Nov 2020.

[3] W. Wang, J. Wang, Y. Luo, X. Wang, and H. Song, "A survey on force sensing techniques in robot-assisted minimally invasive surgery," *IEEE Transactions on Haptics*, vol. 16, no. 4, pp. 702–718, 2023.

[4] I. Boul-Atarass, M. R. Manzanares, A. Padillo-Eguía, J. Racero-Moreno, I. Eguía-Salinas, S. Pereira-Arenas, and R. M. Jiménez-Rodríguez, "Role of haptic feedback technologies and novel engineering developments for surgical training and robot-assisted surgery," *Frontiers in Robotics and AI*, vol. 12, 2025.

[5] G. H. Ballantyne, "Robotic surgery, telerobotic surgery, telepresence, and telementoring," *Surgical Endoscopy*, vol. 16, no. 10, pp. 1389–1402, 2002.

[6] T. B. Mahfoz, "Single-port versus multiport robotic surgery in head and neck procedures: a systematic review and meta-analysis of surgical parameters," *Annals of Surgical Treatment and Research*, vol. 109, no. 5, p. 293, 2025.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[10] M. Sajid, W. Yaseen, and A. U. Khan, "Brain tumor segmentation using deep learning," *VFAST Transactions on Software Engineering*, vol. 11, no. 2, pp. 113–123, 2023.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[12] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International Conference on Artificial Neural Networks*, 2018, pp. 270–279.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[14] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82 031–82 057, 2021.

[15] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[16] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *ICASSP 2020*, 2020, pp. 1055–1059.

[17] Z. Zhou, X. Wang, L. Zhang, and Y. Li, "Deep convolutional neural networks in medical image analysis: A review," *Information*, vol. 16, no. 3, p. 195, 2025.

[18] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[20] G. Jocher, A. Chaurasia, and J. Qiu, "Yolov8: Open-source neural network for object detection, segmentation, and classification," Zenodo, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[21] D. N. Kamtam, J. B. Shrager, S. D. Malla, N. Lin, J. J. Cardona, J. J. Kim, and C. Hu, "Deep learning approaches to surgical video segmentation and object detection: A scoping review," *Computers in Biology and Medicine*, vol. 194, p. 110482, Aug 2025.

[22] A. Q. Aloraibi, A. N. Hasoon, and M. S. Yassen, "Yolov10: Toward a promising improvement of gun detection based on a proposed image enhancement technique," in *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)*. IEEE, 2025, pp. 1–8.

[23] Z. Liu, K. Chen, S. Wang, Y. Xiao, and G. Zhang, "Deep learning in surgical process modeling: A systematic review of workflow recognition," *Journal Title*, 2025, discusses primary focus on workflow and phase recognition with DL in surgery.

[24] G. Boesch. (2024, May) Googlenet explained: The inception model that won imagenet. Accessed: Aug. 30, 2025.