

# Robot-Enabled Machine Learning-Based Diagnosis of Gastric Cancer Polyps Using Partial Surface Tactile Imaging

Siddhartha Kapuria<sup>\*1</sup>, Jeff Bonyun<sup>\*1</sup>, Yash Kulkarni<sup>1</sup>, Naruhiko Ikoma<sup>2</sup>, Sandeep Chinchali<sup>1</sup>,  
and Farshid Alambeigi<sup>1</sup>

**Abstract**—In this paper, to collectively address the existing limitations on endoscopic diagnosis of Advanced Gastric Cancer (AGC) Tumors, for the first time, we propose (i) utilization and evaluation of our recently developed Vision-based Tactile Sensor (VTS), and (ii) a complementary Machine Learning (ML) algorithm for classifying tumors using their textural features. Leveraging a seven DoF robotic manipulator and unique custom-designed and additively-manufactured realistic AGC tumor phantoms, we demonstrated the advantages of automated data collection using the VTS addressing the problem of data scarcity and biases encountered in traditional ML-based approaches. Our synthetic-data-trained ML model was successfully evaluated and compared with traditional ML models utilizing various statistical metrics even under mixed morphological characteristics and partial sensor contact.

## I. INTRODUCTION

Gastric cancer (GC) is the fifth most commonly diagnosed cancer worldwide and the fourth leading cause of cancer-related mortality [1]. A major contributor to this challenge is the fact that a substantial portion — up to 62% — of GC cases are detected at advanced stages, contributing to poorer overall survival rates compared to cases identified at early stages [2]. Upper endoscopy is the primary method for the initial detection of GC lesions as it allows for an inside view of the gastric tract lining where tumors originate. At the advanced GC (AGC) stages, tumors have infiltrated the muscularis propria [3] and can be identified and classified through their morphological characteristics (i.e., their geometry and texture) visible through the images provided by an endoscope. Borrmann classification [3] is a common approach used by clinicians to morphologically classify GC polyps into four types of polypoid (Type 1), fungating (Type 2), ulcerated (Type 3), and infiltrating or Flat (Type 4) (see Fig. 2). Nevertheless, inter-class variance of each type of polyps and solely relying on morphology of the GC polyps in Borrmann classification has resulted in a high-degree of disagreement and inconsistency in decision-making among clinicians [4]. Therefore, long-term specific training and experience is needed to detect GC properly

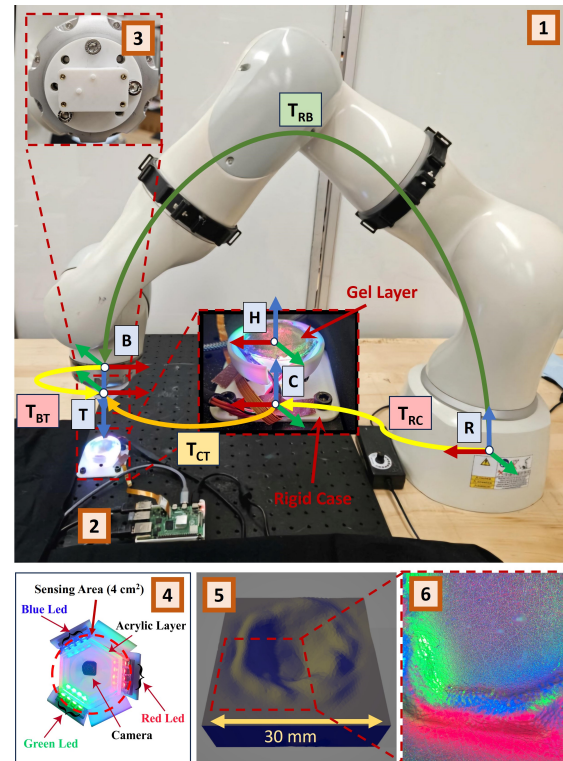


Fig. 1: Experimental Setup including: (1) KUKA LBR Med 14 R820 (KUKA AG); (2) Raspberry Pi 4 Model B; (3) 3D printed mounting plate for the tumor phantoms; (4) Top view of HySenSe sensor showing all components; (5) Example CAD model of synthetic AGC polyp phantom; (6) Example partial textural image output of AGC tumor phantom. Figure also shows the defined reference frames  $R$ : Robot/World,  $B$ : Robot Flange,  $C$ : Camera,  $T$ : Target, and  $H$ : HySenSe base. using endoscopic images and Borrmann classification [5]. Furthermore, similar to many vision-based endoscopic diagnoses (e.g., colonoscopy and laparoscopy), the limited resolution of endoscopic video cameras, visual occlusions, lack of sufficient steerability of the endoscopic devices, and lighting changes within the body make reliable diagnosis of GC polyps even more challenging [6].

To address the above-mentioned limitations, Artificial Intelligence (AI) methods utilizing Machine Learning (ML) have been employed in different modalities, such as histopathological images or endoscopic videos, to detect and classify GC tumors. For example, Li et al. [7] used a custom Deep Learning (DL) based framework for automatic cancer identification from histopathological images. Using the same modality, Huang et al. [8] developed an in-house DL ap-

<sup>\*</sup>Authors had equal contribution to this work

This research was supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA280747

<sup>1</sup>Siddhartha Kapuria, Jeff Bonyun, Yash Kulkarni, Sandeep Chinchali, and Farshid Alambeigi are with Texas Robotics, University of Texas at Austin, Austin, TX, USA. email: {skapuria, jbonyun, kulkarni.yash08, sandeepc}@utexas.edu, and farshid.alambeigi@austin.utexas.edu

<sup>2</sup>Naruhiko Ikoma is with the Department of Surgical Oncology, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, 77030. email: nikoma@mdanderson.org

proach - GastroMIL - for differentiating between cancerous and healthy tissue. For endoscopic videos, Hirasawa et al. [9] used a CNN-based Single Shot MultiBox Detector to successfully detect cancerous lesions. Taking a step further to address the limitations of traditional endoscopy, Xia et al. [10] used a magnetically controlled capsule endoscope coupled with an ML model for *in-vivo* classification of GC.

A common limitation of such ML approaches is the limited availability and access to the large, balanced datasets [11]. This has also been recognized by The American Medical Association, which passed policy recommendations in 2018 for identifying and mitigating bias in data during the testing or deployment of AI/ML-based software to prevent introducing or exacerbating healthcare disparities [12]. DL approaches especially require large amounts of data to be able to generalize over unknown inputs. This is particularly important in the medical domain, where there are several hurdles to obtaining patient data such as time dependence, availability of relevant clinical cases, and privacy concerns [13]. Limited data can lead to (1) spectrum bias and (2) overfitting [13]. Furthermore, in the case of AGC tumors, there is a very high degree of inter-class variance in the morphological characteristics, which can vary considerably amongst patients, hindering the preparation of a rich, well-balanced dataset [4]. This issue of limited data can be partially mitigated through transfer learning, which has now become a staple of modern DL frameworks [14]. However, the limitation of class imbalance persists through this technique. This is important to consider since even if datasets mirror the real-world distribution, rarer cases may not have adequate representation. In such scenarios, where rare cancer cases are both inherently infrequent and underrepresented in the data, ML models may struggle to learn the distinct features associated with these cases.

To address the aforementioned limitations of existing endoscopic procedures, recently we have developed a novel Vision-based Tactile Sensor (VTS) (shown in Fig. 1) and introduced the surface tactile imaging modality for early diagnosis of colorectal cancer (CRC) polyps [15], [16], [17], [18], [19], [20]. As opposed to normal endoscopic images, VTS provides high-resolution  $\sim 50 \mu m$  textural images of the polyps that can potentially improve their classification. However, the size of CRC polyps are smaller than AGC tumors (i.e.,  $\sim 1 \text{ cm} \times 1 \text{ cm}$  versus  $\sim 4 \text{ cm} \times 4 \text{ cm}$ ) [21], [3]. As shown in Fig. 1, since the sensing area of the VTS is limited, only partial textural images of AGC tumors can be captured using VTS, making their data collection and classification very challenging.

To collectively address these limitations and towards developing an ML-based diagnostic assistance for AGC using our VTS, in this work and as our main *contributions*, we propose (i) to use and evaluate, for the first time, the potentials of utilizing the VTS in classifying AGC tumors using their textural features; (ii) a complementary ML-based diagnostic tool that can leverage this new modality to sensitively classify AGC lesions; and (iii) a robot-assisted data collection procedure to ensure the ML model is trained

on a large and balanced dataset. We train the ML models on partial textural data semi-autonomously collected from our unique 3D-printed AGC tumor phantoms. We also use appropriate statistical metrics during evaluation to show that the proposed ML models can reliably and sensitively classify the AGC lesions even under mixed morphological conditions and partial tumor coverage.

## II. MATERIALS AND METHODS

### A. Vision Based Tactile Sensor (VTS)

In this study, we employed our recently developed VTS called HySenSe, as outlined in [22], to acquire high-fidelity textural images of AGC tumor phantoms. As depicted in Fig. 1, the HySenSe sensor comprises: (I) a flexible silicone membrane interacting directly with polyp phantoms, (II) an optical module (Arducam 1/4 inch 5 MP camera) capturing minute deformations in the gel layer during interactions with a polyp phantom, (III) a transparent acrylic plate offering support to the gel layer, (IV) an array of Red, Green, and Blue LEDs for internal illumination aiding depth perception, and (V) a sturdy frame supporting the entire structure. Operating on the principle that the deformations resulting from the interaction between the deformable membrane and the surface of AGC tumors can be visually captured, the HySenSe sensor provides high-fidelity textural images, demonstrating proficiency across various tumor characteristics such as surface texture, hardness, type, and size [17]. This capability is maintained even at extremely low interaction forces, as detailed in [22]. Additionally, due to the arrangement of the LEDs within the sensor, different deformations have different lighting. This means that if the interaction force is fixed, the textural images implicitly encode the stiffness characteristics of the surface in contact as well. Due to these advantages, in our previous works [15], [16], [17], [18], [19], [20], we have successfully used this sensor to differentiate between classes of CRC polyps, which are also distinguished by their morphological characteristics, namely the surface texture presented and the polyp stiffness. However, at this point, it must be noted that CRC polyps and AGC tumors differ greatly in size (i.e.,  $\sim 1 \text{ cm} \times 1 \text{ cm}$  versus  $\sim 4 \text{ cm} \times 4 \text{ cm}$ ). Since building a larger HySenSe sensor is impractical in a clinical setting, we are limited by the area we cover using the sensor. In its current form, the area coverage is around  $4 \text{ cm}^2$  (see Fig. 1.4). Therefore it is impossible to entirely cover an average AGC tumor. To address this issue and cover sufficient amount of data, we opted to employ a semi-autonomous robotic system, to sequentially cover the whole AGC surface and collect data discussed in detail in Section II-C.

### B. Realistic Tumor Phantoms

Towards addressing the limited availability of large, balanced clinical datasets in medical imaging and given this new imaging modality, we opted to meticulously design and manufacture realistic approximations of AGC tumors (see Fig. 2). To design the phantoms, and without loss of generality evaluate the performance of the utilized VTS,

we followed 4 different types of Borrmann’s classification system. The four types specifically are Type I: fungating type; Type II: carcinomatous ulcer without infiltration of the surrounding mucosa; Type III: carcinomatous ulcer with infiltration of the surrounding mucosa; Type IV: a diffuse infiltrating carcinoma (linitis plastica) [3]. It is known that the stiffness of the affected area is more than that of the surrounding regions, which makes it easier for clinicians to differentiate between tumorous sections and healthy tissue [3]. However, this classification has a considerable overlap (especially between Type II and Type III) due to the mixed morphological characteristics of these lesions, making a manual diagnosis through observation difficult [4].

Fig. 2 illustrates a few representative fabricated AGC lesion phantoms and their dimensions. To avoid the issue of data imbalance, in contrast with real patient datasets, we opted to have each tumor class equally represented in the dataset by designing 11 variations of each class (total 44 polyps). As shown in Fig. 2, based on the realistic AGC polyps, the designs first were conceptualized in Blender Software (The Blender Foundation) to make use of the free-form sculpting tool, then imported into Solidworks (Dassault Systèmes) in order to demarcate the different regions with varying stiffness (tumor versus healthy tissue). The high-resolution, realistic lesion phantoms were manufactured using a Digital Anatomy Printer (J750, Stratasys, Ltd) and materials with diverse properties: (M1) Tissue Matrix/Agilus DM 400, (M2) a mixture of Tissue Matrix and Agilus 30 Clear, and (M3) Vero PureWhite. Hardness measurements were obtained using a Shore 00 scale durometer (Model 1600 Dial Shore 00, Rex Gauge Company). M1 has Shore hardness A 1-2, M2 has A 30-40, and M3 has D 83-86. These differing material properties allowed us to make the tumor sections (using M2) stiffer than the surrounding healthy tissue (using M1). M3 was used to print the supporting rigid backplate to be mounted onto the robot flange. Each tumor was printed across a working area of 3 cm × 3 cm, which represents the lower end of AGC tumor sizes.

### C. Experimental Setup and Robotic Data Collection

In our previous works utilizing HySenSe for CRC polyp classification (see [15], [16], [17]), the high-fidelity textural images were captured manually using a setup including a force gauge mounted on a linear stage. This manual and tedious procedure limited our data collection capabilities, which was not efficient for this study. Furthermore, as discussed in Section II-A, AGC tumors are much larger than CRC polyps, making it impossible for the VTS to completely capture the whole textural area of the tumor phantoms. To overcome this limitation, in this work, we used a robotic manipulator to automate the image collection procedure. Using the robotic arm shown in Fig. 1, many different variations of partial contact from different angles were captured to form our dataset, which allowed the trained ML model to be more generalized.

The experimental setup for data collection is illustrated in Fig 1 and consists of the following: **(1) Robot Manipulator:**

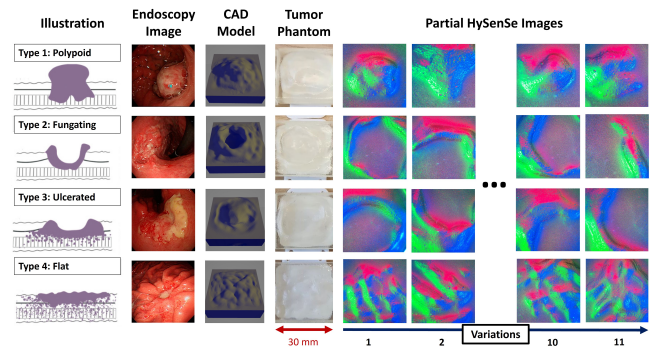


Fig. 2: The first column shows the schematics [23] of 4 types of AGC tumors under the Borrmann classification. The second column indicates the corresponding real clinical endoscopic images [24]. The third and fourth columns show a sample designed CAD model and 3D-printed tumor for each class. Other columns show corresponding VTS outputs.

We used a KUKA LBR Med 14 R820 (KUKA AG) which has seven degrees of freedom (DoF), a large operating envelope, and integrated force sensing. We used ROS as a bridge, with the *iiwa\_stack* project presented in [25], to provide high-level control of the onboard Java environment. **(2) Workspace:** The arm was rigidly attached to a work table. An optical table allows consistent positioning of items in the robot’s coordinate frame. **(3) HySenSe:** The sensor was manufactured by our lab, using the methodology provided in [22], and attached to the optical table. Images are captured by the camera, a 5MP Arducam with 1/4” sensor (model OV5647, Arducam). **(4) Raspberry Pi:** The camera is controlled by a Raspberry Pi 4B over the Arducam’s camera ribbon cable. The Raspberry Pi ran Python software that continuously listened for an external ROS message trigger, which caused it to capture a  $2592 \times 1944$  image and publish it to ROS. **(5) Sample Mount:** An adapting mount was 3D printed in PLA. One side attached to the robot flange. The other side offered two locating pins and four M2 screw holes, which ensure repeatable position and orientation of samples **(6) Polyp Phantom Samples:** The polyp samples were constructed as described above. Each was attached, in turn, to the sample mount using M2 screws. **(7) Command and Control:** An Ubuntu 20.04 system was used to control the ROS components. This computer ran the `roscore` master and the high-level Python command scripts.

To collect textural images with HySenSe, the robotic manipulator was commanded via ROS to position the phantom of interest into random positions with different angles of contact, while maintaining the interaction force under a threshold of 3N using the arm’s internal force sensor. As opposed to our previous manual procedures performed in [17], the only manual step involved was installing each target AGC phantom onto the sample mount on the robot end effector, and the remainder of the process was automated in software, reducing the time and workload required.

### D. Calibration and Registration Procedures

While the apparatus was assembled carefully, there was no way to ensure precise positioning to the sub-millimeter

TABLE I: Hyperparameter Space

Hyperparameter	Range/Options
Learning Rate	(min = 0.001, max = 0.1)
Learning Rate Scheduler	Step Reduce on Plateau 1cycle Cosine Annealing
Optimizer	Stochastic Gradient Descent Adam Adabound
Weight Decay	(min = 0, max = 0.1)

level. This level of accuracy was necessary to successfully automate the action of pressing down an AGC tumor phantom onto the HySenSe gel with a measured and limited force of interaction. Thus, a registration step was performed to find the unknown transformation matrices amongst the different components. We divided our test setup into five frames of reference, visualized in Fig. 1, with the objective of determining the transformation matrices between these frames. Some of these transformations were known. The transformation from robot base ( $R$ ) to robot flange ( $B$ ),  $T_{RB}$ , is provided through the KUKA software. The transformation from the camera frame ( $C$ ) to a known target ( $T$ ),  $T_{CT}$ , was calculated by using a standard checkerboard calibration method. The intrinsic matrix follows the pinhole camera model and the 10-parameter distortion model follows [26] and [27]. The remaining transformations (robot base to the camera frame  $T_{RC}$ , robot flange to the tumor phantom  $T_{BT}$ , and camera frame to the plastic plate that holds the HySenSe gel ( $T_{CH}$ ) were unknown.

Using the checkerboard calibration images as data, we performed  $AX = XB$  calibration using the separable method [28]. This provided us with estimates of  $T_{RC}$  and  $T_{BT}$ . The resulting registration was still noisy, and fine-tuning was done with a simple grid pattern, adjusting the robot position until it was centered and square in the camera image. The final part of the registration was finding  $T_{CH}$  by bringing a flat sample plate down until it was within 1mm of the HySenSe backing plate, with the gel removed, and checking that it was parallel with a feeler gauge.

### E. Dataset and Pre-processing

This semi-automated data collection setup enabled us to collect 50 variations of orientation and contact of the AGC tumor phantoms with the HySenSe in 44 experiments (one for each polyp), leading to a total of 2200 images in the textural image dataset, with 550 unique images in each class. This dataset was then split into training and test sets while ensuring that the split was performed at a tumor level and not the image level. In other words, all textural images belonging to one tumor were kept within the same split. This was done in order to ensure the model was not being evaluated on partially seen data. This resulted in 1600 images from 32 unique tumors in the training set and 600 images from 12 unique tumors in the test set. The output from HySenSe, with original dimensions of  $2592 \times 1944$  pixels,

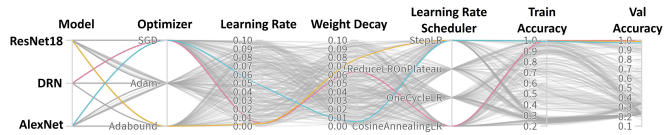


Fig. 3: Hyperparameter search for the three models. Each line encodes a particular configuration.

was resized to a uniform size of  $224 \times 224$  pixels, ensuring consistent input dimensions for all models. To enhance the algorithm’s generalization capability, a range of geometric transformations, including random cropping, vertical and horizontal flips, and random rotations spanning from  $-45^\circ$  to  $45^\circ$ , were incorporated as augmentations. Further to enhance generalizability of data augmentation, we also included Gaussian blur and Gaussian noise with strengths  $\sigma$  ranging from 1 to 256 for blur, and  $\sigma$  from 1 to 50 for noise [19]. The maximums for these values were chosen to represent a level well beyond realistic worst-case scenarios [19]. Each augmentation had an independent occurrence probability of 0.5, contributing to the algorithm’s overall robustness. The images in the holdout test dataset, comprising 600 samples, were only resized to fit as inputs to all models.

### F. Deep Learning Model

In this study, we opted to use the Dilated ResNet architecture for AGC tumor classification. The architecture’s effectiveness stems from its capacity to mitigate issues like exploding gradients [29]. An additional advantage of ResNets lies in their incorporation of skip connections, which helps alleviate the degradation problem associated with the worsening performance of models as complexity increases [29]. Notably, unlike conventional Residual Network architectures, the Dilated ResNet incorporates dilated kernels. Dilations play a crucial role in maintaining feature maps’ spatial resolution during convolutions while expanding the network’s receptive field to capture more intricate details [30]. More details of the architecture of Dilated ResNet can be found in [16]. To verify our model’s performance relative to other commonly used architectures for image classification tasks, we compared the performance metrics described in Sec. II-G for our architecture with ResNet18 [31], and AlexNet [32].

### G. Evaluation

To limit the variability in performance due to the variance in hyperparameters, we used a random search across our hyperparameter space to find a good candidate for each model for comparison. The hyperparameter space in this study included the initial learning rate (LR), the LR scheduler, the optimizer, and weight decay. We considered four possible LR schedulers, namely Step, Reduce-on-Plateau, OneCycle [33], and Cosine Annealing [34]. Three methods for the optimizer were considered, which were Stochastic Gradient Descent (SGD), Adam [35], and Adabound [36]. The full hyperparameter space is provided in Table I. Furthermore, early stopping with a patience of 10 epochs was utilized to cut down on the training time in case the validation loss reached a point of no improvement well

TABLE II: Best Performing Configuration

Hyperparameter	Dilated ResNet	ResNet18	AlexNet
LR	0.06308	0.07825	0.00527
LR Scheduler	Cosine Annealing	Step	Step
Optimizer	SGD	Adabound	SGD
Weight Decay	0.00338	0.00324	0.04663

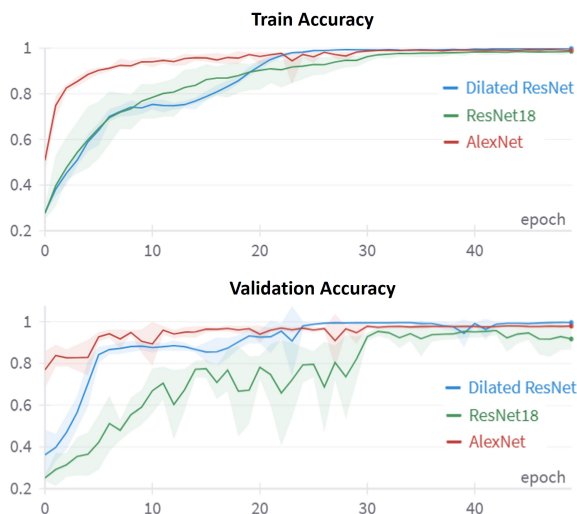


Fig. 4: Stratified 5-fold cross-validation results for the three models. Average accuracy curves are reported, with the shaded region depicting the standard deviation across folds.

before the designated maximum of 50 epochs. 100 random configurations for each model architecture were considered in the random search. The best configuration was chosen such that the trained model minimized the validation set loss while keeping minimal overfit (which is the difference between the training and validation accuracy). Following this step, to ensure generalizable performance, each model was configured using the chosen hyperparameters and trained using stratified 5-fold cross-validation over the training split of data such that the class distribution in each fold remained intact. After verification of generalized performance, the entire training set was used to build the final models to be used for comparative evaluation.

To evaluate model performance, we used the standard evaluation metrics of accuracy ( $A$ ), precision ( $P$ ), Recall ( $Re$ ), F1-score ( $F1$ ), and Area under the ROC Curve ( $AUC$ ). Since this is a multi-class classification problem we reported the macro-averaged values (that is, the metrics are calculated for each class and then averaged). These metrics are crucial in assessing the model’s performance in AGC tumor classification, offering a more comprehensive evaluation than accuracy alone. Accuracy measures overall correctness, but in safety-critical applications like cancer diagnosis, emphasizing true positives and true negatives is vital. Apart from these metrics, we also plotted the confusion matrices for each model to complete the analysis. The entire suite of training and tests was tracked using Weights and Biases [37].

### III. RESULTS AND DISCUSSION

1) *Hyperparameter Search:* Table II lists the best-performing combination of hyperparameters for each model.

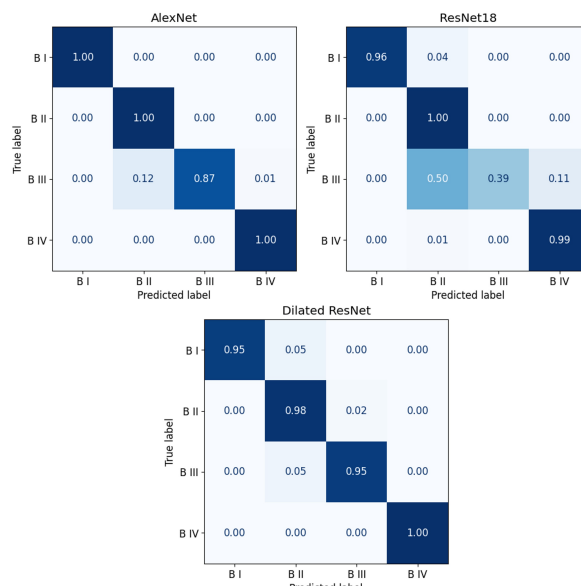


Fig. 5: Normalized confusion matrices for all three models configured with their corresponding best hyperparameters.

For the dilated ResNet, the best LR scheduler was Cosine Annealing while SGD was the best optimizer. The ResNet18 model performed best using the Step LR scheduler and Adabound optimizer. Finally, AlexNet also performed best using the Step LR scheduler coupled with an SGD optimizer. Notably, both the dilated ResNet and ResNet 18 models required an initial learning rate of around 0.06-0.07 and a weight decay of around 0.03. On the other hand, AlexNet performed better with a relatively lower learning rate of 0.005, while having a higher weight decay of 0.047.

2) *Cross Validation:* The average plots of training and validation accuracy for all three candidate models training using stratified 5-fold cross-validation with the hyperparameter configuration selected as described in Sec. III-1 are provided in Fig. 4. Here we observe that ResNet18 has the largest variation in performance across the different folds, making it the most susceptible to changes in training data. The dilated ResNet model has the least variation and therefore we can conclude that it is the most general. Furthermore, the average accuracy across all folds is the lowest for ResNet18 and most for the dilated ResNet. Here again, the performance of AlexNet lies in between the other two models. While AlexNet reaches saturation much earlier than both the other models, its variation in performance across the different folds, and the lower peak accuracy as compared to the dilated ResNet18 hinder it from being the best overall pick.

3) *Model comparisons:* The results on the test set for the three models are summarized in Table III and the corresponding confusion matrices are provided in Fig. 5. As can be seen through our evaluations, the dilated ResNet model proposed in our previous work for colorectal cancer classification is able to outperform both ResNet18 and AlexNet for AGC tumor classification. Observing the drop in testing accuracy for ResNet18, we can conclude that it was overfitting during training. While AlexNet achieved comparable performance in

TABLE III: Performance Results

Metric	Dilated ResNet	ResNet18	AlexNet
Accuracy	<b>0.9667</b>	0.8333	0.9600
Precision	<b>0.9656</b>	0.8685	0.9631
Recall	<b>0.9692</b>	0.8175	0.9550
F1 Score	<b>0.9673</b>	0.8422	0.9591
AUC	<b>0.9990</b>	0.9879	0.9983

terms of the performance metrics, its susceptibility to training data as discussed in Sec. III-2 made it unsuitable for our application. Furthermore, the number of trainable parameters in dilated ResNet is only 2.8M compared to 11.2M and 57M trainable parameters in ResNet and AlexNet respectively. Thus, the dilated ResNet is able to outperform the other two models despite being comparatively lightweight and less complex. Looking at the confusion matrices, we observe that all three models have some difficulties in differentiating between Types II and III. This is expected since both tumor types present similarly on the exterior and there can be an overlap in diagnosis even amongst clinicians [3].

#### IV. CONCLUSIONS

To collectively address the existing limitations on endoscopic diagnosis of AGC tumors, for the first time, we used and evaluated our recently developed VTS for classifying AGC tumors using their textural features. Leveraging a 7 DoF robotic manipulator and unique custom-designed realistic AGC tumor phantoms, we demonstrated the advantages of automated synthetic data collection using the VTS addressing the problem of data scarcity and biases encountered in traditional ML-based approaches. We also proposed and evaluated a complementary ML-based diagnostic tool that can leverage this new modality to sensitively classify AGC lesions. Our synthetic-data-trained ML model was successfully evaluated utilizing appropriate statistical metrics even under mixed morphological characteristics and partial sensor contact. In the future, we plan to test this ML model textural images collected on real patients and perform a Sim2Real study.

#### REFERENCES

- [1] H. Sung, *et al.*, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, pp. 209–249, 2021.
- [2] National Cancer Institute, “Seer cancer stat facts: Stomach cancer,” 2021.
- [3] R. M. Gore, *et al.*, “Stomach malignant tumors,” 2013.
- [4] M. E. Lockhart *et al.*, “Epidemiology of gastric cancer,” 2009.
- [5] Q. Zhang, *et al.*, “Training in early gastric cancer diagnosis improves the detection rate of early gastric cancer,” *Medicine*, vol. 94, 2015.
- [6] L. Ginzburg, *et al.*, “Complications of endoscopy,” *Gastrointestinal endoscopy clinics of North America*, vol. 17 2, pp. 405–32, 2007.
- [7] Y. Li, *et al.*, “Deep learning based gastric cancer identification,” *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 182–185, 2018.
- [8] B. Huang, *et al.*, “Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: A retrospective multicentre study,” *EBioMedicine*, vol. 73, 2021.
- [9] T. Hirasawa, *et al.*, “Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images,” *Gastric Cancer*, vol. 21, pp. 653–660, 2018.
- [10] J. Xia, *et al.*, “Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy,” *Gastrointestinal endoscopy*, 2020.
- [11] S. Fotouhi, *et al.*, “A comprehensive data level analysis for cancer diagnosis on imbalanced data,” *Journal of biomedical informatics*, vol. 90, p. 103089, 2019.
- [12] American Medical Association, “Ama passes first policy recommendations on augmented intelligence,” 2023.
- [13] C. D. Christou *et al.*, “Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology,” *World Journal of Gastroenterology*, vol. 27, pp. 6191–6223, 2021.
- [14] P. Jin, *et al.*, “Artificial intelligence in gastric cancer: a systematic review,” *Journal of Cancer Research and Clinical Oncology*, pp. 1–12, 2020.
- [15] O. C. Kara, *et al.*, “A smart handheld edge device for on-site diagnosis and classification of texture and stiffness of excised colorectal cancer polyps,” *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4662–4668, 2023.
- [16] N. Venkatayogi, *et al.*, “On the potentials of surface tactile imaging and dilated residual networks for early detection of colorectal cancer polyps,” *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4655–4661, 2023.
- [17] O. C. Kara, *et al.*, “A reliable and sensitive framework for simultaneous type and stage detection of colorectal cancer polyps,” *Annals of Biomedical Engineering*, 2023.
- [18] S. Kapuria, *et al.*, “A novel dual layer cascade reliability framework for an informed and intuitive clinician-ai interaction in diagnosis of colorectal cancer polyps,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, pp. 2326–2337, 2024.
- [19] —, “Towards reliable colorectal cancer polyps classification via vision based tactile sensing and confidence-calibrated neural networks,” *2023 International Symposium on Medical Robotics (ISMR)*, pp. 1–7, 2023.
- [20] —, “Enhancing colorectal cancer diagnosis through generative models and vision-based tactile sensing: a sim2real study,” in *SPIE Medical Imaging 2024*, 2024.
- [21] S. ei Kudo, *et al.*, “Colorectal tumours and pit pattern,” *Journal of Clinical Pathology*, vol. 47, pp. 880–885, 1994.
- [22] O. C. Kara, *et al.*, “Hysense: A hyper-sensitive and high-fidelity vision-based tactile sensor,” *ArXiv*, vol. abs/2211.04571, 2022.
- [23] L. Wang, *et al.*, “Comparing single oral contrast-enhanced ultrasonography and double contrast-enhanced ultrasonography in the preoperative borrmann classification of advanced gastric cancer,” *Oncotarget*, vol. 9, pp. 8716–8724, 2015.
- [24] K. Hosoda, *et al.*, “Re-emerging role of macroscopic appearance in treatment strategy for gastric cancer,” *Annals of Gastroenterological Surgery*, vol. 3, pp. 122–129, 2018.
- [25] C. Hennersperger, *et al.*, “Towards mri-based autonomous robotic us acquisitions: a first feasibility study,” *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 538–548, 2017.
- [26] D. Brown, “Decentering distortion of lenses,” *null*, 1966.
- [27] Z. Zhang *et al.*, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- [28] R. Horaud *et al.*, “Hand-eye calibration,” *The International Journal of Robotics Research*, 1995.
- [29] K. He, *et al.*, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [30] F. Yu, *et al.*, “Dilated residual networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 636–644, 2017.
- [31] K. He, *et al.*, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [32] A. Krizhevsky, *et al.*, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84–90, 2012.
- [33] L. N. Smith *et al.*, “Super-convergence: very fast training of neural networks using large learning rates,” in *Defense + Commercial Sensing*, 2018.
- [34] I. Loshchilov *et al.*, “Sgdr: Stochastic gradient descent with warm restarts,” *arXiv: Learning*, 2016.
- [35] D. P. Kingma *et al.*, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2014.
- [36] L. Luo, *et al.*, “Adaptive gradient methods with dynamic bound of learning rate,” *ArXiv*, vol. abs/1902.09843, 2019.
- [37] L. Biewald, “Experiment tracking with weights and biases,” 2020, software available from wandb.com.