GICaps: Leveraging Geometrico-Physical Computer Vision Techniques On Capsule Vision Endoscopy

Anonymous CVPR submission

Paper ID 13677

Abstract

Endoscopy has proved effective for the human body's gastrointestinal anatomic diagnosis and prognosis. Implementing Feature Selection and Feature Engineering on Internal Organs is a tough task that requires complex statistical and numerical analysis of feature vectors across chromatic contours and density considerations of the input images. In the current work, we amplified the train data set chromatic intensities by 65%, improving feature selection efficiency by 25%, and Computer Vision implementation effectiveness by 85%. We employed Computer Vision Modules like YOLO NAS, OrganAMNIST, OrganCMNIST, and TissueMNIST to produce an application that implements predictions on internal organs and tissues of the gastrointestinal tract up to the stomach for examining the proximity of ulcers and other diseases/comorbidities like diseases. Our work produced an overall accuracy of 91.87% compounded as an ensemble accuracy. It was successful in detecting 99.87% of gastrodiseases, producing a probable increase of 37% in the field of health computational sciences.

1. Introduction

Endoscopy images are coarse representations of human organs from the inside, but dealing with them is difficult due to poor illumination conditions. The most challenging aspect of this problem is that tumors/ulcers may often be confused with natural bodily formations/ tissues, leading to a need for a multi-class solution. Chromatic intensities of red and white colors impact the detection quality of the computer vision module even with optimal accuracy. Tears in muscular extensions further aggravate the framework's model training process, overfitting it more towards blood capillaries and injuries which defeats its very purpose of universality in application.

Though multiple libraries are available in the market for developing Computer Vision Models in identifying human-based anomalies through optical intelligence, they are still deprived of the minute intricacies of human muscular-tissue-organ-cardiovascular mixtures thus confusing them among various body parts visible by the camera embedded in the endoscope tube. The models can even confuse blood-rich and blood-poor human parts that vary across regions, races, sex, biological bodies, etc. Surprisingly, The models should be trained across animal bodies for more scalability and enhanced prediction range. From another angle, we can find repeated collisions with human parts as the tube is navigated deeper into the human gastrointestinal tube which significantly distorts the train images as they are captured. Diffractions and Interferences often affect the visual quality of images, repeatedly obscuring essential features behind the glares and reflections, thus hampering an 100% effective feature engineering and further decreasing the efficiency below 75%.

Existing works deal with preliminary assessments of features across input training images, evaluate their related optical relative ratios, and examine possible ways of predicting GI tumors & analyzing probabilities of disease proxemics shortly. Though most of them successfully approached near-optimal performance, there are still some bottlenecks concerning feature selection, engineering, and processing. In other words, data processing takes 65% of the entire data management pipeline, thus requiring necessary steps for building a system that handles these roadblocks. In Addition, issues arise with the way the data is handled along the pathway to the final training data version. During the incremental processing processes, there is ¿0.6 probability for noise additions, training distortions, important feature eliminations, etc.

In the current work, we focused on delivering an ensemble of 3 Computer Vision Modules that are trained extensively on varied illuminations, features, visual orientations, shadow shades, etc., of body organs captured by an embedded camera in a GI Endoscopic tube with a primary focus on 60% of classification, 20% of neural network computations, 5% of testing and 5% of universalization of the developed model. The majority of the process is focused on classification due to the grouping nature and ¿65% similarity patterns

among the human body parts captured.

2. Existing Systems

Though multiple endoscopic image repositories have been built upon endoscopic camera records, Capsule Cameras have recently been invented. Capsule Vision Cameras are specialized cameras that possess a spherical len-body that can produce 3-dimensional views of its captured images. The body parts observed in such lens appear to protrude through space and appear expanded than the original part.

Spherical Vs Conventional Endoscopy Images: Spherical Cameras produce magnified versions of the input images while also scaling their geometrical features. Capsule Vision Cameras also have varied scales of luminations[1-6] since they capture images at different directions and distances and pursue irregular paths during these motions. The light employed also has different luminous intensities, thus differing significantly from conventional endoscopic cameras which use standard lighting systems[7-10].

Computer Vision on Endoscopy Conventional Models like YOLO NAS, OpenCV, CNN, etc., are employed while performing predictions on endoscopic image repositories that need further scrutiny[11-14]. Though conventional systems are largely successful in arriving accurately at results and ultimately making the system intelligent, the current work is novel in terms of how it handles these visual predictions. We employ the use of vector and geometric projections as opposed to the pixellated versions used there. While we also employ pixellated versions in our blood-based feature extraction methodology, it significantly differs from the existing systems' purview of pixel analytics. Existing systems make use of crude analytics on the whole pixel arrangement of the images. The models learn these pixel arrangements but subtly ignore respective chromatic intensities[15-17].

Intelligence to Prediction: There is a simple transition from building intelligent systems to predicting diseases but the predictions revolve not around the disease proximities. They perform based on disease probabilities[18-22] i.e., they help us determine the probability of the given prediction highlighting a disease. Feature extraction is a single-phased process that highlights the disease features solely on chromatic features and ignores luminosity factors. The intelligence to prediction transition is fairly attributed to the numeric metrics possessed by the models across various stages of development. Computer vision intelligence also comprises intellectual information about the training visual characteristics and the disease probability is decided based on that probability only[23-25].

While the existing systems build on the standard pipeline of data processing & computer vision modules, we tested the novel ways of generating geometric and vector projections to the already learned features to learn how they fair across multiple directions. Since capsule vision endoscopy has already enabled 360-degree vision, it has become easy for us to model all the projections efficiently.

3. Methodology

4. Methodology

The process entails a multi-phased procedure with 80% focus on feature-based segregation and 20% on predictions-based analytics. The proposed methodology is novel in terms of how it deals with the data sets of internal perspectives of the GI Tract with a higher focus on bodily feature contour variations i.e., the features and optical characteristics of the GI Tract.

4.1. Data set

We have considered a 10-class data set categorically segregated into 10 diverse GI Diseases, Ailments, Ulcers, and Commorbidities. The data set is further distributed into training and testing channels with an optional validation channel where we employed the use of validation for the diseases that have greater similarities with natural body similarities. We have used a threshold of 85% similarity rate for considering the use of a validation data set while the size of validation is minuscule for the similarity score in between the range 55%-84.9%. The data set has 2,40,000 feature-specific training, testing, and validation data samples with annotations specifying how the data varies from other classifications/features.

4.1.1. Numerical Comptaibility

This is a relatively novel concept where the data elements are evaluated for their suitability for detailed geometric and numeric calculations. Capsule Vision Cameras embedded in an endoscopic tube can view the organs in diverse directions by employing 360 lenses. In our work, we have focussed on geometric projections and hence, we need multi-dimensional views of objects. This requires the image elements captured during forward and backward movements. Additionally, for generic applications, we need different luminosity levels. After a pilot study of diverse data sets, we ultimately have an all-around image repository containing 50000 images.

4.2. Feature Selection and Engineering

The first step of the process relies on how effectively the embedded camera on the endoscopic tube captured images for consideration by the model for feature selection. Certain trade-offs are made in this regard in the direction of capture quality, intensity, and quantity. For larger quantities of input images, the model ignores a subset of unclear, poor-quality, and distorted images, hence such applications move forward in the overall development process. If we consider

Blood-Based			Depth-Based			Density-Based			Luminosity-Based			
Color	Temp.	Key	Act.	Virt.	Key	Dnsty	Conc.	Key	H1	H2	H'	Key
White-Red	0-35	Squamous Epithelial Cells	5	0.3	Biological Projection	0.5-3	0.3	Capillary Clusters	16.098	24.067	2.356	Tissue
Light-Red	36-63	Capillary Clusters	10	0.5	Protruded Part	3-10	0.7	Early Disease Clustering	34.654	54.654	2.546	Ulcer
Bright-Red	64-85	Early Disease Clustering	15	0.7	Start of Disease	10-18	0.8-0.9	Disease Suspect	56.656	45.454	86.547	Benign
Blood-Red	86-100	Disease Suspect	>=20	>0.8	Advanced Stage	>19	>=0.9	54.544	89.900	90.054	Cancer	

Table 1. Feature Master Table: The feature Analysis have been performed based on these feature attributes. Here, Temp. stands for Optical Temperature, Act. and Virt. stand for actual and virtually seen depths. Dnsty is the actual density whereas conc. is virtual concentration. H1 is hyperplane 1, H2 is hyperplane 2, and H' stands for Generated Hyperplane. Key identifies the denotion by respective feature attributes.

the impact of capture quality, minute luminous distortions provided the distortion ratio falls below 0.2:1 of the total image pixel quantity, the obscured parts of the image are predicted via an incrementally trained model, thus ensuring growing learning and applied gradients. In the case of capture intensities, the model is quite sensitive to the luminous temperatures of the input images since it simultaneously captures the chromatic considerations for the input data repositories and the features it uses for training.

4.2.1. Blood-Based Feature Composition

The major portion of the input images contains blood-red and thin-red chromatic compositions with spotted white tissue regions along the surface of the GI tract. At places/junctions of the human organs, the organs and tissues turn visually coarse and dense, thereby making them blood-rich. When we consider the internal bleeding comorbidities through endoscopy, there is a risk for the model confusing the bleeding organ/vessel/capillary/tissue/internal lining with other bloodstreams, thus posing a danger for grave ignorance of ill-stricken features that may cost the patient's life. Hence, to effectively differentiate an ailment from natural human body blood in composition, we have implemented color shade charts as shown in the first segment of the table 1. By implementing the model with certain hyperparameters and filters, we can modify and adapt the model with different shade conditions as tabulated.

4.2.2. Depth-Based Feature Comparisons

The next step in feature selection is to employ the depth measurements through intercepted-wave trigonometric-calculus equations as mentioned in the equation 1. We have also estimated the height of the various features in the Human GI Tract by using solid angle measurements through spherical vector estimations as seen in the equation 2. We have tabulated the diseases identifiable through this approach in the second section of the table 1.

$$\begin{aligned} h'(x) = (&Ah(x) + B'\text{Color}(x) \\ &+ (B + B'x)\text{density}(x)) \\ &\times \text{luminosity}(I) \end{aligned} \tag{1}$$

where h'(x) is the modified height, A is the Amplification factor(based on application), B' is the transformation factor modified based on application, B is the transformation factor based on domain, density(x) is the visual density function, luminosity(x) is the luminous fraction, x is the object under consideration and I is visual intensity.

$$H(x) = \sqrt[3]{\frac{h'(x)}{4\pi}} + b_{spher} \tag{2}$$

where H(x) is the final height, and b_{spher} is the bias. h'(x) is equation 1

4.2.3. Visual Density Based Feature Computations

By analyzing the pixellated versions of the image, we can decipher the visual densities of the captured image and compare them with the thresholds mentioned in the third section of the table 1 for estimating how the given feature is in proximity of a possible disease region/if it is the actual disease region. The numerical estimation is based on the equation

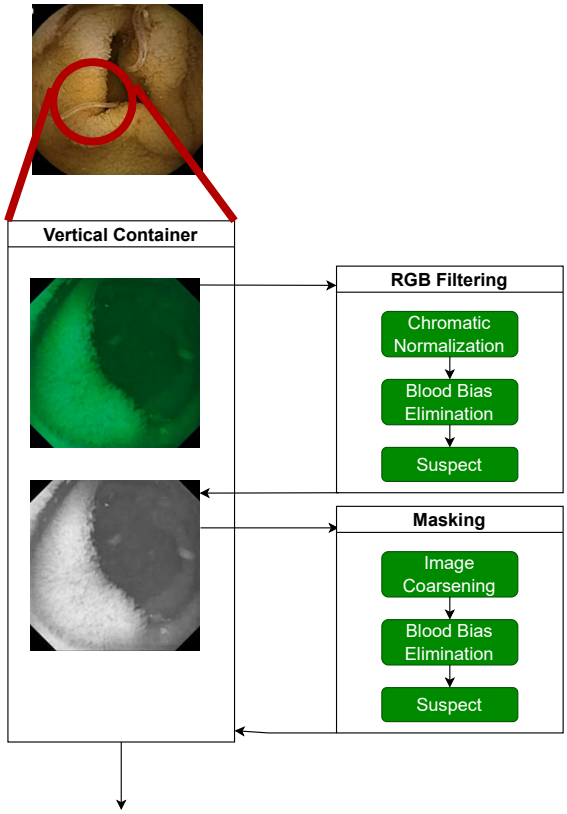

Figure 1. We concentrated on a narrow segment, highlighted in red to process the image as follows: RGB filtering to eliminate blood biases and Black and white masking to assess visual density parameters. The subprocesses are shown within respective boxes shown alongside the main figure.

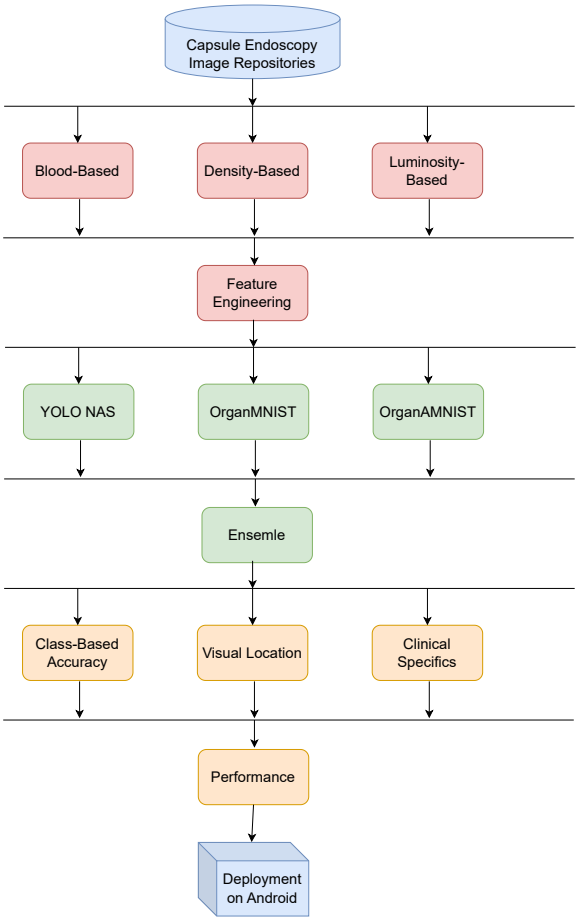

Figure 2. The flow of our schema is presented with merge and forks along the way. Since we are using an ensemble, the accuracy methods employed by us are comparatively complex.

3.

$$d'(x) = (A(d(x) * B'\text{Color}(x)) + (B + B'x)\frac{\text{H}(x)}{4pieh'(x)}) \times \text{luminosity}(I) \quad (3)$$

where h'(x) is the modified height, A is the Amplification factor(based on application), B' is the transformation factor modified based on application, H is the actual depth, d'(x) is actual density, luminosity(x) is the luminous fraction, x is the object under consideration and I is visual intensity.

4.3. Vision-Based Geometrical Evaluations

4.3.1. Depth-Based Calculation Methods

The figure 3 shows Solid Angle Computations through vector-based projections on the object hyperplane H that help estimate height values effectively. Six Projections are employed in the current work which is directed over 6 different hyperplanes H1, H2, H3, H4, H5, and H6 of the part under investigation where the position vectors r1,r2,.... are densely routed to the convoluted regions. The convoluted regions may include steep curves H1, sharp bending H2, and irregular faces H3. The tails of the vectors are extended farther onto another hyperplane H' which resembles the projected face of the object and similarly, 6 such hyperplanes are generated. Another set of position vectors r1',r2',... are projected from H' that further extends to the X-axis of a graphical plane G' to obtain the probability distribution across different hyperplanes. The height is com-

puted based on the equation 4.

$$h'(x) = P(h'(H1) * H(H1)) * (B + B'x)P(h'(H2) * H(H2)) \times \frac{H(H1).H(H')}{|H'|} \times \frac{H(H2).H(H')}{|H'|} + b(H1, H2) \quad (4)$$

where h'(x) is the modified height, A is the Amplification factor(based on application), B' is the transformation factor modified based on application, B is the transformation factor based on domain, density(x) is the visual density function, luminosity(x) is the luminous fraction, x is the Hyperplane, P(x) is the function on Hyperplane, and b(x,y) is a bias function of 2 hyperplanes.

4.3.2. Luminosity Numerical Rubrics

The mirroring effect is pronounced in figure 4 where views are directed opposite to each other to account for different luminous intensities. Hyperplane has different definitions and terminologies here since luminous intensities are considered. There are no reflection points in biomedical computer vision but shadowed faces exist. We have two different interpretations of this effect shown in the figure, one is the conceptual framework while the other is the practical slide. The slide shown in the figure has a total of 36 shadowed faces with multiple convolutions which cannot be solved with in-hand techniques. We have superimposed 5 shadow faces S1,S2,S3,S4, and S5 with projections from 10 different viewpoints V1,V2,V3,V4,V5,V1',V2',V3',V4', and V5'.Rays r1,r2,r3,r4, and r5 are exactly complemented by r1',r2',r3',r4', and r5'. While r1 looks at the face S1 from V1, r1' views the same from V1'. The computations subsequently follow the following equation 5.

$$L"(x, y) = L(x, y) * (B + B'x)L'(x, y) + b_{lumin} \quad (5)$$

where L"(x,y) is luminous intensity at (x,y), L(x,y)&L'(x,y) are respective luminous intensities when viewed from opposite sides, b_{lumin} is luminous bias.

4.4. Ensemble Learning

We have considered an ensemble of 3 Biomedical and Conventional Computer Vision frameworks for modeling the visual components visible from the input training images, informed representations by the classifications in these images, and learning the actual differentiation pointers among the learning gradients during its testing. This process entails the following phases:

4.4.1. YOLO NAS

Neural Architecture Search is very crucial in feature-intensive applications similar to the current case where YOLO NAS with certain modifications to adapt to Healthcare Computer Vision data repositories. We have performed 60% modifications in the data set as detailed in the first segment of the table 2. The rest of the operations reflect the conventional functionalities in the standard YOLO NAS documentation, not explained for simplicity reasons.

4.4.2. OrganMNIST

This is specifically designed for BioComputational Applications and we have performed 15% modifications in its architecture as tabulated in the second section of the table 2.

4.4.3. OrganAMNIST

A slender modification of 0.5% has been made to the previous model for performance purposes and the modifications are elucidated in the third section of the table 2.

4.5. Training, Testing, and Universalization of the Developed Solution

As mentioned earlier, this phase constitutes just 10% of the overall research work, owing to the subtle tedious nature of the preceding tasks.

4.5.1. Training Modalities

Class-Based Segregation: Any minute error/ignorance in multi-class segregation can significantly hamper training epoch accuracy and consequently lead to poor accuracy. The Multi-class segregation has been performed based on the determiners present in the figure 1.
Multi-Class Classification: While Multi-class Segregation of the input data features is important, it is equally important to build a classification system that accommodates the learnings from these classes based on the schema in figure 2.

4.5.2. Testing Mode Selection

Conventional Approach: The model has been tested as other models are with no domain-specific criteria for getting clear descriptive statistics of its performance.
Domain-Specific Approach: We used fine-tuning to further improve the performance of already trained models and implemented domain-specific tests as detailed in the table; 4.

4.5.3. Universalization of Solutions

Adaptability to the Complete GI Tract: The Endoscopy tube is so flexible that it can be made to enter any narrow/shallow passages irrespective of the hindrances posed. Hence, the model has been trained on every reachable part within the GI Tract.
Universal Situations Trained: The model has been trained over all the situations to adapt to all illumination, density, and biological constraints.
All of these factors have been demonstrated in the table 1.

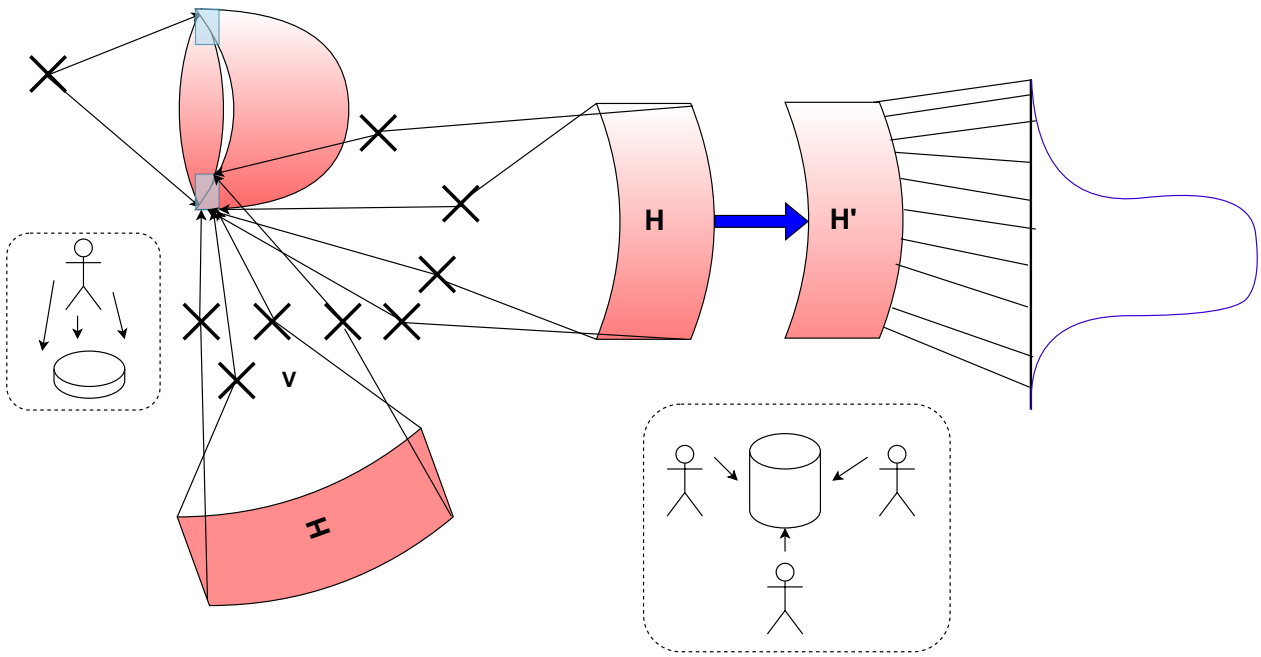

Figure 3. The object can be viewed from multiple sides but for simplicity, we considered two directions. The directions considered in our experimentation are demonstrated in dotted rectangles with viewpoints shown by user symbols. H' Is the projected hyperplane while H is the actual hyperplane. Cross(**X**) is the viewpoint. On projecting points of the hyperplane, we are getting normal distribution as seen on the extreme right.

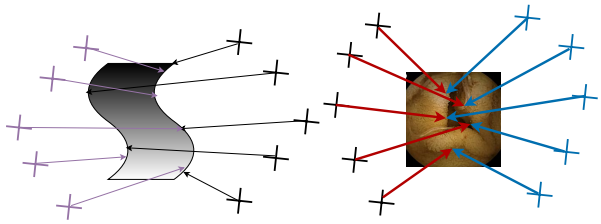

Figure 4. The figure details the mirror view theory. This is a bipartite representation where the opposite of the previously viewed point is seen to scale the model across all the possible illuminations Cross(**X**) has been depicted to represent viewpoints while arrows are showing view directions.

5. Results and Discussions

We can observe a direct relationship between feature engineering and ensemble accuracy. For accuracy estimation of the ensemble, we employed voting classifiers with 80% contribution, model average accuracy at 10% contribution, and ensemble-as-an-entity accuracy at 10% to finally estimate the net accuracy.

5.0.1. Accuracy over the classes

We could find a linear relationship among the classes with diseases that are characterized by a large change in color, demonstrate huge density variations, and cause sufficient contraction/relaxation of the surrounding tissues/organs. In a few cases, inflammations may also be included in this category if they involve a contraction of a large group of epithelial cells or esophagus tissues. The ensemble has successfully identified 9 out of 10 classes with 99.8% accuracy on a general scale.

For the diseases that form indistinct clusters of bruises/spots with their colors approximately matching that of blood follow non-linear deterministic regular curves with their curve gradients and shapes predictable via a standard calculus mechanism. This may fail for the cases where these bruises have the same chromatic values as human blood. The ensemble model has performed here with an accuracy of 79.8%.

For diseases characterized by unnoticeable inflammations, small growths, ulcers, etc., the relation of learning is a non-deterministic curve with no predictable behavior and this needs further techniques to resolve the associated calculations. The accuracy has been reported to be at 77.7%.

5.0.2. Visual Location Results

In our performance benchmarks designed for the current work, we have assigned 17% to Disease Detection and 83% to the Suspected location since location is quintessential for

Modification	**Layer/Step**	**Purpose**
YOLO NAS Modifications		
RGB Filtering	Feature Engineering	Isolate Blood Identifiers
Black and White Masking	Feature Engineering	Density-Based Segregation Apart from Normal Bodily Features
Bounding Box Probability Fitting	Prediction Phase	Based on Contour Distribution at the disease site
Bounding Box Width Fixture	Testing Phase	For Error Computation
OrganMNIST Modifications		
Chromatic Segregation	Feature Engineering	Isolate Blood Components
Fine-Tuning Data Addition	Testing Phase	Improve Performance
OrganAMNIST Modifications		
Chromatic Segregation	Feature Engineering	Isolate Blood Components
Fine-Tuning Data Addition	Testing Phase	Improve Performance
Hyperparameter-Tuning	Training Phase	To ensure optimal performance

Table 2. Modifications for YOLO NAS, OrganMNIST, and OrganAMNIST

Class	**Differentiator**
I-III	Spots
IV-VI	Spotted Clusters
VI-VIII	Scarred Regions
VIII-X	New Growth

Table 3. Classification Attributes

Domain	Differentiator	Example
Cardiovascular	Resulting From Blood Vessels	Internal Bleeding
Muscular	Caused due to Mechanical Damage of Muscles	Muscular Tears
Tissue	In-Depth Bruises& Scars	Tissue Repair
Organs	Change in Secretion Concentrations	Liver Problem

Table 4. OrganMNIST Modifications

further prognosis/treatment. While the models were found to fare well in disease detection, it is critical to examine whether they will be successful in locating the exact region of the disease evidence for the user to further raise cases of patient treatment.

Bounding Box Marker: In our results, we found 97.6% of the total number of bounding boxes to be located at their exact location, 1.4% in the proximity of their exact location, 0.6% in higher probability regions around the exact location, and 0.4% in a completely different location/acts as an outlier.

Spatial Width of the Location: We also observed a tenacious width of the bounding boxes to be away from the error zone with a probability of 0.75, falling within the error zone as 0.10 and the bounding boxes of infinite width to be 0.15.

5.0.3. Numerical Rubrics

As we have discussed previously, Luminosity and Height Estimation techniques have been analyzed via Vector Projection methods with the following observations:

- **Depth Measurement:** The projected hyperplane in fig-

Depth				Luminosity			Blood-Based			Bounding Box			
H1	H2	H'	Res.	Front	Back	Res.	Temp1	Temp2	Res.	Err.	Deg.	Transl.	Res.
35.567	67.890	23.370	T	1.008	2.016	T	23	32	T	1.545	20	1.000	T
34.456	89.900	34.567	F	2.345	4.545	F	32	38	T	1.545	30	1.700	T
54.456	54.422	53.324	T	2.565	1.267	T	12	56	F	0.444	40	0.877	T
79.877	12.344	90.088	F	1.444	8.999	T	45	65	T	0.666	45	0.764	F
45.567	34.455	34.456	T	1.800	9.000	T	56	78	T	2.345	55	3.444	F
65.567	87.632	70.006	F	0.983	0.676	T	45	90	T	15.567	60	13.444	F
72.865	24.566	73.466	F	7.889	3.450	F	46	68	F	7.677	65	4.444	T
32.456	90.788	100.00	F	23.456	23.456	F	12	56	T	80.999	75	2.333	F

Table 5. Results: The first set of attributes: H1 hyperplane, H2 hyperplane, H' generated hyperplane contribute to Depth-Based Results. Front Back are the front and back luminosity intensities in the mirror view procedure. Temp1 and Temp2 are optical temperatures in color-based feature engineering. Visual Location-specific attributes are shown as Error, Degree, Translation, etc., in the fourth segment. Res. are the resultant truth values whether they align with previously set benchmarks.

ure 3 is investigated for how its points vary across space and we found it to be normally distributed. Normal Distributions are simpler to handle but need clear distributive statistics. The results are presented in the first section of the table 5

- **Luminosity Measurement:** The body parts can be modeled across different directions which vary largely in illuminations. Mirror view theory in 5 showcases the different shades of the same part and its numerical results have been demonstrated in the second section of the table 5.

All the results of the table 5 follow the following master equation 6.

$$\begin{aligned} Res. = & f(p1,p2) * f(p3)\& \\ & f(p1,p3) * f(p2)\& \\ & f(p2,p3) * f(p3) \\ & |P(p1 * p2 * p3) \times P(F_{true}) \end{aligned} \tag{6}$$

where Res. is result(a boolean value highlighting if the given set of parameters align with our benchmarks), p1,p2,p3 are attributes, f(x,y) is a boolean function designed according to the application, P(x) is the probability for T of x, F_{true} is the number of possibilities for true to be false, & is boolean and, * is convolution function and | is boolean or.

5.0.4. Numerical Specifics with Clinical Significance

Certain extraneous attributes have been found relevant for the current work which are listed along with their clinical significance in the fourth section of the table 5.

5.0.5. Discussion of the above cases

We can observe a significant rise in the accuracy of health-specific models as we increase the significance of feature selection in building it. This implies that in a standard normal distribution of the weighted health parameters, the feature probabilities that lie in the confidence region play the role of the entire work's performance and its scalability across multiple use cases. Additionally, we can observe that a significant bias towards normal bodily features can be eliminated through deep feature filtering and highly selective feature selection. While optimal intensities also play a role, their effect can be reduced to a minimum by employing incremental growth techniques during the phases of model training which begets the time wasted on training the model for hours. Hence, with efficient feature engineering and effective model training techniques, we can significantly increase the accuracy base of our GI Endoscopic modules by 37%.

6. Conclusion and Future Work

Using Computer Vision Framework to detect and locate GI Tract Internal Diseases and Comorbidities is a simple, yet disruptive technology that can bring significant contributions to the field of AI in Healthcare. A simple camera embedded in an Endoscopic tube can further be used to expand the currently considered data sets to bring in high accuracies and ensure heavy scalabilities. As a part of our future work, we are planning to focus on expanding the data set, covering every possible disease in the internal GI tract up to its end. Since this data set is considerably huge and requires a complex pipeline, we plan to focus our approach on building cost-effective universal models that can give predictions under any biological and technical constraints with an estimated accuracy of 98.9%, similar to the current work.

References

[1] Cao, Q., Deng, R., Pan, Y. et al. Robotic wireless capsule endoscopy: recent advances and upcoming technologies. Nat Com-

mun 15, 4597 (2024). https://doi.org/10.1038/s41467-024-49019-0

[2] Ali, S. Where do we stand in AI for endoscopic image analysis? Deciphering gaps and future directions. npj Digit. Med. 5, 184 (2022). https://doi.org/10.1038/s41746-022-00733-3

[3] Reuangrith J, Scott SA, Kohansal A. Endoscopic ultrasound-guided placement of lumen-apposing metal stent for transgastric drainage of loculated malignant ascites. Therapeutic Advances in Gastrointestinal Endoscopy. 2024;17. doi:10.1177/26317745241289238

[4] Lisen Zhu, Jianan Chen, Huixin Yang, Xinkai Zhou, Qihang Gao, Rui Loureiro, Shuo Gao, Hubin Zhao, Wearable Near-Eye Tracking Technologies for Health: A Review, Bioengineering, 10.3390/bioengineering11070738, 11, 7, (738), (2024).

[5] Chadebecq, F., Lovat, L.B. & Stoyanov, D. Artificial intelligence and automation in endoscopy and surgery. Nat Rev Gastroenterol Hepatol 20, 171–182 (2023). https://doi.org/10.1038/s41575-022-00701-y

[6] Melson J, Trikudanathan G, Abu Dayyeh BK et al. Video capsule endoscopy. Gastrointest. Endosc. 2021; 93: 784–796.

[7] Senger, Sebastian & Lepshokov, Magomed & Tschernig, Thomas & Cinalli, Guiseppe & Oertel, Joachim. (2024). Evaluation of training models for intraventricular neuroendoscopy. Neurosurgical Review. 47. 10.1007/s10143-024-03082-9.

[8] E. G. Lim, Z. Wang, S. Nie, T. Tillo, K. l. Man and N. Zhang, "Moveable wireless capsule endoscopy," 2013 International SoC Design Conference (ISOCC), Busan, Korea (South), 2013, pp. 270-273, doi: 10.1109/ISOCC.2013.6864025.

[9] Handa, Palak & Mahbod, Amirreza & Schwarzhans, Florian & Woitek, Ramona & Gooel, Nidhi & Chhabra, Deepti & Jha, Shreshtha & Dhir, Manas & Gunjan, Deepak & Kakarla, Jagadeesh & Raman, Balasubramanian. (2024). Capsule Vision 2024 Challenge: Multi-Class Abnormality Classification for Video Capsule Endoscopy. 10.48550/arXiv.2408.04940.

[10] Hale, M. F., Sidhu, R., & McAlindon, M. E. (2014). Capsule endoscopy: current practice and future directions. World journal of gastroenterology, 20(24), 7752–7759. https://doi.org/10.3748/wjg.v20.i24.7752

[11] Mukhtorov D, Rakhmonova M, Muksimova S, Cho Y-I. Endoscopic Image Classification Based on Explainable Deep Learning. Sensors. 2023; 23(6):3176. https://doi.org/10.3390/s23063176

[12] Steinmann, R.; Cortegoso Valdivia, P.; Nowak, T.; Koulaouzidis, A. An Overview of the Evolution of Capsule Endoscopy Research—Text-Mining Analysis and Publication Trends. Diagnostics 2022, 12, 2238. https://doi.org/10.3390/diagnostics12092238

[13] Jin, Z., Gan, T., Wang, P. et al. Deep learning for gastroscopic images: computer-aided techniques for clinicians. BioMed Eng OnLine 21, 12 (2022). https://doi.org/10.1186/s12938-022-00979-8

[14] Nguyen, V. X., Le Nguyen, V. T., & Nguyen, C. C. (2010). Appropriate use of endoscopy in the diagnosis and treatment of gastrointestinal diseases: up-to-date indications for primary care providers. International Journal of General Medicine, 3, 345–357. https://doi.org/10.2147/IJGM.S14555

[15] Mourad, M., Kim, J.E., Phillips, S.E. et al. Association of DCI with number of preoperative comorbidities and 30-day outcomes following inguinal hernia repair: an analysis of the ACHQC database. Surg Endosc (2024). https://doi.org/10.1007/s00464-024-11381-x

[16] Axon A.T.R. Fifty years of digestive endoscopy: Successes, setbacks, solutions and the future. Dig. Endosc. 2019;32:290–297. doi: 10.1111/den.13593.

[17] Steinmann, R., Cortegoso Valdivia, P., Nowak, T., & Koulaouzidis, A. (2022). An Overview of the Evolution of Capsule Endoscopy Research-Text-Mining Analysis and Publication Trends. Diagnostics (Basel, Switzerland), 12(9), 2238. https://doi.org/10.3390/diagnostics12092238

[18] Cai, W., Zhang, X., Luo, Y. et al. Quality indicators of colonoscopy care: a qualitative study from the perspectives of colonoscopy participants and nurses. BMC Health Serv Res 22, 1064 (2022). https://doi.org/10.1186/s12913-022-08466-5

[19] Namikawa K, Hirasawa T, Yoshio T, et al. Utilizing artificial intelligence in endoscopy: a clinician's guide. Expert Rev Gastroenterol Hepatol. 2020;14(8):689-706. doi:10.1080/17474124.2020.1779058

[20] Okagawa, Y., Abe, S., Yamada, M. et al. Artificial Intelligence in Endoscopy. Dig Dis Sci 67, 1553–1572 (2022). https://doi.org/10.1007/s10620-021-07086-z

[21] Zha, B., Cai, A., & Wang, G. (2024). Diagnostic Accuracy of Artificial Intelligence in Endoscopy: Umbrella Review. JMIR medical informatics, 12, e56361. https://doi.org/10.2196/56361

[22] Chadebecq, F., Lovat, L.B. & Stoyanov, D. Artificial intelligence and automation in endoscopy and surgery. Nat Rev Gastroenterol Hepatol 20, 171–182 (2023). https://doi.org/10.1038/s41575-022-00701-y

[23] Okagawa Y, Abe S, Yamada M, Oda I, Saito Y. Artificial Intelligence in Endoscopy. Dig Dis Sci. 2022;67(5):1553-1572. doi:10.1007/s10620-021-07086-z

[24] Nagao, S., Tani, Y., Shibata, J., Tsuji, Y., Tada, T., Ishihara, R., & Fujishiro, M. (2022). Implementation of artificial intelligence in upper gastrointestinal endoscopy. DEN open, 2(1), e72. https://doi.org/10.1002/deo2.72

[25] Fujia Guo, Hua Meng, Application of artificial intelligence in gastrointestinal endoscopy, Arab Journal of Gastroenterology, Volume 25, Issue 2, 2024, Pages 93-96, ISSN 1687-1979, https://doi.org/10.1016/j.ajg.2023.12.010.