Reduction of Video Capsule Endoscopy Reading Times Using Artificial Intelligence

Azubuogu (Azu) Anudu, MD[1,2]; Chukwudumebi Uche, DO[2]; Roshan Warman, BS[3], Hunter Morera, MS[2], Nikhil Reddy, MS[4], Ivana Radosavljevic, MS[4], Niketa Patel, PhD[5], Patrick Brady, MD[2], Joe Lezama, MD[5], Dmitri Goldgof, PhD[5], Lawrence Hall, PhD[4], Gitanjali Vidyarthi, MD[5].

[1]Baylor College of Medicine, Houston, TX; [2]University of South Florida Morsani College of Medicine, Tampa, FL; [3]Yale University, New Haven, CT; [4]University of South Florida, Tampa, FL; [5]James A. Haley Veteran Affairs Hospital, Tampa, FL.

Introduction

Video capsule endoscopy (VCE) is an innovation that has revolutionized care within the field of Gastroenterology, but the time needed to read the studies generated has often been cited as an area for improvement. With the advent of artificial intelligence, various fields have been able to improve the efficiency of their core processes by reducing the burden of irrelevant stimuli on their human elements.

- Interpretation of VCE studies can be a daunting task as it requires an individual with expertise to review more than 50,000-70,000 images, with 8 hours of recording, looking for abnormal pathology that is often only present in one or two frames.
- As a result of the limitations of human concentration, VCE readings have a significant miss rate of 5.9% for vascular lesions, 0.5% for ulcers, and 18.9 for neoplasms.
- With the evolution of artificial intelligence (AI), computer-aided diagnosis (CAD) has shown promise in many areas of medicine, including pathology, dermatology, radiology and gastroenterology, where it has been used to decrease observational oversights (i.e. human error).

In this study, we have created and trained a convolutional neural network capable of significantly reducing capsule endoscopy reading times by eliminating normal images while retaining abnormal ones. Our model, a variation of ResNet50, was able to reduce VCE video length by 47% on average and capture abnormal segments on VCE with 100% accuracy as confirmed by reading physicians. Although our model was trained using limited data publicly available from the University of Thessaly, we anticipate that performing this study on a larger scale will yield similar 10 results further supporting the diagnostic use of VCE.

Objectives

- The aim of this study is to use computer-aided diagnosis (CAD) to identify abnormal vascular and inflammatory capsule endoscopy images.
- The secondary goal is to reduce the amount of time spent by physicians reading VCE studies by decreasing the number of frames in a typical capsule endoscopy video.

Methods

- We used publicly available data from the κάψουλα interactive database (KID).
- 3 full length capsule endoscopy videos from the dataset were annotated by a physician with 20 years of GI experience.
- The goal of this work is to reduce the physician time needed to read the VCE videos. Thus, our framework focused on removing confidently predicted normal frames which are not needed for diagnoses.
- We trained a convolutional neural network, ResNet50, using videos from the KID dataset to confidently exclude normal images on VCE, while retaining abnormal ones.
- The CNN was exposed to 3 full length capsule endoscopy videos. A threefold cross-validation scheme was employed whereby the model was trained using two of the aforementioned videos, tested on the third and this was repeated for all possible video combinations.
- We trained our model for nine epochs, but found that many abnormal segments were still being missed. KID videos 1 and 3 had significantly less abnormal frames than KID video 2.
- Our solution was to augment the abnormal frames in these two videos using rotation. These rotations increased the number of abnormal frames significantly.

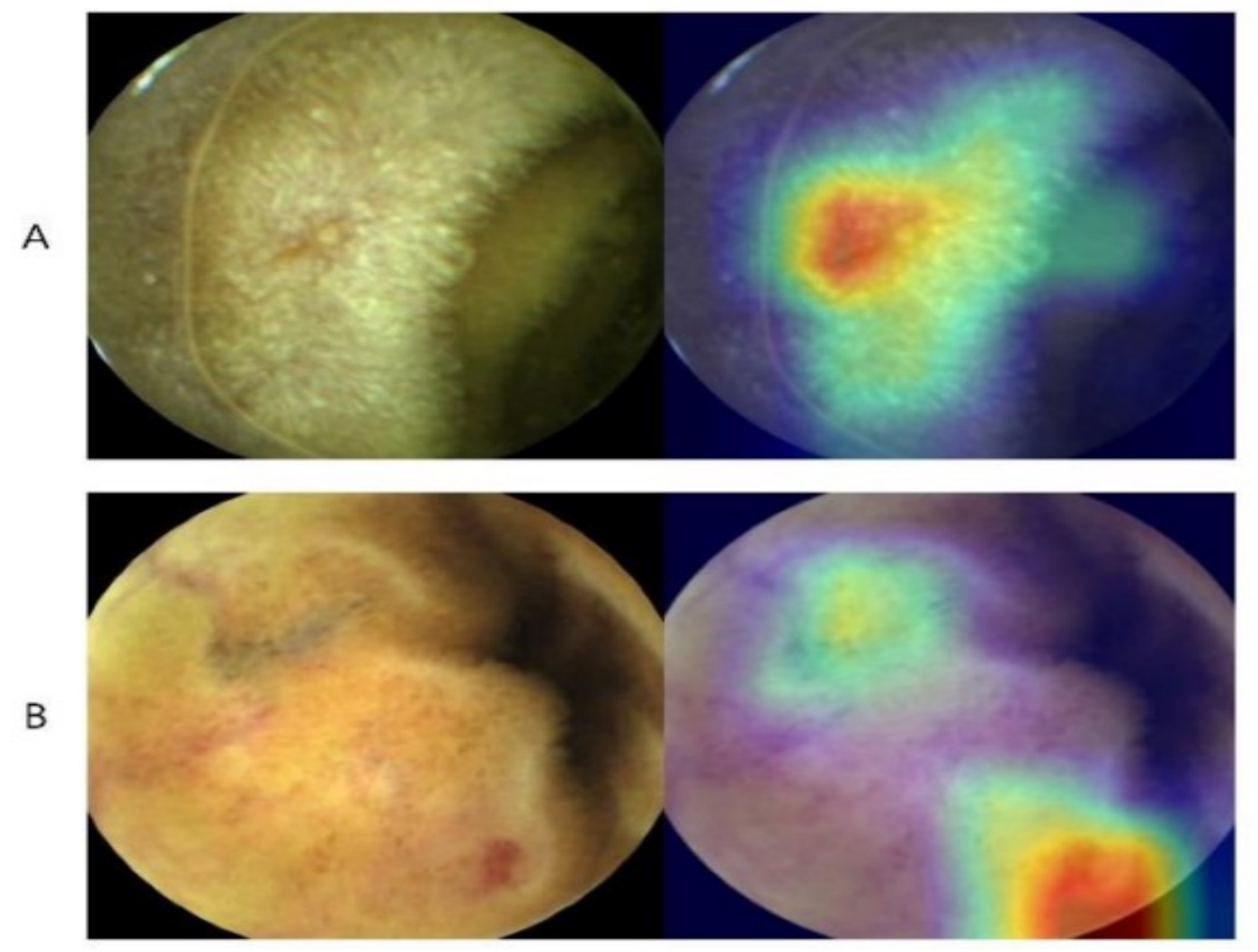


Figure 1. CNN's class activation map (CAM) in capsule endoscopy image reading. This visualized feature map describes how the algorithm predicts each lesion. (A) Erosion with central depression of the mucosa is highlighted as red. (B) Simultaneous detection of both erosion and prominent vasculature.

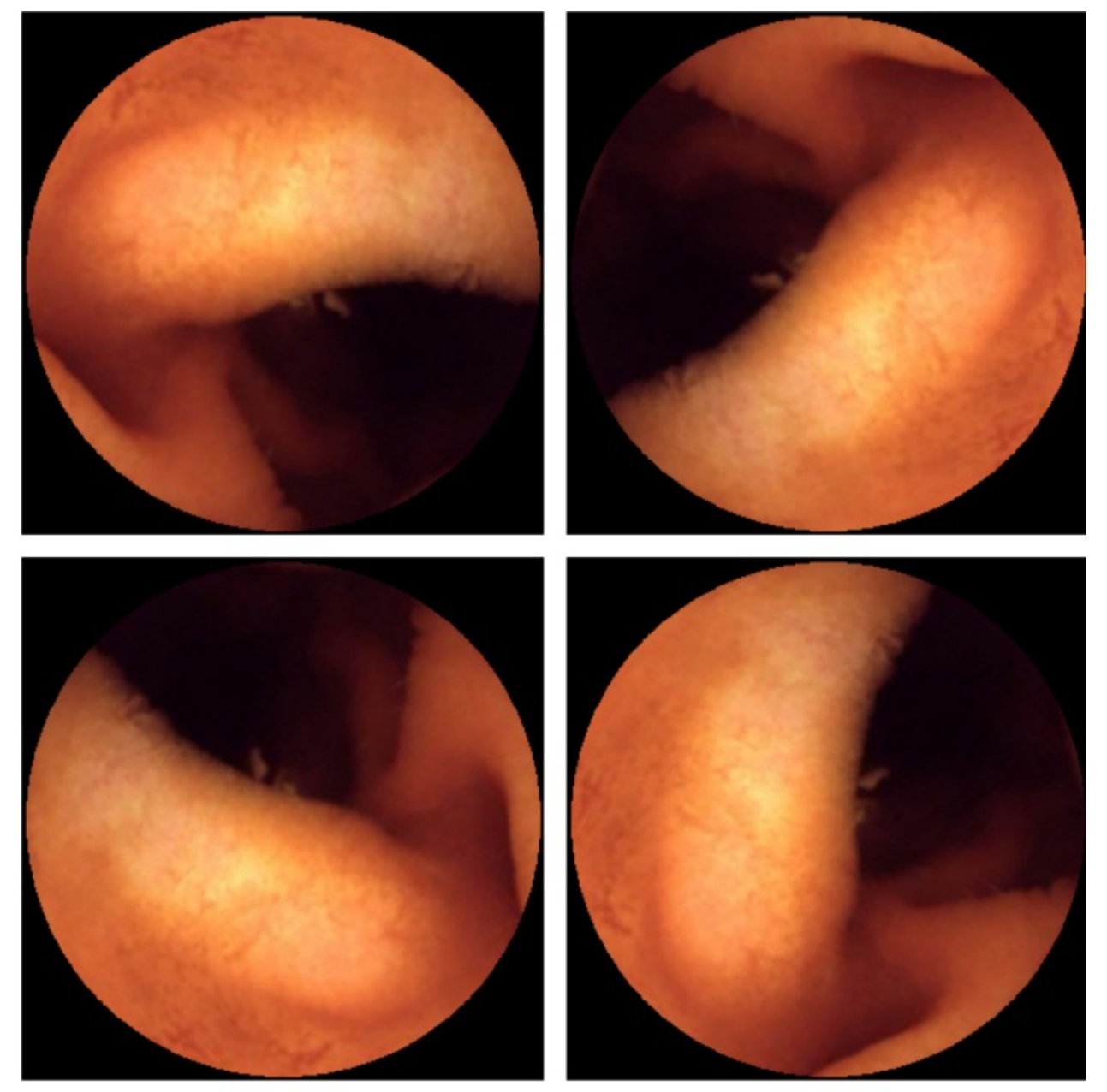
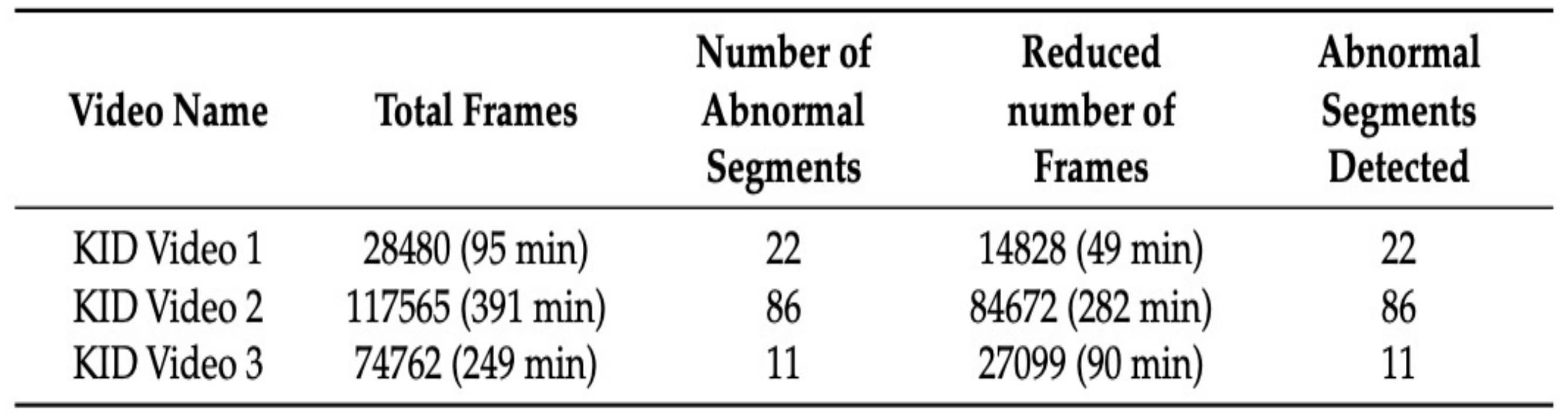


Figure 2. Rotation Augmentation Examples from KID Video

Results

- In our first experiment we trained the deep learning model, for ten epochs, using the final epoch to make predictions on the test videos. This single network was able to identify 21 of the 119 abnormalities, while reducing the number of frames on average by 97%.
- In our second experiment, the ensemble method was able to identify 85 of the 119 abnormal segments, while reducing the number of frames by 75%. However, only 80 of the 119 abnormal segments were detected, and a reduction in number of frames by 76%.
- In our final experiment, we re-trained the networks using the videos with the augmented abnormal 138 frames. The average reduction in number of frames was 47%, while the number of abnormal segments detected is 119 of 119.
- A paired T-test with a confidence interval of 95% showed that this ensemble method with augmented data had a statistically significant improvement in the detection of abnormal segments in the three-fold cross validation.
- Despite working with a small dataset of only two videos for training in each fold, we were able to develop an algorithm that successfully detected 100% of abnormal segments while reducing the reading time for a physician by 47% on average.



Video Name	Total Frames	Number of Abnormal Segments	Reduced number of Frames	Abnormal Segments Detected
KID Video 1	28480 (95 min)	22	14828 (49 min)	22
KID Video 2	117565 (391 min)	86	84672 (282 min)	86
KID Video 3	74762 (249 min)	11	27099 (90 min)	11

Figure 3. Ensemble of Networks Trained on Augmented Videos.

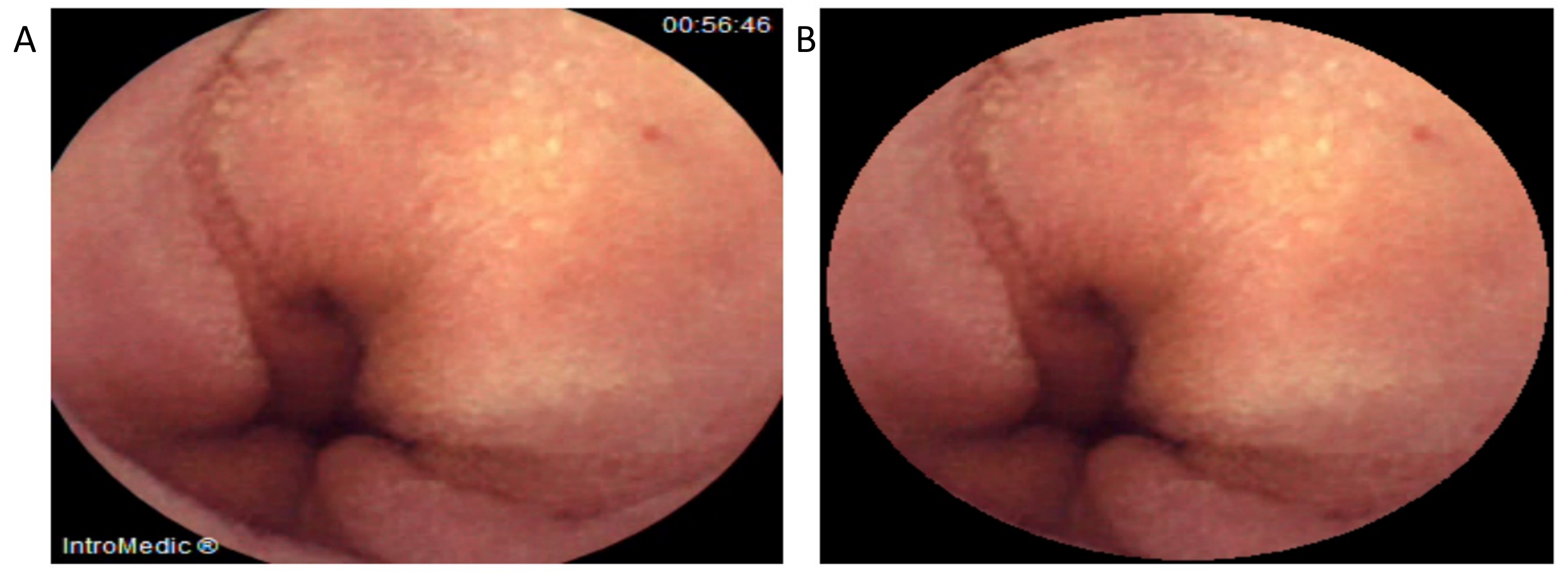


Figure 4. (A) KID video 1 frame (B) KID video 1 frame after pre-processing

Discussion

- The goal of our study is to reduce the physician time needed to read VCE studies by excluding frames classified as normal and as such not needed for diagnosis. This ensures the physician has fewer frames to review thereby potentially increasing diagnostic accuracy.
- Our study used a robust convolutional neural network (CNN) that was trained and tested to classify a frame as normal or abnormal, with the goal of minimalizing the number of false negatives.
- Using the method proposed in our study, we were able to reduce the video length on average by 47%.
- Using this model, we captured frames from all 119 of the abnormal segments, for an accuracy of 100%.
- This approach has the promise of significantly reducing the reading time needed by a physician to review VCE images. This study contributes to the growing field of CNN in the detection of subtle lesions in the small bowel.
- Similar studies have efficiently detected pathology in VCE in shorter time periods, however, our study had the benefit of having the physician review all the abnormal frames, thus ensuring an accurate diagnosis.
- Our study has some limitations which offer opportunity for improvement. Our CNN model was trained only on two videos and tested on one. As such, this limits the scale of our study and the results achieved.
- We hope to improve this by increasing the data available in training our CNN model, with the goal of establishing an algorithm that can delineate between normal and abnormal frames instantly.

Conclusion

Artificial intelligence (AI) is an emerging field in Gastroenterology, particularly in VCE. Our study focused on training a CNN based model to identify subtle inflammatory and vascular lesions, with the goal of developing an algorithm that can quickly differentiate between normal and abnormal frames. Our study results demonstrate the benefit of CNN in processing VCE images. We believe our study lays an excellent foundation for further validation in large multi-center trials.

Contact

Azubuogu (Azu) Anudu
Baylor College of Medicine
Email: azuanudu@gmail.com
Phone: (859)-979-9560

References

1. Byrne, M.F. Artificial intelligence and capsule endoscopy: Is the truly "smart" capsule nearly here? Gastrointest Endosc 2019, 89, 215 195-197.
2. Pogorelov, K. Bleeding detection in wireless capsule endoscopy videos - Color versus texture features. J Appl Clin Med Phys 2019, 217 20, 141-154.
3. Iddan, G. Wireless capsule endoscopy. 2000, 405-417.
4. Kim, S.H. Artificial Intelligence in Capsule Endoscopy: A Practical Guide to Its Past and Future Challenges. Diagnostics (Basel) 2021, 11.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com