Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)
16-18 November 2022, Maldives

Endoscopic image analysis using Deep Convolutional GAN and traditional data augmentation

Muhammad Muzzammil Auzine
Department of Software and Information Systems
University of Mauritius
Reduit, Mauritius
muhammad.auzine2@umail.uom.ac.mu

Maleika Heenaye-Mamode Khan
Department of Software and Information Systems
University of Mauritius
Reduit, Mauritius
m.mamodekhan@uom.ac.mu

Sunilduth Baichoo
Department of Software and Information Systems
University of Mauritius
Reduit, Mauritius
sbaichoo@uom.ac.mu

Nuzhah Gooda Sahib
Department of Software and Information Systems
University of Mauritius
Reduit, Mauritius
n.goodasahib@uom.ac.mu

Xiaohong Gao
Department of Computer Science
Middlesex University
London, England
x.gao@mdx.ac.uk

Preeti Bissoonauth-Daiboo
Department of Software and Information Systems
University of Mauritius
Reduit, Mauritius
Preeti.bissoonauth-daiboo@umail.uom.ac

***Abstract*— One big challenge encountered in the medical field is the availability of only limited annotated datasets for research. On the other hand, medical image annotation requires a lot of input from medical experts. It is noticed that machine learning and deep learning are producing better results in the area of image classification. However, these techniques require large training datasets, which is the major concern for medical image processing. Another issue is the unbalanced nature of the different classes of data, leading to the under-representation of some classes. Data augmentation has emerged as a good technique to deal with these challenges. In this work, we have applied traditional data augmentation and Generative Adversarial Network (GAN) on endoscopic esophagus images to increase the number of images for the training datasets. Eventually we have applied two deep learning models namely ResNet50 and VGG16 to extract and represent the relevant cancer features. The results show that the accuracy of the model increases with data augmentation and GAN. In fact, GAN has achieved the highest accuracy, that is, 94% over non-augmented training set and traditional data augmentation for VGG16.**

Keywords*— *Endoscopic images, DCGAN, traditional data augmentation, deep learning

I. Introduction

As a results of its invasive nature and low survival rate, esophageal cancer is one of the most dangerous cancers in the world. In terms of mortality, it ranks sixth among all cancers. [1]. The oesophagus is a long, hollow tube that moves the swallowed food from the back of a person's throat to his/her stomach to be digested [2]. Oesophageal cancer (OC) usually begins in the cells that line the inside of the oesophagus and it can occur anywhere along the tube [2]. The disease is contracted mostly by elderly people and there were over 500,000 new cases in 2018. Approximately 80% of occurrences worldwide occur in the less developed parts of the world. According to the World Cancer Research Fund International (WCRF), men are twice as likely as women to develop oesophageal cancer. and it varies within different geographic locations. In some places, higher rates of OC may be attributed to tobacco and alcohol use while in others, to particular nutritional habits and obesity [2]. Chances of survival are low primarily because esophageal cancer is typically identified at an advanced stages. According to the WCRF, the highest rates of oesophageal cancer mortality occur in Eastern Asia and Southern Africa in males and in Eastern Asia and Southern Africa in females. The five-year survival rate in the United States is 20%, while it is 10% in Europe.

In 2015, cancer in the digestive system caused the second highest number of fatalities among all sites [3]. To address this problem, early gastric cancer (EGC), screening and early detection of colorectal cancer have been introduced. Thus, it was possible to diagnose organs like breast, colon and intestine where real-time image information is obtained. This process has greatly helped identifying abnormalities, disorders and artefacts. Although there are various possible preventative interventions, none have been shown in prospective studies to reduce the incidence of oesophageal cancer. Population-based screening is impractical due to the low prevalence of oesophageal cancer, the absence of early signs, and the rarity of a hereditary form of the disease [4]. Hence, screening and detecting oesophageal cancer without misinterpreting the artefacts has become a very challenging task.

According to Markus [5], with the success of image processing techniques in the medical industry, Artificial Intelligence (AI) has lately made enormous advancements in autonomously identifying diseases, hence making diagnostics cheaper and more accessible. When applied to endoscopic images, deep machine learning may derive conclusions in a fraction of a second[5] However, the main challenge in this project will be to apply image processing techniques and deep learning techniques in oesophageal endoscopy images to improve the accuracy in detecting pathologies such as esophagitis, polyps and ulcerative colitis which can potentially lead to cancer. Hence, there is a need to develop a feasible and reliable method to alert endoscopists about possible presence of such pathologies [6] and this should be considered as an important and rapidly growing area of research [7]. However, the development of deep learning models requires large annotated datasets, which is a great challenge in the medical field. The creation of these large datasets require extensive input from medical practitioners [8]. Another major issue that is usually encountered in the class imbalance in medical datasets leading to the underrepresentation of some classes of data. This paper addresses these mentioned problems. The paper is organised as follows: Section I, II, III, IV and V.

II. LITERATURE REVIEW

Several works have been conducted in the development of automated detection of esophagus cancer. During an endoscopy surgery, To observe organ surfaces in the body and operate multiple surgical equipment, the surgeon employs a stereoscopic endoscope. Inadequate and irregular light sources with a relatively small field of vision, on the other hand, have become serious obstacles for endoscopic surgery. These circumstances not only impede image processing algorithms, but also have an impact on surgeons who operate in low-light environments where tool visibility is typically poor, greatly increasing the danger of injuring and even puncturing vital organs such as the liver and spleen. Furthermore, low-light areas have a poor signal-to-noise ratio and metrication artifacts caused by quantisation faults. As a result, current image enhancing technologies typically suffer from significant noise amplification in low-light areas. [9]. Janse et al [10] have developed an automated computer aided application for the early detection of esophageal cancer. The authors have then applied random forest as an ensemble classifier for esophageal image classification. Bang et al [11] have reported that deep learning and machine learning techniques are becoming more popular for the diagnosis of oesophagal cancer. These algorithms were developed through an automatic feature selection and self-learning from large endoscopic images datasets. During their study, it was also observed that the performance of these algorithms are hindered by limited datasets.

On the other hand, deep learning models are being investigated. In one research work, Lee et al. [12] uses deep CNN transfer learning (VGG16, InceptionV3, and ResNet-50) to identify 3 classes namely: normal, benign ulcer, and cancer using a custom dataset consisting of 787 images (367 Cancer,200 Normal,220 ulcer) gathered from Gil Hospital. The photos were resized to 224*224 pixel and preprocessed with adaptive histogram equalization (AHE) to eliminate image variances such as brightness and contrast, thereby boosting local contrast and enhancing edge definition in each image region. The accuracy, standard deviation, and AUC of three binary classification tasks using a combination of normal, cancer and ulcer performed on the models were compared in this work. The authors discovered that ResNet-50 had the best accuracy across all three classifications. The accuracy for cases involving normal photos was above 92 percent, whereas the accuracy for cancer vs ulcer was 77.1 percent, presumably due to smaller differences in appearance between cancer and ulcer. ResNet-50 was also reported to be more stable, having the lowest standard deviation. For ulcer vs normal, cancer vs normal, and ulcer vs cancer, the AUC values were 0.97, 0.95, and 0.85, respectively. These high levels of effectiveness across both normal and abnormal cases demonstrate that the proposed deep learning approach may be employed to support practitioners' manual inspection efforts to lessen the probability of missed positives due to repetitive endoscopic frames and diminishing attentiveness.

Xiao et al [13] uses deep CNN based on the UNet++ and Resnet-50 architecture to identify gastritis(AG) and non-atrophic gastritis(non-AG). 6122 images of white light endoscopy (4022 AG + 2100 non-AG) were collected from 456 patients. Two expert physicians randomly segregated the images into the training and test datasets in the proportions of 89 % and 11%, respectively. The binary classification tasks performed on the model reported an accuracy of 83.70%, sensitivity of 83.77% and specificity of 83.75%. Furthermore, the model gave an IOU of 0.648 for the region segmentation of AG by model and the region drawn by expert physicians, and an IOU of 0.777 for incisura segmentation. However, one big problem that many researchers encounter is the limited number of annotated samples available for medical datasets. It is a fact machine learning algorithms and deep learning models require very large training sets [14]. To overcome this issue, researchers are opting for data augmentation techniques. The latter techniques are used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. There are several data augmentation techniques namely: geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning among others [15]. Transformation such as translation, rotation, flip, and scale are among the most commonly used data augmentation methods used on the datasets. A new type of augmentation is the synthetic data augmentation. This technique enables more variability in the dataset to improve the training process. It is composed of two neural network models, a generator and a discriminator. In one work conducted by Frid-Adar et al [14], GAN was applied on liver images for cancer classification. The authors have generated synthetic liver lesions through classic data augmentation and GAN. The results showed that GAN produced better results. In another work, Desai et al [16] have generated geometric/intensity transformations of original images to produce synthetic images for breast cancer. Eventually CNN was used to detect breast cancer. Data augmentation is in great demand in the medical field. There is a need to explore these data augmentation techniques and to evaluate their results.

III. MATERIAL AND METHODS

This section presents the design of the proposed solution and the applications of the Oesophageal cancer detection.

A. *Architecture of the Oesophageal cancer detection application*

Fig. 1 depicts the application's proposed architecture. The pathological findings from the Kvasir-V2 dataset were used in this investigation. The images are first pre-processed to enhance their quality. Enhanced CNNs were developed by adjusting the learning rates at different epochs and fine-tuning the various hyper-parameters. The two augmented datasets, Basic augmentation and GAN augmentation datasets, were then used to analyse and evaluate the performances.

B. *Dataset*

The Kvasir dataset is made up of images that have been annotated and validated by competent endoscopists, with hundreds of images for each class indicating anatomical landmarks, pathological features, or endoscopic operations in

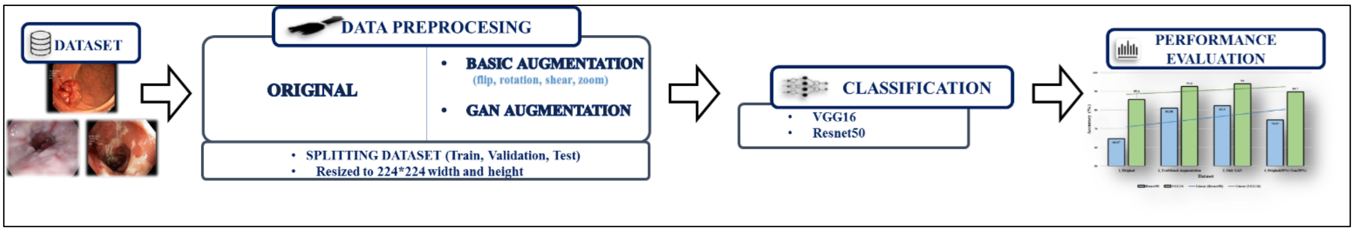

Fig. 1. Architecture of proposed model

the GI tract. Endoscopic equipment was used to acquire the data at Vestre Viken Health Trust (VV) in Norway. The VV consists of 4 hospitals and provides health care to 470.000 people. The dataset is publicly available at: https://datasets.simula.no/kvasir/. In this work, we have considered only the pathological findings which include 1000 images from 3 classes namely esophagitis, polyps, ulcerative colitis. These 3000 images were termed the 'Original' dataset. Firstly, the original dataset was subjected to basic augmentation techniques such as flip, rotation, shear and zoom which resulted in "6000" images and it was termed the 'Basic Augmentation' dataset. Furthermore, 1000 fake images for each class were generated using Generative Adversarial Network (GAN) augmentation technique and was termed the 'GAN augmentation' dataset. In addition, a fourth dataset was formulated using the original and GAN augmented images (50% each).

C. Image data augmentation

Image augmentation is a technique for manipulating existing image data in order to artificially expand the dataset, hence boosting the accuracy of deep learning model prediction. Fig. 2 depicts the image data augmentation approach employed in this work.

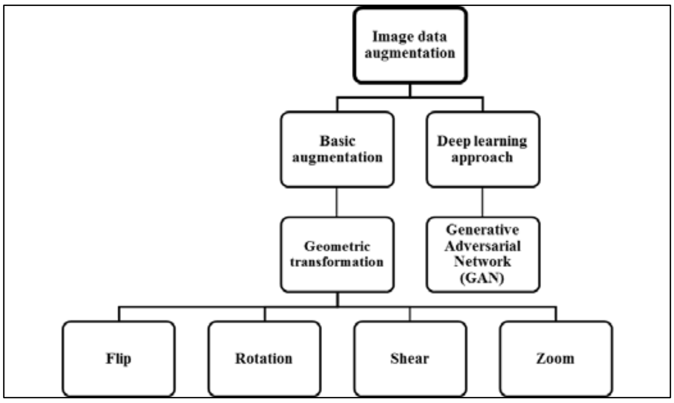

Fig. 2. Image data augmentation techniques

1) Basic augmentation

Geometric transformation is one technique for basic image data augmentation. Geometric transformation is used to transform the shape of an image by altering the pixel position while keeping the pixel values constant [17]. As illustrated on Fig. 2, the augmentation techniques used in this work are flip, rotation, shear and zoom. Flipping is the process of mirroring an image either horizontally or vertically, whilst rotation is the process of rotating images by a particular angle and through a specified axis of rotation. Furthermore, shearing involves shifting an image by an arbitrary amount either horizontally or vertically. Finally, zoom augmentation involves zooming in (enlarge the image) or out (shrink the image) [15].

2) Generative Adversarial Network (GAN) augmentation

GANs are a subtype of generative model framework. The generative model attempts to learn the data distribution implicitly from a series of samples (– for example, images) in order to generate additional samples based on the learnt distribution. We employed deep convolutional generative adversarial networks (DCGAN) suggested by Radford et al [18] in this study.

The model comprises a DCGAN architecture as depicted on Fig. 3 to create the model for both the generator and discriminator. The shape of the input images from the Kvasir dataset is set to 64 x 64. The layers of the Generator and Discriminator use Leaky ReLU activation with the exception of the output layers, which employ the Sigmoid activation function. The Generator and the Discriminator have been constructed using a convolutional neural network. In its network, The Discriminator has 0.2 dropout layers. The DCGAN is built using Adam optimisation with a learning rate of 0.0001 and Binary cross-entropy and was trained over 1000 epochs with a batch size of 40 images provided by the generator. The Generator generates fake images of the esophagitis, polyps and ulcerative colitis classes and gives it to the discriminator to identify which is fake and real.

D. Building classification model

VGG16 and ResNet50 architectures are selected as the base network for the classification purpose. Experiments are conducted on the original, basic augmented, and GAN augmented datasets. The network is trained 10 epochs for the two chosen deep CNNs. To update network weights, a batch size of 32 was selected for each epoch while using Adam optimization with categorical cross entropy as the loss function. One epoch is completed when the entire dataset has been traversed. Softmax is then used to classify the images into the 3 classes namely esophagitis, polyps, ulcerative colitis.

1) ResNet50

He et al [19] proposed the ResNet50 deep convolutional neural network which is a variant of the Resnet model. It consists of 48 Convolution layers along with 1 MaxPool stride size of 2, 1 Average Pool layer and a fully connected layer with softmax function. ResNet50 was trained on the ImageNet dataset. Similarly, this model was also run on the three types of datasets. The architecture of the ResNet50 is illustrated in Fig. 4.

2) VGG16

Simonyan and Zisserman [20] proposed the VGG16 deep convolutional neural network. It has 16 layers, 13 convolutional layers, 3 fully connected layers, and a final softmax classifier. 5 of the13 convolutional layers are max-pooling layers. VGG16 was trained on the ImageNet dataset and employs a 3x3 filter with stride 1 and the same padding for the convolution layers, as well as a 2x2 filter with stride 2 for the maxpool layer. The model was run on three types of datasets. The architecture of the VGG16 is illustrated in Fig. 4.

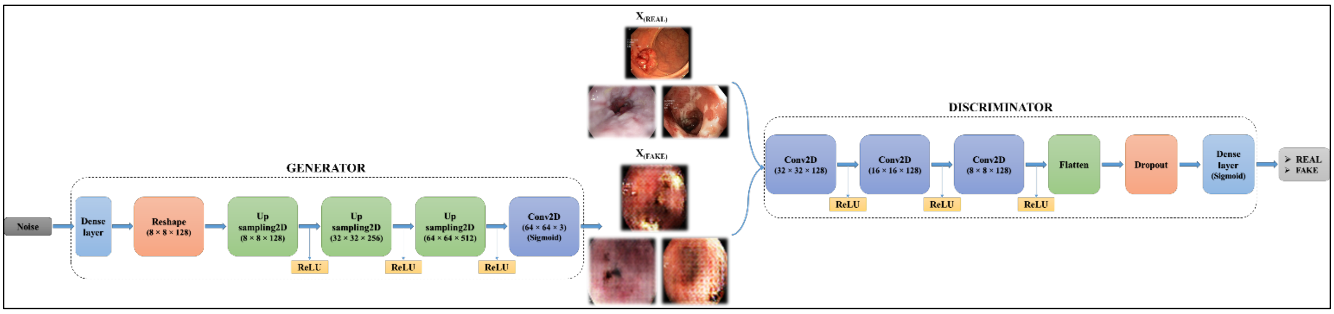

Fig. 3. Architecture of DCGAN

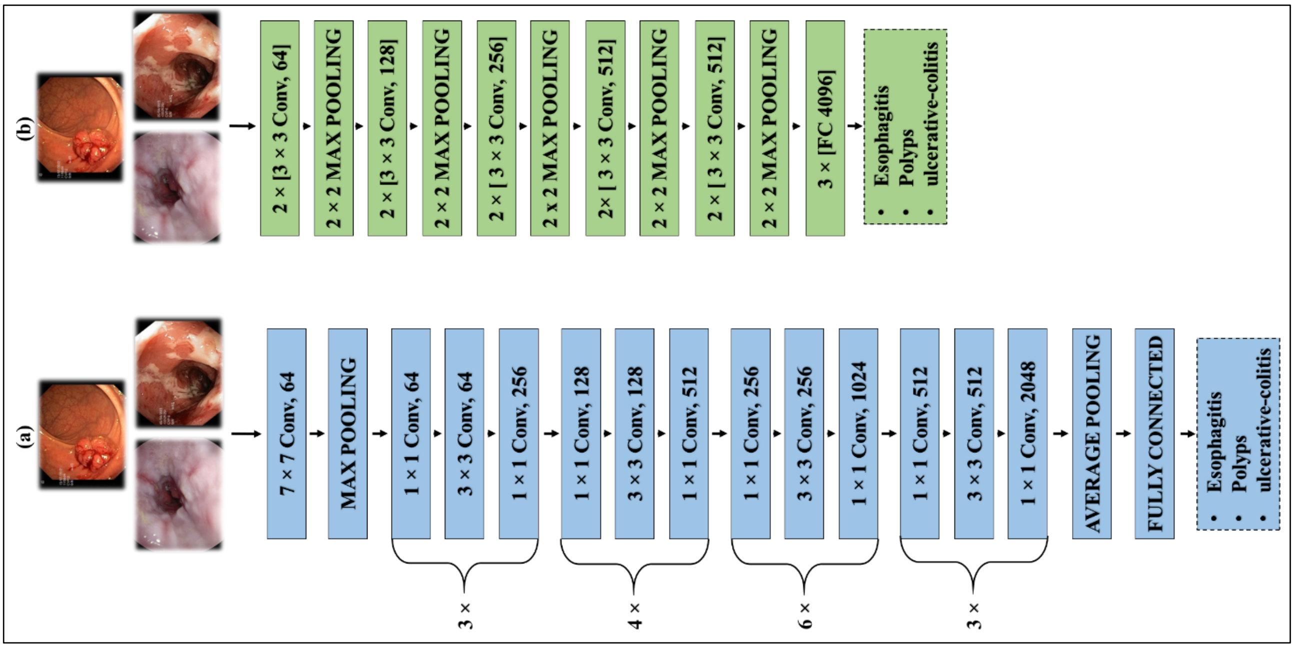

Fig. 4. Architecture of classification models. (a): ResNet50, (b): VGG16

IV. Results & Discussions

This section details the results from all the experiments conducted in this research work. First of all, basic data augmentation was applied on the images. Fig. 5 shows the operations of some basic data augmentation.

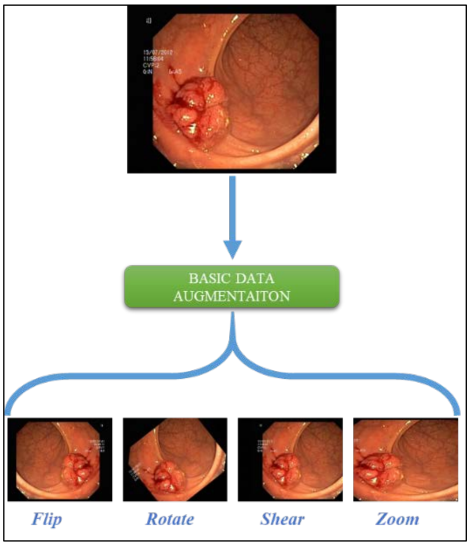

Fig. 5. Basic image augmentation applied on polyp's image

Thus, the 'Basic Augmentation' dataset was formed comprising of 6000 images representing the three classes namely eshophagitis, polyps and ulcerative colitis.

Next, GAN was applied on the original dataset Fig. 6 shows some resulting images of the application of GAN. Eventually the images were used to create the third dataset namely 'GAN augmentation' dataset.

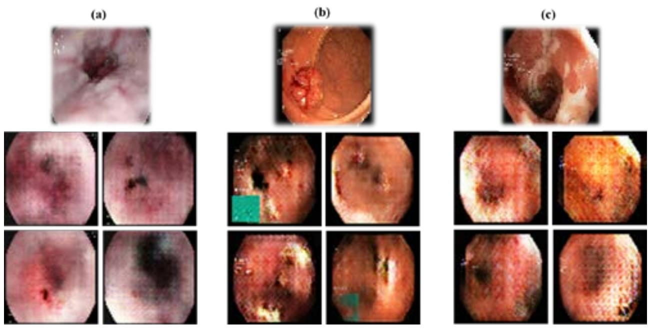

Fig. 6. GAN augmentation. (a): esophagitis, (b): Polyps, (c) ulcerative colitis

Furthermore, a fourth dataset comprising of 50% original and 50% GAN generated images was constructed to be used for the experiment. These four datasets have been used for training and to evaluate the performance of the models. As mentioned earlier, the main aim of the paper is to analyse the impact of data augmentation on the overall performance of the application. Therefore, the two deep learning models are applied on the four datasets. TABLE I. summarises the initial results obtained after the application of the two pre-trained models. The weights were adjusted accordingly. Preliminary studies were to determine whether the models generated satisfactory results on the training, validation, and testing sets. From the performance accuracy generated, it was observed that VGG16 is overfitting. In fact, after a given number of epochs, the validation accuracy began to decline. Similarly, for the Resnet50 model, the training and validation did not generate satisfactory accuracy. The testing performances were likewise poor.

Thus, to optimize the models, the original dataset was subjected to data augmentation. Secondly, Adam optimisation with categorical cross entropy as loss function is further applied to leverage the models. For the deep CNNs chosen, the network is trained across 10 epochs with a batch size of 50. Softmax is then employed for classification. Therefore, the two models are fine-tuned and applied on the Basic Augmentation', 'GAN augmentation' datasets. It was also applied on the fourth dataset consisting of images from the original dataset and that from the GAN datasets

TABLE I. PERFORMANCE OF THE RESNET50 AND VGG16 ON THE DATASETS

Model	**DATASET**	**Accuracy(%)**		
		Training	*Validation*	*Testing*
VGG16	Original	94.83	83.83	85.6
	Basic Augmentation	96.03	93.92	92.6
	GAN augmentation	96.03	93.75	94
	GAN augmentation-Original(50%)-GAN 50%)	92.67	88.75	89.7
Resnet50	Original	69.78	63.17	64.67
	Basic Augmentation	77.22	79.50	81.08
	GAN augmentation	80.67	80.75	82.3
	GAN augmentation-Original(50%)-GAN 50%)	77.03	72.42	74.67

The results demonstrate that data augmentation does have an impact on the overall performance of the esophagus cancer detection application. In addition, it can be seen that VGG16 performs better compared to ResNet50. This was also the case in a work conducted by Ismail and Sovuthy [21], where better performance was achieved for VGG16 over ResNet50 for the detection of breast cancer.

Fig. 7 illustrates a comparison on the testing performance of the two models based on the basic augmentation and GAN augmentation technique.

For the VGG16 enhanced model, an accuracy of 85.6%, 92.6%, 94% and 89.7% were achieved for the original dataset, basic augmented dataset, GAN augmented dataset and mixed dataset respectively.

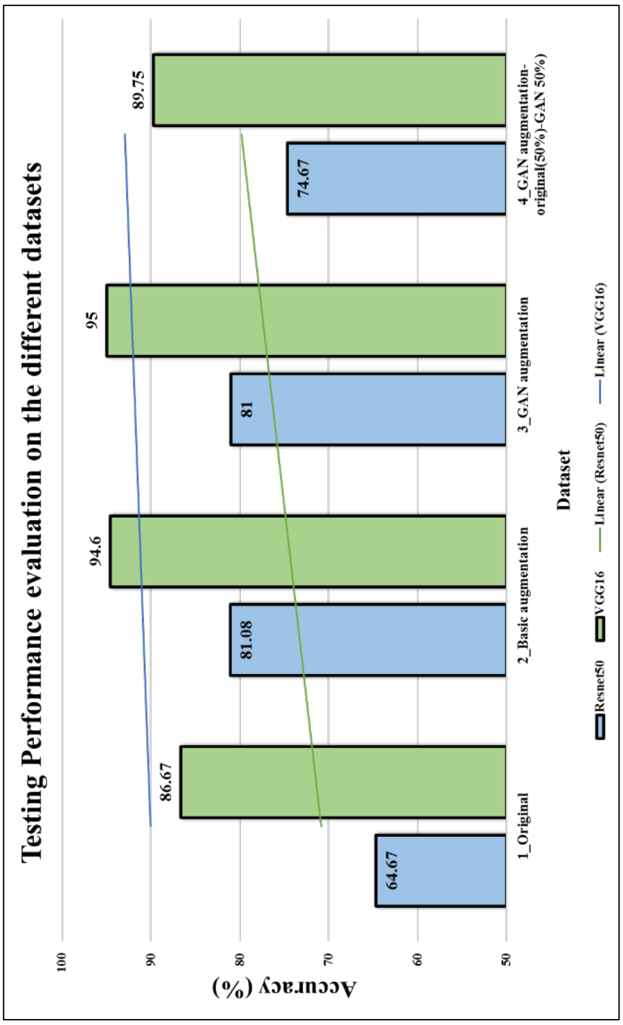

Fig. 7. Testing performance on the four datasets

V. CONCLUSION

Deep learning techniques have been employed in recent years to analyse medical images in a variety of domains, with outstanding results in applications such as image classification and detection. However, the limited number of annotated datasets hinder the performance and accuracy of the model. In some cases, models may be overfitting and thus, affect the whole system. To countermeasure the same problem in the classification of esophagus cancer, we have investigated the different data augmentation techniques. First of all, we have enhanced the pre-trained VGG16 and ResNet50 models and applied them to the original datasets. We have used basic data augmentation on our oesophagus training dataset. Eventually we have developed a GAN model to augment our data sets. The two enhanced deep learning models were applied on the original dataset, basic data augmentation dataset, the GAN dataset and eventually a mixture of original images and GAN images. The results were analysed and discussed. It can be concluded that data augmentation has an effect on the overall performance of the system. However, GAN performs better compared to the basic data augmentation techniques.

REFERENCES

[1] J. Huang *et al.*, 'Global Burden, Risk Factors, and Trends of Esophageal Cancer: An Analysis of Cancer Registries from 48 Countries', *Cancers (Basel)*, vol. 13, no. 1, p. 141, Jan. 2021, doi: 10.3390/cancers13010141.

[2] Mayo Clinic, ‘Esophageal cancer - Symptoms and causes - Mayo Clinic’, 2022. https://www.mayoclinic.org/diseases-conditions/esophageal-cancer/symptoms-causes/syc-20356084 (accessed Aug. 04, 2022).

[3] T. H. Khan, S. K. Mohammed, M. S. Imtiaz, and K. A. Wahid, ‘Color reproduction and processing algorithm based on real-time mapping for endoscopic images’, *SpringerPlus*, vol. 5, no. 1, p. 17, Jan. 2016, doi: 10.1186/s40064-015-1612-4.

[4] B. Dhanddra and R. Hegadi, ‘Abnormality Detection in Endoscopic Images using HSI Segmentation and Curvature Computation’, Jan. 2005.

[5] Markus, ‘Artificial Intelligence in Medicine | The Top 4 Applications’, 2020. https://www.datarevenue.com/en-blog/artificial-intelligence-in-medicine (accessed Apr. 17, 2022).

[6] L. Wu *et al.*, ‘A deep neural network improves endoscopic detection of early gastric cancer without blind spots’, *Endoscopy*, vol. 51, Mar. 2019, doi: 10.1055/a-0855-3532.

[7] Y. Mori, T. M. Berzin, and S.-E. Kudo, ‘Artificial intelligence for early gastric cancer: early promise and the path ahead’, *Gastrointest Endosc*, vol. 89, no. 4, pp. 816–817, Apr. 2019, doi: 10.1016/j.gie.2018.12.019.

[8] M. I. Razzak, S. Naz, and A. Zaib, ‘Deep Learning for Medical Image Processing: Overview, Challenges and the Future’, in *Classification in BioApps: Automation of Decision Making*, N. Dey, A. S. Ashour, and S. Borra, Eds. Cham: Springer International Publishing, 2018, pp. 323–350. doi: 10.1007/978-3-319-65981-7_12.

[9] W. Xia, E. C. S. Chen, and T. Peters, ‘Endoscopic image enhancement with noise suppression’, *Healthc Technol Lett*, vol. 5, no. 5, pp. 154–157, Sep. 2018, doi: 10.1049/htl.2018.5067.

[10] M. H. A. Janse, F. van der Sommen, S. Zinger, E. J. Schoon, and P. H. N. de With, ‘Early esophageal cancer detection using RF classifiers’, vol. 9785, p. 97851D, Mar. 2016, doi: 10.1117/12.2208583.

[11] C. S. Bang, J. J. Lee, and G. H. Baik, ‘Computer-aided diagnosis of esophageal cancer and neoplasms in endoscopic images: a systematic review and meta-analysis of diagnostic test accuracy’, *Gastrointest Endosc*, vol. 93, no. 5, pp. 1006-1015.e13, May 2021, doi: 10.1016/j.gie.2020.11.025.

[12] J. H. Lee *et al.*, ‘Spotting malignancies from gastric endoscopic images using deep learning’, *Surg Endosc*, vol. 33, no. 11, pp. 3790–3797, Nov. 2019, doi: 10.1007/s00464-019-06677-2.

[13] T. Xiao, S. Renduo, W. Lianlian, and Y. Honggang, ‘AN AUTOMATIC DIAGNOSIS SYSTEM FOR CHRONIC ATROPHIC GASTRITIS UNDER WHITE LIGHT ENDOSCOPY BASED ON DEEP LEARNING’, in *Endoscopy*, Apr. 2022, vol. 54, no. S 1, p. OP186. doi: 10.1055/s-0042-1744749.

[14] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, ‘GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification’, *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.

[15] C. Shorten and T. M. Khoshgoftaar, ‘A survey on Image Data Augmentation for Deep Learning’, *J Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[16] S. D. Desai, S. Giraddi, N. Verma, P. Gupta, and S. Ramya, ‘Breast Cancer Detection Using GAN for Limited Labeled Dataset’, in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2020, pp. 34–39. doi: 10.1109/CICN49253.2020.9242551.

[17] D. S. Mathur, D. A. Mathur, and N. Agarwal, ‘A Review of basic Mathematical Transformation used in Image Processing’, *International Journal of Engineering Research & Technology*, vol. 4, no. 12, Apr. 2018, Accessed: May 24, 2022. [Online]. Available: https://www.ijert.org/research/a-review-of-basic-mathematical-transformation-used-in-image-processing-IJERTCONV4IS12009.pdf, https://www.ijert.org/a-review-of-basic-mathematical-transformation-used-in-image-processing

[18] A. Radford, L. Metz, and S. Chintala, ‘Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks’. arXiv, Jan. 07, 2016. doi: 10.48550/arXiv.1511.06434.

[19] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’. arXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385.

[20] K. Simonyan and A. Zisserman, ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: Apr. 29, 2022. [Online]. Available: http://arxiv.org/abs/1409.1556

[21] N. S. Ismail and C. Sovuthy, ‘Breast Cancer Detection Based on Deep Learning Technique’, in *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)*, Aug. 2019, pp. 89–92. doi: 10.1109/EnCon.2019.8861256.