International Journal on Engineering Technology (InJET)

https://kec.edu.np/kec-journal/
Volume 2, issue no. 1, Dec 2024, Page 29-37
Received Date: July 15, 2024
Accepted Date: Oct 17, 2024
ISSN: 3021-940X (Print)
DOI: https://doi.org/10.3126/injet.v2i1.72489

Enhanced Ulcer Detection Using GAN-Augmented Image Processing and Convolutional Neural Networks

Mohan Bhandari[1, *], Darshan Bhusal[2], Kishor Singh[2], Bishnu Bashyal[2]

[1]*Samriddhi College, Lokanthali, Bhaktapur, Nepal*
[2]*Nepal College of Information Technology, Lalitpur, Bagmati, Nepal*

Abstract

Ulcers, especially chronic wounds and pressure sores, create significant healthcare challenges and need early detection for effective treatment. Due to the persistent challenges posed by human error in manual detection and classification processes, there is a need for automated methodologies that can offer robust and reliable solutions. To enhance the accuracy and generalization of the Convolutional Neural Network (CNN), a Generative Adversarial Network (GAN) based image augmentation technique is used in this study to classify the Kvasir dataset. With an initial 1512 ulcer images selected from the dataset, after stable 70 epochs, 70000 images were generated using GAN for 2 classes: Ulcer or Non-Ulcer. To remove the noise from an enhanced batch of images and classify the set of newly generated datasets, GAN-based CNN (G-CNN) was employed to get 99.00% training and 96.04% validation accuracy. The study was compared with conventional CNN which achieved a training accuracy of 98.5% and validation accuracy of 94.34%. A significant improvement has been observed in f1-scores of 0.94 and 0.97 for CNN and the proposed algorithm, respectively. The study is replicable across datasets with a limited number of images, thereby facilitating heightened accuracy levels.

Keywords: Denoising Convolutional Neural Network, Generative Adversarial Networks, Ulcer Classification

1. Introduction

The ulcer has become one of the most fatal diseases, especially in developing countries, and if it is not addressed on time, it may result in death (Ding et al., 2024). The rectification of image processing helps to resolve a wide variety of issues, particularly those related to medical imaging for brain tumors, ulcers, breast cancer, and other conditions (Rajinikanth et al., 2021). Automated early detection and classification techniques can enhance reliable diagnosis and therapy. Deep learning (DL) has established itself as an effective machine learning technology and has shown that it is capable of automating the diagnosis of diseases (Ajagbe and Adigun, 2024). As a result, its usage in medical picture analysis and recognition has been investigated.

Strengthening the skills of physicians and other health-care workers associated with preliminary diagnosis can improve medical image examination. Conventional augmentation methods and enhancement techniques, such as cropping, flipping, rotating, and translating, may not adequately capture the intricate complexity and variability inherent in ulcer characteristics (Xu et al., 2024). As a promising alternative, leveraging Generative Adversarial Network (GAN)based image augmentation offers a solution to overcome the limitations of scarce and diverse ulcer datasets (Thakur et al., 2023).

By generating synthetic ulcer images with realistic characteristics, the proposed study improves the generalization capabilities of the DL model. This ensures that the model can accurately detect ulcers in unseen cases. Addressing the critical need for accurate ulcer detection in medical diagnostics, where traditional methods may struggle due to poor image quality and limited data (Saeedi et al., 2024). Utilizing GANs for

**Corresponding Author*

image augmentation and Convolutional Neural Network (CNN) for denoising can significantly enrich the training dataset and enhance image clarity, leading to improved diagnostic precision and the potential for better patient outcomes.

The resolve the current issues, this study focus:

a. To construct a hybrid learning model (G-CNN), integrating Generative Adversarial Networks (GAN) for generating additional data samples
b. To evaluate the proposed model through comparative analysis using ulcer imaging datasets.

2. Literature Review

Gastric ulcer detection during clinical interventions presents challenges, including diagnosis time and the need for expertise.

A study by Bajhaiya and Unni (2022) explores automatic ulcer detection from wireless capsule endoscopic (WCE) images using transfer learning. DenseNet121, a pre-trained convolutional neural network, is employed for classification, achieving 99.94% accuracy, 100% precision, 97.67% recall, and 98.82% F1-score on the test dataset. These results demonstrate the efficacy of deep learning for rapid and accurate gastric ulcer screening, addressing challenges associated with traditional diagnostic methods.

Yang et al. (2018) proposed a new Computed Tomography (CT) image denoising method using GANs with Wasserstein distance and perceptual similarity. This approach, combining optimal transport theory and perceptual loss, aims to reduce noise while preserving critical information. Experimental results with clinical CT images demonstrate promising outcomes, highlighting the potential of this method.

Auzine et al. (2022) presented a method for identifying ulcers in WCE images by employing a cutting-edge CNN. Improving ulcer detection in WCE images involves employing convolutional layers with pooling layers, two dense layers, and ReLu activation. This study explores different types of ulcers using a large database of WCE images. Optimal performance is achieved by carefully analyzing hyperparameters like batch size, optimizer with learning rate, pooling size, epochs, and dropout. The results indicated that the accuracy of the models improves significantly with data augmentation and GAN. Particularly, GAN achieved the highest accuracy of 94%, surpassing both the non-augmented training set and traditional data augmentation methods for the VGG16 model.

In the study, Klang et al. (2021) illustrated the application of a CNN model for automated assessment of stomach ulcer images to detect malignancy. Their dataset, sourced from Sheba, consisted of 1978 images of gastric ulcers (GU), comprising 1894 benign GU images and 84 malignant ulcer images. The final CNN model achieved an Area Under the Curve (AUC) of 0.91 for detecting malignant ulcers. At a cut-off probability of 0.5, the network demonstrated a specificity of 75% and sensitivity of 92% for identifying malignant ulcers.

Sutton et al. (2022) evaluated various CNN architectures using a dataset comprising 8000 tagged endoscopic still images sourced from hyper-kvasir, the largest multiclass gastrointestinal tract image and video dataset available. They initialized the weights using ImageNet and employed Grid Search to determine optimal hyperparameters through fivefold cross-validation. The DenseNet121 architecture yielded the highest accuracy of 87.50% and AUC of 0.90, outperforming the majority class prediction of “no skill” model which achieved only 72.02% accuracy and an AUC of 0.50.

Tekchandani et al. (2020) address critical challenges in mediastinal lymph node (MLN) classification in lung cancer, where invasive tests carry significant risks and depend heavily on surgical expertise. They emphasize the need for a non-invasive, computer-aided system, highlighting previous work using fully convolutional networks (FCNs) and data augmentation to improve performance. However, FCNs often struggle with overfitting due to high parameter counts. To mitigate this, the authors propose using Generative Adversarial Networks (GANs) for data augmentation and an Inception-based model for efficient hierarchical feature extraction. Their approach achieves outstanding accuracy (94.95%), sensitivity (93.65%), and specificity

(96.67%), demonstrating its effectiveness. Barua et al. (2024) investigated multi-class classification of Corneal Ulcers using deep learning methods on the SUSTech-SYSU dataset, comprising 712 images from Sun Yat-sen University's Zhongshan Ophthalmic Center. Post fluorescein staining, the images are used to enhance CNN models for ECU image classification. A customized hierarchical model is developed, achieving 99% training accuracy and 90% validation accuracy. The dataset is expanded to 7,200 training, 3,000 testing, and 1,800 validation images. Employing categorical cross-entropy loss and the Adam optimizer, the model undergoes hyper-parameter tuning to maximize performance. This work seeks to advance automated Corneal Ulcer classification, aiding ophthalmologists in effective diagnosis and treatment. Table 1. shows the overall comparative study of current state of art methods.

Table 1: List of different SOTA methods

Author	Nature of Dataset	Algorithm Used	Result
Bajhaiya and Unni (2022)	WCE images	DenseNet 121	Accuracy: 99.94% Recall: 97.67%
Auzine et al. (2022)	WCE images	VGG16	Accuracy: 94%
Klang et al. (2021)	Sheba stomach ulcer dataset	CNN	Specificity: 75% Sensitivity: 92% AUCROC: 0.91
Sutton et al. (2022)	Hyper-Kasvir gastrointestinal dataset	DenseNet121	Accuracy: 87.50% AUC: 0.90
Barua et al. (2024)	SUSTech-SYSU dataset	CNN	Accuracy: 99%

Compared with all the currents state-of-art methods, the major problems in traditional CNNs show limited performance in real-world applications due to dataset variety constraints and image noise. Combining GAN-based image augmentation to create realistic synthetic images and Denoising CNNs to reduce noise in real images could significantly enhance model accuracy and robustness, addressing the limitations of conventional CNNs in diverse imaging conditions.

3. Materials and Methodology

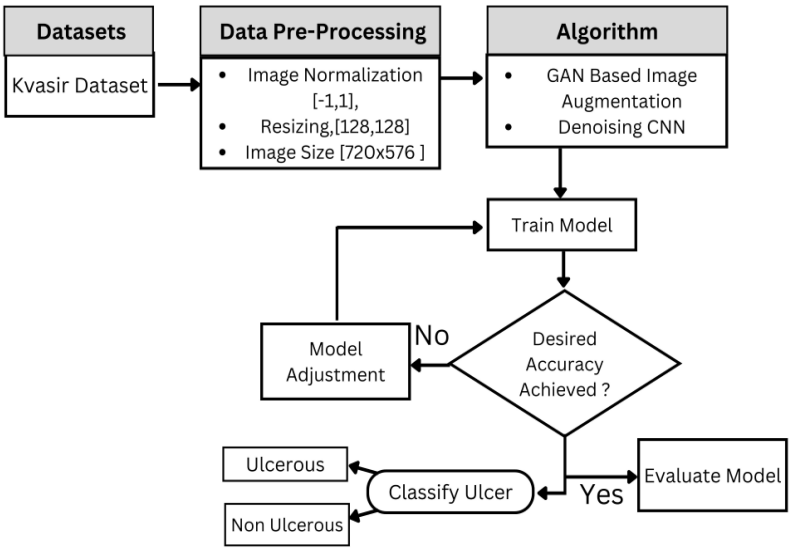


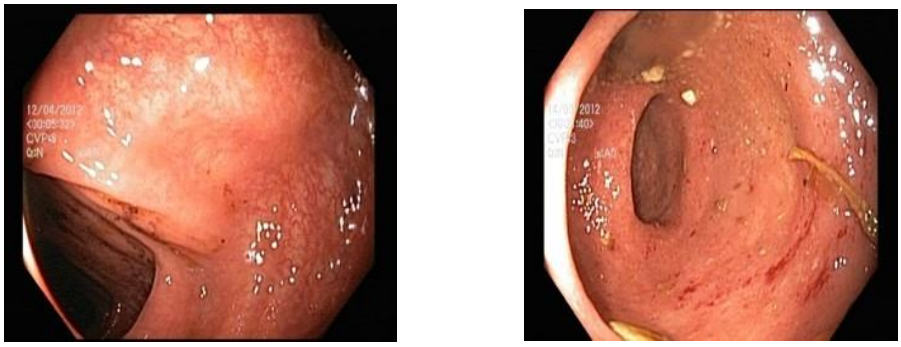
Figure 1: Methodology

The entire process of the study is shown in Figure 1. This study uses the Kvasir dataset to develop a model for classifying ulcerous and non-ulcerous images. The methodology includes image normalization to the range [−1, 1] and resizing to [128×128] pixels for uniform input. To improve model performance, GANs are used for image augmentation, enhancing the diversity of training data. Additionally, a denoising Convolutional Neural Network (CNN) is applied to reduce image noise and retain crucial features. The model

undergoes an iterative training process, involving accuracy evaluation and subsequent adjustments until the desired performance is achieved. The final model is evaluated and used for accurate ulcer classification.

3.1. Dataset

KVASIR dataset (Pogorelov et al., 2017) is used as the primary source of the dataset. A total of 1000 images were used for training whereas 512 images were used for testing. Ulcerous images were collected from the Ulcerative Colitis category of the KVASIR dataset, while normal images were collected from the Cecum category. The sample image for both the categories is shown in Figure 3.



(a) (b)

Figure 2. Sample images for each of the classes. Note that (a) and (b) denote "Ulcer" and "Normal" cases respectively.

3.2. Dataset Preprocessing

The images in the dataset hold images in the range 720 × 576 up to 1920 × 1072. All the images under consideration were resized to 128 × 128 and normalized to fall in the range of [-1, 1].

3.3. Data Augmentation Using GAN

GAN, having generator and discriminator, is trained to create realistic samples and identify them as real or generated for KVASIR dataset in our study. Loss function used in our study is defined as:

$$\text{MinG MaxD V (D, G)} = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad \text{(i)}$$

Here, $E_{x\sim p_{data}(x)}$ is expectation over the real data distribution for KVASIR dataset, $E_{z\sim p_z(z)}$ is expectation over the input noise variables z to the generator, D(x) is discriminator's estimate of the probability that real data instance x is real, G(z) is data generated by the generator from noise z. and D(G(z)) is the discriminator's estimate of the probability that a fake instance is real (Creswell et al.).

3.3.1. Generative Architecture

The GAN generator consists of four transpose convolutional layers with kernel sizes of (4, 4) and Leaky ReLU activation functions. It starts with a dense layer taking a noise vector of length 80. The final layer outputs a 3-channel image with tanh activation. The generative model summary is shown in Table 2.

Table 2. Layer summary of Generative Network

Layer (type)	**Output Shape**	**Param #**
Dense	(None, 16384)	1,327,104
Reshape	(None, 8, 8, 256)	0
Conv2DTranspose	(None, 16, 16, 128)	524,416
Leaky ReLU	(None, 16, 16, 128)	0
BatchNormalization	(None, 16, 16, 128)	512

Conv2DTranspose	(None, 32, 32, 64)	131,136
Leaky ReLU	(None, 32, 32, 64)	0
BatchNormalization	(None, 32, 32, 64)	256
Conv2DTranspose	(None, 64, 64, 32)	32,800
Leaky ReLU	(None, 64, 64, 32)	0
BatchNormalization	(None, 64, 64, 32)	128
Conv2DTranspose	(None, 128, 128, 3)	1,539
Total params: 2,017,891		
Trainable params: 2,017,443		
Non-trainable params: 448		

3.3.2. Discriminative Architecture

The discriminator is a sequential model with two convolutional layers: 64 filters of size (3, 3) and 128 filters of size (3, 3). Leaky ReLU activation functions with alpha=0.2 and dropout layers are applied. It ends with a dense layer using sigmoid activation. The discriminator architecture summary is shown in Table 3. Using the GAN architecture, a total of 70000 images were generated.

Table 3. Layer summary of Discriminator Network

Layer (type)	**Output Shape**	**Param #**
Conv2D	(None, 128, 128, 64)	1,792
Leaky ReLU	(None, 128, 128, 64)	0
Dropout	(None, 128, 128, 64)	0
Conv2D	(None, 128, 128, 128)	73,856
Leaky ReLU	(None, 128, 128, 128)	0
Dropout	(None, 128, 128, 128)	0
Flatten	(None, 2097152)	0
Dense	(None, 1)	2,097,153
Total params: 2,172,801		
Trainable params: 0		
Non-trainable params: 2,172,801		

3.4. Conventional CNN

Table 4 outlines a CNN model architecture, listing each layer's type, output shape, and parameter count. It includes three Conv2D layers (32, 64, and 128 filters), each followed by MaxPooling2D layers. A flattened layer converts the output to a 1D vector (25,088 units). This is followed by a Dense layer (128 units), a Dropout layer, Batch normalization, and a final Dense layer for classification. The model has 3,304,281 parameters in total.

Table 4: Layer Summary of CNN model

Layer (type)	**Output Shape**	**Parameter #**
Conv2D	(None, 126, 126, 32)	896
MaxPooling2D	(None, 63, 63, 32)	0
Conv2D	(None, 61, 61, 64)	18,496
MaxPooling2D	(None, 30, 30, 64)	0
Conv2D	(None, 28, 28, 128)	73,856
MaxPooling2D	(None, 14, 14, 128)	0
Flatten	(None, 25,088)	0
Dense	(None, 128)	3,211,392
Dropout	(None, 128)	0
Batch Normalization	(None, 128)	512
Dense	(None, 1)	129

3.5. Evaluation Metrics

Different evaluation metrics, including accuracy, loss, classification report, and confusion matrix, are computed to assess the model's performance

3.6. Implementation Details

The experiment is performed using an NVIDIA Tesla T4 GPU equipped with 16GB of GDDR6 RAM in Google Colab.

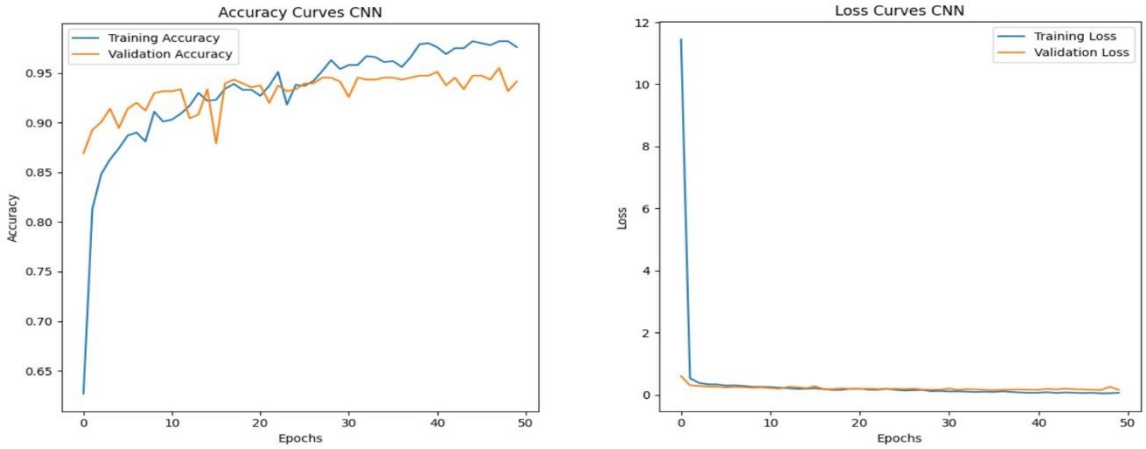
4. Results Analysis

4.1. Conventional CNN

Normal CNN model without augmentation was trained. The loss and accuracy metrics for training are 0.0456 and 98.50%, respectively. For validation, the loss and accuracy metrics are 0.2399 and 94.34%, respectively as shown in Figure 3.

According to Table 5, the Non-Ulcerous class has a precision of 0.91, recall of 0.98, and F1-score of 0.95, while the Ulcerous class has a precision of 0.98, recall of 0.90, and F1-score of 0.94. Overall accuracy is 0.94.

For a total 462 test samples, 228 ulcerous and 208 non-ulcerous images were correctly predicted while rest 26 images were incorrectly predicted as shown in Table 6.



(a) (b)

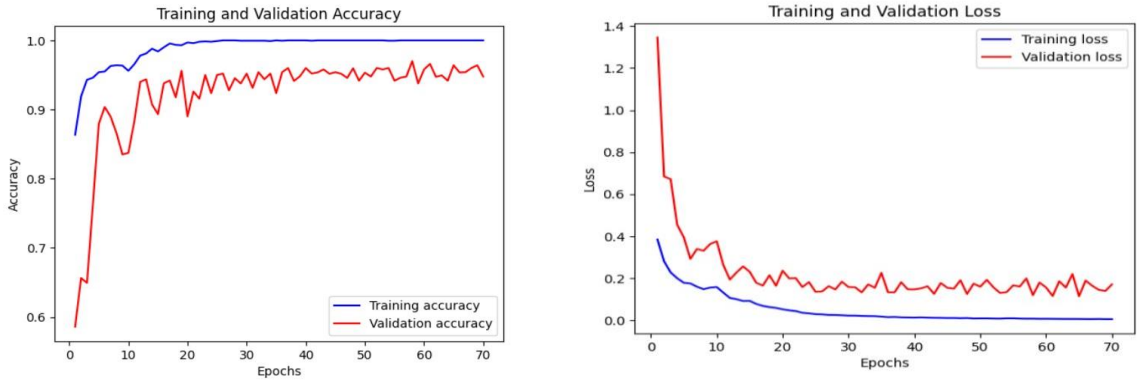
Figure 3: Accuracy and Loss Graph for Conventional CNN. Note that (a) denotes training and validation accuracy whereas (b) denotes training and validation loss

Table 5: Classification Metrics for CNN

Class	**Precision**	**Recall**	**F1-Score**
Non-Ulcerous	0.91	0.98	0.95
Ulcerous	0.98	0.90	0.94
Accuracy			0.94
Macro Average	0.95	0.94	0.94
Weighted Average	0.95	0.94	0.94

Table 6: Confusion Matrix for CNN

		Predicted	
		Ulcerous	Non-Ulcerous
Actual	Ulcerous	228	4
	Non-Ulcerous	22	208



(a) (b)

Figure 4.: Accuracy and Loss Graph for Conventional CNN. Note that (a) denotes training and validation accuracy whereas (b) denotes training and validation loss

4.2. GAN based CNN

Throughout the training process, the model consistently improved with a training loss of just 0.04 and an impressive accuracy of 99.5%. In validation, it showed a bit more loss at 0.1221 but still performed admirably well with an accuracy of 96.04% as shown in Figure 4. Additionally, starting from the 70th epoch, a total of 70,000 images were generated. Table 7 displays classification metrics for GAN-based CNN. The non-ulcerous class achieved precision, recall, and an F1-score of 0.96, 0.98, and 0.97, respectively. ulcerous class showed precision, recall, and F1-score of 0.98, 0.96, and 0.97. Overall accuracy reached 97.0%, with macro and weighted averages also at 97.0%.

Table 7: Classification metrics for GAN-based CNN

Class	**Precision**	**Recall**	**F1-Score**
Non-Ulcerous	0.96	0.98	0.97
Ulcerous	0.98	0.96	0.97
Accuracy			0.97
Macro Average	0.97	0.97	0.97
Weighted Average	0.97	0.97	0.97

Table 8: Confusion Matrix for GAN based CNN

		Predicted	
		Ulcerous	Non-Ulcerous
Actual	Ulcerous	226	5
	Non-Ulcerous	10	221

The confusion matrix (Table 8) shows that out of 462 total images, 447 were correctly classified by the conventional CNN, resulting in an overall accuracy of approximately 96.74%.

5. Discussion

512 images were first taken as testing datasets and 1000 images as training datasets. The total number of images created during the 70 epochs training phase is 70,000. GCNN receives the enhanced collection of images to filter out any unwanted noise and categorize the images as ulcer or non-ulcer. Using the Kvasir dataset, 96.04% validation accuracy and 99.5% training accuracy were attained. The accuracy of 94.34% was attained by the traditional CNN in comparison to the suggested approach. F1-scores of 0.94 and 0.97 for CNN and the suggested method, respectively, show a notable improvement. Ulcer Class maintains better precision, recall and f1-score for around 0.94, while non-ulcerous exhibits a slightly lower value of 0.95. By using GAN, with an overall accuracy of 97.00%, the model is effective in correctly classifying cases for precision, recall, and f1-score. These outcomes demonstrate the GAN's suitability for challenging classification tasks.

6. Conclusions

This study focuses on ulcer diagnosis in the Kvasir dataset by integrating GAN for augmentation. The G-CNN architecture is proposed and fine-tuned for optimal performance, adjusting hyperparameters such as epochs, batch size, optimizer with learning rate, dropout, and pooling size. The proposed algorithm achieves a remarkable 96.04% validation accuracy with high precision, recall, and F1 scores, underscoring the efficacy of data augmentation for small datasets.

Future research should explore GAN performance across diverse domains and data types, and efforts to enhance the robustness and interpretability of GAN-generated classifications are warranted to increase practical utility and trustworthiness. Investigating methods to mitigate biases and uncertainties in GAN-based classifications is essential for advancing their application in real-world scenarios.

References

Ajagbe, S. A., & Adigun, M. O. (2024). Deep learning techniques for detection and prediction of pandemic diseases: A systematic literature review. Multimedia Tools and Applications, 83(2), 5893–5927.

Auzine, M. M., Khan, M. H.-M., Baichoo, S., Sahib, N. G., Gao, X., & Bissoonauth-Daiboo, P. (2022). Endoscopic image analysis using deep convolutional gan and traditional data augmentation. 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 1–6.

Bajhaiya, D., & Unni, S. N. (2022). Deep learningenabled classification of gastric ulcers from wireless capsule endoscopic images. Medical Imaging 2022: Digital and Computational Pathology, 12039, 352–356.

Barua S, Saha S, Bulbuli J, Rahman A, Fuad MI, Dofadar DF, Rahman R. Unlocking Diagnosis Potential: CNN in Multi-Class Classification of Corneal Ulcer. In2024 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) 2024 Jun 29 (pp. 385-390). IEEE.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE signal processing magazine, 35(1), 53-65.

Ding, H., Yu, Z., Yao, H., Xu, X., Liu, Y., & Chen, M. (2024). Global burden and trends of disabilityadjusted life years and mortality for decubitus ulcer: A systematic analysis. International Wound Journal, 21(2), e14604.

Klang, E., Barash, Y., Levartovsky, A., Barkin Lederer, N., & Lahat, A. (2021). Differentiation between malignant and benign endoscopic images of gastric ulcers using deep learning. Clinical and experimental gastroenterology, 155–162.

Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., DangNguyen, D.-T., Lux, M., Schmidt, P. T., Riegler, M., & Halvorsen, P. (2017). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. Proceedings of the 8th ACM on Multimedia Systems Conference, 164–169. https: //doi.org/10.1145/3083187.3083212

Rajinikanth, V., Priya, E., Lin, H., & Lin, F. (2021). Hybrid image processing methods for medical image examination. CRC Press.

Saeedi, M., Gorji, H. T., Vasefi, F., & Tavakolian, K. (2024). Federated vs. central machine learning on diabetic foot ulcer images: Comparative simulations. IEEE Access.

Sutton, R. T., Zaiane, O. R., Goebel, R., & Baumgart, D. C. (2022). Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images. Scientific reports, 12(1), 2748.

Tekchandani H, Verma S, Londhe N. Performance improvement of mediastinal lymph node severity detection using GAN and Inception network. Computer Methods and Programs in Biomedicine. 2020 Oct 1;194:105478.

Thakur, R. S., Chatterjee, S., Yadav, R. N., & Gupta, L. (2023). Medical image denoising using convolutional neural networks. In Digital image enhancement and reconstruction (pp. 115–138). Elsevier.

Xu, Y., Zhi, C., Wang, S., Chen, J., Sun, R., Dong, Z., & Yu, L. (2024). Fabricgan: An enhanced generative adversarial network for data augmentation and improved fabric defect detection. Textile Research Journal, 00405175241237479.

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., & Wang, G. (2018). Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE transactions on medical imaging, 37(6), 1348–1357.