Antral Ulcer Detection Using Deep Learning

Deepa Kandel
Presidential Graduate School
Westcliff University, USA
deepa.kandel@presidential.edu.np

Bishwas Sharma
Kendriya Vidhalaya
Kathmandu, Nepal
bishwassharmavii@gmail.com

Suresh Pokharel
Presidential Graduate School
Kathmandu, Nepal
suresh.pokharel@presidential.edu.np

Vivek Sharma
Grande International Hospital
Dhapasi, Kathmandu
dr_svivek@hotmail.com

***Abstract*—An ulcer is a break on the skin, in the lining of tissues, or on an organ. An antral ulcer is found in the antrum part of the stomach. An antral ulcer can have many complications, such as bleeding, perforation, obstruction, and even cancer, if not treated properly on time.**

In the remote part of Nepal, antral ulcer diagnosis is very challenging due to the scarcity of trained gastroenterologists. Most of the patients visit higher centers, which are located only in urban cities, for proper management. Those patients who cannot afford higher centers, they just take the medicine without proper diagnosis. Those who visit the higher center present very late, so there are more complications of disease. Among them, a few complications lead to serious problems, even the death of the patients.

This research study has been conducted to fulfill this huge gap. Hence, AI-based applications help to eliminate delayed management and early diagnosis and treatment of the patient. It focuses on having a quick diagnosis of antral ulcer to have fast treatment that prevents the serious or complex issue.

Various deep learning models based on Convolutional Neural Networks (CNNs) are used, such as ResNet50, VGG16, MobileNetV2, Inception-ResNet, and CNN. Here, a total of 967 images have been used to train the model, including 485 normal and 482 ulcers. The performance of ResNet50 shows better as compared to other applied models, having a validation accuracy of 96 percent, with precision, recall, and F1-scores all above 96 percent.

Keywords- AI-Based Diagnostic of Antral Ulcer, Automated Antral Ulcer Detection, Gastrointestinal Ulcer Detection using AI

I. Introduction

Antral ulcer is a major gastrointestinal disease that is increasing globally. If it is not cured on time, it can cause serious complications such as gastrointestinal bleeding, perforation and obstruction [1]. It may cause cancer if not treated on time. It should be timely intervention and medication to avoid life-threatening complications, optimize patient care, and streamline health sectors [2].

In remote hospitals of Nepal, proper diagnosis and treatment of antral ulcer is challenging because of lack of specialist physicians and equipments [3]. If not properly managed on time it may lead to so many life threatening complications. In Urban hospitals, there are high volumes of patients, and there are limited resources available even in big center so there may be chances of misdiagnosis and proper management of patients on time. This process is associated with delays, high risks, and inconsistency [4]. Therefore, this AI-powered application seeks to detect ulcers of the antral part in real time with early management and diagnosis of the patients.

The main objective of this research is to have early diagnosis and treatment of the patient on time. More specifically, this research aim to develop a diagnostic model based on AI (deep learning based methods) to detect antral ulcers through endoscopic images [5]. The results may aid scientists and professionals in creating more effective strategies for promoting the application to have medication for similar types of other healthcare problems as well. This study focuses on early diagnosis and proper management of antral ulcer in remote area of Nepal so we can reduce mortality and morbidity which are overlooked by the previous researchers [6]. Hence, this project aims to have early diagnosis and treatment for the patient.

Recent work in AI and in deep learning has demonstrated its capability in analyzing medical images. It has use for specific purposes, pattern recognition, classification, and segmentation tasks involving convolution neural networks (CNNs) [7]. It has seen the usage of AI automation rapidly growing in various fields, such as in dermatology, radiology, ophthalmology, and gastroenterology.

II. Literature Review

Deep learning has transformed medical image analysis by providing powerful computer-aided diagnosis and clinical decision assistance tools. Use of these methods has been effective for detection of a variety of medical conditions like ulcers. such as ulcerative colitis (UC), Crohn's disease and all types of ulcers. UC, Crohn's disease, and antral ulcers are all diseases that are related to the internal abdomen. It may have inflammatory or structural abnormalities in the stomach that can be detected through endoscopic images to determine if it has ulcers or not. Finally, these developments demonstrate that the use of AI in image-based analysis for the diagnosis of disease reduces clinician workload and accelerates early detection in healthcare systems.

Khorasani et al. [8] explained that, with advancements in technology, machine learning (ML) techniques are increasingly being used to process high-dimensional data and identify complex patterns. Ulcerative colitis (UC) is described as an inflammatory disease of the colorectal mucosa, characterized by periods of remission and active inflammation. A machine learning pipeline incorporating feature selection was implemented to identify UCs. Multiple datasets of gene expression have been taken, and dimensionality reduction using the projection technique for identifying 32 discriminative genes. It has applied support vector machine (SVM) to train a classifier that had a perfect classification of active UC cases.

Raju et al. [9] investigated colorectal cancer, reporting that it ranks third in prevalence and affects both genders. It is one of the most prevalent and deadliest malignancies in the world. The risk factor associated with colorectal cancer is the formation of polyps, those growths of the mucosa that may develop into cancer without any intervention. The existing clinical practice is based on endoscopic examination. It has been mentioned that gastroenterologists use their hands to observe the colon and retrospectively analyze obtained images to determine and count polyps. The study employs a region-based convolutional neural network (R-CNN) to classify and segment polyps with high performance. The research has explained about the work of Kempegowda Institute of Medical Sciences (KIMS), Bangalore, India. It is the first open-source endoscopic dataset with both polyps (180 images) and ulcers (232 images) that has been tested on R-CNN models using the Detectron2 framework in a dataset with Intersection over Union (IoU) scores reaching 0.63.

Lin et al. [10] have conducted research to detect gastrointestinal (GI) diseases, including bleeding detection. Gastrointestinal bleeding is associated with conditions such as peptic ulcers, colorectal cancer, and other digestive system pathologies, where timely intervention and prompt medication can improve patient outcomes. Wireless Capsule Endoscopy (WCE) generates large quantities of GI images, enabling the development of automated models for both classification and localization of bleeding. Multi-Task Learning (MTL) models have been applied to classification, detection, and segmentation for having high optimization and minimizing computational costs. MTL has also been utilized in medical imaging applications such as lesion detection, tumor segmentation, and disease classification. As per the past few research mentioned, various techniques are used to enhance the strong performance of deep learning models in medical imaging. Stochastic weight averaging stabilizes convergence and improves generalization by averaging model weights during training. Use of Test-Time Augmentation (TTA) helps to increase inference robustness by averaging predictions over augmented test inputs. Ensemble methods are also used to perform better, particularly on high-resolution medical imaging. Overall, as described in the literature, MTL has a bright future in medical imaging but also limitations.

Huang et al. [11] have conducted research on the systematic review on ulcerative colitis and Crohn's disease using machine learning (ML) approaches. The study performed an extensive search in six major databases. It has been claimed that most of the studies done were retrospective. The commonly used technique is random forest, and the majority of studies were on the evaluation of the model using ML.

Hudu et al. [12] conducted a critical analysis of AI applications in the diagnosis, prognosis, and management of diseases. The study highlights both the strengths and rapid transformation of AI as well as its limitations. Several challenges using AI are mentioned, such as ethical concerns and implementation. The research also discusses AI-driven predictive modeling and real-time tracking as well as critical issue identification such as data bias, lack of algorithm transparency, and regulatory compliance. Due to these critical parts, it may threaten the reliability and fairness of using AI in healthcare systems. Furthermore, the study emphasizes that AI research should not only focus on developing models but also focus on ethical governance and interdisciplinary collaboration.

III. Methods and Materials

A. Datasets

The endoscopic image dataset was compiled from multiple publicly available medical sources. The images are verified by the domain expert. To enhance data diversity and improve model robustness, standard image augmentation techniques including rotation, width and height shifting, zooming, horizontal flipping, and brightness adjustment, were applied. This process expanded the original dataset to a total of 967 images, which were used for model training and evaluation. It has 485 normal antral images and 482 antral ulcer images which is shown in Figure 1. The dataset has a balanced distribution between the two classes, with nearly equal numbers of normal and ulcer images.

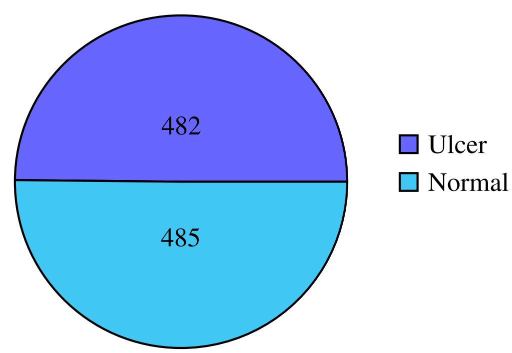

Fig. 1: Dataset Distribution Pie Chart

B. Preprocessing

The study uses JPG format images, which have been resized to 224×224 pixels. Figure 2 presents representative samples of ulcer and normal images. The ulcer images exhibit visible ulcerative characteristics, while the normal images show no such abnormalities.

(a) Ulcer Image 1

(b) Ulcer Image 2

(c) Normal Image 1

(d) Normal Image 2

Fig. 2: Ulcer and Normal Medical Images

Preprocessing included normalization and augmentation to improve model performance. The preprocessing of images was done through image resizing to 224 x 224 pixels, conversion to arrays, and normalization [13].

The processed images were separated into a training and testing set in an 80:20 proportion.

C. Evaluation Criteria

In this study, ulcer images are considered as positive (+) cases, while normal images are treated as negative (–) cases.

The dataset was used to train and evaluate a set of deep learning models: ResNet50, VGG16, MobileNetV2, Inception-ResNet, and CNN. Each model was applied to identify antral ulcers in endoscopic images. Before training, all images went through the same preprocessing steps, such as resizing, normalization, and splitting into training and testing sets. The models used ReLU in the hidden layers and a Sigmoid function in the final layer for binary classification.This approach allowed for a comparative analysis of model performance in terms of accuracy, precision, recall, F1-score, and ROC-AUC for binary classification [14].

D. Confusion Matrix

It helps to determine how well the model handles the positive class for ulcer detection [15]. The confusion matrix contains True Positives, False Positives, True Negatives, and False Negatives. The table below shows how well the model performs across the classes for binary classification. It provides a visualization of positive and negative values with the corresponding labels. The confusion matrix is defined as follows:

TABLE I: Confusion Matrix for Binary Classification of Antral Ulcers

Actual / Predicted	**Ulcer (+)**	**Normal (-)**
Ulcer (+)	TP	FN
Normal (-)	FP	TN

where:

- TP (True Positive): Ulcer images correctly classified as ulcer.
- TN (True Negative): Normal images correctly classified as normal.
- FP (False Positive): Normal images incorrectly classified as ulcer.
- FN (False Negative): Ulcer images incorrectly classified as normal.

The performance metrics are computed using the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is defined as the percentage of predictions that the model correctly classifies.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision measures the proportion of correctly identified positive cases among all cases predicted as positive.

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Recall or Sensitivity measures the ability of a model to identify all relevant cases (true positives) from all actual positive instances.

$$\text{F1-score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

F1 score is the harmonic mean of precision and recall, as given by the formula below. It is very useful for maintaining a balance between precision and recall.

IV. Experiment

The model used transfer learning with pretrained weights using ResNet50, MobileNetV2, Inception ResNet, CNN, and VGG16, including 967 endoscopic images (485 normal and 482 ulcer images). The Python code has been used to implement and train the model based on CNN and performed training logs capturing accuracy, loss, and performance metrics over 20 epochs. The model was trained on an 80:20 split for training and testing, without cross-validation or external validation due to dataset limitations. Data augmentation was applied to improve model generalization. The preprocessing was applied to reduce dataset leakage. The techniques included rotation (±20°), width and height shifting (range 0.1), zooming (range 0.1), horizontal flipping, and brightness adjustment (range 0.8–1.2).

The same preprocessing, training, and evaluation pipeline was applied to all models to consistently assess their performance.The model uses pretrained weights, with frozen base layers, followed by a global average pooling layer, dense layers with 128 and 64 units, dropout layers with rates of 0.25 and 0.1, ReLU activation in dense layers, and a sigmoid output layer for binary classification. The model was trained for 20 epochs using the Adam optimizer (learning rate 0.000035) with a batch size of 8, and binary cross-entropy was adopted as the loss function. Early stopping and other architectural modifications were not applied in this implementation. Noise, vertical flipping, aggressive color jittering, and CLAHE were not used in the datasets.

It has plotted the accuracy curves on the training and testing to identify how the model converges and performs over the epochs. The confusion matrix has calculated which help to analyze the relationship between actual and predicted classes that determine true positives, true negatives, false positives, and false negatives to ensure the consistency of the CNN models [16]. Furthermore, the model's performance was compared with other CNN architectures using confusion matrix analysis [17].

V. Result and Analysis

It was observed that the ResNet50 model achieved high performance in classifying normal and ulcer images.

The ROC curve shown here identifies how well the model can differentiate between positive and negative cases. The area under the curve (AUC) is 0.991; it means it has excellent classification performance [18].

VI. Observations

The confusion matrix of the ResNet50 model shows 94 true positives, 94 true negatives, 3 false positives, and 3 false negatives, indicating a low likelihood of false positives. The model achieved a training accuracy of 0.97 and a testing accuracy of 0.969. The classification report further demonstrates strong performance for both classes, with a precision of 0.969 for positive cases.

The recall of the model is 0.969, indicating that 0.96% of positive cases were correctly identified. This shows that only a small number of cases were missed, with FN = 3. The F1-score is 0.969, demonstrating the overall strength of the classification model. The results indicate very good learning on the training data as well as strong generalization on unseen data. The small difference between training and testing accuracy suggests that the model did not overfit.

	Ulcer +	**Normal -**
Ulcer +	**94**	**3**
Normal -	**3**	**94**

Fig. 3: Confusion matrix of the ResNet50 evaluated on the test dataset for antral ulcer classification.

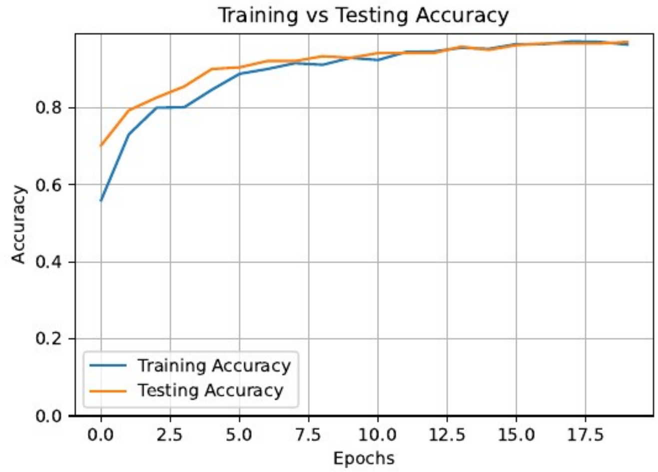

Fig. 4: Training and testing accuracy curves for ResNet50

Overall, the evaluation of the ResNet50 model shows that it is a powerful, real-time, and trustworthy model. In general, the findings indicate that the model is capable of identifying normal and antral ulcer regions with high precision.

The model was trained for 20 epochs, and the accuracy curves indicate progressive and efficient learning. In the first epoch, the ResNet50 model achieved a training accuracy of 54.7% and a testing accuracy of 73.1%, demonstrating good generalization on unseen data. As training continued, both training and testing metrics improved consistently. By epoch 5, the model reached 85.8% training accuracy and 90.9% testing accuracy, indicating strong early convergence.

After epoch 5, the model continued to improve steadily. By epoch 10, the training accuracy reached 92.9% while the testing accuracy attained 92.5%, showing that the model was effectively learning the patterns without losing generalization. Overall, these results indicate that the model is well-trained, reliable, and demonstrates strong and consistent performance.

It was observed that the ResNet50 model performs better on the normal antrum class. The classification report indicates an overall accuracy of 0.969 for both normal and ulcer classes, demonstrating consistent performance. The precision and recall are both 0.969, indicating that the model produces very few false predictions and correctly identifies nearly all actual cases. The F1-score is also 0.969, reflecting the model's overall strong classification performance.

It means there is balanced performance between precision and recall. The area under ROC curve shows that model has perfect performance. It has ideal in separation between the classes.Overall, the model of this AI powered application have strong and reliable performance in accurately diagnosis between ulcer and normal images.

Using this application, ulcers can be detected by analyzing endoscopic images through a computer running the AI system, with patient data securely maintained according to ethical guidelines. The images are preprocessed before being analyzed by the AI model. Challenges include limited infrastructure,

TABLE II: Results of Various Model Predictions

Label	TP	FP	FN	TN	F1 Score	Sensitivity	Specificity
ResNet50	94	3	3	94	96.91%	96.91%	96.91%
VGG16	86	7	11	90	90.53%	88.66%	92.77%
Inception-ResNet	80	9	17	88	86.02%	82.47%	90.70%
CNN	85	12	23	74	78.00%	78.00%	86.07%
MobileNetV2	93	4	4	93	94.00%	94.50%	95.86%

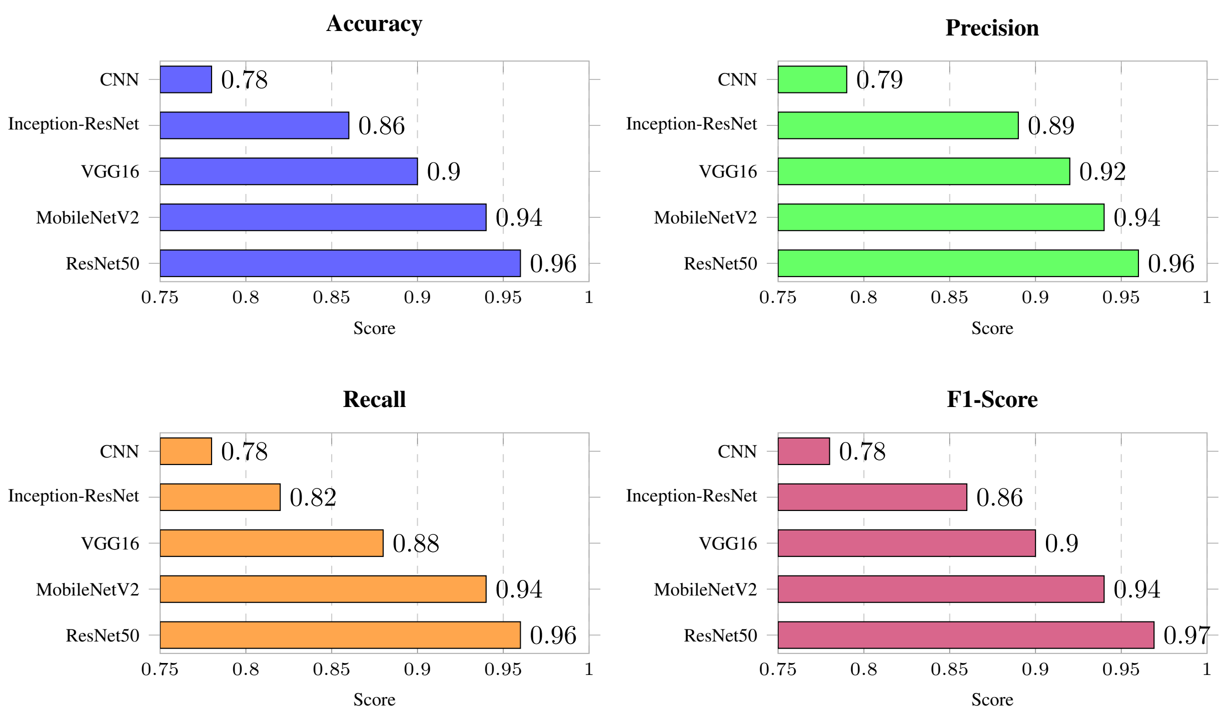

Fig. 5: 2×2 performance comparison of deep learning models using Accuracy, Precision, Recall and F1-score

older devices, staff training, and maintaining data security.

VII. Comparison with Previous Research

Yacob et al. [19] conducted research on the detection of atrophic gastritis (AG), which is caused by Helicobacter pylori infection. It is a serious condition that can lead to cancer if not treated in time. The study used deep learning models for accurate detection.

Swathi et al. [20] conducted research on identifying affected stomach regions using deep learning segmentation methods such as U-Net and SegNet. It also used a classification model to detect antral gastritis from endoscopic images.

Behera et al. [21] investigated the preliminary diagnosis of antral gastritis using endoscopic images with CNNs. The study emphasized the use of CNN-based visual explanation methods to assist medical endoscopists in identifying affected regions and guiding early-stage treatment.

VIII. Recommendations and Limitations

Transfer learning based CNNs are applied for automated antral ulcer detection. However, additional work is needed to optimize the adaptation of pre-trained models when larger datasets are used. However, a limitation is that when trained on larger datasets, further work is needed to determine how the pre-trained model can be effectively adapted, which could help improve the model's overall performance.

In this study, it has not addressed the severity of ulcers or other gastrointestinal disorders concerning the antral part, as it is based on binary classification only. This may limit its clinical applicability. Therefore, future work should focus on multi-class detection and expanding the dataset.

IX. Conclusions and Future Works

Under various models of CNN methods used here, it has shown that the ResNet50 architecture performs effectively

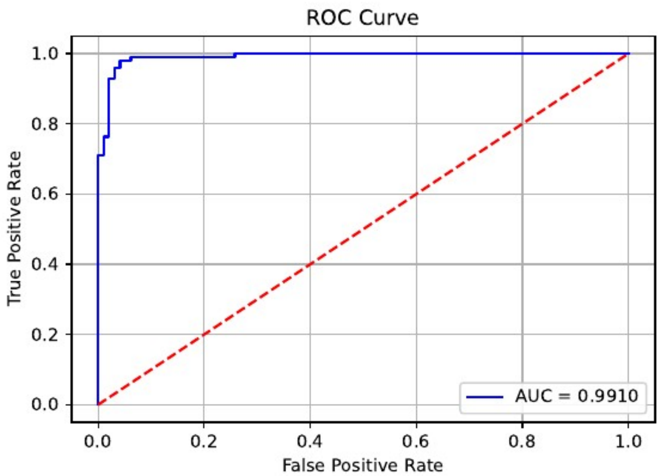

Fig. 6: Receiver Operating Characteristic (ROC) curve for the ResNet50

in classifying normal and ulcerated antral images from endoscopic data. This model has high accuracy and strong performance in terms of precision, recall, and F1-score. It indicates that this approach can serve as valuable support for gastroenterologists for the diagnosis of ulcers, reducing manual assessment and enhancing the overall efficiency of the diagnostic process. Furthermore, the study provides a systematic evaluation of multiple CNN models to assess their performance in detecting antral ulcers.

The distal part of the stomach is a common site for gastric ulcers. Since the antrum is the distal part of the stomach, antral ulcers are selected for diagnosis. The practical applicability of this methodology could be further enhanced through future studies involving larger datasets, multi-class classification, and integration into real-world clinical workflows.

This research can be further extended by expanding the size of the dataset used. To achieve better generalization, it has to train on more images from diverse demographic groups.The model should also be compared with various pretrained models to evaluate where its accuracy performs best.

It is possible to consider multi-class classification so that the severity of ulcers can be divided into mild, moderate, and severe, providing more clinical information. Accuracy can be further improved by training on a greater number of labeled images for multi-class classification. Additionally, applying a real-time detection system in endoscopic devices would further enhance clinical usability.

REFERENCES

[1] A. Mouffak, M. Salihoun, F. Bouhamou, M. Acharki, I. Serraj, and N. Kabbaj, "Upper gastrointestinal bleeding in patients with chronic renal failure: What are the particularities," *Saudi J Med Pharm Sci*, vol. 10, no. 7, pp. 505–508, 2024.

[2] Z.-N. Ye, L.-G. Huang, R. Zhang, W.-R. Xie, L.-H. Wu, L. Li, H. H.-X. Xia, and X.-X. He, "Identification of antralization-specific factors in peripheral blood and gastric mucosa of patients with upper gastrointestinal symptoms: A prospective study," *Cancer Screening and Prevention*, vol. 4, no. 3, pp. 137–147, 2025.

[3] L. R. Celestin, "Antral activity and symptom periodicity in duodenal ulceration," *Gut*, vol. 8, pp. 318–324, August 1967, accessed via NCBI PMC. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1552539/

[4] R. Parajuli, D. Bohara, M. Kc, S. Shanmuganathan, S. K. Mistry, and U. N. Yadav, "Challenges and opportunities for implementing digital health interventions in nepal: a rapid review," *Frontiers in digital health*, vol. 4, p. 861019, 2022.

[5] S. Debnath, A. Khurana, M. Senbagavalli, S. Naik, J. Chandra Patni, P. K. Mishra, and J. Kishore, "Sustainable ai for diabetic foot ulcer detection: a deep learning approach for early diagnosis," *Discover Applied Sciences*, vol. 7, no. 9, p. 1012, 2025.

[6] A. Harry, "Revolutionizing healthcare: the transformative role of artificial intelligence in the health sector," *BULLET: Jurnal Multidisiplin Ilmu*, vol. 2, no. 2, pp. 326–335, 2023.

[7] C. Lei, Y. Jiang, K. Xu, S. Liu, H. Cao, and C. Wang, "Convolutional neural network models for visual classification of pressure ulcer stages: Cross-sectional study," *JMIR Medical Informatics*, vol. 13, p. e62774, 2025.

[8] H. M. Khorasani, H. Usefi, and L. Peña-Castillo, "Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning," *Sci. Rep.*, vol. 10, no. 1, p. 13744, Aug. 2020.

[9] C. Raju, S. Datar, K. Hari, K. Vijay *et al.*, "Applied deep learning to identify and localize polyps from endoscopic images," *arXiv preprint arXiv:2301.09219*, 2023, preprint, available under CC BY 4.0 license. [Online]. Available: https://arxiv.org/abs/2301.09219

[10] Y.-F. Lin, B.-C. Qiu, C.-M. C. Lee, and C.-C. Hsu, "Divide and conquer: Grounding bleeding areas in gastrointestinal images with two-stage model," arXiv:2412.16723, 2024.

[11] J. Huang, X. Zhu, Y. Ma, Z. Zhang, J. Zhang, Z. Hao, L. Wu, H. Liu, H. Wu, and C. Bao, "Machine learning in the differential diagnosis of ulcerative colitis and crohn's disease: a systematic review," *Translational Gastroenterology and Hepatology*, vol. 10, p. 56, 2025.

[12] S. A. Hudu, A. S. Alshrari, E. J. I. Abu-Shoura, A. Osman, and A. O. Jimoh, "A critical review of the prospect of integrating artificial intelligence in infectious disease diagnosis and prognosis," *Interdisciplinary Perspectives on Infectious Diseases*, vol. 2025, no. 1, p. 6816002, 2025.

[13] D. Novita, H. Hayurani, E. K. Sutedja, F. R. Pratomo, A. D. Saputra, Z. Ramadhanti, N. Abutani, M. R. Triandi, A. M. Guferol, A. A. Pravitasari *et al.*, "A mobile application lukaku as a tool for detecting external wounds with artificial intelligence," *Intelligence-Based Medicine*, vol. 11, p. 100200, 2025.

[14] D. Hartama, A. P. Windarto, P. Alkhairi, W. Rosdiana *et al.*, "Improving stock market forecasting: Comparative insights into cnn architectures with a focus on relu activation." *IAENG International Journal of Computer Science*, vol. 52, no. 3, 2025.

[15] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," *African Journal of Biomedical Research*, vol. 27, no. 4S, pp. 4023–4031, 2024.

[16] M. Kahveci and L. Uğur, "Prediction and stage classification of pressure ulcers in intensive care patients by machine learning," *Diagnostics*, vol. 15, no. 10, p. 1239, 2025.

[17] D. W. Girmaw and G. B. Taye, "Mobilenetv2 model for detecting and grading diabetic foot ulcer," *Discover Applied Science*, vol. 7, p. 268, 2025, published 27 March 2025. [Online]. Available: https://doi.org/10.1007/s42452-025-06745-4

[18] N. Fitriah and S. Sriani, "Classification of foot wound severity in type 2 diabetes mellitus patients using mobilenetv2-based convolutional neural network," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2163–2170, 2025.

[19] Y. M. Yacob, H. Alquran, W. A. Mustafa, M. Alsalatie, H. A. M. Sakim, and M. S. Lola, "H. pylori related atrophic gastritis detection using enhanced convolution neural network (cnn) learner," *Diagnostics*, vol. 13, no. 3, p. 336, 2023.

[20] G. Swathi and S. Umapathy, "Antral gastritis segmentation based on deep learning models in endoscopic images," in *2025 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*. IEEE, 2025, pp. 1–6.

[21] L. K. Behera and S. N. Tripathy, "Antral gastritis preliminary diagnosis by endoscopic image analysis using deep learning approach," in *Computing, Communication and Intelligence*. CRC Press, 2025, pp. 62–66.