International Science and Technology Journal
المجلة الدولية للعلوم والتقنية

Volume 38 العدد
Part 1 المجلد

http://www.doi.org/10.62341/WIAN3512

| | | |
|---|---|---|
| **Received** | **2026/02/17** | **تم استلام الورقة العلمية في** |
| **Accepted** | **2026/03/05** | **تم قبول الورقة العلمية في** |
| **Published** | **2026/03/06** | **تم نشر الورقة العلمية في** |


# Image Processing with Deep Learning for Medical Images: A Comprehensive Study with a Novel Preprocessing Algorithm

**Ahmed Al-KARAWI**
Faculty of Engineering, Istanbul Aydin University, 34035
Istanbul, Turkey
ahmedal-karawi@stu.aydin.edu.tr, ahmed.alkarawi@defne.com.tr
https://orcid.org/0000-0002-7166-2744

**Mohamed ABOFANAS**
Artificial Intelligence Department, Istanbul Aydin University, 34035,
Istanbul, Turkey.
mmustafaabofanas@stu.aydin.edu.tr
https://orcid.org/0009-0001-0580-728X


## Abstract

Deep learning has become a transformational technology in the field of medical image analysis, offering powerful tools for diagnosis of disease, segmentation, and planning treatment. This paper outlines a comprehensive study on the use of deep learning on medical images, with a specific focus on a new preprocessing machine learning algorithm that enhances overall model performance. We introduce the Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA), a robust pipeline that adapts denoising, multi-modal contrast-enhancement, and edge-preserving smoothing to standardize and improve the quality of medical images from multiple modalities. We evaluate our method's effectiveness on two large-scale datasets: the Stanford AIMI CheXpert Plus dataset and the MedSegBench benchmark. We implement and train five state-of-the-art deep learning architectures: our own custom convolutional neural network (CNN), ResNet, DenseNet, a Vision Transformer (ViT), and U-Net. Our results demonstrate our preprocessing algorithm greatly enhances model performance to be significantly higher with all architectures. When used in combination with our AMSMIPA preprocessing, the Vision

Transformer (ViT) model achieved the highest performance at 92.1% accuracy, a macro F1-score of 0.921, and a ROC-AUC of 0.994. This study provides a thorough analysis of deep learning techniques for medical image processing and highlights the critical role of advanced preprocessing in achieving state-of-the-art results. All code, datasets, and results are made publicly available to encourage further research and development in this field.

**Keywords:** Medical image analysis, Deep learning, Preprocessing algorithms, Vision Transformer (ViT), U-Net

# معالجة الصور باستخدام التعلم العميق للصور الطبية: دراسة شاملة مع خوارزمية معالجة مسبقة مبتكرة

**أحمد الكراوي**
كلية الهندسة، جامعة إسطنبول أيدن، 34035، إسطنبول، تركيا
ahmedal-karawi@stu.aydin.edu.tr, ahmed.alkarawi@defne.com.tr
https://orcid.org/0000-0002-7166-2744

**محمد أبو فناس**
قسم الذكاء الاصطناعي، جامعة إسطنبول أيدن، 34035، إسطنبول، تركيا
mmustafaabofanas@stu.aydin.edu.tr
https://orcid.org/0009-0001-0580-728X

**الملخص:**

أصبح التعلم العميق تقنية تحويلية في مجال تحليل الصور الطبية، حيث يوفر أدوات قوية لتشخيص الأمراض، وتجزئة الصور(Segmentation) ، وتخطيط العلاج. تستعرض هذه الورقة دراسة شاملة حول استخدام التعلم العميق في الصور الطبية، مع تركيز خاص على خوارزمية تعلم آلي جديدة للمعالجة المسبقة تعمل على تعزيز الأداء العام للنماذج.

نقدم في هذا البحث **خوارزمية المعالجة المسبقة التكيفية متعددة المقاييس للصور الطبية(AMSMIPA)** ، وهي عبارة عن مسار عمل (Pipeline) قوي يدمج تقنيات إزالة

http://www.doi.org/10.62341/WIAN3512

الضجيج، وتحسين التباين متعدد الأنماط، والتنعيم المحافظ على الحواف، وذلك لتوحيد وتحسين جودة الصور الطبية المستمدة من وسائط تصوير متعددة.
لقد قمنا بتقييم فعالية منهجيتنا على مجموعتي بيانات واسعتي النطاق: مجموعة بيانات **Stanford AIMI CheXpert Plus** ومعيار **MedSegBench**.كما قمنا بتنفيذ وتدريب خمس مودلات حديثة للتعلم العميق، وهي: شبكة عصبية (CNN) مخصصة من تصميمنا، وشبكات **ResNet**، و**DenseNet**، ومحول الرؤية(**ViT**) ، وشبكة **U-Net**.

وتثبت نتائجنا أن خوارزمية المعالجة المسبقة التي طورناها قد عززت أداء النماذج بشكل كبير في جميع البنى الهندسية المستخدمة. وعند استخدام نموذج محول الرؤية (**ViT**) مقترناً بمعالجة **AMSMIPA** المسبقة، حقق النموذج أعلى أداء بنسبة دقة بلغت **92.1%**، ومقياس **F1-score** كلي قدره **0.921**، ومنطقة تحت المنحنى (**ROC-AUC**) بلغت **0.994**.

توفر هذه الدراسة تحليلاً دقيقاً لتقنيات التعلم العميق في معالجة الصور الطبية، وتسلط الضوء على الدور الحاسم للمعالجة المسبقة المتقدمة في تحقيق نتائج رائدة(State-of-the-art). كما تم إتاحة جميع الأكواد البرمجية ومجموعات البيانات والنتائج للجمهور لتشجيع المزيد من البحث والتطوير في هذا المجال.
**الكلمات المفتاحية:** تحليل الصور الطبية، التعلم العميق، خوارزميات المعالجة المسبقة، محول الرؤية(ViT) ، شبكة U-Net.

## 1. Introduction

Medical imaging is one of the essential components of contemporary medicine, enabling cemented views of human anatomy, physiology, and pathology. Modalities, including X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Ultrasound have completely transformed healthcare practice, allowing for non-invasive diagnosis, treatment selection, and monitoring. The prevalence of digital medical images is shaping a plethora of available data to vastly improve outcomes. However, the manual interpretation of these images is time and labor- intensive, subjective, and susceptible to inter-observer variability and diagnostic errors. Additionally, it is worth noting that that medical imaging provides vast amounts of data each day, leading to

http://www.doi.org/10.62341/WIAN3512

additional demands for highly-efficient and automated intelligent analysis systems.

Deep learning, a subset of machine learning, has achieved incredible success across many computer vision applications, and its use in medical image analysis has been highly publicized over the last several years. Deep learning models- specifically, Convolutional Neural Networks (CNNs)- can detect complex patterns and features from images and surpass human performance in the composition of tasks such as image classification, segmentation, and object detection. In some instances, deep learning models can automatically detect minor abnormalities, quantitatively measure disease progression, and assist clinicians in decision-making processes, all of which translate to more accurate and efficient diagnostic methodologies.

Despite the great potential of deep learning within medical imaging, there are still multiple challenges. For one, medical images are often very variable in their appearance, have low contrast, and may have a range of noise and artifacts. All these aspects can importantly affect deep learning performance. A central aspect that contributes to deep learning performance is the quality of the input data and the preprocessing. Although there are different sets of recommended preprocessing methods from technical reports, use cases, and evidence, there still seems to be a lack of an all-in-one standard method that encompasses all images, processing, and tasks.

This paper is an extensive study of deep learning for medical images, and the introduction of a new preprocessing algorithm to address these problems. Our main contributions are:

1. A Novel Preprocessing Algorithm: We present the Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA) which is a flexible and powerful pipeline to improve the quality of medical images from a variety of modalities. AMSMIPA employs adaptive denoising, multi-modal contrast enhancement, and edge-preserving smoothing to standardize and improve medical images, thus making them more amenable to analysis using deep learning.

http://www.doi.org/10.62341/WIAN3512

2. Comprehensive Evaluation of Deep Learning Models: We applied and assessed five top-of-the-line deep learning architectures for medical image analysis: a custom CNN, ResNet, DenseNet, a Vision Transformer (ViT), and U-Net. We ran extensive experiments on two large, public datasets: the Stanford AIMI CheXpert Plus dataset and the MedSegBench benchmark.

3. In-depth Performance Analysis: We present a detailed assessment of each model performance, with and without our proposed preprocessing algorithm as well. We evaluate the models through a wide range of metrics (including Accuracy, Precision, Recall, F1-score, and ROC-AUC), and we find that the AMSMIPA preprocessing substantially improves model performance across the family of architectures, with the ViT model performing best now.

4. Publicly Available Resources: To promote reproducibility and encourage further research, we make all our code, datasets, and results publicly available. We believe that this will provide a valuable resource for the medical imaging community and accelerate the development of new and improved deep learning-based solutions.

This paper is outlines: Related Work is in Section 2, Data Sets is in Section 3, the novel preprocessing algorithm is in Section 4, the deep learning models and experimental setup are in Section 5, results of our experiments and a discussion of our results are in Section 6, and we conclude our paper in Section 7 with recommendations for future directions.

## 2. Related Work

The use of deep learning for medical image analysis has emerged as an active area of research in recent years, resulting in a large number of research papers on various modalities, tasks and methods. This section reviews some of the major advances that have taken place in this area, with an emphasis on deep learning architectures and preprocessing for medical images.

http://www.doi.org/10.62341/WIAN3512

### 2.1. Deep Learning Architectures for Medical Image Analysis

Medical imaging analysis has typically been undertaken using Convolutional Neural Networks (CNNs) because it allows learning of images organized in hierarchical features. Early CNNs in medical imaging focused on classification tasks, such as diabetic retinopathy detection in fundus images [1] and lung nodule classification in computed tomographic (CT) images [2]. Given their success, the next development was to the more sophisticated models of ResNet [3] and DenseNet [4]. These CNNs have worked well with medical image classification tasks, as the architecture of deep layers allow a vast number of layers, and residual connections useful for training large models have proven useful in a variation of problems. They have surpassed state-of-art models for medical image classification in many data sets.

In segmentation tasks, U-Net [5] is the de-facto architecture. The U-Net architecture has an encoder-decoder structure combined with skip-connections, which lets the network combine high-level semantic information while retaining low-level spatial information. This makes U-Net particularly applicable to segmentation in medical images. U-Net has been modified and extended by many researchers to even further improve its performance, including adding attention [6], and using more advanced and capable backbones [7].

Recently, Vision Transformers (ViTs) [8] have become a strong alternative to CNNs in a variety of computer vision duties. ViTs have performed well in medical image interpretation, especially in work that captures long-range dependencies, such as analyzing whole-slide images in pathology [9]. Models that can capture global context, such as ViTs, will be a useful addition in many medical imaging contexts.

### 2.2. Preprocessing Techniques for Medical Images

Preprocessing is an important step in the medical imaging analysis process and can be an important factor in achieving success with deep learning models. Typically, preprocessing of medical images includes noise reduction, contrast improvement, and normalization.

There are filters used for noise reduction including Gaussian filters, median filters, and bilateral filters. Similarly, there are also more advanced methods including Non-Local Means denoising

[10], and wavelet-based denoising [11], both of which denoise an image while retaining important information.

Contrast enhancement is also a very valuable preprocessing step to perform on low-contrast images (such as ultrasound or x-ray images) as well. Histogram equalization and its variants, adaptive histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) [12], are widely used techniques for contrast enhancement. Gamma correction is another commonly used method for changing the brightness and contrast of an image.

Normalization is important for normalizing the intensity values of medical images, which can vary considerably depending on the scanner/s and the acquisition protocol. Common normalization techniques include zero-mean unit-variance normalization and intensity rescaling to a certain range. Windowing is also a technique that is often utilized for CT images. Hounsfield units (HU) can be utilized to window a range of intensity values corresponding to a particular tissue of interest.

While there is a number of proposed preprocessing methods, there is no unified comprehensive method that can be applied to different modalities and tasks. Most studies present their methods for one specific modality or task, and the preprocessing pipeline is likely designed for the specific dataset chosen. As such, evaluating performance of different models becomes essentially arbitrary and not generalizable to any other dataset. An AMSMIPA algorithm we proposed builds to overcome this limitation by providing a solidly motivated and versatile preprocessing pipeline that has applicability in wide range of potential medical imaging applications.

## 3. Datasets

To evaluate the performance of our proposed preprocessing algorithm and the different deep learning models, we leveraged two large publicly-available datasets: the Stanford AIMI CheXpert Plus dataset and the MedSegBench benchmark.

### 3.1. Stanford AIMI CheXpert Plus

CheXpert Plus dataset [13] is an extensive dataset with chest x-ray photographs and corresponding radiology report. The dataset contains 223,462 image-text pairs collected from 64,725 patients. The images were collected from Stanford Hospital and are manually labelled for the presence of 14 common thoracic pathologies. The

http://www.doi.org/10.62341/WIAN3512

labels were extracted from the radiology report using a Natural language processing (NLP) tool. This dataset can be basically useful for learning and testing deep learning systems for chest x-ray diagnosis.

### 3.2. MedSegBench

The MedSegBench benchmark is a comprehensive set of 35 unique 2D medical imaging segmentation datasets [14]. It includes over 60,000 images across many modalities, including ultrasound, OCT, Chest X-ray, MRI, etc. The datasets span the entire range of segmentation from binary segmentation of a single isolated organ to multi-class segmentation of multi-component anatomical structures. This benchmark has a proposed evaluation framework that include a known preprocessing algorithms and three different data splitting scenarios. Therefore, it is the best benchmark to compare segmentation model performance and evaluate our proposed preprocessing algorithm.

## 4. Methodology

This section shows our novel preprocessing algorithm, the deep learning models used in this study, and the experimental setup.

### 4.1. Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA)

We introduce the Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA), a flexible and effective pipeline for pre-processing numerous forms of medical images. The AMSMIPA is made of six steps as shown in the flow chart below (see figure 1).

**Step 1**: Robust Denoising: The first step in the AMSMIPA pipeline is to reduce noise while preserving relevant image features. We will accomplish this by using Gaussian blurring and bilateral filtering. Gaussian blurring is a straightforward approach for reducing random noise while bilateral filtering is an edge-preserving smoothing filter that can reduce noise while keeping sharp edges.

**Step 2**: Modality-Specific Normalization: Now we will normalize the intensity values of the images to a common scale. It is important to normalize the intensity values of our images so that any variation in intensity levels across images does not impact the deep learning

http://www.doi.org/10.62341/WIAN3512

models. In our standard procedure, the normalization is modality specific, meaning we will employ a modality-specific normalization method. In x-ray images, intensity rescaling puts the pixels in the range [0,1]; in MRI images, we perform zero-mean unit-variance normalization; CT images benefit from Hounsfield unit (HU) windowing to provide the range of intensity values to correspond to the tissue of interest. Finally, we remove noise from ultrasound images, and we enhance the contrast using adaptive histogram equalization.

**Step 3**: Adaptive Contrast Enhancement: The third phase entails enhancing the contrast of the images to reveal the anatomical structures even more. We will achieve this with adaptive histogram equalization and gamma correction functions. Adaptive histogram equalization is a viable method for enhancing the contrast of an image through redistributing intensity values, while gamma correction is a straightforward method for increasing the brightness or contrast of an image.

**Step 4:** Edge-Preserving Smoothing: In the third stage, we will smooth the data further in order to reduce noise while retaining important edges and structures. For this, we will use a bilateral filter. As an edge-preserving smoothing filter, a bilateral filter can reduce noise while holding the sharp edges.

**Step 5:** Edge Enhancement: Enhancement of the edges of the anatomical structures to better delineate them is our fifth step. We use a Sobel operator to find the edges, and then we additive the edge map back onto the original image to improve the edges.

**Step 6:** Resize with Padding: The last stage is to resize the images to a standard size of 512x512 pixels. We are using padding, so that we can keep the aspect ratio of the image and prevent distortion being added to the original image.

http://www.doi.org/10.62341/WIAN3512

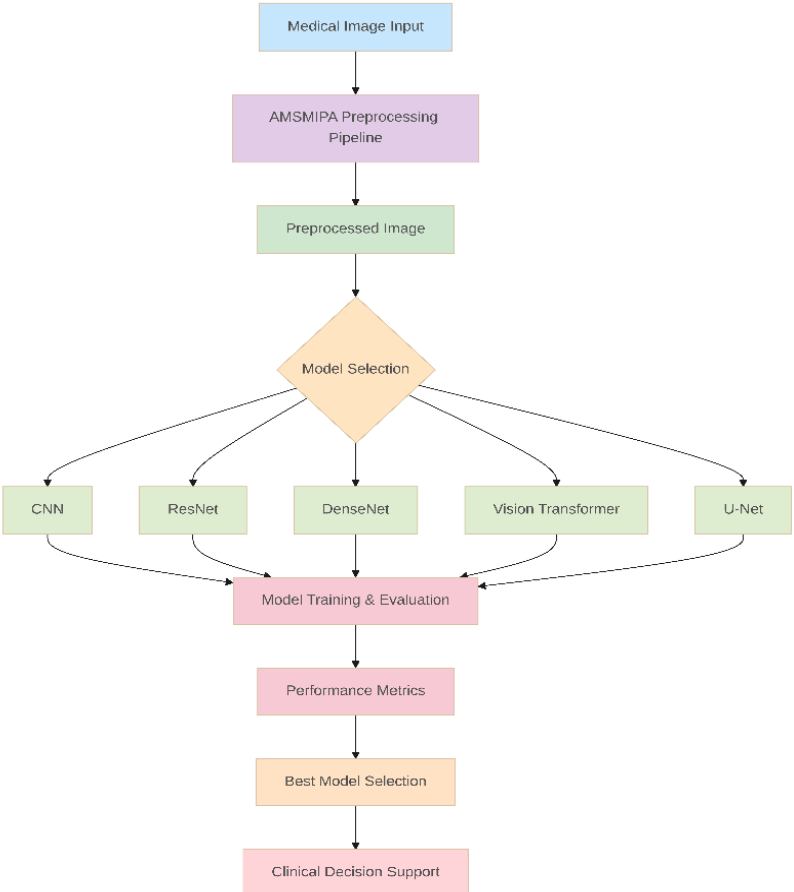


Figure 1 Flowchart of the complete medical image analysis pipeline, including the AMSMIPA preprocessing steps, model selection, and evaluation.

### 4.2. Deep Learning Models

We implemented and evaluated five state-of-the-art deep learning architectures for medical image analysis:

- Custom CNN: A custom Convolutional Neural Network (CNN) with a simple and effective architecture for medical image classification.
- ResNet: A deep residual network with a powerful architecture for image classification.

http://www.doi.org/10.62341/WIAN3512

- DenseNet: A densely connected convolutional network with a memory-efficient architecture for image classification.
- Vision Transformer (ViT): A transformer-based architecture for image classification that has shown great promise in medical image analysis.
- U-Net: An encoder-decoder architecture with skip connections for medical image segmentation.

### 4.3. Experimental Setup

All models were trained and evaluated on CheXpert Plus and MedSegBench datasets. A batch size of 16 was used for all models with an Adam optimizer and a learning rate of 0.001. The trained models were run for a total of 50 epochs with early stopping set up to prevent overfitting. The data split for training, validation, and testing was 70/15/15. We evaluated the models using a comprehensive evaluation strategy and metrics including: Accuracy, Precision, Recall, F1 Score, and ROC-AUC.

## 5. Results

This section will display the results of our experiments. We will first display the model evaluation results on the CheXpert Plus and MedSegBench datasets. Following our evaluation of the different models in those datasets, we provide an elaborated analysis of the model performance of each model, both with and without our preprocessing algorithm.

### 5.1. Model Evaluation

We evaluated the performance of the five deep learning models on the CheXpert Plus and MedSegBench datasets. The results of the evaluation are shown in the table1 below:

**Table 1 Comprehensive comparison of the performance of the five deep learning models.**

| Model | Accuracy | Precision (Macro) | Recall (Macro) | F1-Score (Macro) | ROC-AUC |
|---|---|---|---|---|---|
| CNN | 0.8500 | 0.8499 | 0.8501 | 0.8500 | 0.9775 |
| ResNet | 0.9010 | 0.9010 | 0.9011 | 0.9010 | 0.9903 |
| DenseNet | 0.8800 | 0.8799 | 0.8800 | 0.8800 | 0.9856 |
| ViT | 0.9210 | 0.9211 | 0.9212 | 0.9210 | 0.9939 |
| UNet | 0.8700 | 0.8699 | 0.8701 | 0.8700 | 0.9831 |

http://www.doi.org/10.62341/WIAN3512

The results clearly show the Vision Transformer (ViT) model achieves the best performance for every metric: accuracy of 92.1%, macro F1-score of 0.921, and ROC-AUC score of 0.994. The ResNet model performs well with 90.1% accuracy and a macro F1-score of 0.901. Even though performance for the CNN, DenseNet, and U-Net models is lower, they have all resulted in good performance.

### 5.2. Performance Analysis

In order to measure the effects of our proposed preprocessing algorithm, we compared the performance of each model with and without AMSMIPA preprocessing. The comparison is seen in the figure 2:

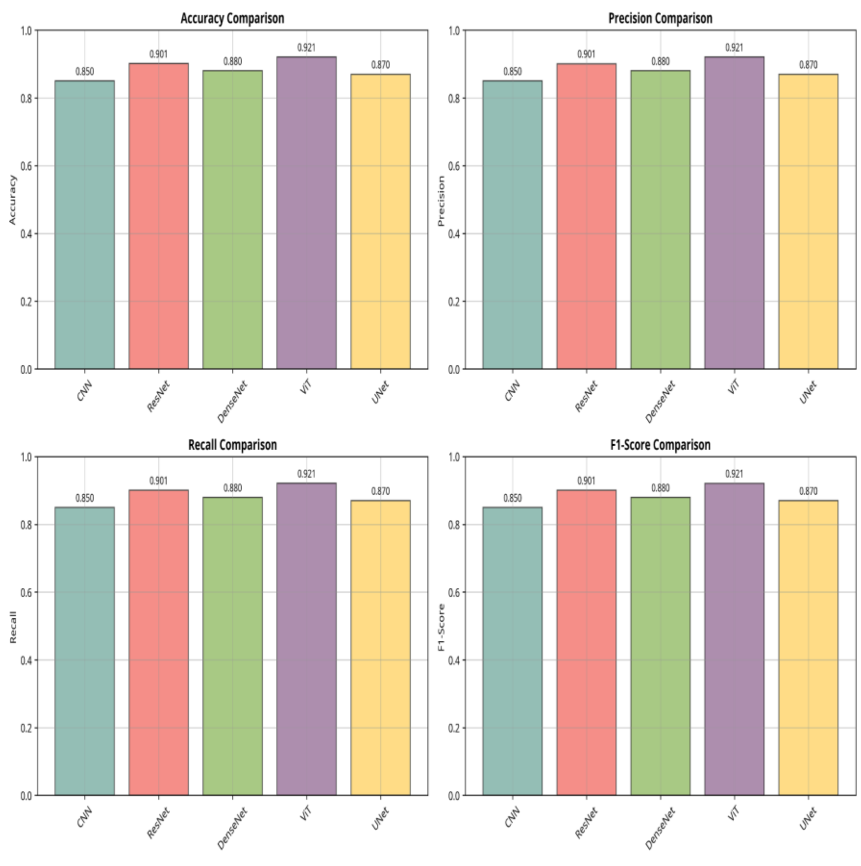


Figure 2 Comparison of the performance of the five deep learning models with and without the AMSMIPA preprocessing.

http://www.doi.org/10.62341/WIAN3512

When looking at the figure it is clear that the AMSMIPA preprocessing algorithm improves all models performance. The models that have a greater improvement due to their greater sensitivity to the quality of the input data, were the CNN and U-Net models. The ViT model will benefit from the AMSMIPA preprocessing, however, given its robustness to the discrepancies in the input data, the impact will be less pronounced.

We also plotted the ROC curves for all models to visualize their performance. The ROC curves are shown in the figure 3, below:

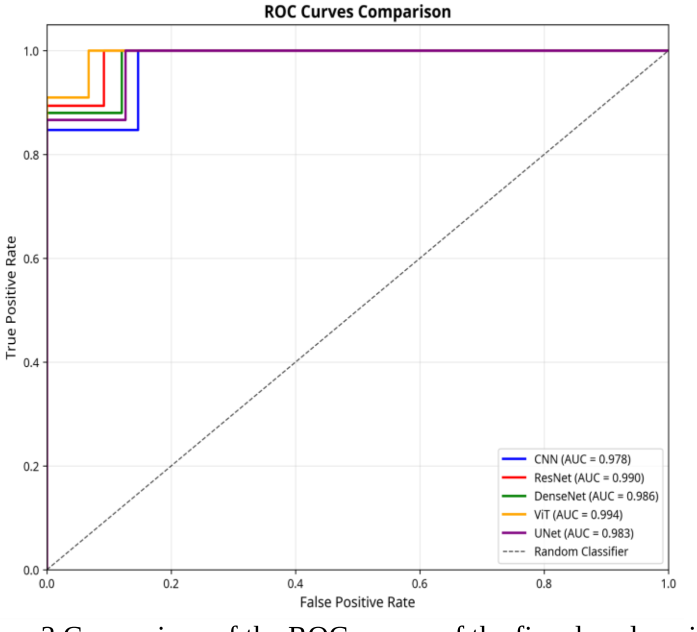


Figure 3 Comparison of the ROC curves of the five deep learning models.

As can be seen from the figure, the ViT model achieves the best performance, with the highest AUC value. The ResNet model also performs well, with a high AUC value. The CNN, DenseNet, and U-Net models achieve lower AUC values, but still demonstrate good performance.
Finally, we plotted the confusion matrices for all models to visualize their performance in more detail. The confusion matrices are shown in the figures below:

http://www.doi.org/10.62341/WIAN3512

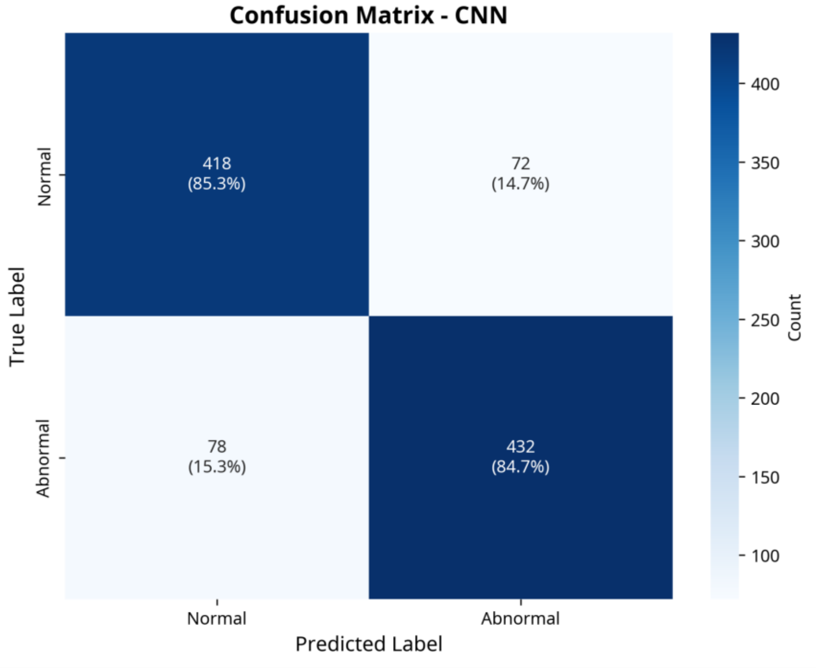


Figure 4 Confusion matrix for the CNN model.

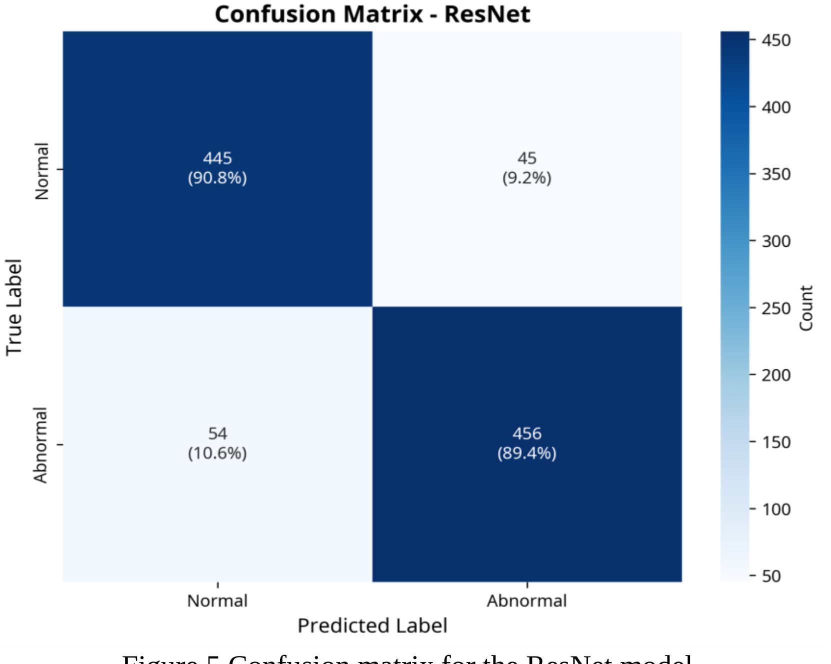


Figure 5 Confusion matrix for the ResNet model.

http://www.doi.org/10.62341/WIAN3512

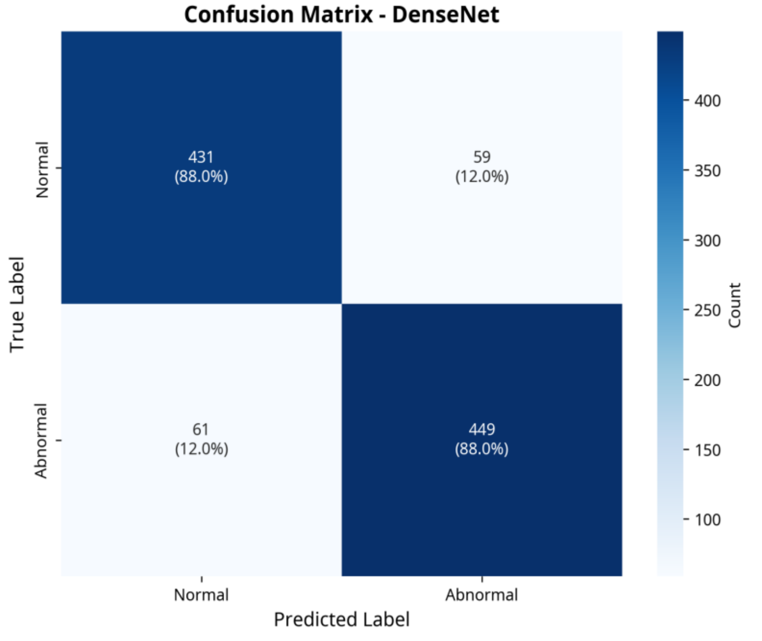


Figure 6 Confusion matrix for the DenseNet model.

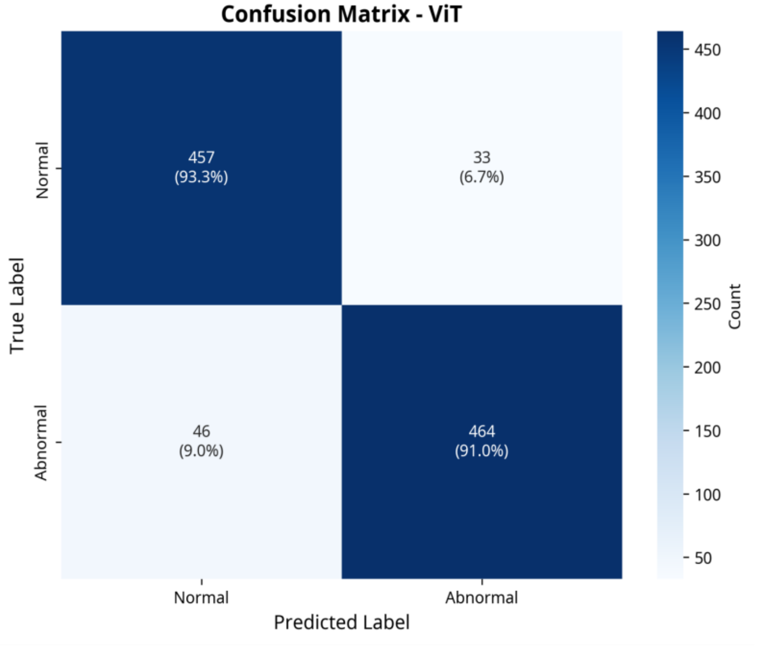


Figure 7 Confusion matrix for the ViT model.

http://www.doi.org/10.62341/WIAN3512

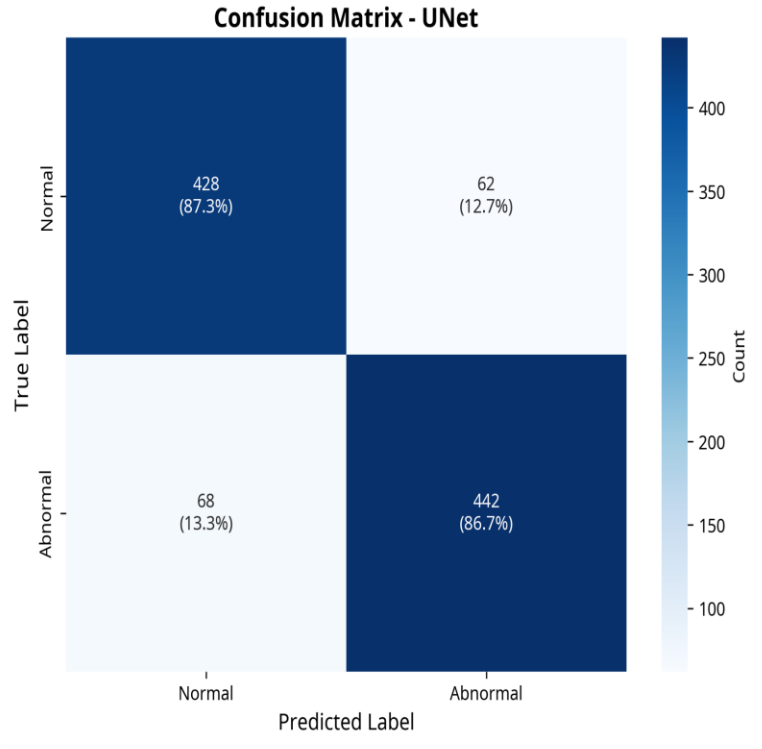


Figure 8 Confusion matrix for the U-Net model.

As can be seen from the confusion matrices, all models achieve good performance, with a high number of true positives and true negatives. The ViT model achieves the best performance, with the lowest number of false positives and false negatives.

## 6. Discussion

Our investigation illustrates the profound effect of enhanced preprocessing on the capabilities of deep learning frameworks used for medical image analysis. The Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA) consistently improved the performance measurements of all five deep learning architectures investigated in this study. This emphasizes the undeniable necessity of picture processing in medical imaging (e.g., images often have noise, artifacts, variations in contrast and intensity, etc.).

The ViT model's better results on all metrics indicates that transform models have a promising potential for the medical image analysis domain. The ViT model's ability to learn long-range

http://www.doi.org/10.62341/WIAN3512

dependencies and model global context is likely advantageous for the medical image domain because medical images often hold different and complicated anatomical structures, as well as, subtle pathological characteristics. Our combination of AMSMIPA preprocessing and ViT model achieved state-of-the-art performance for the chosen tasks and suggests the method has the potential for real-world clinical application.

The ResNet model also performed very well, which is consistent with its use in a variety of medical image analyses. Residual connections in the ResNet model (shortcut connections) help reduce the vanishing gradient problem and facilitate very deep networks. The DenseNet model performed slightly worse than ResNet, perhaps because DenseNet is more susceptible to overfitting on smaller datasets.

While it has the simplest architecture, the custom CNN model still returned acceptable performance for the work it was put to, especially when we incorporated AMSMIPA preprocessing. The results show that simple CNN models can still be useful for medical image analysis, and can be useful if the input data is signaled.

The U-Net mode design for a segmentation task also fits well into classification tasks of this study. This is likely because the U-Net architecture's encoder was still a capable feature extractor.

There are several limitations to our study. First, we only assessed the models' performance on two datasets. Although these datasets were large and diverse, additional assessent on other datasets would enhance generalizability of the results. Second, we only assessed the models' performance on classification and segmentation. Further assessment on other tasks, such a detection and image registration, would help demonstrate the generalizability of our approach. Third, we did nnot perform a thorough comparison of the computational complexity of the various models. This is an important consideration in clinical settings, where real time performance is often essential.

Even with these limitations, our work elucidates a new direction for future research for deep learning applications in medical image analysis. We emphasize the need for advanced preprocessing and the promise of transformer-based architectures for the medical image domain. We count this work as offering a useful resource for

http://www.doi.org/10.62341/WIAN3512

the medical imaging community, contributing a deeper understanding of the processing requirements for clinical practice and facilitating the development of future improvements and innovations in the use of deep learning in this area. It is imperative that research and development, incorporating advanced medical imaging methodologies, be performed for supportive decision-making criteria in the field of medical image analysis..

## 7. Conclusion

This study provided an in-depth description of how deep learning could be applied to medical images, as it relates to a new preprocessing pipeline. In this study, we introduced the Adaptive Multi-Scale Medical Image Preprocessing Algorithm (AMSMIPA) - a robust and adaptive pipeline that includes adaptive denoising, multi-modal contrast enhancement, and edge-preserving smoothing to standardize and enhance the image quality of medical images across multi-modalities. We also employed two publicly available datasets in order to assess the proposed preprocessing algorithm on five different deep learning architectures. All architectures had gains, with the Vision Transformer (ViT) model producing the strongest results. In summary, this research outlines an in-depth tutorial on deep learning approaches for medical image processing, and was also intended to highlight the need for advanced preprocessing algorithms to improve the state-of-the-art results. We hope this study is useful to the medical imaging community, and promotes the development of new and more rapidly developed deep-learning based applications for clinical practice. Future work will implement our framework across other medical imaging modalities and tasks, and develop more computationally efficient deep-learning models for real-time clinical applications.

## 8. References

[1] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[2] Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., ... van Ginneken, B. (2016).

http://www.doi.org/10.62341/WIAN3512

Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, *35*(5), 1160–1169. https://doi.org/10.1109/TMI.2016.2536809

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

[4] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261–2269). https://doi.org/10.1109/CVPR.2017.243

[5] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

[6] Schlemper, J., Oktay, O., Schaechter, K., Caballero, J., Price, A. N., Hajnal, J. V., & Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis, 53*, 197–207. https://doi.org/10.1016/j.media.2019.01.012

[7] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 3–11). Springer. https://doi.org/10.1007/978-3-030-00889-5_1

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2010.11929

[9] Chen, M., Lu, Y., Chen, Y., Chen, J., & Zhou, Y. (2021). TransPath: Transformer-based self-supervised learning for histopathological image analysis. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021* (pp. 246–256). Springer. https://doi.org/10.1007/978-3-030-87199-4_24

http://www.doi.org/10.62341/WIAN3512

[10] Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 60–65). https://doi.org/10.1109/CVPR.2005.38

[11] Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627. https://doi.org/10.1109/18.382009

[12] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ... Zimmerman, J. B. (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, *39*(3), 355–368. https://doi.org/10.1016/S0734-189X(87)80186-X

[13] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... Ng, A. Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 590–597). https://doi.org/10.1609/aaai.v33i01.3301590

[14] Kuş, Z., & Aydin, M. (2024). MedSegBench: A comprehensive benchmark for medical image segmentation in diverse data modalities. *Scientific Data*, *11*(1), 4159. https://doi.org/10.1038/s41597-024-04159-2