Proceedings of Machine Learning Research vol 245:1–9, 2024 2024 International Conference on Machine Learning and Intelligent Computing

Enhancing Anomaly Detection in Medical Imaging: Blood_UNet with Interpretable Insights

Hao Wang* 2106020521@ST.BTBU.EDU.CN
Institute of Problem Solving, Beijing Technology and Business University, School of Computer and Artificial Intelligence (School of Cyberspace Security), Beijing, China

Kexin Cao 2206160101@ST.BTBU.EDU.CN
Institute of Problem Solving, Beijing Technology and Business University, School of Computer and Artificial Intelligence (School of Cyberspace Security), Beijing, China

Editors: Nianyin Zeng and Ram Bilas Pachori

Abstract

Researching anomaly detection in medical imaging, particularly in blood samples, is crucial for enhancing diagnostic accuracy. This article is to devise a robust method using Blood_UNet, a modified U_Net architecture integrated with Lesion Enhancing Network (LEN) and Shape Model (SHAP), to improve anomaly detection precision. Specifically, the research involves preprocessing the BloodMNIST, training the Blood_UNET model, and interpreting its predictions using LEN and SHAP. Experimental results on BloodMNIST showcase the efficacy of identifying anomalies within blood samples. This study highlights the importance of leveraging advanced techniques like LEN and SHAP in medical diagnostics, contributing to better patient care and healthcare efficiency.

Keywords: Blood_UNet, LEN, SHAP, anomaly detection, medical imaging

1. Introduction

Medical image analysis is an important research direction in the field of medical imaging. This study aims to explore the important task of identifying and classifying diseases by analyzing medical images based on the MedMNIST dataset (Yang et al., 2023). The potential value of this model in disease diagnosis, treatment planning, and clinical decision-making. With the development of neural networks, the accurate classification and diagnosis of medical images are more precise (Anwar et al., 2018), helping doctors improve diagnostic efficiency and accuracy. It is hoped to provide new enlightenment and solutions for research and clinical practice in the field of medical image classification.

In the field of medical image classification, researchers have proposed various methods and techniques to process and analyze medical images. Deep learning technology, especially convolutional neural networks (CNN), has become one of the mainstream methods (Kayalibay et al., 2017). By adjusting CNN and other models on X-ray and CT scan modes, the identification and detection of chest pathology are now well achieved, and good classification results are achieved (Shin et al., 2016). Feature Pyramid Visual Transformer (FPViT) (Liu et al., 2022), which fully covers the analysis of the MedMNIST subset, makes up for the limitations. With the continuous development of technology, more and more researchers have begun to pay attention to interpretability and model stability (Holzinger et al., 2019) . Nowadays, interpretability is combined with various models to help better process or analyze the medical field (Tjoa and Guan, 2021). In summary, the current research trend is to combine deep learning technology and interpretability methods to improve the accuracy and credibility of medical image classification.

The main objective of this study is to develop an effective medical image segmentation model Blood_UNet architecture, because blood contains rich physiological and pathological information, reflecting the overall health status of the human body. Moreover, blood sampling is relatively easy, and non-invasive, and the cost of obtaining blood data is relatively low, making it a crucial data source for medical research, disease diagnosis, and treatment (Woolf, 1955). So blood is perfect for all medical research. This article aims to improve diagnostic accuracy and streamline medical decision-making processes for every relationship with blood disease. First, this paper utilizes the Blood_UNet model because it is effective in handling medical image segmentation tasks with complex structures and varying textures, especially blood-related images. Second, attention mechanisms are integrated into the Blood_UNet to enhance feature extraction and focus on relevant regions, improving segmentation accuracy. Third, conduct comprehensive analyses and comparisons of different model variants to evaluate their predictive performance on the BloodMNIST dataset. Additionally, this article proposes a novel approach to leverage interpretability techniques such as Lesion Enhancing Network (LEN) and Shape Model (SHAP) (Heimann and Meinzer, 2009) to provide insights into model predictions, aiding medical professionals in understanding and verifying segmentation results. The experimental results demonstrate the effectiveness of this method, highlighting the importance of attention mechanisms in improving segmentation accuracy and the utility of interpretability techniques in enhancing model transparency and trustworthiness. This model attention is paid to addressing challenges such as data imbalance, noise, and variability in medical imaging data. Techniques such as data augmentation, transfer learning, and model ensemble are employed to enhance the robustness and generalization capability of the proposed method. This study contributes to advancing the field of medical image analysis by providing a reliable segmentation framework tailored for the BloodMNIST dataset, offering a reliable and efficient tool for medical professionals to assist in disease diagnosis and treatment planning, ultimately facilitating more accurate and interpretable diagnoses in clinical practice. All MedMNIST can be seen here https://medmnist.com/.

2. Methodologies

2.1. Dataset Description and Preprocessing

The dataset used in this study is BloodMNIST (Yang et al., 2023), as shown in Figure 1. It is a subset of the MedMNIST dataset. This dataset contains images from ten different modalities from medical imaging, such as chest X-ray, mammogram, etc. Also includes 2D and 3D. This dataset is a fairly complete real medical image dataset, used in real life for medical image analysis tasks such as classification and segmentation. This article uses data preprocessing methods such as resampling, image resizing, data enhancement and normalization. It is converted into a grayscale image and undergoes Sobel edge detection, morphological closing operation, Otsu calculation and threshold segmentation to segment the cell nucleus from the background and generate a mask image (Xu et al., 2011).

2.2. Proposed Approach

This research aims to develop a deep learning model for medical image analysis to help doctors diagnose and treat diseases more accurately. The new architecture Blood_Net architecture is combined with an attention mechanism to develop a powerful blood cell classification and segmentation

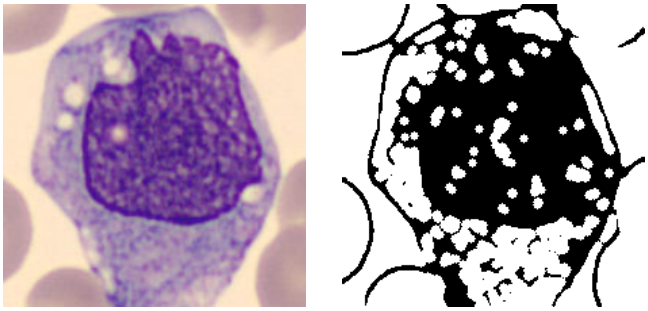

Figure 1: Blood dataset and its masked image. (Photo/Picture credit: Original).

framework and with interpretability technologies such as LEN and Shap to improve the model's interpretability and predictive performance. The Blood_UNet model is used to segment medical images, while LEN and Shap are used to explain the model's decision-making process and extract image features. It starts with data preprocessing and then feature extraction using the Blood_UNet architecture enhanced with attention mechanism. The extracted features are then fed into an auxiliary model for additional analysis and interpretation. This article built an end-to-end deep learning framework including Blood_UNet, LEN and Shap. Next, the model adopt the BloodMNIST data set as training and validation data, and use Dice Loss as the training target. During the training process, the model optimized the model through the Adam optimizer and evaluated the performance of the model on the validation set. Finally, this research use visualization techniques and metric evaluation to explain the model's prediction results and feature extraction process. In summary, this study aims to provide a deep learning framework to improve the accuracy and interpretability of medical image analysis. The pipeline as show in Figure 2.

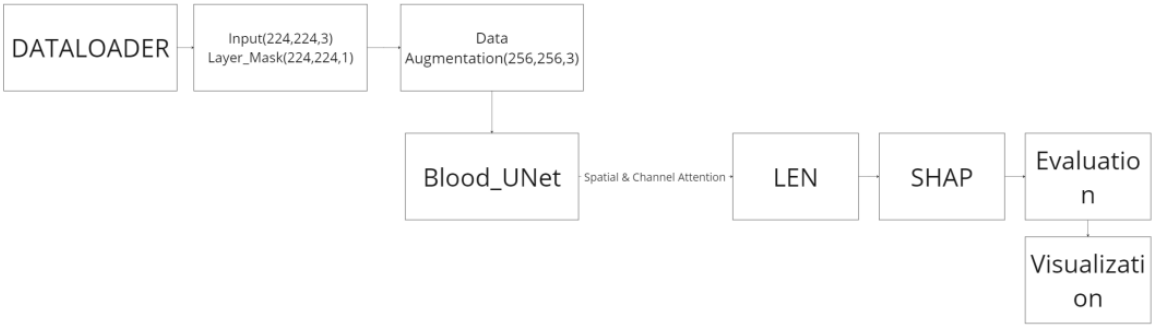

Figure 2: The pipeline of the model (Photo/Picture credit: Original).

2.2.1. COMBINATION OF BLOOD_UNET, LEN AND SHAP

This article with the core of the framework, is a modified U-Net (Ronneberger et al.) architecture with attention mechanisms – Blood_UNet. The basic architecture is shown in Figure 3. Specifically tailored for blood cell classification and segmentation in medical images. This model combines the strengths of the U-Net architecture for semantic segmentation with attention mechanisms to improve feature representation and localization. The Blood_UNet architecture consists of an encoder-

decoder structure, where the encoder captures the context information from the input image, and the decoder reconstructs the segmented output. This architecture's ability to capture both local and global features effectively. Additionally, spatial and channel attention mechanisms are integrated to enhance the model's ability to focus on informative image regions, thereby improving segmentation accuracy. The model extracts features at different scales, allowing it to capture both fine details and global context, crucial for accurate blood cell segmentation. Skip connections between encoder and decoder layers facilitate the propagation of spatial information, enabling precise segmentation of blood cells. Incorporating attention mechanisms helps the model focus on relevant image regions, enhancing segmentation accuracy in complex medical images.

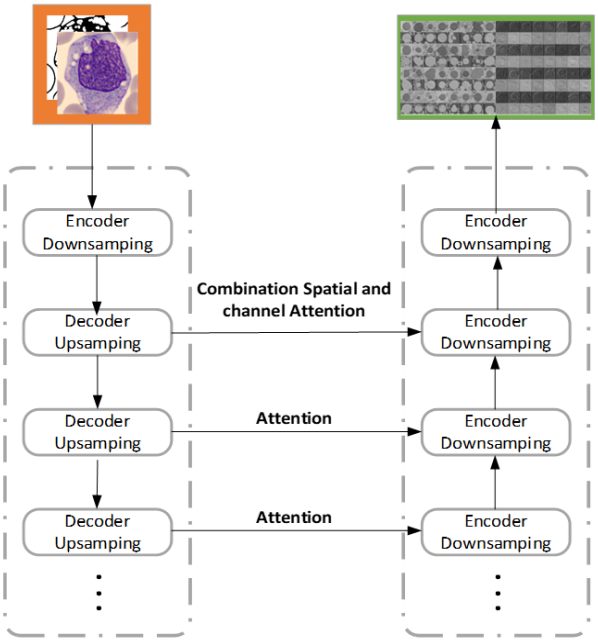

Figure 3: The most basic model (Photo/Picture credit: Original).

The paper introduces LEN and SHAP for analyzing the prediction results of a modified Blood_UNet model in vessel segmentation tasks. LEN is a type of local interpretability model designed to explain the predictions of individual samples or local regions. Aiding in understanding the model's decision-making process in different areas. It achieves interpretation by constructing an auxiliary model within the local region of input data.SHAP is a global interpretability method based on the concept of Shapley values, aiding in understanding the importance and influence of different features, and explaining model predictions by computing the contribution of features. By combining LEN and SHAP interpretability methods, the paper aims to improve the understanding and interpretability of the model for vessel segmentation tasks.

First, after obtaining predictions from the Blood_UNet model for individual samples or local regions, use LEN to analyze the prediction results of individual samples or local regions to obtain local explanations. Then, use SHAP to analyze the overall prediction behavior of the model to obtain global explanations. Finally, integrate both interpretability methods to enhance a comprehensive understanding of the model's prediction results. This comprehensive analysis helps identify

specific features that are significant at local and global levels, thereby improving the predictive interpretability and credibility of the Blood_UNet model in vessel segmentation tasks.

The combination of spatial and channel attention gates selectively highlights informative features while suppressing irrelevant information, improving the model's segmentation performance. The feature maps are weighted in channel and spatial dimensions to further improve the model's attention to different locations and features in the image. During training, the model is optimized using a Dice Loss function, to minimize segmentation errors. Hyperparameters such as learning rate and batch size are fine-tuned to achieve optimal performance.

By accurately segmenting blood cells, this model can help clinicians diagnose blood-related diseases more effectively and accurately. Figure 4 shows the joint architecture of this model. Furthermore, the model's modular design can be easily integrated with other auxiliary models for further analysis and interpretation, enhancing its usefulness in medical image analysis tasks.

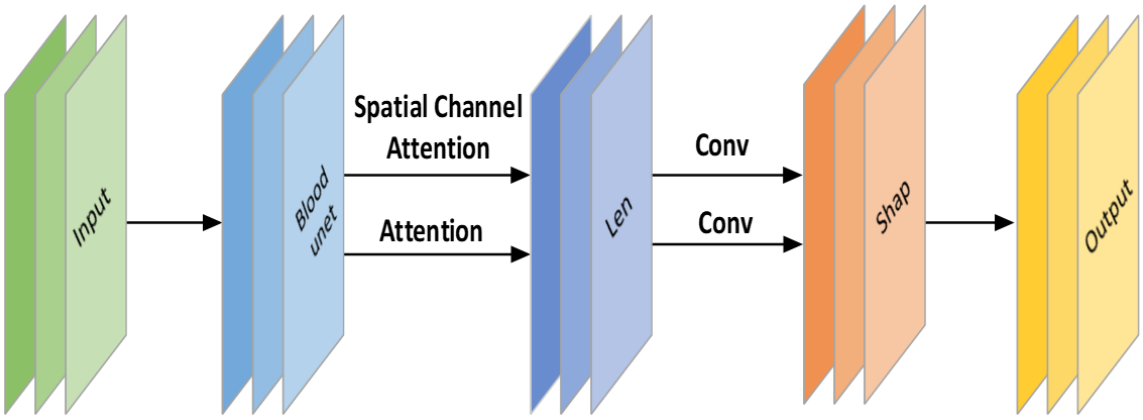

Figure 4: Combine results (Photo/Picture credit: Original).

2.2.2. LOSS FUNCTION

$$\text{Dice} = \frac{2 \cdot \times |\text{A} \cap \text{B}|}{|\text{A}| + |\text{B}|} \tag{1}$$

A represents the prediction area of the model, B represents the real label area and $|\,.\,|$ represents the size of the set. The Dice coefficient represents the degree of overlap between the model prediction area and the real label area, and the value range is from 0 to 1. A value of 1 means complete overlap, and a value of 0 means no overlap at all. DiceLoss is usually defined as 1 minus the Dice coefficient, so the lower the loss value, the better the model's performance.

DiceLoss is a loss function for image segmentation tasks that is more robust in dealing with class imbalance and aims to measure the degree of overlap between model predictions and true labels. It is based on the Dice coefficient, which is a commonly used metric to evaluate segmentation accuracy. DiceLoss is more sensitive to the pixel-level accuracy of predictions and therefore provides a better measure of a model's ability to predict subtle structures.

The DiceLoss loss function is used to measure the similarity between the model prediction results and the real labels during the model training process to guide the update of model parameters.

2.3. Implementation Details

There data augmentation techniques, including random horizontal flipping and rotation, were applied to improve model robustness. The experiments were performed on a Windows environment using Python 3.10. Model training was executed using PyTorch on a GPU setup. A dynamic learning rate scheduler was employed during training to optimize effectiveness. Regularization and prevention of overfitting were achieved through weight decay. The Adam optimizer facilitated efficient gradient descent, while gradient clipping was employed to control gradient size and prevent explosion. Training duration was determined by specifying the number of epochs.

3. Results and Discussion

As shown in Figure 5. By generating a heat map and superimposing it on the original image, the picture visually observes where the model focuses on the image. The heat map shows the degree of activation of different areas by the model, and the depth of the color indicates the degree of attention the model pays to that area. By overlaying the heat map onto the original image, these results can clearly show in which areas the model has a stronger response and thus infer which parts of the image the model is focusing its attention on. If the highlighted area of the heat map is located in the center of the image, it can be inferred that the model pays more attention to the features in the center of the image.

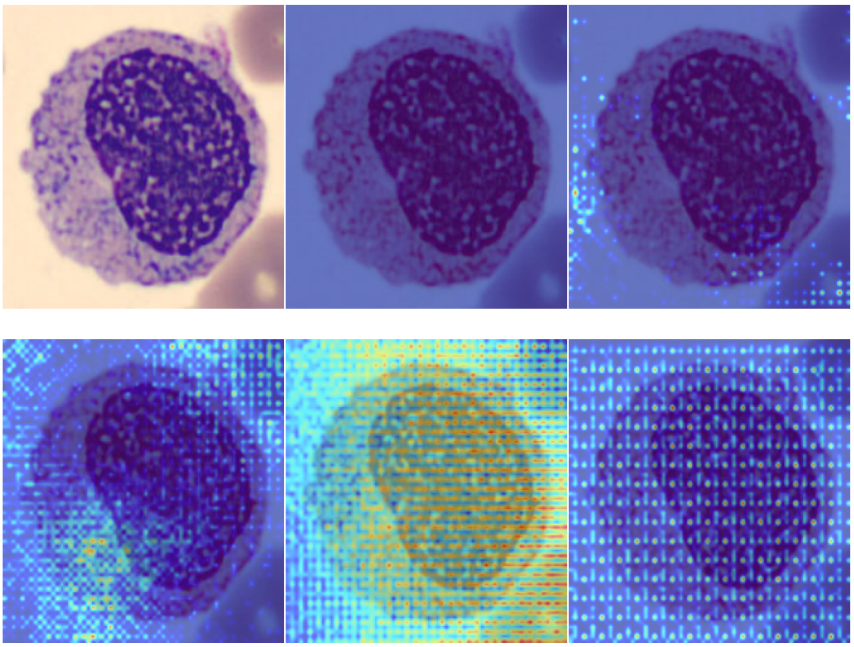

Figure 5: Heat map visualization results (Photo/Picture credit: Original).

The Figure 6 shown. By observing the feature maps of each layer, these pictures can understand the abstract representation of the input image by the model at different levels. These feature maps reflect the abstract features learned by the model at different levels, such as edges, textures, shapes, etc. By analyzing the changes in the feature map, this study can infer the model's understanding and abstraction capabilities of the input image at different levels and then optimize the model's structure

and parameter settings. If the model pays more attention to the edge parts of the image, may mean that the edge features are critical to the model's prediction results.

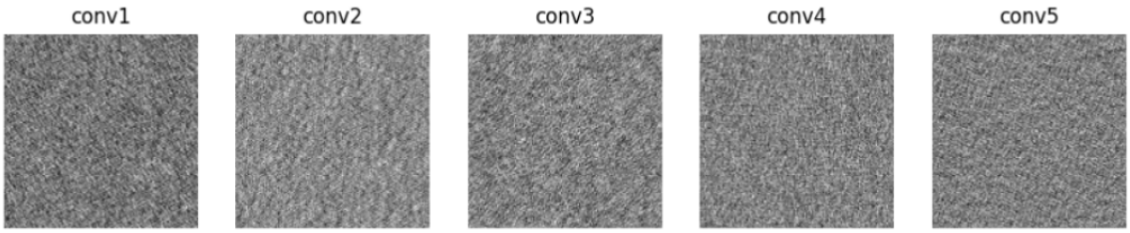

Figure 6: Feature map visualization results (Photo/Picture credit: Original).

Figure 7 shows the interpretation process of model prediction decisions. By visualizing the model's predictions and analyzing their basis, the model's predictions are mapped into the space of the input data to demonstrate the model's degree of focus or importance for different areas. The rules or patterns can be observed, which play a significant role in improving the confidence of the model. And identify the important features that the model focuses on during the prediction process, so as to be able to understand the prediction logic of the model. LEN can explain the model's prediction results through local linear approximation, while SHAP can provide a more global assessment of feature importance.

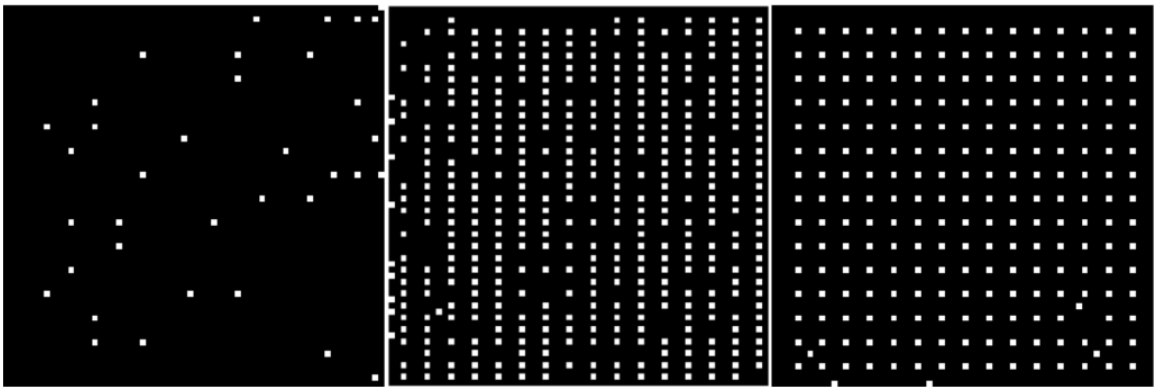

Figure 7: Prediction results (Photo/Picture credit: Original).

As shown in Figure 8. By visualizing the output results, we can intuitively understand the performance of the model on different output tasks, and analyze and compare the output results of the model. This comparison helps to discover the advantages and disadvantages of the model on different tasks and guides the improvement and optimization of the model. A SHAP model performs better on certain samples, the study may mean that it has stronger generalization or learning capabilities for a specific type of task. By deeply analyzing the differences in model output, we can propose targeted improvement strategies to improve model performance and effectiveness.

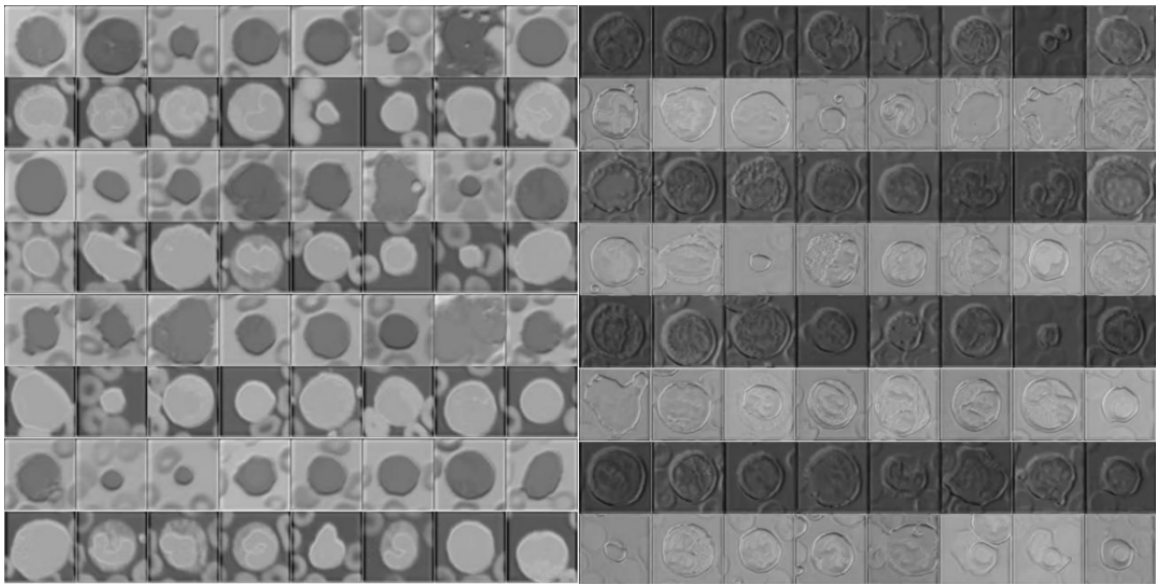

Figure 8: LEN and SHAP visualization results (Photo/Picture credit: Original).

4. Conclusions

This study introduces a new method for analyzing medical images, with a particular focus on anomaly detection in blood samples. Because of the crucial role of blood, the proposed method adopts a deep learning architecture, specifically the modified Blood_UNet model, which is tailored to handle complex medical image analysis. This model combines the attention mechanism and feature fusion technology, and finally the combination of LEN and SHAP interpretability, through extensive experiments and research on the BloodMNIST data set to enhance its performance in detecting subtle anomalies. Evaluate and demonstrate its efficacy in accurately identifying abnormal areas in blood samples. Future work will provide more in-depth analysis of specific types of abnormalities, such as rare cell types or abnormal cell distributions, to further refine the model's functionality and applicability in clinical settings.

References

Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, and Muhammad Khurram Khan. Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42(11):226, 2018. doi: 10.1007/s10916-018-1088-1.

Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, 2009. doi: 10.1016/j.media.2009.05.004.

Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4):e1312, 2019. doi: 10.1002/widm.1312.

Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. Cnn-based segmentation of medical imaging data, 2017.

Jinwei Liu, Yan Li, Guitao Cao, Yong Liu, and Wenming Cao. Feature pyramid vision transformer for medmnist classification decathlon. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. doi: 10.1109/IJCNN55064.2022.9892282.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241. Springer International Publishing.

Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016. doi: 10.1109/TMI.2016.2528162.

Erico Tjoa and Cuntai Guan. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11):4793–4813, 2021. doi: 10.1109/TNNLS.2020.3027314.

B. Woolf. On estimating the relation between blood group and disease. *Ann Hum Genet*, 19(4): 251–3, 1955. doi: 10.1111/j.1469-1809.1955.tb01348.x.

Xiangyang Xu, Shengzhou Xu, Lianghai Jin, and Enmin Song. Characteristic analysis of otsu threshold and its applications. *Pattern Recognition Letters*, 32(7):956–961, 2011. doi: 10.1016/j.patrec.2011.01.021.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2 - a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023. doi: 10.1038/s41597-022-01721-8.