Developing a Fundamental Understanding of Explainable Artificial Intelligence using Texture Analysis

Md Masum Billah[1], Pragati Manandhar[1], Sarosh Krishan[1], Alejandro Cedillo Gamez[1], Kurt K. Benke[2], Sepinoud Azimi[1,3], Janan Arslan[4]

Affiliations: 1. EMJMD Program, EDISS, Faculty of Science and Engineering, Information Technology, Åbo Akademi University, 20500, Turku, Finland.
2. School of Engineering, University of Melbourne, Parkville, Victoria 3010, Australia.
3. Department of Information and Communication Technology, Technology, Policy and Management, Delft University of Technology, Delft, 2600GA, The Netherlands.
4. Sorbonne Université, Institut du Cerveau—Paris Brain Institute—ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013 Paris, France.

Contact Email: Md.Billah@abo.fi

Introduction

Concerns about the opacity of AI models in cancer diagnosis have led to the emergence of Explainable AI (XAI) focused on creating transparent and interpretable medical AI technologies. Current XAI techniques like Class Activation Maps (CAMs), SHAP, and Quantus provide only partial understanding. This project proposes a novel approach using texture analysis in cancer datasets to enhance AI model interpretability. By emphasizing specific texture features in medical images during training, the goal is to elucidate how AI models make predictions and improve trustworthiness. The project aims to bridge the gap between AI predictions and human understanding in cancer diagnosis through rigorous analysis of texture features and their correlation with AI predictions.

Methodology

The overall pipeline is illustrated in Figure 1.

1. TrainSegmentation Model: Utilized U-Net and DeepLabv3 on CBIS-DDSM dataset for precise tumor delineation.

2. Texture Feature Extraction: Employ GLCM, LBP, and aTEM methods to extract texture features from raw images and segmentation model feature maps.

3. Correlation Analysis: Correlation of texture energy maps from model layers with extracted features, identifying influential texture types and features across different model levels.

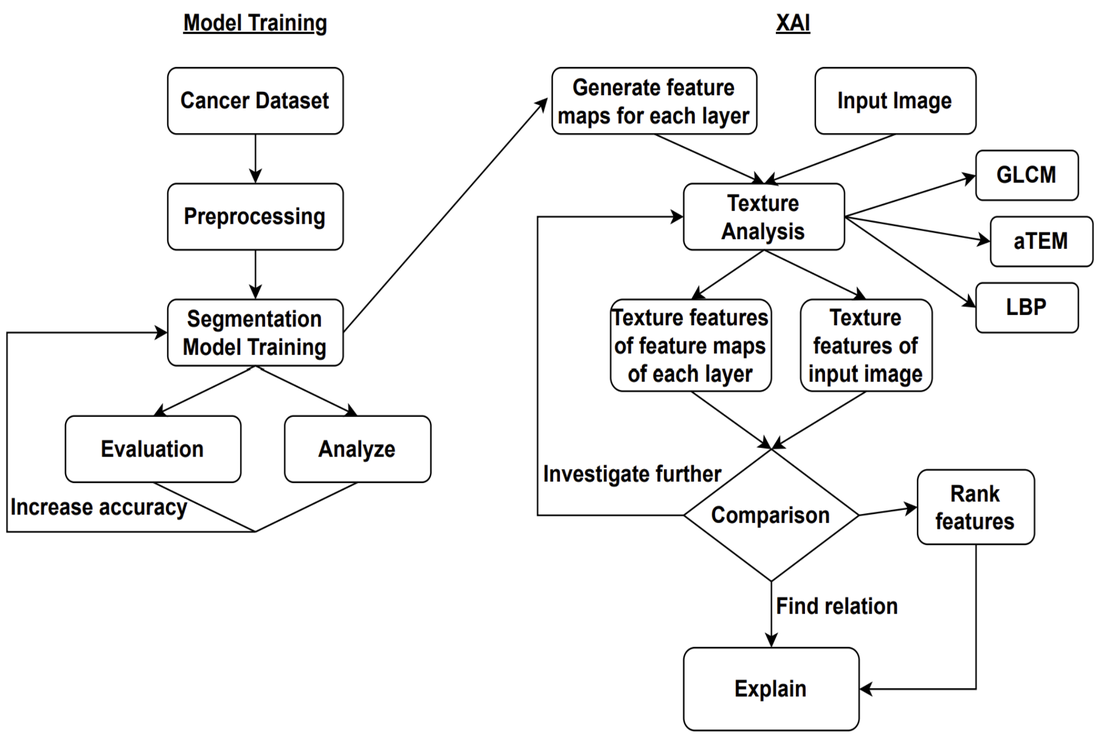


Figure 1: XAI Pipeline.

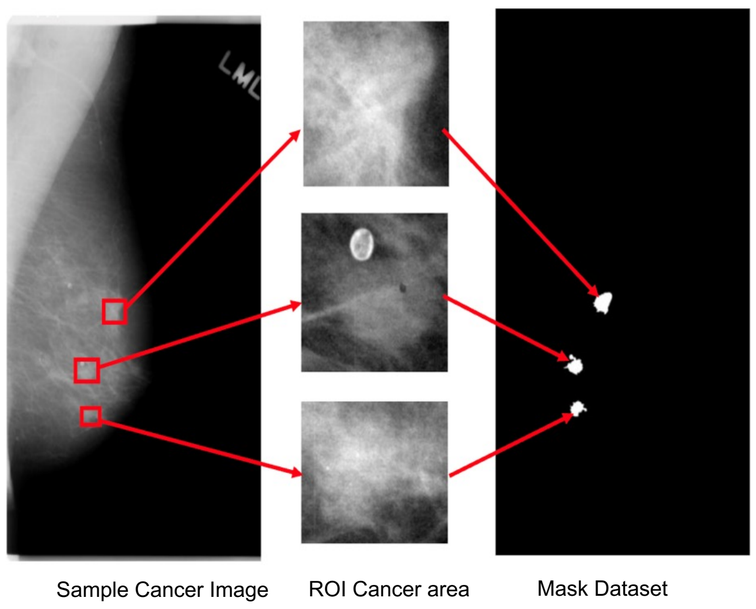


Figure 2: Sample breast cancer image from CBIS-DDSM dataset.

Preliminary Results and Analyses

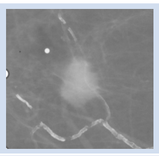
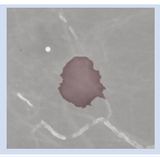
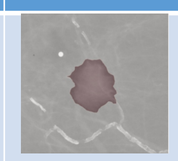
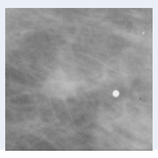
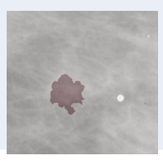
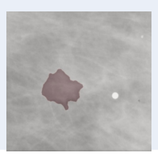
Model	Input Image	Ground Truth	Predicted Mask	mIoU
Deeplabv3				93.0
UNet				77.82

Table 1: Model performance for input images

We obtained feature maps across four layers of our model for an input image, as illustrated in Figure 3, depicting the first 10 feature maps for each layer. This visualization highlighted the evolution of features as input images traversed different layers. Our focus was on texture features influencing the model's mask predictions, with a specific emphasis on GLCM features. Extracting 13 GLCM features from the input image, including ASM, Contrast, Correlation, Variance, IDM, Homogeneity, Sum Entropy, Entropy, Difference Entropy, IMC1, IMC2, MCC, and Autocorrelation, allowed for a detailed analysis of influential factors in our model's decision-making process.

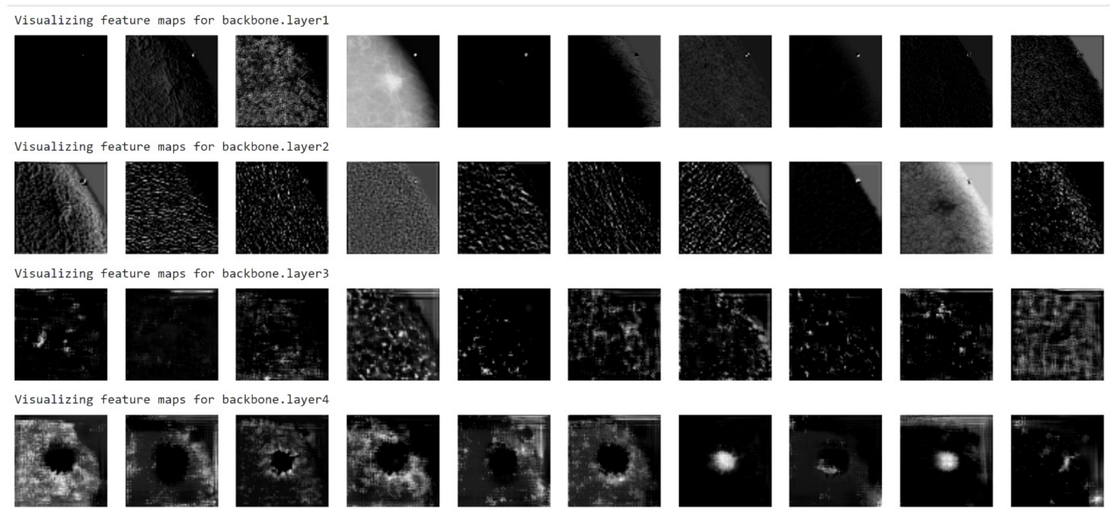


Figure 3: Layer-wise feature maps (First 10 for each layer)

Position	First Approach	Second approach
Layer 1		
First	IMC1 (Score: 0.00116)	IMC1 (Score: 0.33411)
Second	ASM (Score: 0.11807)	ASM (Score: 0.41204)
Third	Autocorrelation (Score: 0.13050)	Autocorrelation (Score: 0.41806)
Layer 2		
First	IMC1 (Score: 0.00084)	IMC1 (Score: 0.33389)
Second	ASM (Score: 0.13291)	ASM (Score: 0.42194)
Third	Autocorrelation (Score: 0.16339)	Autocorrelation (Score: 0.44226)
Layer 3		
First	IMC1 (Score: 0.00265)	IMC1 (Score: 0.33510)
Second	IDM (Score: 0.25426)	IDM (Score: 0.50284)
Third	Autocorrelation (Score: 0.31222)	Autocorrelation (Score: 0.54148)
Layer 4		
First	IMC1 (Score: 0.000755)	IMC1 (Score: 0.3338)
Second	Autocorrelation (Score: 0.02375)	Autocorrelation (Score: 0.3491)
Third	ASM (Score: 0.0745)	ASM (Score: 0.3830)

Table 2: Comparison of 13 GLCM Features Extracted from 2048 Feature Maps in Each Layer of the Model Using Average Absolute Differences and Various Distance Metrics.

Further Works

Our work continues with the successful implementation of the GLCM method for texture feature extraction in our cancer diagnosis segmentation model. Ongoing efforts involve integrating the aTEM method and progressing with our correlation approach to link texture energy maps with analysis outcomes. These additions aim to deepen our understanding of influential features across different layers, enhancing the interpretability and accuracy of our AI system in cancer diagnosis.

Conclusion

Our project significantly improves transparency in cancer diagnosis through AI. Our robust segmentation model, trained on a relevant dataset, incorporates advanced texture analysis and ongoing correlation efforts, ensuring more interpretable and reliable medical outcomes. This collaborative approach marks a significant step in bridging the gap between AI predictions and human understanding in cancer diagnosis.

References

1. Laws. (1980). Textured Image Segmentation.
2. Ertugrul. (2014). Adaptive Texture Energy Measure Method. Doi: 10.11648/j.ijiis.20140302.11
3. Ronneberger, Fischer, & Thomas Brox. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597
4. Chen, Papandreou, Schroff, & Adam. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv:1706.05587
5. Scott M. Lundberg, Su-In Lee. A Unified Approach to Interpreting Model Predictions. Part of Advances in Neural Information Processing Systems 30 (NIPS 2017).
6. J. Smith et al. Quantus: An Explainable AI Toolkit for Responsible Evaluation of Neural Network Explanations and Beyond . arXiv:2202.06861.
7. Lee, R., Gimenez, F., Hoogi, A. et al. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 4, 170177 (2017). https://doi.org/10.1038/sdata.2017.177