Improving Medical Image Identification with Quantum Machine Learning

PURDUE UNIVERSITY | The Data Mine

PURDUE UNIVERSITY | The Data Mine

INTRODUCTION

This project integrates quantum machine learning and data augmentation techniques to enhance image classification within the INbreast dataset. Our research aims to explore the real-world application of quantum computing in the medical industry., facilitating more accurate image identification and, ultimately, improving on current medical imaging diagnostics. Current diagnostic methods are expensive and time consuming as they are done by hand by specialists in the field. Our model aims to streamline this process by identifying regions of interest (ROI's) which can then be sent to the specialist for review reducing time spent on images and reducing cost of labor.

RESEARCH METHODOLOGY

- INBreast dataset Review and Presentation
- Data Augmentation: artificially increases diversity of training dataset through rotation and contrast adjustment; improves diagnostic accuracy by improved generalization by models
- Quantum Information Processing (QIP) research/UNet research and Review
- U-NETS: convolutional neural network specifically for biomedical image segmentation with higher precision consisting of an encoder-decoder structure.
- Quantum Filters: filter and process quantum information, resulting in clarification and de-blurring of images
- Training convergence with and without quantum filters applied: this changes the depth of the images and placement of objects within the image. This means, altering the images in order to expose unique features about them that otherwise may not be observable in the raw data.
- Testing and Debugging: trying different parameters in augmentation and selecting appropriate test cases based on the outputs as well as debugging existing methods, code blocks to ensure error-free application of filters. Parameters that we view would be regarding noise, resolution, and remanent artefacts in the images. The INBreast dataset is relatively noiseless and is high quality, thus cleaning up artefacts in the data using masks is essential to debugging.

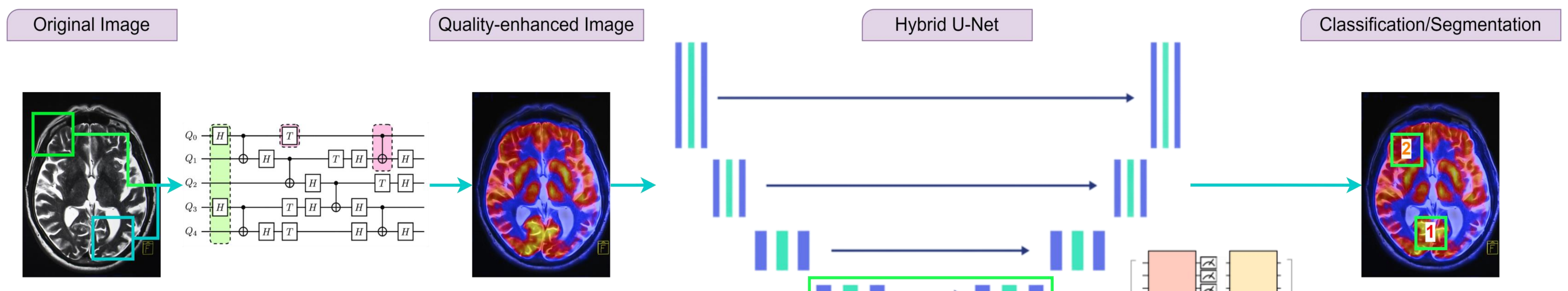

Figure 1. Our pipeline takes MRI data and with the use of quantum filters enhances the image quality to make predictions on tumor segmentation and classification with the use of U-NET and CNN.

FUTURE GOALS

We are committed to exploring the untapped potential of accelerating analysis through quantum filters and refining our data augmentation techniques. Our ongoing research into tensor network compression and the development of generative models promises to further elevate model efficiency and accuracy, bringing us closer to our vision of reducing misdiagnoses.

Goals:

- Continue exploration of quantum filters as a means to enhance the training dataset for our pipeline and model
- Expand our classical data augmentation techniques by implementing more advanced or diverse ways to modify and enhance our data using a wider range of transformations to generate new data points from our existing dataset.
- Explore tensor network decomposition to decrease the complexity of the model and reduce training time
- Tune all of these parameters to optimize speed and accuracy for our model pipeline

Identified Clusters: 2

Identified Clusters: 4

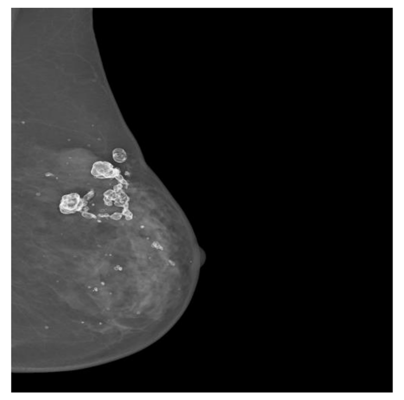

Figure 4.

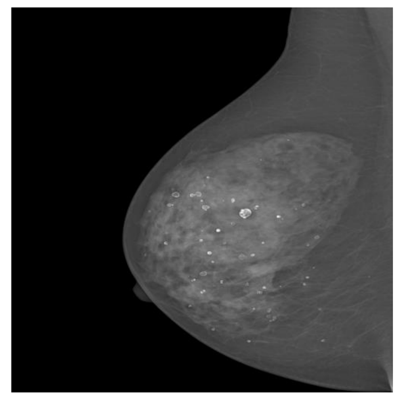

Figure 5.

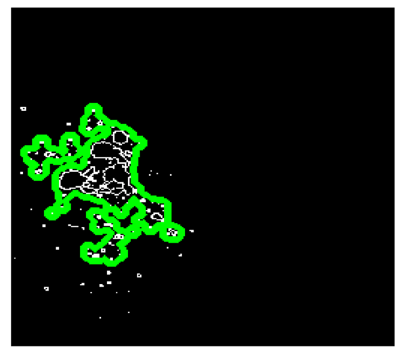

Figure 6.

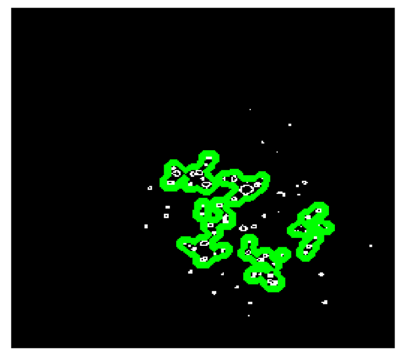

Figure 7.

BACKGROUND

Leveraging the INbreast dataset, renowned for its clinical application in mammogram analysis, we encountered the challenge of its complexity and variability in tumor appearance. To address these, our project utilizes data augmentation techniques to enhance dataset quantity and model performance, enabling and expanding upon current research in the intersection between quantum computing and its real-world applications.

MOTIVATION

Annually, over 12 million adults are misdiagnosed in the United States alone, a statistic that underscores the urgent need for diagnostic precision. Our project is driven by the potential of quantum computing and data augmentation to surmount the limitations of classical machine learning and current identification methods, offering a leap forward in the generalization capabilities of models across complex imaging data.

CONCLUSIONS

So far, we have been able to produce working results for the preprocessing and data augmentation sections of the project.

Data Augmentation:

2 Types of Data Augmentation:

- Rotation of Image, which rotates images 90, 180, and 270 degrees.
- Contrast Limited Adaptive Histogram Equalization (CLAHE), which alters contrast of the image.

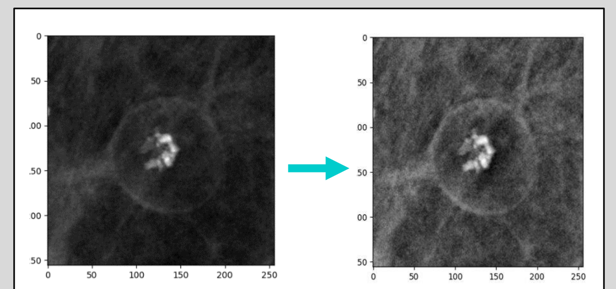

Figure 2. Example of CLASHE augmentation.

Mask Creation:

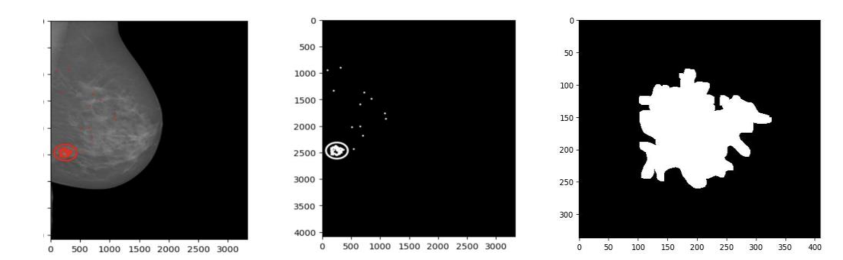

Figure 3. Process of identifying & zooming into the region of interest

Image 1: Unprocessed mammogram containing region of interest.
Image 2: Identified clusters using k-nearest neighbors' method.
Image 3: Create mask of region of interest (ROI) and zoom in.

Limitations:

- There exists very limited amounts of annotated mammograms, allowing for less data to be run through the model, impacting performance.
- Quantum hardware constraints exist as well as quantum computers are not readily available.
- The complexity of the models themselves can limit one's ability to use it efficiently as it would require a large learning curve.

Biases:

- The main biases exist in the data that we train the model with.
- Because we don't know for sure that the data represents the entire population, accounting for age, ethnicity, or a particular type of breast cancer, the model may not accurately work for all demographics.

Figures 4 and 5 show the identification of clustered regions of interest (ROI). The ROIs are segmented and highlighted for efficient interpretation for medical specialists, shown in figures 6 and 7.

ACKNOWLEDGEMENTS

- Lauren Dalder- Corporate Partner Advisor
- Shawn Yuxun- Teaching Assistant, Mentor
- Mark Daniel Ward – Executive Director

REFERENCES

Baccouche A, Garcia-Zapirain B, Castillo Olea C, Elmaghraby AS. Connected-UNets: a deep learning architecture for breast mass segmentation. NPJ Breast Cancer. 2021 Dec 2;7(1):151. doi: 10.1038/s41523-021-00358-x. PMID: 34857755; PMCID: PMC8640011.
Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Acad Radiol. 2012 Feb;19(2):236-48. doi: 10.1016/j.acra.2011.09.014. Epub 2011 Nov 10. PMID: 22078258.

Anna Nierzwick, Anuraag Kolli, Armann Kokan, Elijah Van Grootheest, Bhaskaram (KB) Sarma, Navin Bhat, Tejas George, Yana Gomon

The Data Mine Corporate Partners Symposium 2024