UR-124 BreastNet: Breast Cancer Survival Prediction With Deep Learning

Abstract

In the United States, 12% of women are diagnosed with breast cancer in their lifetime, and it is the second leading cause of cancer-related death in women. Early detection and screening can improve the patient's life expectancy of 10 years on average. Unfortunately, breast cancer can be challenging to detect, since it can appear anywhere in the breast tissue. Detecting the cancer before it has progressed into aggressive stages can give patients more options and save thousands of dollars in medical costs. Magnetic Resonance Imaging (MRI) produces high quality images that can be used by Radiologists as a diagnostic tool for breast cancer. Additionally, advances in Machine Learning technology enable computers to learn from high-quality labeled datasets to achieve human-like ability in well-defined problem area.

Research Question(s)

The main goal of BreastNet is to predict the 5-year survival rate of breast cancer patients. We use a Convolutional Neural Network (CNN) to predict survival based on pre-processed MRI data, and a Recurrent Neural Network (RNN) to predict survival based clinical data. By implementing dual models based upon different data sources, we can create a robust classification system that makes data-driven decisions based upon the most information possible. We trained the CNN using 12254 sagittal slices from 340 MRI volumes for 222 patients, and we trained the RNN using a language-vector model with vocabulary tokenization. Additionally, we employ methods such as stochastic resampling to guard against class imbalance bias. We apply a disease-specific survival censor to enable the models to use binary classification for each input tensor. Since breast cancer is a disease that drastically affects people's lives, there is a great need for technology that can help patients make informed decisions about the treatment plan they wish to pursue.

Introduction

When patients are diagnosed with breast cancer, it is important for them to understand how serious the disease is, and they rely on experts like doctors and radiologists to interpret a wide range of medical data in their record, which includes high quality images as well as clinical data and patient history. By understanding the risk score of high and low risk, the patients are able to make the right decisions about their future treatment plans. CNNs have the ability to make use of medical images from a variety of sources to aid in the diagnosis and prognosis of diseases like cancer. Since breast cancer is so common in women, there are large publicly available datasets that contain a large number of Magnetic Resonance Images (MRI). The I-SPY clinical trial was used to train BreastNet, and it is a publicly available dataset which is available through The Cancer Imaging Archive.

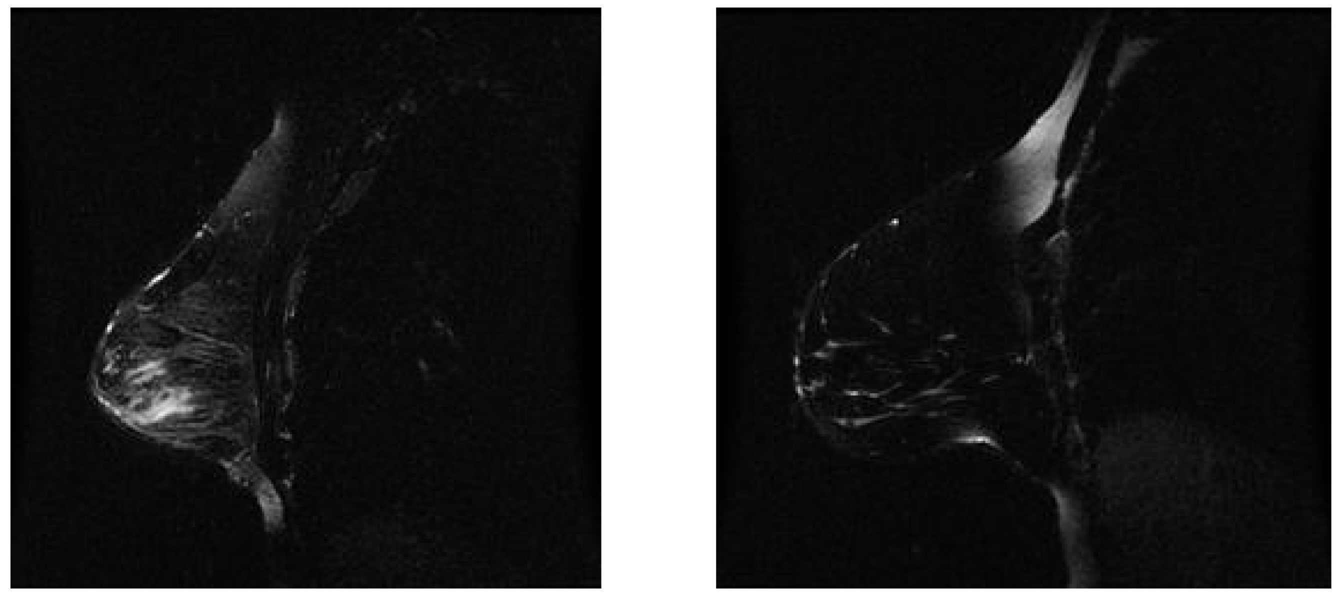

Fig.1-2 Two-dimensional Sagittal cross-sectional images of the MRI scans from the I-SPY1 ACRIN Clinical Trial, obtained from The Cancer Imaging Archive

Materials and Methods

We used MRI images and Clinical Data from the ACRIN 6657 Clinical Trial, which is publicly available on The Cancer Imaging Archive. We analyzed the image data using a Convolutional Neural Network (CNN), and we analyzed the clinical data using a fully connected Artificial Neural Network (ANN). We split this data into a training, validation, and testing partitions. We used the data to train the ResNet101 CNN model which then classified whether a patient in the image was likely to survive or not after 5 years. We trained the network for 100 epochs on 90% of our dataset. Finally, 10% was used for cross-validation and 10% was used for testing were used from the dataset.

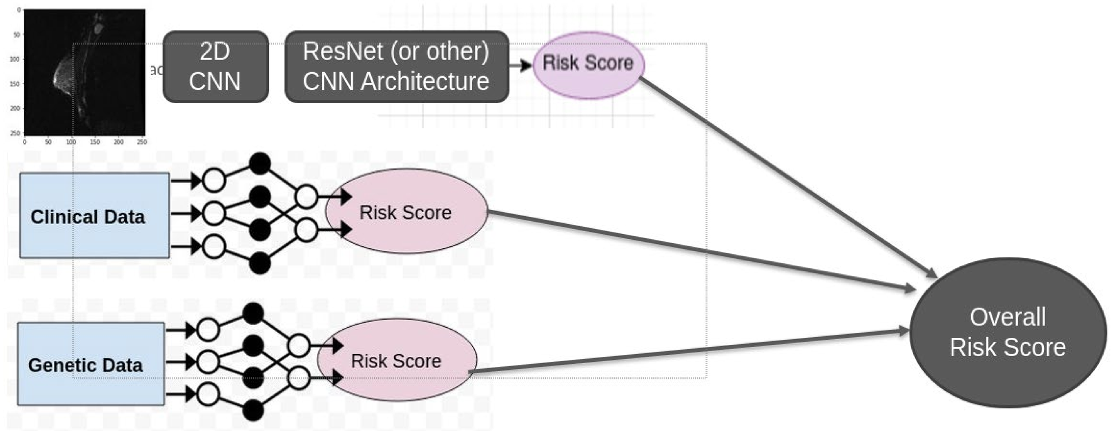

Fig.7 BreastNet architecture that makes use of the full implementation of networks that accept clinical, image, and genetic datasets since each sub-model is specialized for each data type. This framework could be used with ensemble learning techniques like Boosting and Bagging

Results

The ResNet CNN accuracy score of about 98% between our training, testing, and validation sets. After training it for 100 epochs we got an AUC score of 0.9986. This is the AUC for when positive is when someone survives within the 5 years. Then we also had an F1 score of 0.9980.

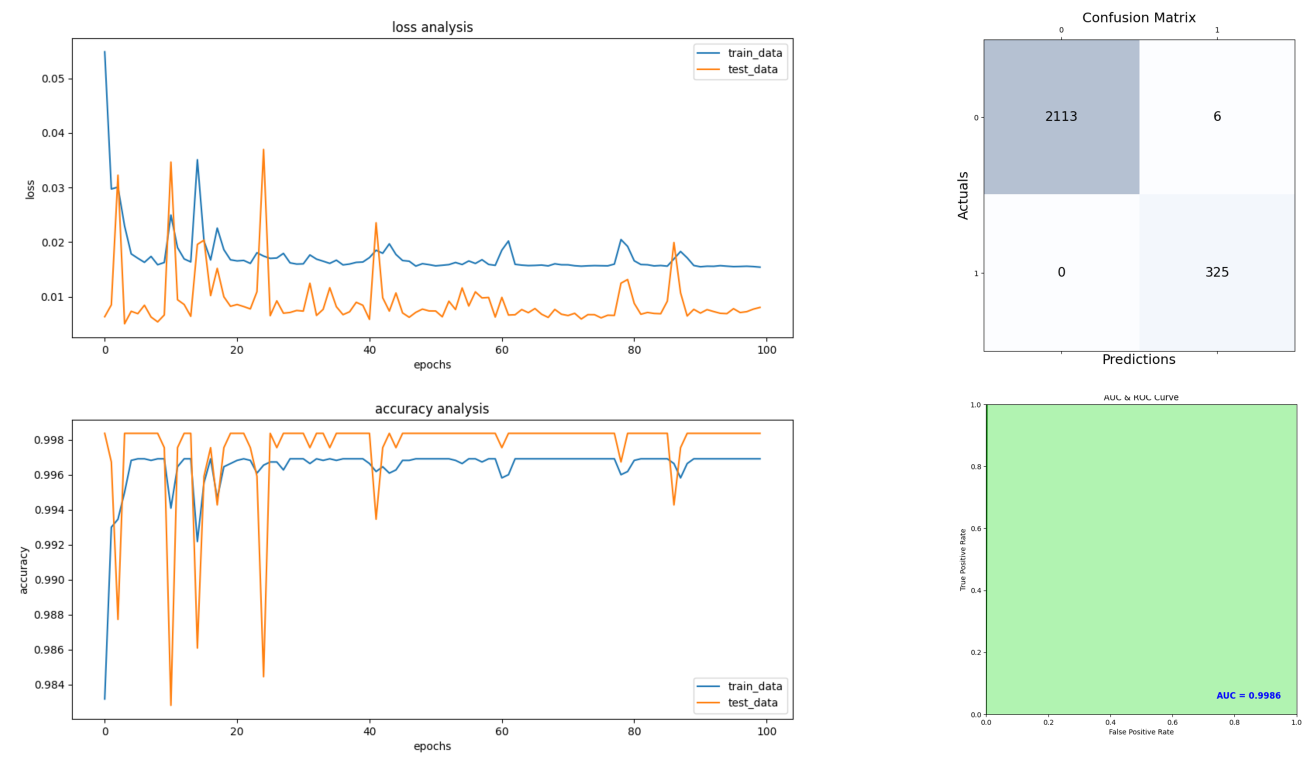

Fig.3-6 Top left: Loss function over each training epoch. Bottom left: Accuracy function over each training epoch. Top right: Confusion Matrix. Bottom Right: Receiver Operating Curve, with AUC=0.9986

Conclusions

BreastNet is a model that contributes to both the medical field and data science. In particular, some of the data preprocessing work can be valuable to other data scientists who work on similar problems in biostatistics. The inclusion of a program that is able to convert a DICOM format, which is highly isolated and mostly used by the medical community alone, into the numpy format, can make it easier for people who primarily manipulate python-native data formats.By creating a network that directly manipulates 3D numpy arrays from directories, BreastNet paves the way for general purpose computational intelligence to be used to model 3 dimensional image data in a simple, straightforward way BreastNet contributes to a growing body of studies that use machine learning to help cancer patients understand their diagnosis and make informed treatment decisions \cite{review}. By understanding whether they have a high or low risk of death within the next five years, they will have agency to maximize their quality of life. By focusing on specific results that patients find most important when learning about their breast cancer diagnosis, BreastNet can help make a positive impact on people's lives; one that goes beyond the simple results the program generates.

References

[1] "How common is breast cancer?: Breast cancer statistics,"2021. [Online]. Available: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html

[2] David Newitt, Nola Hylton, on behalf of the I-SPY 1 Network and ACRIN 6657 Trial Team. (2016). Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.HdHpgJLK

[3] P. Mukherjee, M. Zhou, E. Lee, A. Schicht, Y. Balagurunathan, S. Napel, R. Gillies, S. Wong, A. Thieme, A. Leung, and O. Gevaert, "A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomographyimage datasets,"Nature Machine Intelligence, vol. 2, no. 5, pp. 274–282, 2020.

[4] "Biomarkers to guide treatment for metastatic breast cancer," 2019, [Online]. Available: https://www.cancer.net/research-and-advocacy/asco-care-and-treatment-recommendations-patients/biomarkers-guide-treatment-metastatic-breast-cancer

[5] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin,"Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features,"IEEE Access, vol. 7, pp.105 146–105 158, 2019.

[6] S. U. C. A. Katzman, J.L., "Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network, "BMC Medical Research Methodology, vol. 18,no. 24, 2018. [Online]. Available: https://doi.org/10.1186/s12874-018-0482-1

[7] L.-H. Cheng, T.-C. Hsu, and C. Lin, "Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction,"Scientific Reports, 2021

[8] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society. Series B (Methodological), vol.34, no.2, pp.187–220, 1972. [Online]. Available: http://www.jstor.org/stable/2985181

[9] X. Zhou, Y. Li, and W. Liang, "Cnn-rnn based intelligent recommendation for online medical pre-diagnosis support,"IEEE/ACMTransactions on Computational Biology and Bioinformatics, vol. 18,no. 3, pp. 912–921, 2021.

[10] G.R.XuY,"On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalizationperformanceofsupervisedlearning."JournalofAnalysisand Testing, vol. 2, p. 249–262, 2018. [Online]. Available:https://doi.org/10.1007/s41664-018-0068-2

Contact Information

Ryan Deem's LinkedIn: https://www.linkedin.com/in/rtdeem/
Ryan Deem's email: rdeem@students.kennesaw.edu
Cora Meador LinkedIn: https://www.linkedin.com/in/cora-meador/
Cora Meador email: cmeador6@students.kennesaw.edu

Acknowledgments

We would like to thank our faculty advisor Dr. Mohammed Aledhari, the KSU College of Computer Science and Software Engineering, and the Office of Undergraduate Research for supporting our work

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Authors: Ryan Deem and Cora Meador
Advisor: Dr. Mohammed Aledhari