UR-46 BreastNet; A CNN Model for Breast Cancer Prognosis

Abstract

Cancer that is detected in its early stages can give patients more options and save thousands of dollars in medical costs. Some of the most recent developments in computer science and machine learning are in the biomedical field, especially individualized healthcare. There is also an increase in the demand for telehealth options, reducing healthcare costs. With the help of computational technology, medical practitioners will be able to process data more quickly, which will allow more patients to have access to reliable treatment. Besides, systematic processes for interpreting various data types (such as clinical features, genetic information, and medical images) can identify trends that a human eye would not detect.

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.

Research Question(s)

How well can we predict the 5 year life expectancy of Breast cancer based on patients' MRI and clinical data?
We used MRI and clinical data from publicly available datasets found through The Cancer Imaging Archive. The ACRIN I-SPY clinical trial included MRI images and survival information that was used to model a binary classifier based on censored survival data. Data was processed into training and test directories, each containing two survival classifications.

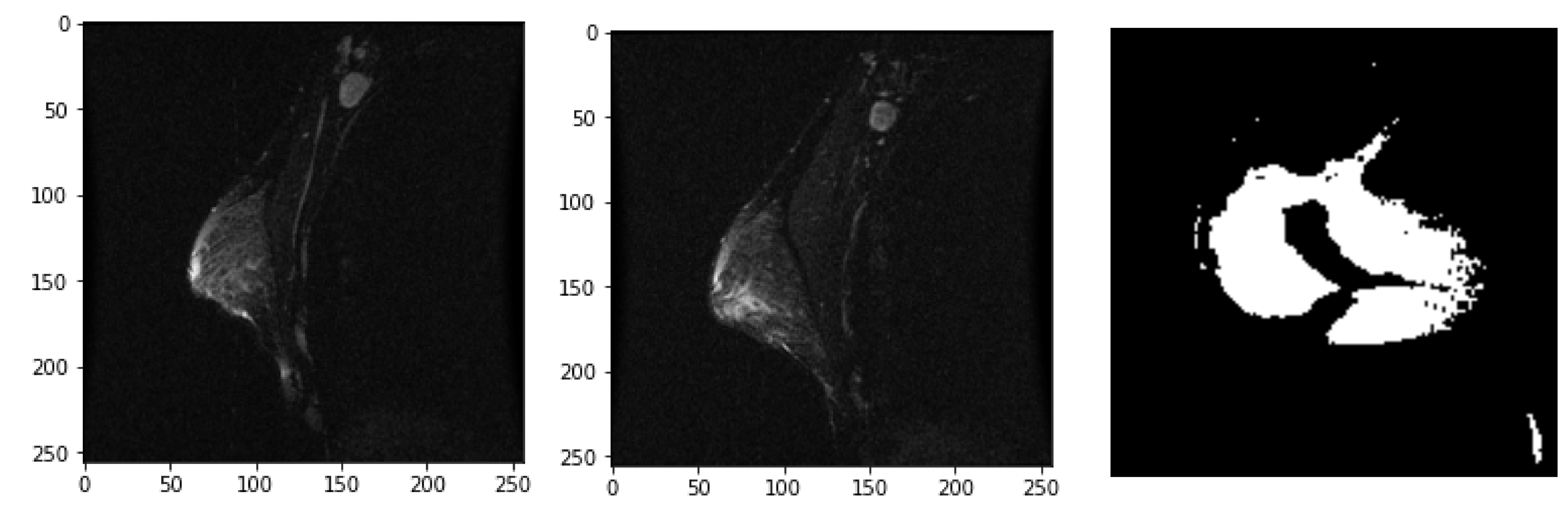
Materials and Methods

The initial datasets were composed of folders containing slices of DICOM data that were organized by patient, time relative to treatment, and view type (eg. Scout, Sagittal, etc). Since the model was constructed using Keras and Tensorflow libraries, we needed a data that was in a standardized format for quick array processing. We created a preprocessing program using the pydicom library to convert a DICOM directory from a series of 2D Sagittal view images to a 3D numpy array. We also processed a separate "region of interest" (ROI) image, which was provided by radiologists, into a sparse 3D numpy array. The ROI could be used as a bitmap, which can be used to focus the training of the model on image details directly related to the breast cancer tumors, which can reduce some of the noise found in the high-dimensional data. Some of the benefits of using 3D numpy arrays include the fact that it can be sliced in any number of dimensions, and that it can be easily processed by python scripts, which can open the problem of medical image analysis to include computer scientists who use python as their primary tools of choice.

Model

We created a prototype for a model that took in 3D MRI images containing different scans of breast cancer. The model is a shallow 3D Convolutional Neural Network that accepts numpy array inputs with the shape (1 x 256 x 256 x 36).

Data Preprocessing



We created a prototype for a model that took in 3D MRI images containing different scans of breast cancer. The model is a shallow 3D Convolutional Neural Network that accepts numpy array inputs with the shape (1 x 256 x 256 x 36).

Challenges

Some of the major challenges related to this project include the fact that Keras does not natively support 3D image preprocessing. Some of these steps must be done. As a result, we needed to use custom libraries to generate data to train the network.

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

Conclusions

BreastNet contributes to a growing body of studies that use machine learning to help cancer patients understand their diagnosis and make informed treatment decisions. 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.

Acknowledgments

We would like to thank our advisor Mohammed Aledhari for his support for our work, and th the KSU College of Computer Science and Software Engineering

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] P. Mukherjee, M. Zhou, and E. Lee et al., "A shallowconvolutional neural network predicts prognosis of lungcancer patients in multi-institutional computed tomographyimagedatasets,"p.274–282,2020.[Online].Available:https://doi.org/10.1038/s42256-020-0173-6
[3] "Breast mri for screening," Sep 2016. [Online]. Available:https://www.breastcancer.org/symptoms/testing/types/mri/screening
[4] D.NewittandN.Hylton,"Multi-centerbreastdce-mri data and segmentations from patients in the i-spy1/acrin6657trials,"accessed:2021.[Online].Available:https://doi.org/10.7937/K9/TCIA.2016.HdHpgJLK
[5] E. R. e. a. KC, Fontham ETH, "Breast cancer screeningforwomenataveragerisk,"2015.[Online].Available:doi:10.1001/jama.2015.12783
[6] "Biomarkerstoguidetreatmentformetastaticbreastcancer,"2019,accessed:19-Mar-2021.[Online].Available:https://www.cancer.net/research-and-advocacy/asco-care-and-treatment-recommendations-patients/biomarkers-guide-treatment-metastatic-breast-cancer
[7] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, andJ. Schmidhuber, "Fast image scanning with deep max-pooling con-volutional neural networks," in2013 IEEE International Conferenceon Image Processing, 2013, pp. 4034–4038.
[8] S. U. C. A. Katzman, J.L., "Deepsurv: personalized treatmentrecommender system using a cox proportional hazards deepneural network,"BMC Medical Research Methodology, vol. 18,no. 24, 2018. [Online]. Available: https://doi.org/10.1186/s12874-018-0482-1
[9] "Neoadjuvant chemotherapy for breast cancer: Functionaltumorvolumebymrimagingpredictsrecurrence-freesurvival—results from the acrin 6657/calgb 150007 i-spy 1trial,"Radiology, vol. 279, no. 1, 2015. [Online]. Available:https://pubs.rsna.org/doi/10.1148/radiol.2015150013
[10] P. Herent, B. Schmauch, P. Jehanno, O. Dehaene, C. Saillard,C. Balleyguier, J. Arfi-Rouche, and S. J´egou, "Detection and char-acterization of mri breast lesions using deep learning,"Diagnosticand Interventional Imaging, vol. 100, no. 4, pp. 219–225, 2019.
[11] P. Harish and S. Baskar, "Mri based detection and classificationof brain tumor using enhanced faster r-cnn and alex net model,"Materials Today: Proceedings, 2020.
[12] A. Farooq, S. Anwar, M. Awais, and S. Rehman, "A deep cnn basedmulti-class classification of alzheimer's disease using mri," in2017IEEE International Conference on Imaging Systems and Techniques(IST), 2017, pp. 1–6.
[13] Z. A. Corbin and S. Nagpal, "Leptomeningeal Metastases,"JAMAOncology, vol. 2, no. 6, pp. 839–839, 06 2016. [Online]. Available:https://doi.org/10.1001/jamaoncol.2015.3502
[14] J.-C. J. et. al., "Clinical features and outcome of leptomeningealmetastasis in patients with breast cancer: a single centerexperience,"Cancer Chemotherapy and Pharmacology, 2013. [Online].Available:https://link.springer.com/article/10.1007%2Fs00280-013-2185-yciteas
[15] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin,"Breast cancer detection using extreme learning machine based onfeature fusion with cnn deep features,"IEEE Access, vol. 7, pp.105 146–105 158, 2019.
[16] Shih-Chung B Lo, Huai Li, Yue Wang, L. Kinnard, and M. T.Freedman, "A multiple circular path convolution neural networksystem for detection of mammographic masses,"IEEE Transactionson Medical Imaging, vol. 21, no. 2, pp. 150–158, 2002.
[17] A. P. Pawlovsky and M. Nagahashi, "A method to select a goodsetting for the knn algorithm when using it for breast cancerprognosis," inIEEE-EMBS International Conference on Biomedicaland Health Informatics (BHI), 2014, pp. 189–192.
[18] Naveen, R. K. Sharma, and A. Ramachandran Nair, "Efficientbreast cancer prediction using ensemble machine learning mod-els," in2019 4th International Conference on Recent Trends on Elec-tronics, Information, Communication Technology (RTEICT), 2019, pp.100–104.
[19] D. R. Cox, "Regression models and life-tables,"JournaloftheRoyalStatisticalSociety.SeriesB(Methodological),vol.34,no.2,pp.187–220,1972.[Online].Available:http://www.jstor.org/stable/2985181
[20] S. M. P. A. Sakshi Indolia, Anil Kumar Goswami, "Conceptualunderstanding of convolutional neural network- a deep learningapproach,"Procedia Computer Science, vol. 132, pp. 679–688, 2018.[Online]. Available: https://doi.org/10.1016/j.procs.2018.05.069
[21] G.R.XuY,"Onsplittingtrainingandvalidationset:Acomparativestudyofcross-validation,bootstrapand 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
[22] K. T. Shorten, C., "A survey on image data augmentation for deeplearning,"Journal of Big Data, vol. 6, 2019. [Online]. Available:https://doi.org/10.1186/s40537-019-0197-0
[23] M. T. Chung, N. Quang-Hung, M. Nguyen, and N. Thoai, "Usingdocker in high performance computing applications," in2016IEEE Sixth International Conference on Communications and Electron-ics (ICCE), 2016, pp. 52–57.
[24] "Pydicom,"accessed:2021.[Online].Available:https://pydicom.github.io/
[25] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The numpyarray: A structure for efficient numerical computation,"Computingin Science Engineering, vol. 13, no. 2, pp. 22–30, 2011.[
26] D. W. B. M. P. D. S. H. M. D. M. O. L. P. D. C. D. L. M. P. D. E.M. M. e. a. Dr. Debbie Saslow PhD Dr. Carla Boetes MD, PhD,"American cancer society guidelines for breast screening with mrias an adjunct to mammography,"American Cancer Society, 2009.[Online]. Available: https://doi.org/10.3322/canjclin.57.2.75
[27] M. K. et. al, "Efficacy of mri and mammography for breast-cancerscreening in women with a familial or genetic predisposition,"The New England Journal of Medicine, 2004. [Online]. Available:https://www.nejm.org/doi/full/10.1056/NEJMoa031759
[28] E. L., "Mris more accurate than mammograms but expensive."Canadian Medical Association Journal, 2004. [Online]. Available:https://doi.org/10.1503/cmaj.1041508
[29] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard,"A study of the behavior of several methods for balancingmachine learning training data,"SIGKDD Explor. Newsl.,vol. 6, no. 1, p. 20–29, Jun. 2004. [Online]. Available:https://doi.org/10.1145/1007730.1007735
[30] H. Kvamme, Ørnulf Borgan, and I. Scheel, "Time-to-eventprediction with neural networks and cox regression,"Journalof Machine Learning Research, vol. 20, no. 129, pp. 1–30, 2019.[Online]. Available: http://jmlr.org/papers/v20/18-424.html
[31] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Under-standing deep learning requires rethinking generalization," 2017.
[32] K. M. W. J. G. T. Wolbers M, Blanche P, "Concordance forprognostic models with competing risks,"Biostatistics, 2014.[Online]. Available: https://doi.org/10.1093/biostatistics/kxt059
9
[33] G. T. Blanche P, Kattan MW, "The c-index is not proper for theevaluation oft-year predicted risks,"Biostatistics, 2019. [Online].Available: https://doi.org/10.1093/biostatistics/kxy006
[34] G.MKandK.P.K.J.,"Understandingsurvivalanalysis,"Int J Ayurveda Res., 2010. [Online]. Available:https://doi.org/10.4103/0974-7788.76794
[35] S. L. A. W. A. e. a. Murtaza, G., "Deep learning-based breast cancerclassification through medical imaging modalities: state of the artand research challenges,"Artificial Intelligence Review, pp. 1655–1720, 2020. [Online]. Available: https://doi.org/10.1007/s10462-019-0971

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

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