BACKGROUND

A. Related Work

Classification is one of the most important and essential tasks in machine learning and data mining. About a lot of research has been conducted to apply data mining and machine learning on different medical datasets to classify Breast Cancer. Many of them show good classification accuracy [1]. There are many deep learning and machine learning techniques available for breast cancer detection and prediction. Some of most used deep learning techniques for breast cancer detection are Convolutional Neural Network, Recurrent Neural Network and some pre-trained model Alex Net, Google Net, VGG16, VGG19, ResNet. Some most used datasets available for training and testing are Mammogram image, SEER, UCI, WBCD [2]. The author Abdullah-Al Nahid developed Deep Learning Model utilizing a restricted Boltzmann machine that mainly used back propagation algorithm for classify histography images [3]. Vikas Chaurasia and Saurabh Pal compared the performance of supervised learning classifiers such as Naïve Bayes, SVM-RBF kernel, RBF neural networks, decision trees to find the best classifier in breast cancer datasets [4]. A few investigations are led on clinical informational collections utilizing different classifiers and element determination strategies. Quite a bit of the examination on breast cancer related datasets can be obtained on the internet. Majority of the datasets show great order of accuracy [5]. Sivaprakasam et al. [6] compared the performance of C4.5, Naïve Bayes, Support Vector Machine and K- Nearest Neighbor (KNN) to find the best classifier and SVM turns out to be the most accurate with an accuracy of 96.99%. In machine learning the feature selection is the process of choosing a subset of relevant attributes from various candidate subsets, and it is a fundamental to build a training model. Feature selection plays a vital role in creating an effective predictive model. There are several benefits on applying the feature selection methods and some of the benefits are they provide more effectiveness in training the machine learning algorithm, and they reduce the complexity of a model and makes it easier to interpret, improve the accuracy of a model if the right dataset is chosen [7].

B. Overview

Overview

Breast cancer is a devastating disease, with high mortality rates around the world.[8] Breast cancer is the second most diagnosed cancer in women, affecting one in every eight women in the U.S. The goal of screening exams for early breast cancer detection is to identify breast abnormalities as early as possible. If breast cancer is found early, there are more treatment options and a better chance for survival. Women whose breast cancer is detected at an early stage have a 93 percent or higher survival rate in the first five years.[12]

Screening mammography is estimated to decrease breast cancer mortality by 20 to 40 percent.[8] Mammograms are currently the best screening tool to detect breast cancer early but reading and interpreting them is a visually challenging task, error prone for even experienced radiologists," said Dr. Vijayaraghavan, associate professor of radiology, who co-authored the retrospective study with lead author Bill Lotter, PhD, chief technology officer and co-founder of DeepHealth. "We want to improve the health of women in Massachusetts with reliable tools that assist clinicians."[12]]However, there are such limitations to these screenings. To address these limitations, researchers have explored using deep learning in mammography. [8] Researchers have come up with various deep learning techniques to overcome these limitations in mammography. But they still struggle with obtaining large amounts of annotated data for training these deep learning models. [8]

Cancer has become the most concern point in biomedical informatics. Breast cancers are the second highest recorded cancers from the 14 different types. There are different techniques and algorithms which are used for the prediction of breast cancer. Handcrafted feature extraction [9] (Humoment, haralick textures and color Histogram) techniques are used for the extraction of the characteristics like texture, shape, and color of the histopathological images [9]. In this Classification we compare the deep neural network with BC histopathology images for the better accuracy results. To predict breast cancer, support vector machines (SVMs) with an attribute extraction method were presented in Akay (2009) with 99.51% prediction accuracy. Abdel-Zaher and Edlib (2016) presented a two-phase scheme for breast cancer classification using Wisconsin Breast Cancer dataset [9]. Data augmentation methods are used for cropping, flipping for extraction of histopathological images [9]. The main issue is the over fitting. we can also predict the breast cancer by using computer aided diagnosis which learns the nature of cancer tissues through deep brief network algorithms. In this we get these results from the Mammograms and screening, and we are going to compare with the algorithms and other techniques. In this Mass ROI'S and Whole Mass ROI'S techniques [10] to detect the mass of the effected cancer tissue. The chosen characteristics are utilized to train and evaluate standard classifiers like Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), [10] and Neural Network (NN). On the contrary, the suggested DBN-based CAD system offers extensive capabilities [10]. In Technique A (Mass ROIs), we gathered 600 ROIs and used 300 in training and 300 in testing. ROIs were distributed evenly according to class type in both training and testing. In Technique B, we gather 168 ROIs (i.e. Whole Mass ROIs) and use half of them in training and the other half in testing [10]. These techniques are used for overall accuracy, specificity, and sensitivity for the mass of the breast cancer tissues. We also used the globally aware multiple instance classifier (GMIC) [11]. Here we used the retrieve-roi algorithm to utilize the localization information from saliency maps. The other algorithm greedily \ [11] selects information patches of input images. We reported the test performance of the top-5 GMIC heatmap models that achieved the highest validation [11] AUC on identifying breasts with malignant lesions. The top-5 GMIC ResNet-18-heatmap models achieved a mean AUC of 0.927 ± 0.04 / 0.792±0.008 in identifying breasts with malignant/benign lesions, outperforming the vanilla GMIC models (0.913±0.007 / 0.791 ± 0.005). The ensemble of the top-5 GMIC-ResNet18-heatmap models [11] achieved an AUC of 0.931/0.80 in identifying breasts with malignant/benign lesions matching the performance of vanilla GMIC models (0.930/0.80)

References:

1. https://www.sciencedirect.com/science/article/pii/S2667305322000072

2. https://www.researchgate.net/publication/305703509_Automatic_Computer-Aided_Diagnosis_of_Breast_Cancer_in_Digital_Mammograms_via_Deep_Belief_Network

3. https://deepai.org/publication/an-interpretable-classifier-for-high-resolution-breast-cancer-screening-images-utilizing-weakly-supervised-localization

C. Convolutional Neural network for deep learning

To perform classification or segmentation on large complex images, a common strategy involves the use of a classifier in sliding window fashion to recognize local patches on an image to generate a grid of probabilistic outputs [14]. After, the segmentation and classification are done we perform a set of process on the results that we obtained from them. Such methods have been used to detect metastatic breast cancer using whole slide images of sentinel lymph node biopsies [15]and to segment neuronal membranes in microscopic images [16]. The drawback of the above approach is that requires a series of steps which involves being executed distinctly. Hence, we propose a method to combine the two steps into a single step for training on the whole images [13].

Since, we will be using two data sets, on the first dataset we will use whole image classifiers and develop patches. Then we will train the classifier using two steps. The first step was to train a patch classifier where we e compare the networks with pre-trained weights using the ImageNet database to those with randomly initialized weights. The second step was to train a whole image classifier converted from the patch classifier. [14] Using max pooling and two FC layers we can test the model. The next thing would be to use a heatmap for all the convolutional networks. For the next dataset we will use transfer learning approach and check its efficiency and effectiveness

References:

[1] Asri, H., Mousannif, H., Moatassime, H. A., & Noel, T. (2016, May 12). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science. Retrieved April 5, 2022, from https://www.sciencedirect.com/science/article/pii/S1877050916302575

[2] N. Khuriwal and N. Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 98-103, doi: 10.1109/ICACCCN.2018.8748777.

[3] Abdullah-Al Nahid, Aaron Mikaelian and Yinan Kong, Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation, Biomedical Research Volume 29, Issue 10, (2018).

[4] V. Chaurasia and S. Pal, “Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability,” vol. 3, no. 1, pp. 10– 22, 2014.

[5] Y. Khourdifi and M. Bahaj, "Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification," 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), 2018, pp. 1-5, doi: 10.1109/ICECOCS.2018.8610632.

[6] Y. Christobel, A., & Sivaprakasam, “An empirical comparison of data mining classification methods,” Int. J. Comput. Inf. Syst., vol. 3, no. 2, pp. 24–28, 2011.

[7] Dhahri, Habib; Al Maghayreh, Eslam; Mahmood, Awais; Elkilani, Wail; Faisal Nagi, Mohammed (2019). Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. Journal of Healthcare Engineering, 2019(), 1–11. doi:10.1155/2019/4253641

[8]**Deep Learning May Detect Breast Cancer Earlier than Radiologists URL:** https://healthitanalytics.com/news/deep-learning-may-detect-breast-cancer-earlier-than-radiologists

[9] https://www.sciencedirect.com/science/article/pii/S2667305322000072

[10] https://www.researchgate.net/publication/305703509_Automatic_Computer-Aided_Diagnosis_of_Breast_Cancer_in_Digital_Mammograms_via_Deep_Belief_Network

[11] https://deepai.org/publication/an-interpretable-classifier-for-high-resolution-breast-cancer-screening-images-utilizing-weakly-supervised-localization

[12] EARLY DETECTION IS KEY
Url: https://www.carolmilgardbreastcenter.org/early-detection

[13] **Mammography expert finds deep-learning artificial intelligence may improve cancer detection**

URL: Mammography expert finds deep-learning artificial intelligence may improve cancer detection.

[14] Shen, L., Margolies, L.R., Rothstein, J.H. et al. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci Rep 9, 12495 (2019). https://doi.org/10.1038/s41598-019-48995-4.

[15] Dayong Wang et al. “Deep Learning for Identifying Metastatic Breast Cancer”. In: (June 18, 2016). arXiv: 1606.05718 [cs, q-bio]. URL: http://arxiv.org/abs/1606.05718 .

[16] Dan Ciresan et al. “Deep Neural Networks Segment Neuronal Membranes in Electron Mi- croscopy Images”. In: Advances in Neural Information Processing Systems 25. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 2843–2851. URL: http://papers.nips. cc / paper / 4741 - deep - neural - networks - segment - neuronal - membranes - in - electron-microscopy-images.pdf