I. Overall Objective

Breast cancer is the most frequent cancer diagnosis in women worldwide[1]. According to a literature review published in the journal Nature,

> Early breast cancer — that is, cancer that is contained in the breast or that has only spread to the axillary lymph nodes — is considered curable. Improvements in multimodal therapy have led to increasing chances for cure in ~70–80% of patients. By contrast, advanced (metastatic) disease is not considered curable using currently available therapeutic options.[1]

Breast cancer's rate of incidence combined with its prospects in early stage and inoperability in late stage place a unique emphasis on early diagnosis. The earlier an accurate diagnosis is made, the greater a patient's likelihood of survival.

Diagnostic methods are numerous and multivariate. Mammography is a popular screening method in developed countries but its implementation varies widely among developing nations (e.g., sub-Saharan and east African countries).[1] We choose instead to focus on histological stains – photographic images of cell samples from biopsies. These images may be processed by a machine learning (ML) model and provide additional diagnostic information for breast cancer pathologists.

Our deliverables will be entirely software-based. We will use biopsy images from Fox Chase Cancer Center to train our ML model. We will investigate, document and describe the training algorithm used to produce the model and evaluate its efficacy (diagnostic accuracy and computational performance) using similarly sampled biopsy images.

The uniqueness of our approach relies on frame-level diagnosis. In addition to classifying the image as a whole, our classifier will segment each test image into a uniform grid and label grid squares as malignant or non-malignant. This will allow pathologists to interpret biopsies faster and with greater accuracy. If successful, our software will have a significant advantage over similar ML-based methods that merely classify entire images. The marginal utility of this feature will determine our software's market capitalization.

II. Background

Today, if breast cancer is not caught and treated within a small time-window, the patient's prognosis is either invasive surgery or death. That makes the best treatment for this disease proactive screening to catch early. Currently, pathologists employ a variety of methods to screen and diagnose breast cancer. The first line of defense is population screening via mammograms.[1] This is a method that uses low doses of x-rays to detect early indicators of malignancy, so the malignancy can be treated using non-invasive options such as personalized drugs. Unfortunately, mammographic screening may produce false positive or false negative results due to breast tissue density.[2]

Another method is Magnetic Resonance Imaging (MRI), which is generally safer for patients who require frequent imaging, since MRI screening does not rely on x-rays and ionizing radiation (unlike mammograms). While mammographic images are insensitive to tumors in dense tissue, MRI images are easier to differentiate between white matter and gray matter, aiding in accurately detecting benign and malignant lesions. Moreover, MRI technicians are able to create detailed visualizations of the tissue. However, this screening method is very expensive and requires a lot of time and specialized personnel to use MRIs efficiently and effectively.[3] MRIs and mammograms are part of a class of screening methods known as tomographic imaging. When a patient is screened using tomographic methods, their next step is a needle biopsy.[1][4]

Needle biopsies use a less invasive surgical method compared to traditionally cutting into breast tissue, leading to a lower rate of re-excision. According to oncologist C. H. Weaver of the University of Pennsylvania, “Of the 3481 women who underwent needle biopsy, 23% had to have a breast re-excision compared with 92% of the 2650 women who underwent surgical biopsy.”[5] In this procedure, a needle is inserted into a suspicious part of the breast (lump, scar tissue, foreign object, or anything deemed suspicious). The tissue is removed, stained with hematoxylin and eosin dyes to highlight cell geometry, inspected under microscope and photographed for reference. The disadvantage of a needle biopsy is that it does not remove significant malignant material as a traditional biopsy would. To prevent metastasis of the tumor, it’s essential when performing a needle biopsy that time-to-diagnosis is short.

The aforementioned tomographic and biopsy images require significant time to be reviewed by a breast cancer pathologist. Unfortunately, there is more demand than supply for pathologists with the relevant skills and training. One method of minimizing time-to-treatment is to preprocess patient data using machine learning methods. One study of Bangladeshi patients suggests a clear advantage of using the XGBoost training algorithm to model classification of histological data from biopsies, citing the highest precision (0.94), recall (0.95), and F1 scores (0.96) of a survey including decision trees, random forests, logistic regression, and naive Bayes classifiers.[6] The results were promising but the dataset “... was limited in size and diversity, potentially limiting the generalizability of the findings.” In addition, the study did not evaluate newer methods such as deep learning or ensemble algorithms.

III. Needs Statement

Creating an urgency system utilizing machine learning, we can provide hospitals with a theoretical risk factor. While these biopsies will still need to pass through a human's hands, the hospitals could utilize our output to have doctors self-check.

IV. Implications of Project Success

Pre-existing machine-learning models do not presuppose the possibility that models cannot be improved by integrating new methods of processing data. Our goal is to find unique ways of training and classifying data and thereby contribute some increase in computational performance or diagnostic accuracy to existing software. We aim to improve the speed and accuracy of cancer detection in parts of the world where health infrastructure and access to expensive screening methods (such as MRIs) are limited.

If our goal were met, patients could receive more accurate diagnoses within shorter time-frames and receive earlier treatment. As a result, the mortality rate of breast cancer patients would diminish. Future innovations may build on our findings to solve other related problems, whether it may be an algorithm to detect early-stage cancer in the lungs or other tissue.

These possible outcomes will contribute to the UN’s Sustainable Development Goals (SDG) [7] such as:

- (1) No poverty: Economically accessible screening methods reduce the financial burden on individuals and nations.
- (3) Good health and well-being: Because patients will receive a diagnosis quicker, they can get treated faster, requiring less time and therapy usage to get better.
- (8) Decent work and economic growth: The symptom burden of therapy in late-stage treatment is often debilitating.
- (9) Industry, innovation, and infrastructure: Our work may open more paths to other innovations. One such path could be improving cancer detection in a different part of the body, thereby improving standards in healthcare.
- (10) Reduced inequalities: Marginalized peoples are less likely to be financially prepared for expensive screening methods.

Especially important to note is that the marginal utility of ML models to breast cancer pathologists is greater in developing countries where cancer burden is higher. The symptom, financial and psychosocial burden of cancer affect not only individuals but also the collective material conditions of entire groups of people [UN SDG 1].

Symptoms of therapy greatly inhibit patients' quality of life [UN SDG 3]. Endocrine therapy causes frequent hot flashes; aromatase inhibitors produce joint pain; chemotherapy causes nausea, fatigue, infertility, cardiotoxicity, neuropathy and cognitive dysfunction.[1]

Financial and psychosocial burdens of therapy introduce additional hardships–

> Lost employment and cost of care can be economically challenging [UN SDG 8], and dealing with a potentially fatal diagnosis (including relying on friends and family to help with, for example, transportation and home responsibilities) can be emotionally challenging.[1]

These burdens collectively and disproportionately affect impoverished nations with poor health infrastructure lacking the capacity for early diagnosis [UN SDG 10].

V. Citations

[1]

N. Harbeck et al., "Breast cancer," Nature Reviews Disease Primers, vol. 5, no. 1, Sep. 2019, doi: https://doi.org/10.1038/s41572-019-0111-2.

[2]

A. Khalid et al., "Breast Cancer Detection and Prevention Using Machine Learning," Diagnostics, vol. 13, no. 19, p. 3113, Jan. 2023, doi: https://doi.org/10.3390/diagnostics13193113.

[3]

"Magnetic Resonance Imaging (MRI)," www.nibib.nih.gov. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri.

[4]

L. Irwig, P. Macaskill, and N. Houssami, “Evidence relevant to the investigation of breast symptoms: the triple test,” The Breast, vol. 11, no. 3, pp. 215–220, Jun. 2002, doi: https://doi.org/10.1054/brst.2001.0409.

[5]

Dr. C. H. Weaver M. D., “ Ask the Doctor: What Should I Know About a Breast Biopsy?,” Ask The Doctor, Aug. 07, 2023. https://news.cancerconnect.com/breast-cancer/ask-the-doctor-what-i-should-know-about-a-breast-biopsy. (accessed Aug. 30, 2024)

[6]

T. Islam et al., “Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI,” Scientific Reports, vol. 14, no. 1, Apr. 2024, doi: https://doi.org/10.1038/s41598-024-57740-5.

[7]

United Nations, “The 17 Sustainable Development Goals,” United Nations, 2015. https://sdgs.un.org/goals.