

**MSc in Artificial Intelligence and Machine Learning**

**CS6472 – Research Methods and Specification**

**Assignment 1: Semester 2 AY 23/24 - Literature Review (Essay)**

**Research Topic:** Leveraging Machine Learning for Automated HER2 Scoring in Breast Cancer Pathology.

**Student Name:** Siddharth Prince

**Student Id:** 23052058

**Date of submission:** 14/04/2024

**Literature Review**

**State of the Art**

Breast cancer is one of the most prevalent cancer types among women globally (Sung *et al.*, 2021). As with any cancer type, early and accurate diagnosis is critical for improving patient outcomes. Given that cancer pathology is a manual and time-consuming process, machine learning (ML) techniques are making significant strides in the field of breast cancer diagnosis, offering promising avenues to complement traditional methods. ML is being explored for various aspects of breast cancer diagnosis. A recent systematic review by Nemade, Pathak and Dubey, (2022) considered 162 research publications for the time period of 2015-2021 and investigated the use of machine intelligence techniques in breast cancer diagnosis. The review found that ML algorithms demonstrated promising results in tasks such as computer-aided diagnosis (CAD) using mammograms, ultrasound image analysis, and even gene expression profiling.

One of the key areas where ML is showing promise is in HER2 (Human Epidermal Growth Factor Receptor 2) analysis. HER2 is a protein overexpression associated with aggressive breast cancer. The traditional approach to assess HER2 status is to use immunohistochemistry (IHC) tests. However, IHC analysis can be very subjective and prone to inter-observer variability. The following quote from Pathmanathan & Bilous (2012) describes this issue well.

[However,] several studies have demonstrated significant rates of discordance between IHC and ISH methods of testing and also differences between centralised facilities performing the test versus peripheral testing facilities. The implications of this are that IHC is subject to a large number of variables throughout the entire course of the testing phase, from tissue fixation and processing through to test interpretation by the pathologist. (Pathmanathan and Bilous, 2012)

Deep learning approaches that employ deeper neural network architectures such as VGG19, ResNet, etc. are more capable at handling image processing tasks such as this. They offer an alternative to manual HER2 analysis, with studies demonstrating high accuracy. For instance, a study by Che et al. (2023) developed a deep learning model for HER2 recognition in breast cancer based on immunohistochemical pathology digital slide scans. Their model achieved a high accuracy score of 95.77% at the patch level and an f1 score of 83.09%, not only indicating excellent performance in differentiating HER2-positive from HER2-negative, but also capability to accurately classify cases that fall under the +1 (low expression) and +2 (equivocal and usually warrants additional pathology tests) categories cases. This suggests that deep learning has the potential to improve the objectivity and consistency of HER2 analysis.

Another study employed transfer learning and decision-level fusion to achieve high accuracy of 97% when using the mode statistical aggregation of the classified slid patches in HER2 molecular marker scoring (Tewary and Mukhopadhyay, 2021). The study also compared the effectiveness of 5 deep learning models; VGG16, VGG19, ResNet50, MobileNetV2 and NASNetMobile with VGG19 performing the best. These findings suggest that ML is already a valuable tool across different aspects of the breast cancer diagnostic pipeline.

Machine learning is emerging as a powerful tool in the fight against breast cancer. These works of research are promising and provide a solid base to build up on.

**References**

Che, Y. *et al.* (2023) ‘Immunohistochemical HER2 Recognition and Analysis of Breast Cancer Based on Deep Learning’, *Diagnostics*, 13(2). Available at: <https://doi.org/10.3390/diagnostics13020263>.

Nemade, V., Pathak, S. and Dubey, A.K. (2022) ‘A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques’, *Archives of Computational Methods in Engineering*. Springer Science and Business Media B.V., pp. 4401–4430. Available at: <https://doi.org/10.1007/s11831-022-09738-3>.

Pathmanathan, N. and Bilous, A.M. (2012) ‘HER2 testing in breast cancer: an overview of current techniques and recent developments’, *Pathology*, 44(7), pp. 587–595. Available at: <https://doi.org/10.1097/PAT.0b013e328359cf9a>.

Sung, H. *et al.* (2021) ‘Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries’, *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249. Available at: <https://doi.org/10.3322/caac.21660>.

Tewary, S. and Mukhopadhyay, S. (2021) ‘HER2 Molecular Marker Scoring Using Transfer Learning and Decision Level Fusion’, *Journal of Digital Imaging*, 34(3), pp. 667–677. Available at: <https://doi.org/10.1007/s10278-021-00442-5>.