**Capstone Project**

**Project Proposal**

**By**

**M Zaker Humayoon AMIN**

**Hamda Abdi**

# **INTRODUCTION**

With an estimation of over 2.3 million new cases diagnosed globally each year, Breast cancer (BC) is considered one of the second deadliest cancer disease [1], [2]. Breast cancer is a medical condition in which abnormal breast cells grow and develop spontaneously to lead to the formation of tumors. Strategies to counter the growing number of breast cancers require a fundamental strengthening proper early diagnosis and treatments that will mitigate the risk of cancer recurrence. The treatment comprehend surgery, radiation therapy, and medications. With proper combination of the treatment, the risk associated with BC can be reduced. According to the World Health Organization, by the end of 2020, approximately 7.8 million women alive was registered with a history of breast cancer in the last 5 years[3]. The latter statistical number can furthermore be increase with a proper early diagnosis[4], [5]. Recently, a new approach to traditional diagnostics has been introduced. Which is a form of machine learning inspired by the human learning system, has recently emerged as a simple and high accuracy predictive technique for predicting. The latter approach learns to find a pattern by mapping input to output data. Combining Machine learning approach to the detection of breast cancer, a capable of diagnosing breast cancer up to 12 months earlier than those using conventional clinical procedures is noted [6]

A project is proposed following the latter thematic. The project converges breast cancer detection with machine learning (ML). The specific goal is to develop a predictive model that can accurately classify breast masses as either benign or malignant based on features extracted from digitized fine needle aspirate (FNA) images.

# **Relevance to Sustainable Development Goals (SDGs)**

This project aligns with the United Nations SDGs, specifically SDG 3 - Good Health and Well-being. Early detection of breast cancer is crucial for improving health outcomes and reducing mortality rates. By leveraging machine learning to enhance diagnostic accuracy, the project contributes to achieving SDG 3 targets related to preventing and treating non-communicable diseases.

# **LITERATURE**

* "Breast Cancer Analysis and Prognosis Using Machine Learning" by Shetty, Simitha (2020)
  + **Summary:** This master's thesis addresses the prediction of breast cancer severity using Machine Learning algorithms to provide early-phase identification and treatment. The study balances datasets using Synthetic Minority Over-Sampling Technique and compares seven models, including Support Vector Machine, Naive Bayes, Logistic Regression, K-Nearest Neighbour, Classification and Regression Tree, Artificial Neural Network, and Extreme Gradient Boosting. Notably, K-Nearest Neighbour and Artificial Neural Network demonstrated good performance, achieving high sensitivity rates on both datasets.
  + **Link:** [Thesis Link](https://norma.ncirl.ie/id/eprint/4284)
  + **Reference:** Shetty, S. (2020). *Breast Cancer Analysis and Prognosis Using Machine Learning* (Doctoral dissertation, Dublin, National College of Ireland).
* "Random Forest Classifier Combined with Feature Selection for Breast Cancer Diagnosis and Prognostic" by Cuong Nguyen et al. (2013)
  + **Summary:** Published in the Journal of Biomedical Science and Engineering, this article proposes the use of a Random Forest classifier combined with feature selection for breast cancer diagnosis and prognosis. The study, conducted by Cuong Nguyen and Yong Wang, emphasizes the importance of feature selection in enhancing the performance of the classifier. The research contributes insights into optimizing the diagnosis and prognostic aspects of breast cancer using machine learning techniques.
  + **Reference:** Nguyen, C., Wang, Y., & Nguyen, H. N. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. doi: [10.4236/jbise.2013.65070](https://doi.org/10.4236/jbise.2013.65070)

These literature reviews provide a comprehensive overview of machine learning applications in breast cancer analysis, covering different models, techniques, and considerations for improving diagnostic and prognostic outcomes.

# **Describe Your Data**

The dataset for this project is the Breast Cancer Wisconsin (Diagnostic) dataset, sourced from the UCI Machine Learning Repository. It comprises features computed from FNA images, including radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The dataset is well-structured, with 569 instances and no missing values.

# **Approach (Machine Learning or Deep Learning)**

A goal to employ a machine learning approach for this project is maintained. Given the relatively smaller size of the dataset and the interpretability requirements in medical contexts, machine learning models such as Support Vector Machines, Random Forests, and Logistic Regression can be effective. These models can provide insights into the importance of different features in making diagnostic predictions. The choice is justified to ensure transparency and explainability, critical factors in medical decision-making.

This breast cancer detection project aims to enhance the accuracy of diagnoses, ultimately contributing to improved health outcomes and supporting the broader goal of achieving good health and well-being (SDG 3).

1. **References**

[1] S. Łukasiewicz, M. Czeczelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanisławek, “Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review,” *Cancers 2021, Vol. 13, Page 4287*, vol. 13, no. 17, p. 4287, Aug. 2021, doi: 10.3390/CANCERS13174287.

[2] S. Paul, P. P. Solanki, U. P. Shahi, and S. Srikrishna, “Epidemiological Study on Breast Cancer Associated Risk Factors and Screening Practices among Women in the Holy City of Varanasi, Uttar Pradesh, India,” *Asian Pac J Cancer Prev*, vol. 16, no. 18, pp. 8163–8171, 2015, doi: 10.7314/APJCP.2015.16.18.8163.

[3] “Breast cancer,” Jun. 2023. Accessed: Nov. 12, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[4] “KNOWLEDGE SUMMARY EARLY DETECTION: BREAST HEALTH AWARENESS AND EARLY DETECTION STRATEGIES.”

[5] L. Tabár *et al.*, “Early detection of breast cancer rectifies inequality of breast cancer outcomes,” *J Med Screen*, vol. 28, no. 1, pp. 34–38, 2021, doi: 10.1177/0969141320921210.

[6] M. Nasser and U. K. Yusof, “Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction,” *Diagnostics*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/DIAGNOSTICS13010161.