Reliable detection of breast cancer using GLCM and LBP

Introduction:

Despite decades of research and advances in technology across many fields, cancer remains a formidable challenge. Significant progress has been made in diagnostic techniques and treatment approaches, but important gaps still exist particularly in the fight against cancer. According to the World Health Organization (WHO), breast cancer is both the most common cancer and the leading cause of cancer‐related death among women worldwide.

According to recent studies, breast cancer accounts for more than 29 % of all new cancer cases in the USA and over 24 % in Europe. However, early detection remains a major global challenge. The quality of diagnostic images and the expertise of medical personnel are critical to identifying tumours at a treatable stage. Unfortunately, many deaths result from late diagnoses, when available treatments are far less effective.

To this day, breast cancer remains a significant global health challenge, particularly among women. However, one major positive development is the critical role of mammography in detection and diagnosis. Despite the challenges, mammographic imaging has proven essential in assessing a patient’s condition, determining whether the cancer is in an early or critical stage, and guiding timely treatment decisions to prevent further spread. Early detection continues to have a profound impact on improving treatment outcomes and overall patient health.

Although mammograms play an important role in the early detection of breast cancer, they have proven to be insufficient on their own, even today. A wide range of advanced techniques and imaging technologies are currently used to detect abnormal breast tissue; however, these methods still have limitations and cannot guarantee consistent early diagnosis. While these technologies have contributed to earlier treatments and improved survival rates, the medical community continues to seek more reliable solutions. As the fields of Artificial Intelligence (AI) and Machine Learning (ML) continue to evolve and advance across many sectors, including healthcare, there is growing interest in leveraging these technologies to develop more accurate, efficient, and accessible tools for the early detection and diagnosis of breast cancer.

Despite the considerable progress made with Artificial Intelligence (AI) and Machine Learning (ML) in recent years, several key challenges still persist. One major issue is the heterogeneity of medical images, which plays a critical role in the development of protocols and in the effectiveness of feature extraction techniques. While technology has certainly made the detection process faster and more accessible, concerns remain particularly among medical professionals about the reliability of results generated by AI systems. A significant factor contributing to this distrust is the "black box" nature of many AI models; the internal decision-making processes and data extraction methods are often opaque and lack transparency. This lack of explainability makes it difficult for doctors to fully understand or verify how certain diagnostic conclusions are reached, leading to hesitation in fully adopting complex AI-based diagnostic tools in clinical practice.

One of the key challenges in developing effective AI models for cancer detection lies in several critical factors. First, the size of the dataset is essential—larger datasets generally lead to more reliable and generalizable models. Equally important is the balance within the dataset, particularly in the case of breast cancer, where there must be an adequate and proportional representation of both benign and malignant cases. This balance is crucial for training models that can accurately differentiate between the two categories. Additionally, the quality and reliability of the annotations used in both the dataset and the model training process play a vital role. Poor or inconsistent labelling can significantly impact the model’s performance, leading to misclassifications and reduced trust in the system's results. Therefore, maintaining high standards in dataset preparation and annotation is fundamental for building accurate, trustworthy, and clinically useful AI systems.

Despite the presence of many challenges in this field, researchers have proposed various solutions to improve the reliability and performance of AI models. One common approach involves refining and adjusting the datasets, ensuring accurate labeling and consistent annotations to support model training. To address classification tasks in breast cancer detection, researchers have increasingly turned to advanced techniques such as Convolutional Neural Networks (CNNs). CNNs are widely recognized as powerful end-to-end classifiers that operate in multiple stages, aiming to achieve an optimal balance between accuracy, interpretability, and efficiency. However, despite their proven effectiveness, CNNs remain highly complex, requiring a well-designed and carefully structured architecture to perform reliably. In addition to CNNs, other traditional machine learning methods—such as Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Naïve Bayes, and Random Forests—are also commonly used. While these methods are considered reliable, they too present significant challenges in terms of complexity, parameter tuning, and scalability, especially when dealing with large and diverse medical imaging datasets.

Proposed method which is LBP and GCLM