**REVIEW PAPER**

**Abstract**:

Breast cancer is prevalent and potentially a life-threatening disease that demands early detection for effective treatment and improved patient outcomes. Machine learning and deep learning techniques aid in breast cancer prediction by training algorithms on medical images, identifying specific features and classifying disease presence or absence.The project entails a multi-stage process. First, a diverse and well-curated dataset of medical images is collected, encompassing both cancerous and non-cancerous cases. Through meticulous preprocessing and data augmentation, the dataset's quality and diversity are optimized, preparing it for subsequent analysis. Various deep learning architectures, including Convolutional Neural Networks (CNNs) and possibly more specialized models for medical image analysis, are explored to extract intricate patterns and features from the images.Upon determining the optimal model, the system is deployed to classify new, unseen medical images. The model's predictions aid healthcare professionals in making informed decisions about breast cancer diagnosis and treatment planning.

**Introduction:**

Cancer is an incurable illness where there is uncontrolled growth of cells in a particular part of the body to form lumps, which we call tumors. According to the WHO, the most common type of cancer is breast cancer among women. It is predicted that one out of three women die because of breast cancer. However, if diagnosed early, they can be saved from their seemingly inevitable death.

For the purpose of diagnosis, there are various imaging modalities, which itself is an evolving field. Some of the imaging modalities include Mammography, Magnetic Resource Imaging (MRI), pathological tests, etc. Among these methodologies, Images from histopathology are regarded as the best and have shown improved accuracy of diagnosis. Via these histology images, pathologists can look at microelements and the tissue's structure, hence identifying cancer. These images help them to distinguish between normal cells, benign cells, and malignant cells. Nowadays we have automated solutions, from traditional machine learning practices to current deep learning practices.

To date, there have been various rule–based models and ML models made for the purpose of classification of cancerous to non–cancerous ones. Deep learning models work like the human brain like a human nervous system is made up of neurons, in a similar fashion, deep learning neural networks seem to mimic this capability to make machines perform tasks like humans do. If we talk about data from images, in particular, CNN (Convolutional Neural Network) stands out the most.

CNNs are Deep Learning modalities that are used to perform various tasks on images such as object recognition, image segmentation, face recognition, etc. In the biomedical field, due to their ability to learn advanced and complicated features from the image based on the data they are provided, these models have shown remarkable progress in computer vision. That is why researchers use these modalities to perform various tasks including classification of histology images. Now there have been various CNN models like VGG16, VGG19, Resnet v2, AlexNet, GoogleNet, etc. Some of these models already have prior training on ImageNet dataset which is one of the largest collections of images and hence they yield high accuracy on various computer vision operations and can also be directly used instead of training these models from scratch. This is called transfer learning. Transfer learning makes it possible to get high accuracy even on systems with low computation power since these systems are trained on huge datasets so they do not need to be trained that much.

**Literature Survey:**

This section explains 8 research papers published previously regarding the work done earlier and how they approached this problem. In 2020, there was a study performed by Ghulam, et al. [1] where different imaging modalities such as Mammography, MRI, Ultrasound, Tomography, and Histopathology were compared as to which would be better for image diagnosis through various parameters, and it was found that Mammography and Histopathology images were the way to go here. In 2018 in a paper published by Meriem, et al. [2] two machine learning models were compared namely, Naive Bayes (NB) classifier and K Nearest neighbor (KNN) classifier. It was found that KNN showed a better accuracy of 97.51% over the NB classifier which showed a degree of accuracy of 96.19%. Another work that was published in the same year was done by Sarosa, et al.[3]. Here, binary classification of mammography images was performed by using GLCM (Gray-level co-occurrence matrix) and SVM (support vector machine) combined (SVM). However, accuracy obtained was 63.03% and specificity was 89.01%. In 2021, a survey was conducted by Zebari, et al.[4]. Here, 118 publications published in 2018-2021 were reviewed. The studies showed different stages of CAD and highlighted research gaps. In another paper published in 2018 by Nawaz, et al.[5], a deep learning convolutional neural network DenseNet was introduced. This model obtained an accuracy of 95.4% in the task of multi-classifying breast cancer images from BreakHis dataset. Another paper published by Obaid, et al. published in 2018[6] compared 3 ML models namely SVM, KNN, and Decision Tree were used to perform classification on a Wisconsin Breast Cancer (Diagnostic) dataset. In the comparison, it was found that SVM outperformed both other models having an accuracy of 98.1%. In 2018, another paper was written by Khourdifi, et al.[7]. In this paper, 4 ML models namely, Random Forest, Naïve Bayes, SVM, and KNN were compared for classification and it was found that SVM outperformed them all with an accuracy of 97.9%. The above studies solidify SVM as the best model for image classification among traditional ML models. In 2021, a study conducted by Vijayakumar, et al.[8] to explain deep neural networks for breast cancer prediction. This paper used multiple activation functions to understand DNN layer-wise. Ayelet et al. [9] undertook a significant endeavor by training a DL algorithm on a dataset consisting of 9611 mammograms. Their objective was to predict malignancy and differentiate between normal and abnormal screenings. However, it's worth noting that the limitations of this study include the fact that mammography technology has advanced since then, rendering the dataset somewhat outdated. Nevertheless, their DL model achieved a commendable 77.3% specificity and 87% sensitivity.Similarly, Buda et al. [10] focused their efforts on applying DL techniques to the detection of mass and architectural abnormalities in digital breast tomosynthesis (DBT) examinations. They utilized a dataset comprising 16,802 DBT examinations, which were divided into four groups for training and testing purposes.However, this dataset was collected from a single institution, which may introduce bias. Their DL model achieved a sensitivity rate of 65%.Shahidi et al. [11] conducted a comparative study involving DL models for the classification of histopathology images. They implemented models such as ResNeXt, Dual Path Net, SENet, and NASNet for binary and multi-class classification on the BACH and BreakHis datasets. However, it's crucial to recognize that newer models have emerged since the time of this survey. Notably, Inception Res-Net demonstrated superior performance compared to other models in their study.

**Existing System:**

The current existing system for breast cancer diagnosis predominantly relies on established methods such as mammography, histopathology, and clinical evaluation. Mammography, a widely adopted screening tool, utilizes X-ray imaging to detect breast abnormalities, yet it may exhibit limitations in sensitivity and specificity. Histopathological analysis, involving the visual examination of tissue samples, is employed for confirmation and grading of cancer but can be subjective. Clinical evaluation by healthcare professionals complements these methods. However, challenges persist, including the potential for misdiagnosis, interobserver variability, and limited access to expert pathologists. The existing system, while pivotal in breast cancer diagnosis, can benefit from advancements such as machine learning-based models to enhance accuracy, objectivity, and efficiency in cancer detection and classification.

**Proposed System:**

The Breast Cancer Histopathological 400X (BreakHis 400X) dataset from Kaggle, which contains 1693 microscopic biopsy pictures of breast tumors, will be employed in the proposed approach.Both benign and malignant tumor samples can be found in the dataset.To enable an unbiased assessment of the model's performance, the dataset will be divided into training,validation, and test sets.The planned system's backbone architecture will be DenseNet201.It has shown outstanding performance in a variety of computer vision applications, making it suitable for medical picture analysis, including the categorization of breast cancer.A number of criteria, including accuracy, precision, and recall, will be used to assess the proposed system.These metrics offer a thorough evaluation of the model's capability to distinguish between benign and malignant breast cancers.DenseNet201 being a deeper and more densely connected architecture compared to Inception-V3 offers Improved Classification Accuracy. Its ability to capture intricate patterns and feature reuse can lead to improved breast cancer classification accuracy.It’s connectivity pattern allows for efficient feature extraction and propagation throughout the network. leading to faster training times and inference.It offers Reduced Risk of Overfitting with its dense connections,especially when dealing with relatively small medical datasets like the BreakHis 400X dataset.We can also develop an ensemble model with existing and proposed models for higher accuracy.

**Conclusion:**

There are various models present in Machine Learning which we can use for breast cancer classification. We analysed the models and studied their characteristics. We also studied different research papers in order to choose our model(Densenet201) or technique.We compare different CNN models on our dataset  and analyse their accuracy.The proposed project harnesses the power of machine learning and deep learning techniques to revolutionize breast cancer detection and diagnosis. By leveraging curated datasets of histopathology images,meticulous preprocessing, and advanced data augmentation, the project aims to create a sophisticated deep learning model capable of accurately classifying breast cancer cases.

**References:**

[1] Murtaza, Ghulam, et al. "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges." Artificial Intelligence Review 53.3 (2020): 1655-1720.

[2] Amrane, Meriem, et al. "Breast cancer classification using machine learning." 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). IEEE, 2018.

[3] Sarosa, Syam Julio A., Fitri Utaminingrum, and Fitra A. Bachtiar."Mammogram breast cancer classification using gray-level co-occurrence matrix and support vector machine." 2018 international conference on sustainable information engineering and technology (SIET). IEEE, 2018.

[4] Zebari, Dilovan Asaad, et al. "Systematic review of computing approaches for breast cancer detection based computer aided diagnosis using mammogram images." Applied Artificial Intelligence 35.15 (2021): 2157-2203.

[5] Nawaz, Majid, Adel A. Sewissy, and Taysir Hassan A. Soliman."Multi-class breast cancer classification using deep learning convolutional neural network." Int. J. Adv. Comput. Sci. Appl 9.6 (2018): 316-332.

[6] Obaid, O. Ibrahim, et al. "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer." International Journal of Engineering & Technology 7.4.36 (2018): 160-166.

[7] Khourdifi, Youness, and Mohamed Bahaj. "Applying best machine learning algorithms for breast cancer prediction and classification." 2018 International conference on electronics, control, optimization and computer science (ICECOCS). IEEE, 2018.

[8] Vijayakumar, K., Vinod J. Kadam, and Sudhir Kumar Sharma. "Breast cancer diagnosis using multiple activation deep neural network." Concurrent Engineering 29.3 (2021): 275-284.ent care.

[9] Akselrod-Ballin, Ayelet, et al. "Predicting breast cancer by applying deep learning to linked health records and mammograms." Radiology 292.2 (2019): 331-342.

[10] Buda, Mateusz, et al. "A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images." JAMA network open 4.8 (2021): e2119100-e2119100.

[11] Shahidi, Faezehsadat, et al. "Breast cancer classification using deep learning approaches and histopathology image: a comparison study." IEEE Access 8 (2020): 187531-187552.