Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 12 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

A HYBRID DATASET-BASED ENSEMBLE STRATEGY FOR EFFICIENT BREAST CANCER DETECTION

[1]Muhammad Sajid Maqbool*, [2]Nosheen Fatima, [3]Rubaina Nazeer, [4]Dr. Naeem Aslam, [5]Faisal Abbas, [6]Unaiza Sumra, [7]Muqadas Nadeem

[1, 4, 6]Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan.
[2, 3]Department of Software Engineering, National University of Modern Languages, Multan, Pakistan.
[5]Department of Artificial Intelligence, NFC Institute of Engineering and Technology, Multan, Pakistan.
[7]Department of Computer Science & Technologies, Emerson University Multan, Pakistan.

***Corresponding Author:** (sajid.maqbool@nfciet.edu.pk)

Article Info

Abstract

Breast cancer is especially dangerous for women because it kills and hurts a lot of people. Because of this, there is a need for an algorithm that can spot the first signs of breast cancer essential. One of the most common types of cancer in women is breast cancer. Its spread among people is a major worry all over the world. To save the patient's life, it is very important to find the disease and treat it right away. Last year, more than 2.3 million women were told they had breast cancer, and about 0.7 million of them died. Manually diagnosing the disease isn't very good, and most of the time, it's almost impossible to find severe cancer early, which means that the patients die. Machine learning is a key part of figuring out what early signs of breast cancer to look for.

In this paper Machine Learning (ML) and Deep Learning (DL) methods can be used to find breast cancer in early stage and accurately to save lives. In this work breast cancer predicted by using a hybrid model to combine ML and DL techniques. Firstly, a DL method Scale-Invariant Feature Transform (SWIFT) is used to extract the meaning full information from the image dataset and then different machine learning models such as (K-Nearest Neighbors (KNN), Random Forest (RF) and Support Vector Machine (SVM)) are applied to classify the malignant and benign cases. Secondly, these models classified the instances into two classes. The main focus of this work is to figure out whether a breast cancer is benign or malignant based on different specific characteristics taken from a group of images from normal and breast cancer patients. The results of contracted ML models compared on the basis of accuracy, precision, Recall and F1-Score. This study used Principal Component Analysis to find the optimized features from the set of extracted features for improving the accuracy of the used models. This is done by using multiple Machine Learning algorithms and choosing the classification model with the highest accuracy.

Keywords:
Machine Learning, Detection of Breast Cancer, Computer Vision, Deep Learning, Classification of Breast cancer.

1. Introduction

The frequent sickness in ladies globally and a prime worldwide is the breast cancer. It is essential to locate the condition early for improved remedy results because it impacts thousands and thousands of ladies and their families. In order to growth survival fees and decrease mortality quotes related to breast most cancers, timely and particular detection is critical. The discipline of clinical imaging and analysis, especially the identity of breast most cancers, has seen a revolution in current years because to enhancements in deep mastering and gadget learning methods. These techniques have the ability to identify and categorize breast most cancers lesions by means of extracting full-size records from medical photographs [1]. A branch of system mastering referred to as deep getting to know uses a couple of-layered synthetic neural networks to method and extract patterns from huge amounts of complicated information. An instance of a deep gaining knowledge of model is convolutional neural networks (CNNs), that have validated wonderful effectiveness in photo processing programs which include breast cancer detection. CNNs are an amazing choice for assessing breast most cancers snap shots because they are able to automatically research and extract pertinent information from mammograms or different imaging modalities. In addition to getting used for characteristic extraction and breast cancer screening, device getting to know algorithms have also been used in greater conventional techniques like guide vector machines (SVMs) and random forests. These techniques use statistical models and handmade functions to identify breast most cancers-related patterns in medical imaging. Although they're now not as complicated as deep studying fashions, they have shown success in a few conditions and may be used as alternatives while there is a lack of schooling data [2].

In order to diagnose breast cancer, the usage of deep studying and machine studying methods, characteristic extraction is a vital step. It entails changing unprocessed input facts—inclusive of ultrasound or mammography snap shots—into a concise illustration that captures the vital details for categorization. The capabilities that can be retrieved from breast lesions encompass things like their texture, shape, and depth. These strategies can distinguish among benign and malignant tumors via extracting discriminative features, which enables radiologists make specific diagnosis. The mixture of deep mastering and device studying methods has notably progressed breast cancer detection and characteristic extraction. These strategies offer robust gear for photo analysis, which facilitates in the early detection and analysis of breast most cancers [3]. Healthcare vendors can enhance patient effects and make a difference in the combat against breast cancer through utilizing the capabilities of synthetic intelligence and automated characteristic extraction. Automation and Efficiency: ML and DL algorithms provide automation within the detection of breast most cancers, lowering the workload on scientific personnel and boosting efficiency. Without the want for human participation, these algorithms can examine scientific pictures and extract pertinent data, releasing radiologists to pay attention on interpretation and choice-making. This would possibly hasten the diagnosing technique and hasten the begin of remedy.

Machine mastering (ML) and deep gaining knowledge of (DL) processes maintain good sized significance for the identification of breast cancer for some of motives. Early Detection: Raising breast most cancers survival charges requires early detection. By inspecting clinical pictures like mammograms, ultrasounds, and MRIs, ML and DL techniques can resource within the early detection of breast most cancers [4]. These strategies can identify minor anomalies and assist radiologists make unique diagnoses, enabling set off treatment and better patient outcomes. ML and DL algorithms can have a look at massive quantities of breast most cancers data and apprehend complicated styles that would be difficult for a human to observe [5]. Healthcare vendors can get extra specific and straightforward diagnostic equipment by using these algorithms. This can lessen the possibility of fake negatives or false positives inside the identification of breast cancer, assure suitable remedy techniques, and restrict the need for pointless remedies. Automation and Efficiency: ML and DL algorithms provide automation in the detection of breast most cancers, decreasing the workload on scientific team of workers and boosting efficiency. Without the need for human participation, those algorithms can examine clinical pix and extract pertinent records, releasing radiologists to concentrate on interpretation and decision-making [5]. This would possibly hasten the diagnosing method and hasten the start of remedy.

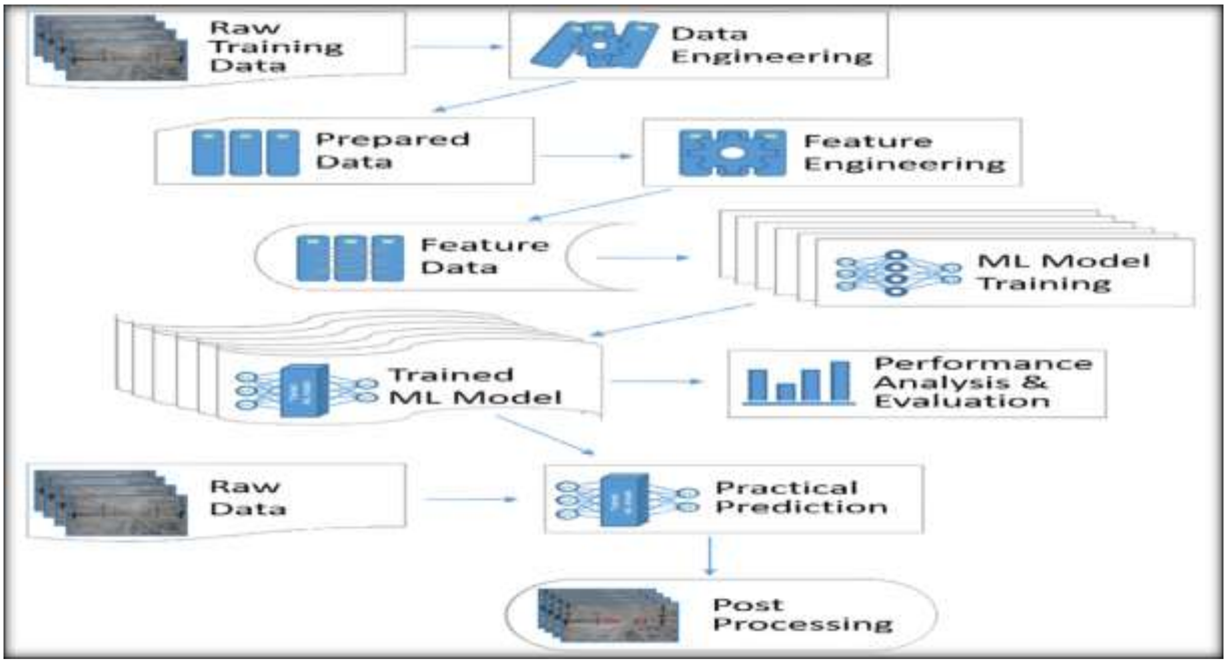

Figure 1 Graphical overview of introduction

The problems of the previous study are discussed as, unavailability of efficient diagnosis of serious /dangerous breast cancer in early stages, less usability of standard automatic techniques to minimize the death rate, less availability of datasets that are understood by machine learning algorithms, need for breast cancer detection based on hybrid method [6]. The objectives to reduce the above mentioned problems are, to diagnose severe breast cancer in its early stages, to apply standard automatic techniques to minimize the death rate in women by detecting the breast cancer in its early stages, to find and select image dataset then transform the dataset using SIFT to easily understood by machine learning algorithms, to apply Principal Component Analysis (PCA) to select the optimized Feature for improve the training and Testing accuracy of Machine Learning models, to evaluate several machine learning models and comparative analysis based on accuracy and confusion matrix.

The purpose of this study is to demonstrate how machine learning (ML) and deep learning (DL) techniques can be utilized to reliably and quickly detect breast cancer in order to save lives. In this study, a hybrid model that combines ML and DL approaches used to predict the occurrence of breast cancer. Prior to applying several machine learning models, such as KNN and random forest, to classify the malignant and benign cases, the Scale-Invariant Feature Transform (SWIFT) DL approach is used to extract the complete meaning from the image's dataset. The major goal of this research is to distinguish between benign and malignant breast cancers using a variety of distinct characteristics extracted from a collection of photos from healthy individuals and breast cancer patients. On the basis of the models' precision, recall, accuracy, and F1-Score, the results will be compared. To increase the accuracy of the models we utilize, we use Principal Component Analysis to extract the optimum features from the data. To do this, various machine learning methods will be used, and the categorization model with the highest accuracy will be selected.

2. Literature Review

In [2] findings were satisfactory. The program can extract the margins of the entire skull image with better information, more precise locating, and noise resistance. The other one [7] employed morphological procedures and image-based smoothing to detect depth boundaries in 2016. The calculation took 0.01 seconds to complete. The method is unsuitable for real-time applications despite the noise reduction. The suggested method makes use of image filtering to apply a Gaussian filter on mammography and MRI images to reduce noise. The tumorous area is then removed. We separated the MRI and mammography images using an edge detection and edge detection method. Calculating the volatility of the final image allowed for a comparison analysis to be done in order to find the optimal thresholding and edge detection technique [8].

An Overview of Deep Learning Methods for Breast Cancer Diagnosis is presented by [9] Using Multi-Image Modalities with certain strategies like, "Routine mammograms, which are a low-cost, effective method for calcification identification, are one of the strengths of mammography, which is used as the first wave of imaging for the diagnosis of early-stage breast cancer, MRI, which has strengths such as MRI, which is used to prevent high risk for breast cancer patients and for those with high sensitivity hereditary breast cancer. MRI is ideal for examining and screening breasts, evaluating the size of suspicious lumps region identified in mammograms, and other tasks, in order to avoid unnecessary biopsy, have the strength that may be used as a biopsy guideline, and to find masses, ultrasound is used to differentiate between thick and soft scar tissue, malignant and benign tissue, as well as to avoid unnecessary biopsy".

"Among others [10] used methods like, a gadget for detecting infrared radiation from objects with temperatures above absolute zero is a thermo-camera, Artificial Neural Networks (ANN) are an attempt to imitate biological brain networks using a network of connected, smaller computing units called neurons, Analysis of Images. The majority of research on IR-thermal imaging uses an infrared database.

During the pre-processing stage, noise is removed from the input cytology images. This operation was carried out by a robust filtering technique. Using his to-sigmoid data and fuzzy clustering, this was accomplished. Using artificial intelligence, diagnose breast cancer: [11] conducted a detailed analysis of the literature using computer vision and machine learning approaches. Given that the data are MRI images and that CNN is regarded as the best in computer vision challenges, CNN was mentioned as the algorithm for tumor diagnosis and subtype categorization. The dataset that is most frequently utilized is the cancer Genome Atlas. For researchers, it offers a lot of examples. It provides some general information and clinical data to each participant. Analysis of Histopathological Images for the Detection of Breast Cancer, Detection and classification of breast cancer using the Cubic SVM, K-nearest neighbor and fuzz KNN with 50/50 training and testing validation were used by [12] to achieve detection performance of 98.25 and 98.83%, respectively.

[13] Conducted a deep learning-based study to detect breast cancer on early stages using MRI. Despite having a high sensitivity for spotting breast cancer, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) frequently results in pointless biopsies and patient workup. In order to tailor patient care for patients receiving DCE-MRI and increase overall breast cancer detection accuracy, we deployed a deep learning (DL) approach. Our system attained an area under the receiver operating characteristic curve (AUROC) on the internal test set (n = 3936 exams) of 0.92 (95% CI: 0.92 to 0.93)". In a reader research that was conducted retrospectively, there was no statistically significant difference between the DL system and five board-certified breast radiologists (mean AUROC, +0.04 in favor of the DL system; P = 0.19). When radiologists' forecasts were combined with DL's predictions, their performance improved [mean AUPRC (area under the precision-recall curve), +0.07]. They used numerous datasets from Poland and the US to show how the DL technique is generalizable. Using a Polish dataset, a reader research revealed that the DL system was just as resistant to distribution shift as radiologists. We noticed consistent outcomes in subgroup analysis across various cancer subtypes and patient demographics. With the aid of decision curve analysis, we were able to demonstrate how the DL system can decrease pointless biopsies in the vicinity of clinically meaningful risk thresholds. In up to 20% of all patients with BI-RADS category 4 lesions, this might result in avoiding biopsies that produced benign results. Finally, we conducted an error analysis, looking into instances where DL predictions were primarily wrong. This exploratory work lays the groundwork for the application and future investigation of DL-based breast models.

A study of machine learning based system to detect breast cancer is introduced by [14]. In this study they explain that According to the Breast Cancer Organization, breast cancer is one and only one of the most dangerous types of viruses that are active in the biosphere for females. It helps to maintain breaths by using experimental expert methods to identify this malignancy in its early stages. Based on the cancer.net plan, hereditary illnesses were linked to individual funnels for an additional 120 types of cancer. AI training methods are primarily used to find breast cancer. The WBC (Wisconsin Breast Cancer) record was used to break down the breast cancer using the adaptive ensemble voting system that was anticipated. Our goal is to associate and explain how the logistic algorithm and CNN may be utilized to find breast cancer despite the condensed factors. There are still 2 categories of tumors to be located here. Malignant tumors are cancerous tumors, whereas benign tumors are non-cancerous growths.

[15] Conducted a study of deep learning-based system for diagnosis of breast cancer in early stage. In this study they stated that Finding breast cancer as soon as possible is crucial. This publication introduces a new approach to categorizing breast cancer using deep learning and a few segmentation approaches. For the purpose of categorizing benign and malignant mass tumors in breast mammography pictures, a new computer-aided detection (CAD) approach is suggested. Data augmentation, then, is a technique for enlarging the input data by creating new data from the original input data. The form of data augmentation used in this case is rotation, which has several different variations. When manually clipping the ROI from the mammography, the DCNN architecture with new training achieves an accuracy of 71.01% [16]. The samples acquired using both segmentation procedures had an area under the curve (AUC) that was highest at 0.88 (88%) overall. Furthermore, the accuracy of the DCNN is raised to 73.6% when employing the

samples from the CBIS-DDSM. As a result, the SVM accuracy rises to 87.2% with an AUC of 0.94 (94%). When compared to earlier research conducted under similar settings, this AUC value is the greatest. All of the above studies detect the breast cancer using ML and DL techniques. Some researchers used ML models to detect the breast cancer using the textual data and some researcher using DL method to detect the disease from image dataset. The Proposed Model is focused on using a hybrid methodology that combine the characteristics of ML and DL methods by extracting the meaningful features from the images of patient using DL models and then classify the disease using ML models.

3. Proposed Framework

3.1. Dataset

The used dataset contains two class value B (Benign) and M (Malignant). The Break His database is used in our study that contains macroscopic biopsy images of benign and malignant breast tumors. For this dataset, there are distinct folders for training and test data that created, and each file has a different screen image. If you are more interested in the dataset, there is included a

Table 1 Image Distribution

Images	No of images
Used for Training	1148
Used for Testing	545
All	1693

Table 1 shows that total of 1693 images will be use in this study in which 1148 images are used for the training of models and 545 images are used for the testing of models.

3.2. Methods

The Proposed Framework will contain four phases that is depicted in figure. In this proposed methodology first we will select a Breast Cancer image dataset from the Kaggle repository ("https://www.kaggle.com/datasets/for deration/breakhis-400x") and then clean the dataset by removing the blur images and then labeled the images, Second, the different feature will apply extraction techniques (SIFT) and feature selection techniques to choose the optimized feature on the basis of their covariance values and then split the dataset by using two techniques (Taring &Testing split and Cross-Validation), thirdly will apply some ML models on the optimized dataset and finally evaluate the ML models on the basis of Precision Recall F1-Score and accuracy and also confusion matrix graphs are constructed.

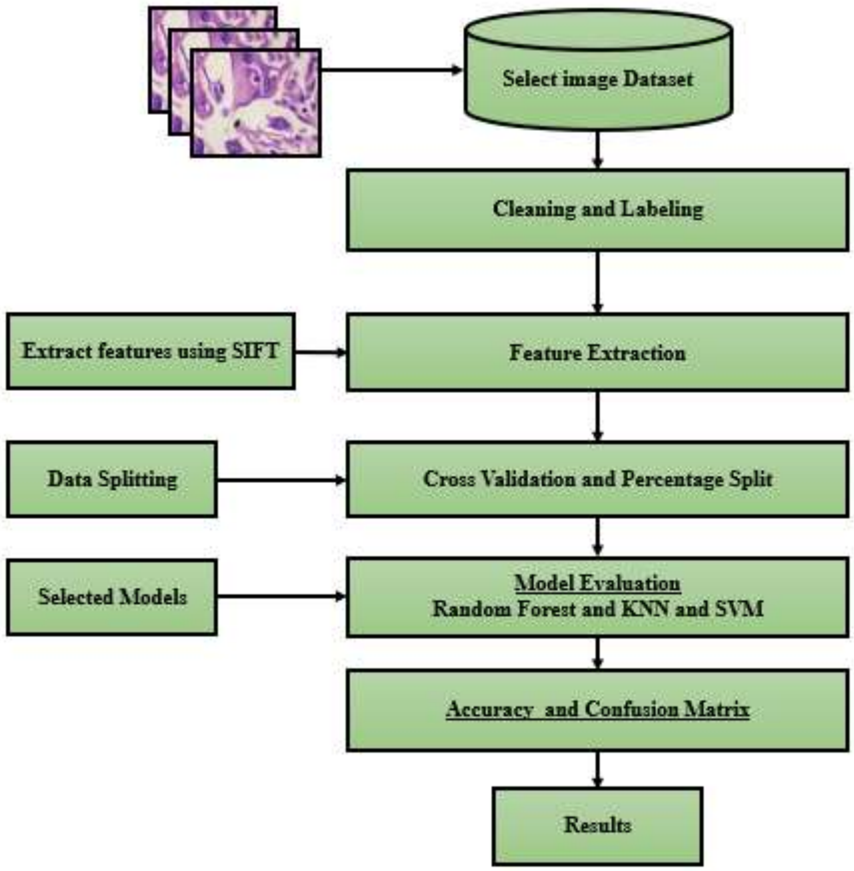

Figure 2 Proposed Framework

The steps of the proposed methodology are given below:

- Cleaning and Labeling
- Feature Extraction
- Data splitting
- Model Evaluation

3.3. Cleaning and Labeling

Image cleaning, also known as image preprocessing or image enhancement, refers to the set of techniques and processes applied to an image to improve its quality, remove noise, and make it more suitable for further analysis or visualization. Cleaning techniques can include in our research work are given in the below section:

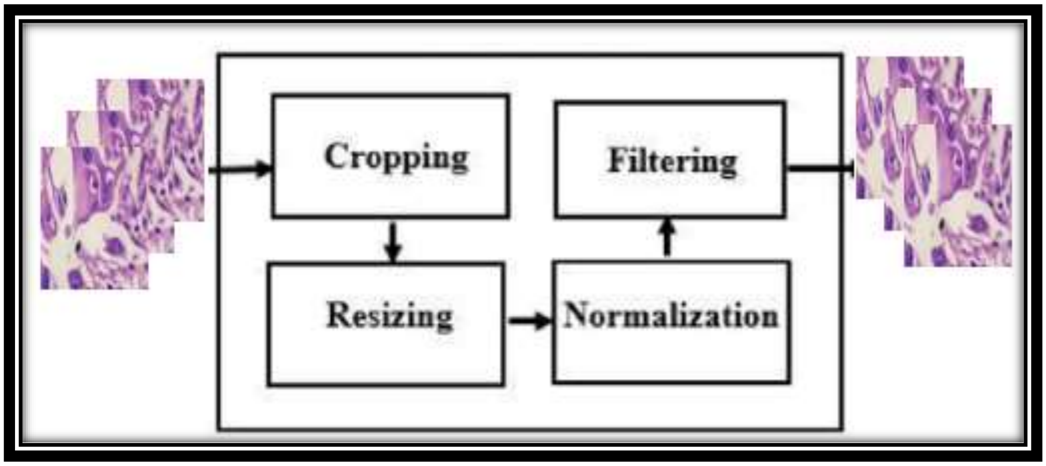

Figure 3 Image Preprocessing

- Noise Removal: Removing undesirable noise from the photo that might intrude with analysis. This would possibly involve making use of filters like Gaussian blur, median filter out, or denoising algorithms like bilateral filtering.
- Contrast Enhancement: Adjusting the contrast and brightness of an image to bring out important features and make details more visible.
- Image Smoothing: Reducing sharp transitions and details in an image to create a smoother version, often achieved using techniques like blurring or averaging.
- Image Sharpening: Enhancing the rims and information in an photo to enhance its readability.
- Normalization: Scaling pixel values to a common range to ensure constant analysis across snap shots.
- Resizing and Cropping: Changing the dimensions of the image or selecting a specific region of interest.

Labeling is vital for training supervised learning models. In supervised learning, pictures need to be paired with corresponding labels to educate the version what the items appear to be and assist it make predictions on new, unseen information. Labeling, in the context of image processing, involves assigning semantic labels or identifiers to particular areas or objects inside an image. Labeling is frequently utilized in duties consisting of object detection, picture segmentation, and category. The process of labeling may be guide or automatic.

3.4. Feature Extraction

The method of selecting and representing distinct data from a images in a form that is better appropriate for ML as feature extraction within the context of images processing. A big amount of records can be contained in a image, not all of which may be pertinent for a given hobby. By capturing the maximum crucial developments or styles in a picture at the same time as minimizing its complexity, function extraction can help.

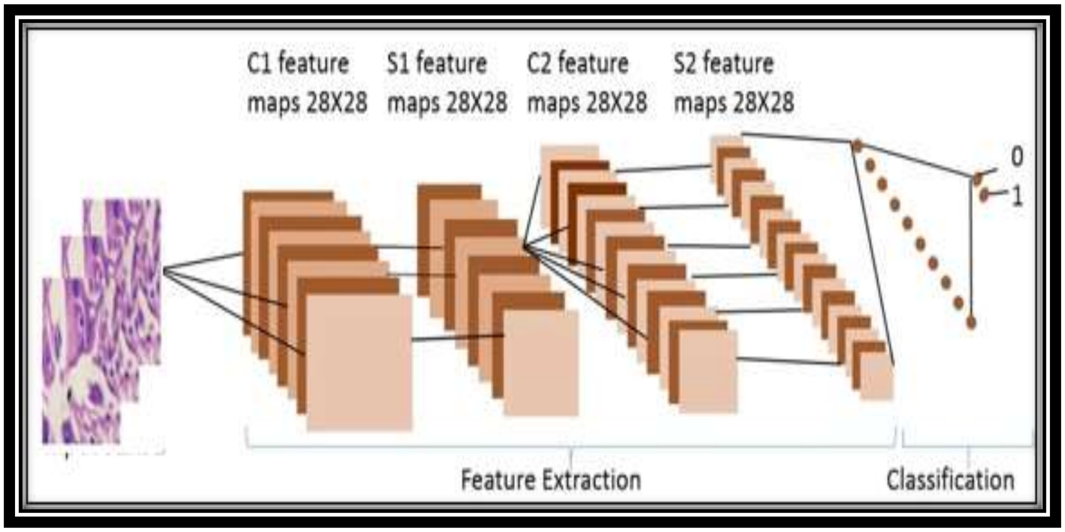

Figure 4 Feature Extraction Method

"Scale-Invariant Feature Transform (SIFT) is a popular feature extraction algorithm utilized in computing imaginative and prescient for detecting and describing neighborhood functions in pictures, regardless of their scale, rotation, or lighting fixtures adjustments". The evaluation of ways SIFT works for function extraction:

Key points Localization: "Low-assessment key points and key points which might be too near the photograph boundary, in which right gradient computation is tough, are removed from the key points that were comprised of the DoG pyramid".

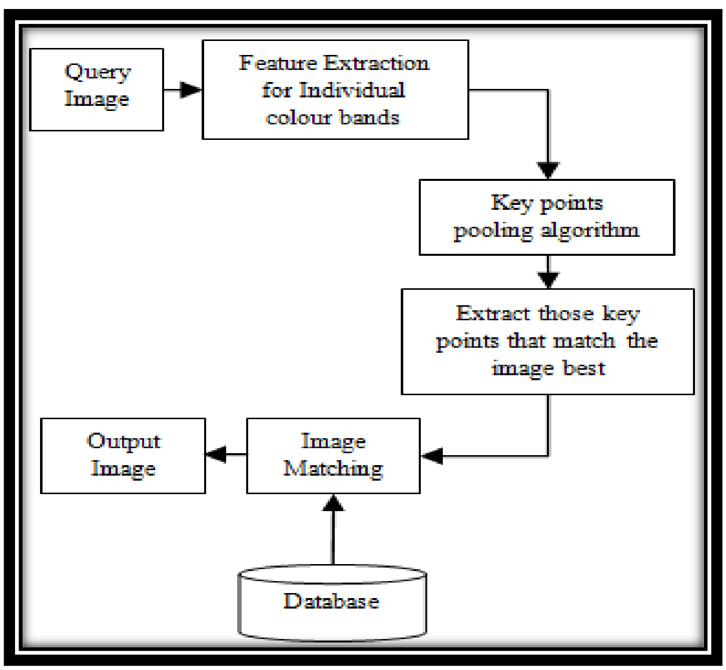

Figure 5 SWIFT Working Diagram

3.5. Data splitting

After extraction of features the next step is to divide the dataset into training and testing for the evaluation of selected dataset. Eighty percent records are used for the training of machine learning models (Random Forest, K Nearest Neighbor and Support Vector Machine) and 20 percent for the testing of models. Cross Validation is also used for the evaluation of models by using 10K-Folding. Other data splitting methods are also used but they do not give better results as compare to the 80 percent training and 20 percent testing.

3.6. Model Evaluation

There are three famous machine learning models using in this proposed study, Random Forest, Support Vector Machine and K-Nearest Neighbors. This study evaluated builder machine learning models (KNN, RF, SVM) on the basis of precision, accuracy, recall and f1 score.

4. Implementation and Results

This study using Colab tool to build three machine learning classifiers in python language. Accuracy, Classification Report and confusion matrix of all the classifiers are calculated. Random Forest (RF) Model is used with n_estimators of 100, maximum depth of tree with value six and random state of 42. RF is evaluated on other attributes like n_estimators of (10, 50, 70, 150 etc.) but not give better results as compare to the n-estimators of 100, so n-estimators of 100 is used in our RF model evaluation. RF model is evaluated on different data distribution method (80:20 percent, 30:70 percent and cross validation of 10K folding) but 80 and 20 percent division for training and testing gives highest result as compare to others.

The second model that is applied on the extracted feature is the Support Vector Machine (SVM) and different functions parameters of SVM (Linear, Sigmoid, and Radial Based Function/rbf) is used to evaluate the model by using sklearn library of python language. Sigmoid Function of SVM giving better results as compare to linear and rbf. Like RF model, SVM is evaluated on different data distribution method (80:20 percent, 30:70 percent and cross validation of 10K folding) but 80 and 20 percent division for training and testing gives highest result as compare to others. Third model that is applied on the extracted feature is K Nearest Neighbor with K neighbors of 3. Different other neighbor values are also check to improve the accuracy but 3 gives the highest value as compare to 2,4,5 etc.

For experiment purposes in the proposed study the specifications regarding hardware and software used including Windows 10 Pro Operating system, having 500GB HDD with Memory power of 16GB and the generation of system is 11^{th} with 5 cores, the tool used for experiments are Google Colab using python language.

4.1. Accuracy

This study trains all models on 80 percent of dataset and test on remaining 20 percent dataset. Table 4 shows the training and testing accuracies of classifiers. RF classifier gives training accuracy of 98.12%

and testing accuracy of 94.63%, SVM classifier gives training accuracy of 82.36% and testing accuracy of 83.08%, KNN classifier gives training accuracy of 96.90% and testing accuracy of 81.78%.

Table 2 Training and Testing Accuracy of Models

Algorithm	Training	Testing
RF	98.12	94.63
SVM	82.36	83.08
KNN	96.90	81.78

Random forest performs better on both training and testing dataset as compared to other classifiers but SVM gives poor results. Figure 9 chows the graph of all three models with both training and testing accuracy values and distinguish the models from other two models. Graph shows that model RF perform well.

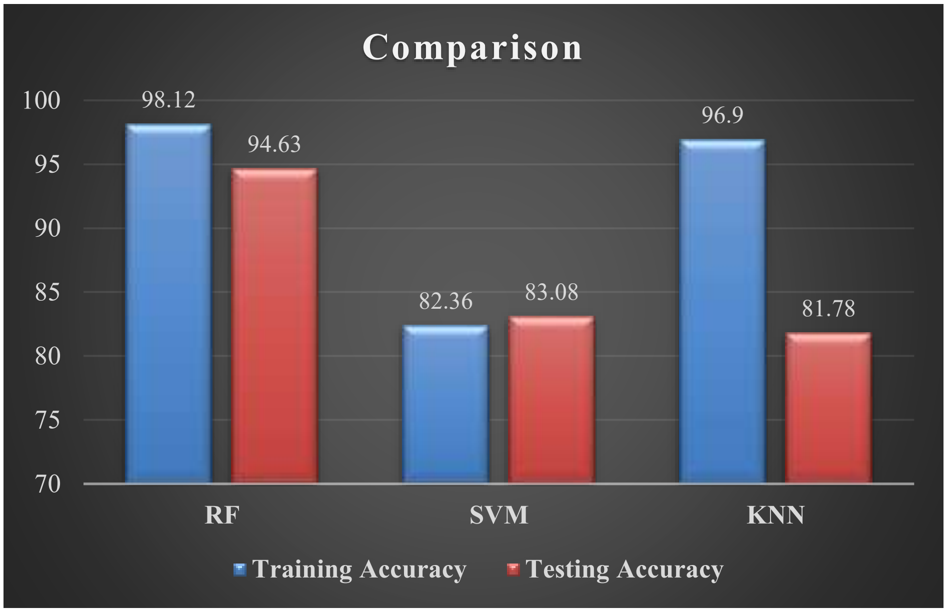

Figure 6 Comparison of Models

4.2. Classification Report

The classification report of each class is calculated by every classifier and gives result of precision, recall and f1-score of class Malignant and Benign. Table 5, 6 and 7 shows the classification reports of classifiers for each class individually.

Table 3 RF Classification Report

Class	Precision	Recall	F1-Score
Benign	82	85	88
Malignant	81	72	76
Average	82	80	81

Table 3 shows that Random Forest Model gives precision of 82, 81 and 82 percent of class Benign, Malignant and average. Class malignant gives recall of 72 percent and f1score of 76 percent. Benign class having and recall of 85 percent and F1-score of 88 respectively

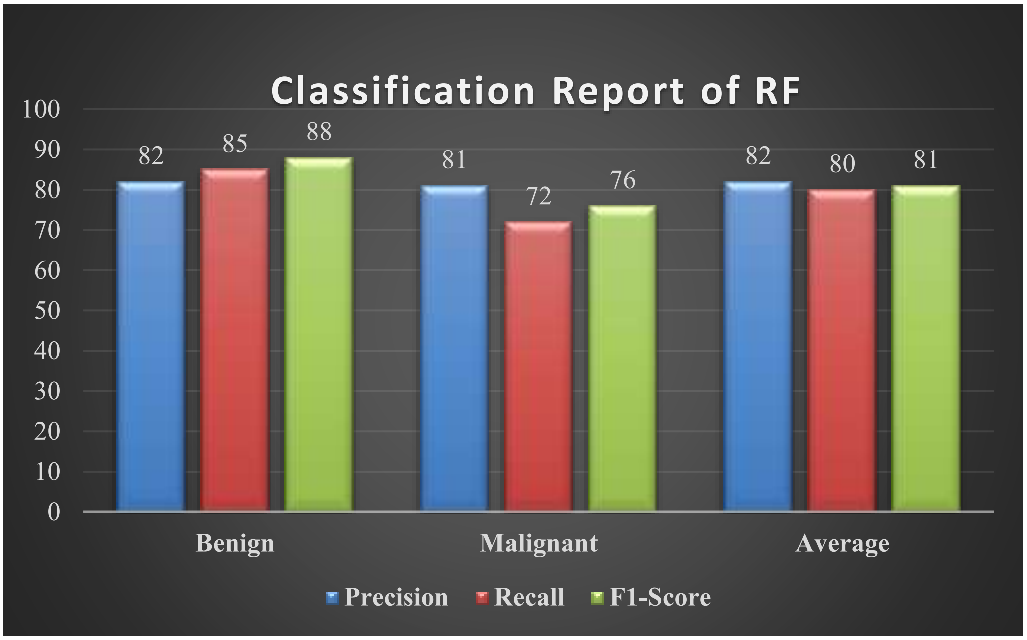

Figure 7 Classification Report of RF

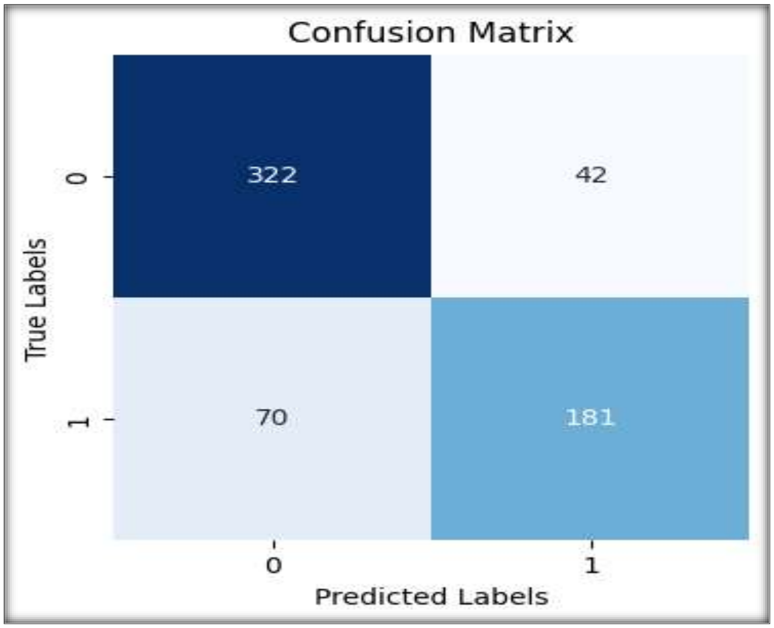

Figure 8 RF Graph

Table 4 SVM Classification Report

Class	Precision	Recall	F1-Score
Benign	83	90	86
Malignant	83	74	78
Average	83	82	82

Table 4 shows the confusion matrix results of SVM that gives precision of 83, 83 and 83 percent of class Benign, Malignant and average. Class malignant gives recall of 74 percent and f1score of 78 percent. Benign class having and recall of 90 percent and F1-score of 86 respectively.

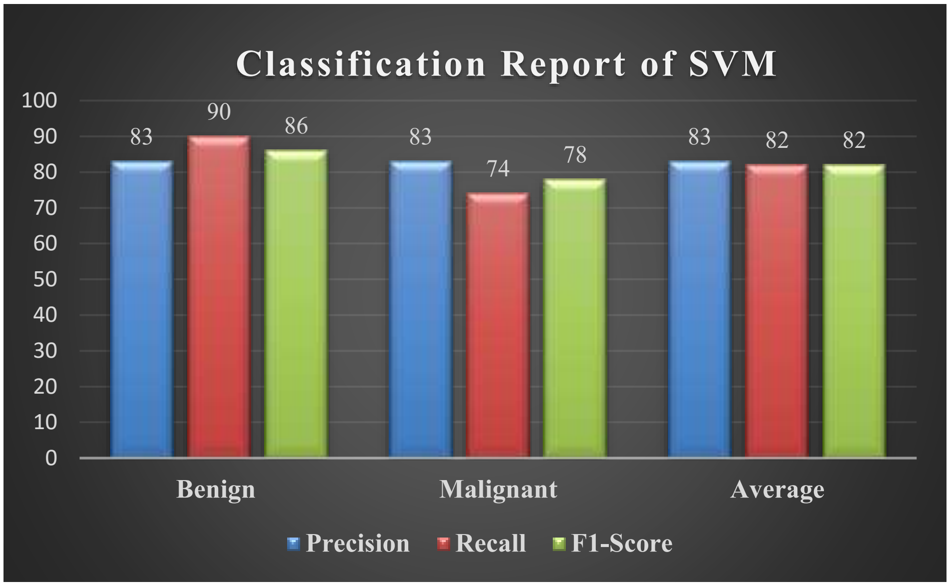

Figure 9 Classification Report of SVM

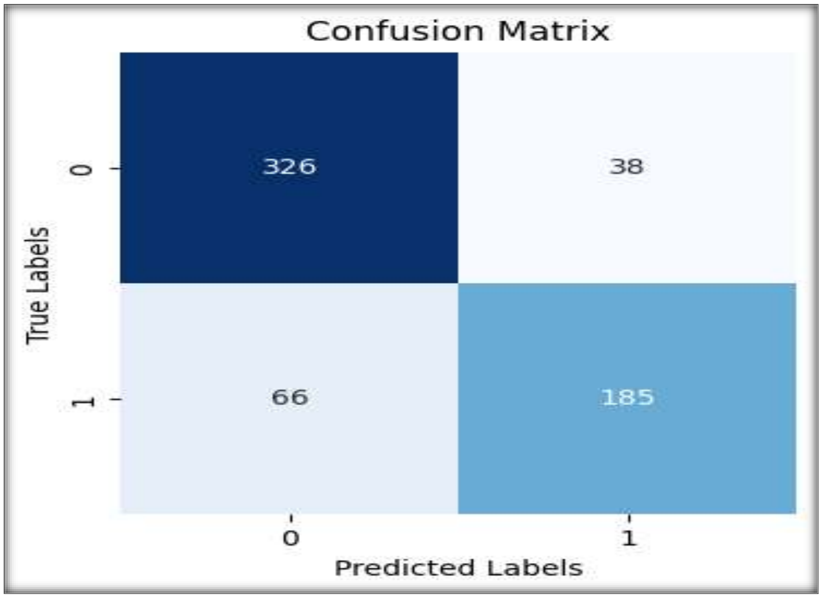

Figure 10 SVM Graph

Table 5 KNN Classification Report

Class	Precision	Recall	F1-Score
Benign	82	88	85
Malignant	81	72	76
Average	82	80	81

Table 5 shows that KNN gives precision of 82, 81 and 82 percent of class Benign, Malignant and average. Class malignant gives recall of 72 percent and f1score of 76 percent. Benign class having and recall of 88 percent and F1-score of 85 respectively

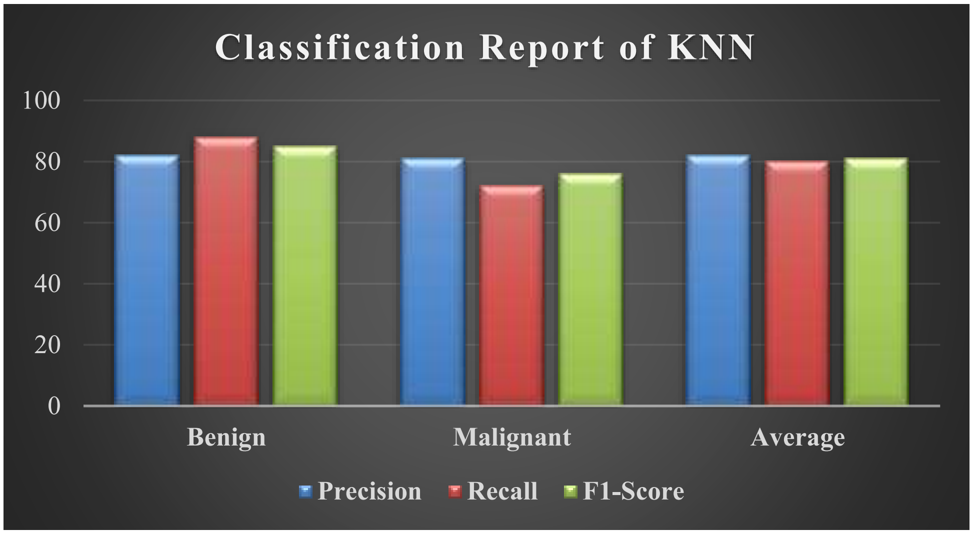

Figure 11 Classification Report of KNN

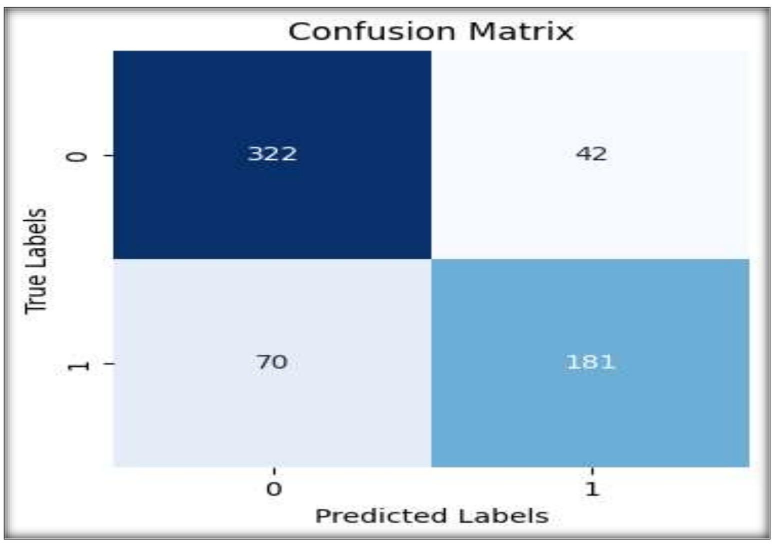

Figure 12 KNN Graph

5. Conclusion

Detecting early signs of breast cancer became an imperative task, given its prevalence among women globally. Identifying the illness promptly was crucial for saving lives, as breast cancer ranked among the most common cancers in women. Last year alone, more than 2.3 million women received a breast cancer diagnosis, and sadly, approximately 0.7 million lost their lives to it. Manual disease diagnosis showed poor accuracy, often failing to detect aggressive cancer at an early stage, resulting in patient fatalities. Therefore, employing machine learning to identify early breast cancer warning signs became imperative. The primary objective of this study was to showcase the application of machine learning (ML) and deep learning (DL) techniques to efficiently and reliably detect breast cancer for the purpose of saving lives. In this study, a hybrid model was used to combined the ML and DL approaches to predict the images of breast cancer. To unlock the full potential of the image dataset, the Scale-Invariant Feature Transform (SWIFT) DL technique was initially used to extract meaningful information. Subsequently, various machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machine and Random Forest, were applied to classify cases into malignant and benign categories. The central aim of this research was to differentiate between benign and malignant breast cancers by leveraging distinct features extracted from a dataset comprising images of healthy individuals and breast cancer patients. Results were compared based on the models' precision, recall, accuracy, and F1-Score.

References

[1] Ashrafi Zadeh, M., Zarrabi, A., Bigham, A., Taheri Azam, A., Saghari, Y., Mirzaei, S., ... & Shen, M. (2023). (Nano) platforms in breast cancer therapy: Drug/gene delivery, advanced nanocarriers and immunotherapy. Medicinal Research Reviews.

[2] Britt, K. L., Cuzick, J., & Phillips, K. A. (2020). Key steps for effective breast cancer prevention. Nature Reviews Cancer, 20(8), 417-436.

[3] Debien, V., De Caluwé, A., Wang, X., Piccart-Gebhart, M., Tuohy, V. K., Romano, E., & Buisseret, L. (2023). Immunotherapy in breast cancer: an overview of current strategies and perspectives. NPJ Breast Cancer, 9(1), 7.

[4] Fahad Ullah, M. (2019). Breast cancer: current perspectives on the disease status. Breast Cancer Metastasis and Drug Resistance: Challenges and Progress, 51-64.

[5] Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., ... & Siegel, R. L. (2022). Breast cancer statistics, 2022. CA: a cancer jnal for clinicians, 72(6), 524-541.

[6] Kashyap, D., Pal, D., Sharma, R., Garg, V. K., Goel, N., Koundal, D., ... & Belay, A. (2022). Global increase in breast cancer incidence: risk factors and preventive measures. BioMed research international, 2022.

[7] Katsura, C., Ogunmwonyi, I., Kankam, H. K., & Saha, S. (2022). Breast cancer: presentation, investigation and management. British Jnal of Hospital Medicine, 83(2), 1-7.

[8] Kumar, A., Sushil, R., & Tiwari, A. K. (2019). Comparative study of classification techniques for breast cancer diagnosis. Int. J. Comput. Sci. Eng, 7, 234-240.

[9] Kunkler, I. H., Williams, L. J., Jack, W. J., Cameron, D. A., & Dixon, J. M. (2023). Breast-conserving surgery with or without irradiation in early breast cancer. New England Jnal of Medicine, 388(7), 585-594.

[10] Obaid, O. I., Mohammed, M. A., Ghani, M. K. A., Mostafa, A., & Taha, F. (2018). Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. International Jnal of Engineering & Technology, 7(4.36), 160-166.

[11] Turner, N. C., Oliveira, M., Howell, S. J., Dalenc, F., Cortes, J., Gomez Moreno, H. L., ... & Rugo, H. S. (2023). Capivasertib in Hormone Receptor–Positive Advanced Breast Cancer. New England Jnal of Medicine, 388(22), 2058-2070.

[12] Yedjou, C. G., Sims, J. N., Miele, L., Noubissi, F., Lowe, L., Fonseca, D. D., ... & Tchounwou, P. B. (2019). Health and racial disparity in breast cancer. Breast cancer metastasis and drug resistance: challenges and progress, 31-49.

[13] Zagami, P., & Carey, L. A. (2022). Triple negative breast cancer: Pitfalls and progress. NPJ Breast Cancer, 8(1), 95.

[14] Sahni, P., & Mittal, N. (2019). Breast cancer detection using image processing techniques. In Advances in Interdisciplinary Engineering: Select Proceedings of FLAME 2018 (pp. 813-823). Springer Singapore.

[15] Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Rehman, K. U. (2020). A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. IEEE Access, 8, 165779-165809.

[16] Al Husaini, M. A. S., Habaebi, M. H., Hameed, S. A., Islam, M. R., & Gunawan, T. S. (2020). A systematic review of breast cancer detection using thermography and neural networks. IEEE Access, 8, 208922-208937.

[17] Nassif, A. B., Talib, M. A., Nasir, Q., Afadar, Y., & Elgendy, O. (2022). Breast cancer detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine, 127, 102276.

[18] Maqbool, M. S., Hanif, I., Iqbal, S., Basit, A., & Shabbir, A. (2023). Optimized feature extraction and cross-lingual text reuse detection using ensemble machine learning models. Journal of Computing & Biomedical Informatics, 5(01), 26-40.

[19] Abid, K., Aslam, N., Fuzail, M., Maqbool, M. S., & Sajid, K. (2023). An efficient deep learning approach for prediction of student performance using neural network. VFAST Transactions on Software Engineering, 11(4), 67-79.

[20] Kanwal, F., Abid, M. K., Maqbool, M. S., Aslam, N., & Fuzail, M. (2023). Optimized classification of cardiovascular disease using machine learning paradigms. VFAST Transactions on Software Engineering, 11(2), 140-148.

[21] Basit, A., Hanif, I., Maqbool, M. S., Qayyum, W., Hasnain, M. A., & Nazeer, R. (2023). Cross-Lingual Information Retrieval in a Hybrid Query Model for Optimality. Journal of Computing & Biomedical Informatics, 5(01), 130-141.

[22] Tuz Zuhra, Fatima, and Khalid Saleem. "Towards Development of New Language Resource for Urdu: The Large Vocabulary Word Embeddings." ACM Transactions on Asian and Low-Resource Language Information Processing 24.8 (2025): 1-14.

[23] Hasnain, M. A., Ali, S., Malik, H., Irfan, M., & Maqbool, M. S. (2023). Deep learning-based classification of dental disease using x-rays. Journal of Computing & Biomedical Informatics, 5(01), 82-95.

[24] Rafiqee, M. M., Qaiser, Z. H., Fuzail, M., Aslam, N., & Maqbool, M. S. (2023). Implementation of efficient deep fake detection technique on videos dataset using deep learning method. Journal of Computing & Biomedical Informatics, 5(01), 345-357.

[25] Malik, F., Fuzail, M., Aslam, N., Sarwar, R., Abid, K., Maqbool, M. S., & Yousaf, A. (2024). A Hybrid Machine Learning Model to Predict Sentiment Analysis on X. Journal of Computing & Biomedical Informatics, 6(02), 64-79.

[26] Fazal, U., Khan, M., Maqbool, M. S., Bibi, H., & Nazeer, R. (2023). Sentiment analysis of omicron tweets by using machine learning models. VFAST Transactions on Software Engineering, 11(1), 67-75.

[27] Alanazi, S. A., Kamruzzaman, M. M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting breast cancer detection using convolutional neural network. Journal of Healthcare Engineering, 2021.

[28] Trabelsi Ben Ameur, S., Sellami, D., Wendling, L., & Cloppet, F. (2019). Breast cancer diagnosis system based on semantic analysis and Choquet integral feature selection for high risk subjects. Big Data and Cognitive Computing, 3(3), 41

[29] Dar, R. A., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Computers in biology and medicine, 106073.

[30] Witowski, J., Heacock, L., Reig, B., Kang, S. K., Lewin, A., Pysarenko, K., ... & Geras, K. J. (2022). Improving breast cancer diagnostics with

[31] Nallamala, S. H., Mishra, P., & Koneru, S. V. (2019). Breast cancer detection using machine learning way. Int J Recent Technol Eng, 8(2-3), 1402-1405.

[32] Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ, 7, e6201.

[33] Mambou, S. J., Maresova, P., Krejcar, O., Selamat, A., & Kuca, K. (2018). Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors, 18(9), 2799.

[34] Hamed, G., Marey, M. A. E. R., Amin, S. E. S., & Tolba, M. F. (2020). Deep learning in breast cancer detection and classification. In Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020) (pp. 322-333). Springer International Publishing.

[35] Bazazeh, D., & Shubair, R. (2016, December). Comparative study of machine learning algorithms for breast cancer detection and diagnosis. In 2016 5th international conference on electronic devices, systems and applications (ICEDSA) (pp. 1-4). IEEE.

[36] Sharma, A., Kulshrestha, S., & Daniel, S. (2017, December). Machine learning approaches for breast cancer diagnosis and prognosis. In 2017 International conference on soft computing and its engineering applications (icSoftComp) (pp. 1-5). IEEE.

[37] Hung, P. D., Hanh, T. D., & Diep, V. T. (2018, December). Breast cancer prediction using spark MLlib and ML packages. In Proceedings of the 2018 5th International Conference on Bioinformatics Research and Applications (pp. 52-59).

[38] Sindhwani, N., Rana, A., & Chaudhary, A. (2021, September). Breast cancer detection using machine learning algorithms. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

[39] Abid, S., Akbar, S., Hassan, S. A., & Gull, S. (2022, October). Automatic Detection of Breast Cancer through Mammogram Images. In 2022 International Conference on IT and Industrial Technologies (ICIT) (pp. 1-7). IEEE.