

# A Novel Feature Fusion for the Classification of Histopathological Carcinoma Images

Salini S Nair<sup>1</sup>, M. Subaji<sup>2\*</sup>

School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, India<sup>1</sup>  
Institute for Industry and International Programmes, Vellore Institute of Technology, Vellore, India<sup>2</sup>

**Abstract**—Breast cancer is a significant global health concern, demanding advanced diagnostic approaches. Although traditional imaging and manual examinations are common, the potential of artificial intelligence (AI) and machine learning (ML) in breast cancer detection remains underexplored. This study proposes a hybrid approach combining image processing and ML methods to address breast cancer diagnosis challenges. The method utilizes feature fusion with gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), and histogram features, alongside an ensemble learning technique for improved classification. Results demonstrate the approach's effectiveness in accurately classifying three carcinoma classes (ductal, lobular, and papillary). The Voting Classifier, an ensemble learning model, achieves the highest accuracy, precision, recall, and F1-scores across carcinoma classes. By harnessing feature extraction and ensemble learning, the proposed approach offers advantages such as early detection, improved accuracy, personalized medicine recommendations, and efficient analysis. Integration of AI and ML in breast cancer diagnosis shows promise for enhancing accuracy, effectiveness, and personalized patient care, supporting informed decision-making by healthcare professionals. Future research and technological advancements can refine AI-ML algorithms, contributing to earlier detection, better treatment outcomes, and higher survival rates for breast cancer patients. Validation and scalability studies are needed to confirm the effectiveness of the proposed hybrid approach. In conclusion, leveraging AI and ML techniques has the potential to revolutionize breast cancer diagnosis, leading to more accurate and personalized detection and treatment. Technology-driven advances can significantly impact breast cancer care and management.

**Keywords**—Breast cancer; machine learning; artificial intelligence; feature extraction; ensemble classifier

## I. INTRODUCTION

Cancer is one of the complex and devastating diseases that continues to pose significant challenges to global healthcare systems and individuals worldwide [1]. It is one of the most dreadful diseases that is not easily curable. In the body, aberrant cells develop and spread out of control, which is a term used to describe a set of disorders called Cancer [2]. In the modern world methods like computerized tomography (CT) scans, magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, etc. are used to detect this disease [3]. Breast cancer stands out among the numerous types of cancer as one of the commonest and worrisome forms, impacting millions of people every year [4]. Breast cancer often affects the breast tissue and frequently begins in the milk-producing glands (lobules) or the ducts that

supply milk to the nipple. Mammography, a low-dose X-ray examination of the breast, is the most common technique used for the common detection of breast abnormalities [5]. In addition to this clinical breast examination is another method performed by professionals to detect it. The application of artificial intelligence and associated approaches is still not well practiced for this goal, despite the fact that there are numerous computerized automated procedures utilized for the diagnosis and detection of breast cancer [6]. Manual work that has to be done to diagnose and detect even after this automated process is still cumbersome since it demands intelligent decision-making [7]. The introduction of AI-ML on it will be the solution to it, where doctors do not need to manually examine and diagnose the disease [8]. The advancement of technology has enabled different kinds of methods to detect cancer which mainly include Liquid biopsy, Genome profiling, Image techniques, Metabolomics, Optical techniques, and finally AI -ML techniques [9]. Although these cutting-edge techniques have substantially improved cancer detection, their use may differ depending on the type and stage of the disease, the accessibility of resources, and the state of the healthcare system. Moving forward AI-ML techniques have the potential to be further honed and improved by future research and technological developments, which could ultimately result in earlier cancer diagnosis, better treatment outcomes, and higher overall survival rates for cancer patients [10]. Here we use traditional image processing and machine learning techniques in a hybrid way to realize the detection module. The major advantages of using this technique are early detection, improved accuracy, helping to suggest better personalized medicine, faster and more efficient analysis, integration and multimodal data, and continuous learning and improvement [11]. The benefits of utilizing AI-ML algorithms for cancer diagnosis, as described above, are generally very applicable to the particular situation of breast cancer.

The proposed approach is driven by a comprehensive set of motivations and potential benefits that promise to significantly advance the field of histopathological image classification. Its core aim is to elevate the accuracy and robustness of this critical task. To achieve this, the approach combines three distinct feature extraction techniques: GLCM, which captures pixel-level spatial relationships; LBP, designed to characterize intricate texture patterns; and histogram features, which provide a global view of intensity distribution within the images. By amalgamating these diverse features, the approach seeks to create a holistic representation of the carcinoma images, enabling the model to capture both local nuances and global context, thus enhancing classification

\*Corresponding Author.

accuracy. Histopathological carcinoma images are notoriously diverse due to variations in tissue preparation, staining, and imaging conditions. Therefore, another vital motivation is to bolster the model's resilience to such variability. The fusion of GLCM, LBP, and histogram features offers a multi-faceted approach to understanding these images, making it more adaptable to different staining protocols and equipment, ultimately resulting in a more reliable diagnostic tool.

Moreover, the approach combats overfitting—a common challenge in machine learning—by employing an ensemble of classifiers. Ensemble methods aggregate the decisions of multiple classifiers, reducing the risk of the model memorizing noise in the training data and improving its generalization performance. This becomes crucial in histopathological image classification, where datasets can be limited in size and prone to noise. Class imbalance is yet another challenge in this domain, with some carcinoma subtypes having fewer samples than others. The fusion technique, coupled with appropriate strategies like weighted voting, can help address these class imbalance issues, ensuring that the model's performance is not skewed towards the majority class, which can be critical for effective clinical diagnosis.

The combined use of different feature types also enhances the interpretability of classification results. Researchers and clinicians can gain insights into which aspects of the images are most influential in making the classification decisions. This not only provides transparency in the model's decision-making process but also aids in building trust in its recommendations. Additionally, the versatility of this approach extends to its potential for transferability. By fusing diverse features and leveraging ensemble classifiers, it can potentially be applied to related image classification tasks within the medical domain, paving the way for broader applicability and impact.

In essence, the given method is a forward-thinking approach that aims to improve classification accuracy, increase model robustness, and enhance the overall performance of histopathological carcinoma image classification. By integrating multiple feature extraction methods and harnessing the power of ensemble classifiers, this approach holds great promise in delivering more accurate and reliable cancer diagnoses, thus contributing significantly to the field of medical image analysis and ultimately benefiting patients and healthcare providers.

These techniques could enhance the precision, effectiveness, and personalization of breast cancer diagnosis and treatment, improving patient outcomes and assisting doctors in their decision-making [12]. In this proposed work we have used a feature fusion for extracting robust features and ensemble learning [13] for better classification performance on classifying the three classes of carcinoma images said ductal, lobular, and papillary. We have used features like GLCM, LBP [14] and Histogram [15]. The novel approach outperformed existing techniques even without using any computational heavy deep learning technique. The remaining portion of the paper is described as essential

preliminaries, detailed methodology, obtained results analysis and discussions, conclusion and the future work.

## II. LITERATURE REVIEW

Alqudah et al. [16] proposed a new sliding window technique for local feature extraction from 25 sliding windows for each image. They used the LBP for features extraction of each window, support vector machine (SVM) to classify the windows, and to find the final class based on the majority voting technique. For the categorization of breast cancer, Gour et al. [17] introduced ResHist, a 152-layered convolutional neural network based on residual learning. They extracted discriminative features from the histopathological images and used the data augmentation technique to enhance the model's performance. Gandomkar et al. [18] have proposed classifying hematoxylin-eosin stained breast digital slides of 81 patients resulting in 7786 images in all. They demonstrated a system known as MuDeRN, which stands for "MULTI-category classification of breast histopathological image using DEep Residual Networks." A deep residual network (ResNet) of 152 layers has been trained to categorize patches from the images in the first stage of the project, which comprises of two stages. Second, the images classified as malignant and benign were classified into four subtypes. Using a meta-decision tree, the authors combined the outputs of ResNet's processed images in different magnification factors. Multiscale generalized radial basis function (MSRBF) neural networks were recommended by Beltran-Perez et al. [19] for the extraction and categorization of image features. Three steps make up the architecture described in this work: first, an input-output model is derived from the image; second, high-level image features are extracted from the model; and third, a module for classification is intended to forecast breast cancer. An approach based on deep convolutional neural network that supports 16 layers (VGGNet-16) has been proposed by Kumar et al. [20] who also assessed how well the fused framework performed in comparison to other classifiers like the support vector machine and random forest. They increase the data size using data augmentation.

Li et al. [21] have evaluated histological images using convolutional neural network (CNN) architecture for classification. In order to improve feature information, authors proposed densely-connected-convolutional network (DenseNet) as the fundamental building block and interspersed it with the squeeze-and-excitation network (SENet) module. Vo et al. [22] have proposed data augmentation approaches to improve classification performance in addition to increasing the diagnosis effectiveness of biopsy tissue utilizing hematoxylin and eosin-stained images. To improve classification performance in the situation of a small number of breast cancer images and imbalanced training data, they have presented an ensemble of deep convolutional neural networks (DCNNs) trained to extract visual features from multiscale images and used gradient boosting tree classifiers. Whereas Saxena et al. [23] have proposed a hybrid ML model to solve the class imbalance problem. They created the kernelized weighted extreme learning machine and the pre-trained ResNet50 for breast cancer classification using histological image. Alom et al. [24] state that the Inception Recurrent Residual

Convolutional Neural Network (IRRCNN) is assessed for breast cancer classification at the image, patient, and patch levels. Boumaraf et al. [25] have put forward the deep neural network ResNet-18, and transfer learning helps to avoid overfitting and boost the training speed on histopathological images. Furthermore, they used global contrast normalization (GCN) to strengthen the approach and three-fold data augmentation to enhance the model. On histopathology images, Burçak et al.'s deep convolutional neural network [26] presents a method for automatically identifying and categorizing malignant areas. For quicker backpropagation learning, they computed the network's starting weight and updated the model parameters using a variety of algorithms, including stochastic gradient descent (SGD), nesterov accelerated gradient (NAG), adaptive gradient (AdaGrad), root mean squared propagation (RMSprop), AdaDelta, and Adam. A graphics processing unit with compute unified device architecture (CUDA) support is utilized in parallel computing architecture for quick processing. Xie et al. [27] investigated to extract expressive features from images of breast cancer's histopathology. They suggested Inception\_V3 and Inception\_ResNet\_V2 deep convolutional neural networks that have been developed using transfer learning strategies. Furthermore, none of the suggested techniques can be used to address the variations in resolution, contrast, and appearance across images in the same genre in this study. Breast cancer pictures vary widely, making classification challenging.

Jiang et al. [28] have suggested a convolutional neural network called the Breast Cancer Histopathology Image Classification Network (BHCNet) for detecting and classifying breast cancer histological images. Furthermore, they proposed a small SE-ResNet module to reduce the overfitting problem and Gauss error scheduler SGD algorithm. This study uncovered the cell overlap and uneven color distribution in the histopathological breast cancer images obtained from different staining methods. To address the imbalanced class problem, Han et al. [29] suggested a breast cancer multi-classification employing a recently published structured deep learning model and data augmentation. Kumar et al. [30] proposed the contrast-limited adaptive histogram equalization approach to enhance microscopic biopsy images, and for segmentation,  $k$ -means clustering is used. Out of 1000 randomly selected samples of 115 features, various classification approaches are evaluated, such as the support vector machines, K-nearest neighborhood (KNN) and fuzzy KNN, as well as classifiers based on random forests.

Sheikh et al. [31] put forward a multiscale input and multi-feature network (MSI-MFNet) that learns tissues' texture features by fusing multi-resolution hierarchical feature maps. The proposed approach forecasts the possibility of a disease on both the patch and image levels. Using the structural and statistical data from the images, Nahid et al. [32] proposed novel deep neural network (DNN) approaches. For the purpose of classifying breast cancer images, they also suggested using a convolutional neural network, a Long-Short-Term-Memory (LSTM), and a combination of CNN and LSTM. Once they had extracted the features from the novel

DNN model, they used Softmax and SVM layers to make decisions. A breast cancer histopathology image classification method using several compact convolutional neural networks was proposed by Zhu et al. [33]. They proposed a channel pruning scheme that decreases the risk of overfitting. The different data partition and composition-based models were assembled to enhance the model's ability to classify the data. The graph convolutional network developed by Gong et al. [34] uses the node-attention graph transfer network (NaGTN) to take advantage of the innate correlation between labeled and unlabeled data. In order to undertake the extraction of knowledge for the target domain, this approach uses a fully labeled source domain. Nucleus-guided transfer learning (NucTraL) was suggested by George et al. [35] as a technique for classifying breast tumors. Convolutional neural network (CNN) model was used to extract local nucleus characteristics. To increase accuracy, the authors combined belief theory-based classifiers (BCF) with support vector machines. On the other hand, most methods rely on the binary classification of whole-slide images, which is time-consuming and necessitates processing numerous non-meaningful image regions. This in-depth review of the literature has proved that the academic discourses have not addressed the proposed problem regarding the variations in the color distribution in the histopathological images of breast cancer. Most of the methods address only binary classification problems.

### III. PRELIMINARIES

#### A. Gray-Level Co-Occurrence Matrix

Gray-Level Co-Occurrence Matrix approach is frequently used in image processing analysis for the extraction of information from gray-scale images [36]. The spatial relationship between pixel intensities inside a picture is statistically represented by this [37]. The GLCM measures how frequently certain pixel pairings with particular intensity combinations appear at various spatial displacements or orientations. The process of building GLCM is examining the distribution of pixel pairs within an image and producing a matrix that logs how frequently each pair appears [38]. Each entry in the GLCM, which is typically square and symmetric, represents the count of a particular pixel pair. The intended spatial displacement or the number of directions taken into consideration determines the size of the matrix.

#### B. Local Binary Patterns

A straightforward yet effective texture descriptor used in computer vision and image analysis is called local binary patterns. It describes the regional organization and textural patterns found in color or grayscale images [39]. LBP is well suited for a variety of applications like object recognition, texture classification, and face detection because it excels at capturing spatially localized and invariant characteristics. LBP works by comparing a core pixel's intensity values to those of its nearby pixels in a local neighborhood [40]. The effectiveness of LBP's computations is one of its benefits. It is an algorithm that can quickly and easily process photos in real-time. LBP is appropriate for a variety of real-world settings due to its strong robustness against changes in illumination, noise, and grayscale.

### C. Histogram Features

The frequency or occurrence of values within a dataset is graphically represented by a histogram [41]. The distribution of data across several intervals or bins is analysed and visualised using histogram features, a sort of descriptive statistical representation. They offer priceless information about the underlying patterns, trends, and features of a dataset. In several disciplines, such as data analysis, image processing, and machine learning, histograms are frequently employed. Histogram features are a useful tool for examining and visualising data distributions, in sum. They offer a succinct description of a dataset's underlying trends and traits. Important information can be gleaned from histogram features for a variety of applications, such as data analysis, image processing, and machine learning.

## IV. METHODOLOGY

This section describes the methodology of the proposed work. The Fig. 1 depicts overall architecture which consists of several sub-stages.

### A. Dataset

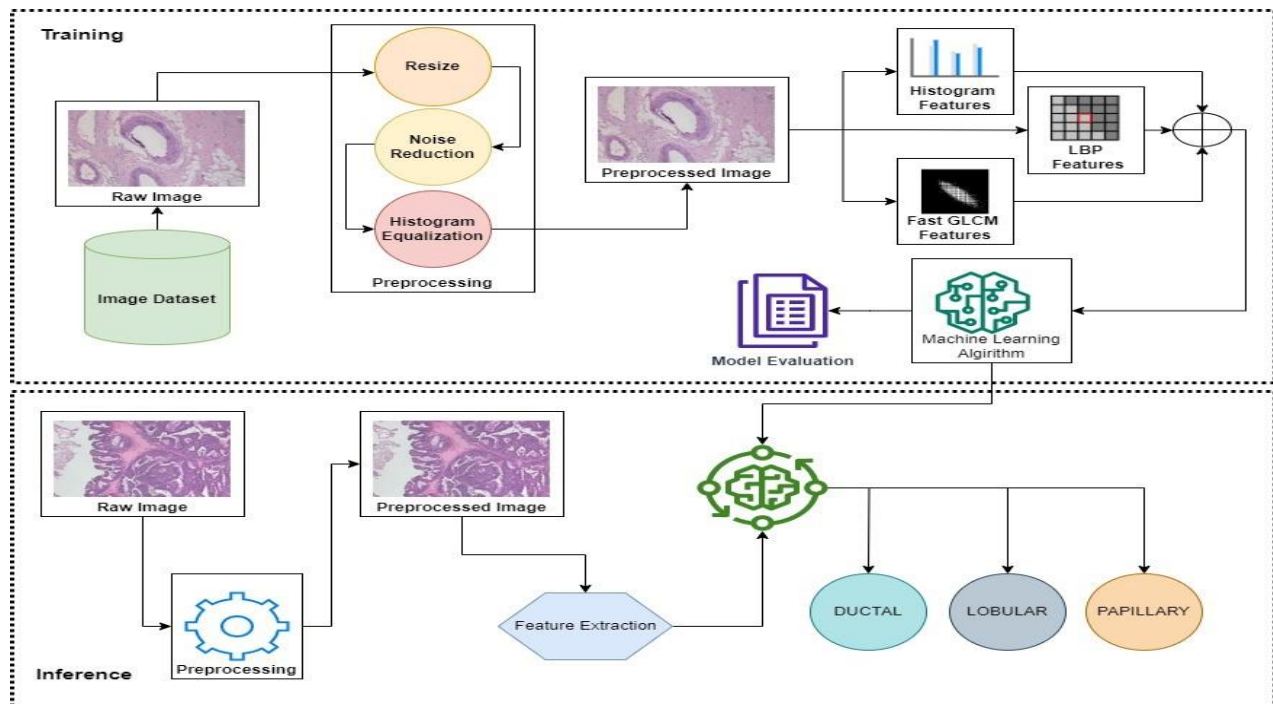


Fig. 1. Architecture.

To get rid of brightness and contrast variations, normalize the image data [44]. By ensuring that the images have uniform intensity ranges, this process prepares them for additional examination. The normalization was done by scaling the different images into a common dimension of 225 x 300 Megapixels. Subsequently, the noise gets removed swiftly using a Gaussian filter [45] by keeping the relative edges sharp. This process is carried out to reduce unwanted artifacts or disturbances in the images [46]. Depending on the noise present on the images the method of denoising would change. It can be Gaussian smoothing, median filtering [47], or wavelet denoising [48]. The final process is histogram

In the present study, we have used images from the Breast Cancer Histopathological (BreakHis) dataset, which includes 7909 images of 82 patients at four different magnifications [42]. The data set was gathered from Brazil's Pathological Anatomy and Cytopathology (P&D) Lab. Images have a dimension of 700 460 pixels, are in the PNG (Portable Network Graphics) format, and are in the 3-channel RGB (Red-Green-Blue) format with an 8-bit depth per channel. Hematoxylin and eosin were used to stain the biopsy samples for breast tissue on slides. The dataset contains information on malignant and benign breast tumors. We have only considered a subset consisting of malignant tumors in the BreakHis dataset. The subset chosen is based on the suggestions of domain experts. It contains the most commonly found cancer patterns from different patients.

### B. Pre-processing

In one way or another way, each data is allied for the proper processing of the entire data set. Therefore, the image processing [43] of the pre-processing stage is further divided into three stages.

equalization [49] to boost the contrast of the image and the visibility of key details. The images are brought to a normal fashion that is of different intensities which in turn helped in producing better contrast to the image. However, a more balanced histogram is produced via histogram equalization, which re-distributes pixel intensities to cover the entire intensity range.

### C. Feature Engineering

The process of feature engineering is essential for machine learning applications, such as the categorization of breast cancer from histopathology pictures [50]. To enhance a

prediction model's performance, important and instructive elements from raw data must be extracted. Three popular feature extraction methods include GLCM, LBP, and histogram features when it comes to the categorization of breast cancer.

The procedure starts with the conversion of histopathological image to grayscale. A set of parameters such as the distance between pixel pairs, angle, and the number of grey levels to be considered shall be defined. Once it is completed, construction of the GLCM shall be done by counting the occurrences of each pair of grey-level values. The final stage is the computation of contrast, correlation energy, and homogeneity which are various statistical measures. Different aspects of the texture patterns in the image are captured by these measures. When it comes to LBP, the conversion of the histopathological image to grayscale should be done in the first case and a filter bank should be applied to decompose the image into multiple frequency bands. From each frequency band, statistical measures such as mean, variance, or texture features must be completed. Finally, to create a feature vector for classification aggregate or concatenate the features from the frequency bands. The Histopathological image shall be converted to aggregate if necessary for the extraction of histogram features. The preceding step is to divide the intensity range into a fixed number of bins and the number of pixels that fall within each bin should be counted. Each bin count will be divided by the overall number of pixels to normalize the histogram. Later, mean, variance or skewness can be calculated if required. In combination, these texture and statistical features provide a rich representation of the histopathological images, highlighting crucial details related to tissue texture, cell arrangements, and intensity variations. Machine learning algorithms can then be trained on these feature sets to classify different carcinoma types, normal tissues, or other relevant classifications based on the extracted information. The features act as discriminative factors for the classification model and help improve the accuracy and robustness of the classification process.

#### D. Ensemble Classifier

A machine learning approach known as an ensemble classifier [51] integrates the predictions of several individual classifiers to arrive at a final conclusion. It is especially helpful for tackling complex issues where a single classifier

would not produce good outcomes. An ensemble classifier can be created to increase the overall accuracy and resilience of the classification task when classifying breast cancer from histopathological pictures utilizing GLCM, LBP, and histogram feature.

Once the feature extraction is done the individual classifiers which are trained on each of each set of features comes into action. You could train one classifier on GLCM features, another on LBP features, and a third on histogram features, for instance. Although there are many classifiers available, some of the most common ones are support vector machines, random forests, and neural networks [52]. The next step is construction of ensemble. Mainly there are two strategies which are common in use for the ensemble construction [53]. Voting is the first one. Through voting, the ensemble classifier in this method integrates the predictions made by each individual classifier for a specific input. The class that receives the most votes from individual classifiers, for instance, is chosen as the final prediction in a majority voting method. The second strategy is weighted averaging. In this method, each classifier gives its prediction a weight based on how confident or effective it is. Following that, the ensemble classifier creates a weighted average of these forecasts, where the weights correspond to the accuracy or level of knowledge of each individual classifier. Once all these are done prediction and decision making are the last steps of classification. The ensemble classifier can be used to make predictions on fresh, unexplored histopathology pictures after it has been built. Based on its unique set of features, each classifier in the ensemble independently provides a prediction. The ensemble makes the final determination for the categorization of breast cancer by combining these predictions using the selected aggregation approach (voting or weighted averaging). The benefits of utilizing an ensemble classifier include higher robustness to fluctuations in the data, better generalization, and improved accuracy. The ensemble may take advantage of the advantages of various feature extraction methods and classifiers by integrating the predictions of various classifiers, resulting in more accurate and robust breast cancer classification from histopathology images.

#### V. RESULT ANALYSIS AND DISCUSSION

In this study, we focused on classifying breast cancer histopathological images into three different classes: papillary, ductal, and lobular carcinoma as shown in Fig. 2.

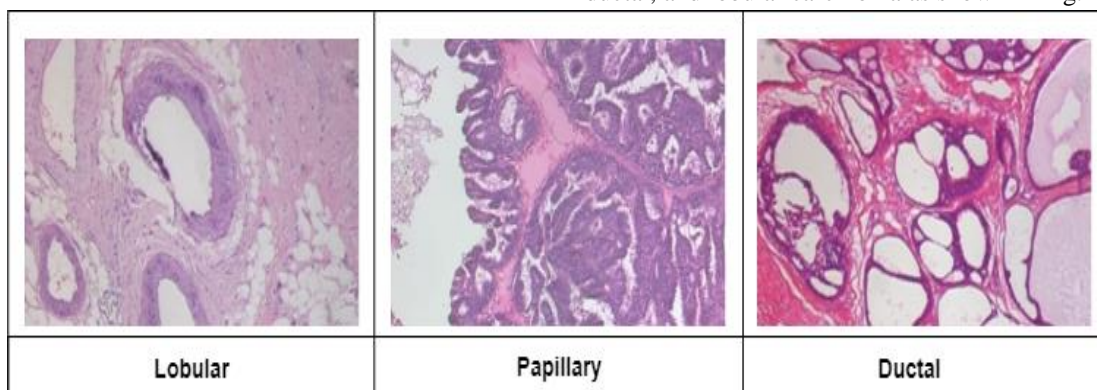


Fig. 2. Carcinoma classes.

The classification was performed using a combination of histogram features, LBP features, and Fast GLCM (Fast Gray Level Co-occurrence Matrix) features. Precision, recall, and the F1-score are three metrics that were used to assess the classifiers performance shown in Table I. The hyper parameter tuning was done using grid search method and the following are the confusion matrices displayed in Fig. 3.

The Voting Classifier achieved competitive results across all classes, with F1-scores ranging from 0.87 for ductal carcinoma to 0.94 for lobular carcinoma. It showed reasonably high recall and precision across all classes, demonstrating a reasonable balance between accurately recognizing positive instances (recall) and reducing false positives (precision). Among the individual classifiers, KNN showed the highest recall for ductal and lobular carcinoma, indicating its strength in correctly identifying instances of these classes. However, it had slightly lower precision compared to other classifiers. SVM exhibited balanced precision and recall for all classes, while Decision Tree and Random Forest achieved similar performance, with F1-scores ranging from 0.82 to 0.87. Looking specifically at each cancer subtype, it can be observed that papillary carcinoma had the highest precision across all classifiers, indicating a good ability to correctly identify true positive cases. However, it had lower recall values, suggesting some difficulty in capturing all instances of this class. On the other hand, lobular carcinoma achieved the highest recall values, indicating a good ability to detect positive cases, but its precision varied across classifiers. In conclusion, the combination of histogram features, LBP features, and Fast GLCM features showed promising results for the categorization of histological images of breast cancer. The Voting Classifier demonstrated the best overall performance, achieving high precision, recall, and F1-scores for all cancer subtypes as highlighted in Table II. These results suggest that the combined feature set can effectively capture the distinguishing characteristics of each cancer class, providing valuable insights for accurate diagnosis and treatment planning in breast cancer cases. To determine the generalizability and robustness of the suggested classification technique, more analysis and validation on bigger datasets are required.

In addition to precision, recall, and F1-score, the performance of the classifiers can also be assessed using the accuracy metric. Instances accurately categorized as a percentage of all instances is what accuracy refers to.

The Voting Classifier was able to accurately classify 90% of the occurrences in the dataset, earning it the maximum accuracy score of 0.90. This classifier outperformed the individual classifiers and demonstrated the best overall

performance. Random Forest also performed well with an accuracy of 0.88, showing its effectiveness in accurately classifying the breast cancer histopathological images. SVM achieved an accuracy of 0.86, indicating a relatively high level of accuracy in its predictions. KNN showed an accuracy of 0.82, while Decision Tree had the lowest accuracy of 0.75 among the classifiers considered. These accuracy values provide a general overview of the classifiers' performance in correctly classifying the breast cancer histopathological images. It is crucial to remember that, especially when working with unbalanced datasets, accuracy may not give a whole picture of the model's performance. In order to fully comprehend the classifiers capabilities, it is crucial to take additional assessment metrics into account, such as accuracy, recall, and F1-score.

TABLE I. EVALUATION METRICS

Class	Voting Classifier			KNN			SVM			Decision Tree			Random Forest		
	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score
Ductal	0.83	<b>0.92</b>	0.87	0.71	0.96	0.82	0.85	0.88	0.87	0.83	0.73	0.78	0.83	0.92	0.87
Lobular	0.89	<b>1.00</b>	0.94	0.84	0.94	0.89	0.85	1.00	0.92	0.75	0.91	0.82	1.00	0.94	0.89
Papillary	0.93	<b>0.92</b>	0.95	0.82	0.81	0.80	0.89	0.86	0.86	0.70	0.62	0.66	0.93	0.74	0.82

TABLE II. TEST ACCURACY

Classifier	Random Forest	KNN	SVM	Decision tree	Voting Classifier
Accuracy	0.88	0.82	0.86	0.75	<b>0.90</b>



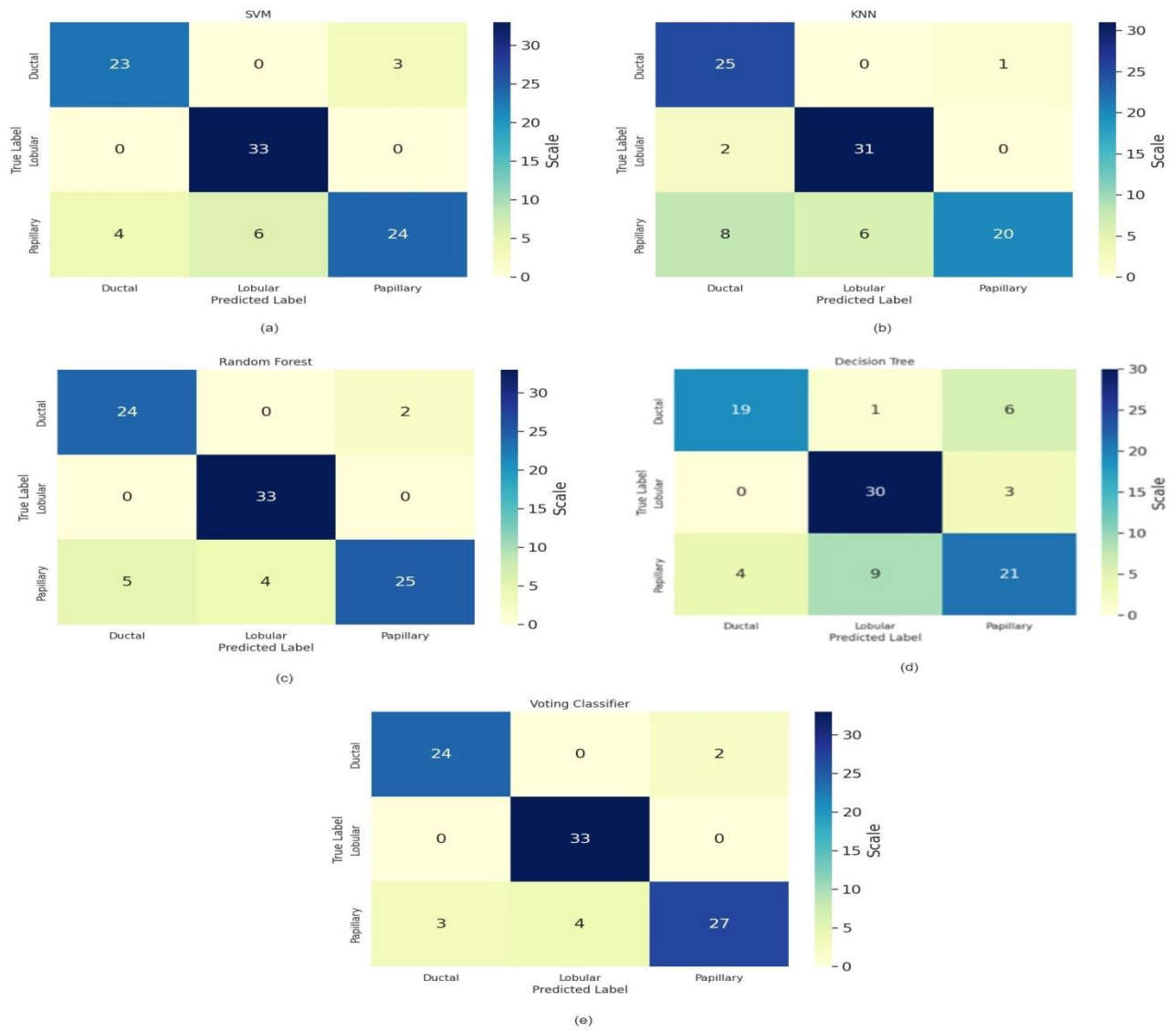


Fig. 3. Confusion matrix of different classifiers.

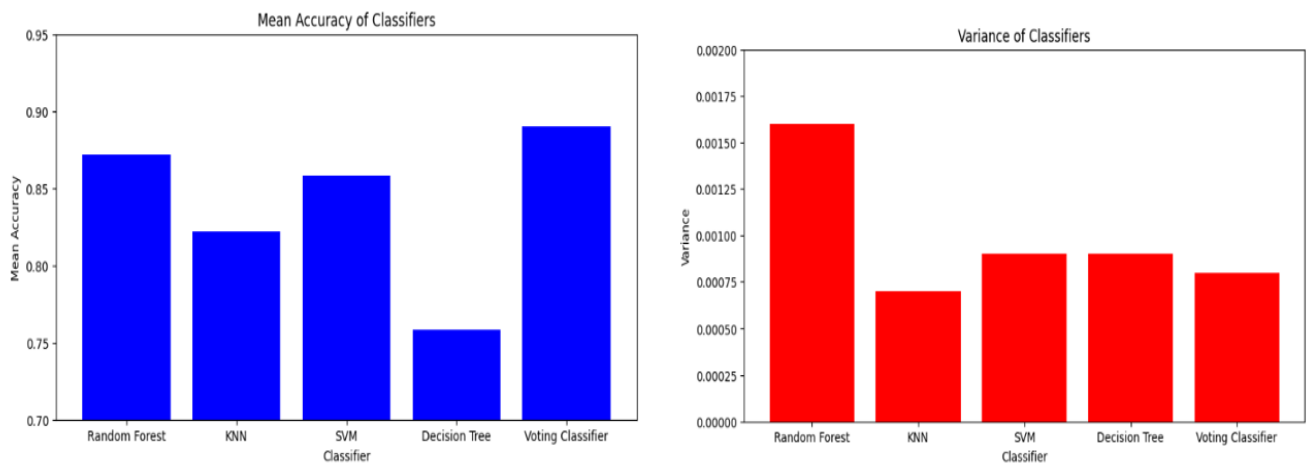


Fig. 4. Mean and variance cross validation accuracy of classifiers.

Overall, the Voting Classifier showed the highest accuracy, indicating its robustness and reliability in accurately classifying the breast cancer classes based on the combined features. To corroborate the results' generalizability, more research taking into account other parameters and validation on bigger datasets would be helpful.

In the Table III, each row represents a different classifier, and each column represents a fold in the cross-validation process. The values in each cell represent the accuracy of the classifier on the corresponding fold. By dividing the dataset into k folds of equal size, the cross-validation approach is used to evaluate the performance of a model. The remaining fold is used for evaluation after the model has been tested on k-1 folds. Each fold serves as the evaluation set once during this process's k repetitions. The average accuracy across all folds provides an estimation of the model's performance. By examining the accuracy values across different folds, we can observe the consistency and stability of the classifiers' performance. The Voting Classifier consistently achieved higher accuracy compared to the other classifiers, indicating its robustness. Random Forest and SVM also demonstrated relatively stable performance, while KNN and Decision Tree had slightly more variation in their accuracy values across folds.

TABLE III. K-FOLD VALIDATION RESULTS

Classifier	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Variance
Random Forest	0.86	0.88	0.90	0.85	0.87	0.872	0.0016
KNN	0.82	0.83	0.80	0.85	0.81	0.822	0.0007
SVM	0.85	0.87	0.84	0.86	0.88	0.858	0.0009
Decision Tree	0.75	0.77	0.73	0.78	0.76	0.758	0.0009
Voting Classifier	0.88	0.90	0.87	0.89	0.91	<b>0.89</b>	<b>0.0008</b>

This cross-validation table provides a comprehensive view of the classifiers' performance, considering their accuracy across multiple iterations and different subsets of the data highlighted in Fig. 4. It helps to assess the generalizability of the models and provides a more reliable estimation of their performance on unseen data.

## VI. CONCLUSION

Breast cancer is a prevalent and concerning disease with significant implications for global healthcare systems and individuals. Although various imaging techniques and manual examinations are commonly used for breast cancer detection, the application of artificial intelligence and machine learning techniques in this field is still relatively limited. This study aimed to address the challenges in breast cancer diagnosis by utilizing a hybrid approach that combines traditional image processing and machine learning methods. The proposed method incorporated feature fusion using GLCM, LBP, and histogram features, along with an ensemble learning approach

for improved classification performance. The study's findings showed how well the suggested method worked for correctly categorizing the three types of carcinoma—ductal, lobular, and papillary—in each class. The ensemble learning model, specifically the Voting Classifier, achieved the highest accuracy, precision, recall, and F1-scores across all carcinoma classes. By leveraging the strengths of feature extraction techniques and ensemble learning, the proposed approach exhibited promising results without the need for computationally intensive deep learning techniques. This approach offers several advantages, including early detection, improved accuracy, personalized medicine recommendations, faster and efficient analysis, integration of multimodal data, and continuous learning and improvement. Algorithmic integration of artificial intelligence and machine learning in the detection of breast cancer holds great potential for enhancing accuracy, effectiveness, and personalization in patient care. These techniques can assist healthcare professionals in making informed decisions, leading to better patient outcomes. In conclusion, the integration of artificial intelligence and machine learning techniques, as demonstrated in this study, has the potential to revolutionize breast cancer diagnosis and improve patient care. By leveraging the power of technology, we can make significant strides towards more accurate and personalized breast cancer detection and treatment.

## VII. FUTURE WORK

Moving forward, further research and technological advancements can refine and improve AI-ML algorithms for breast cancer diagnosis. These developments may contribute to earlier detection, better treatment outcomes, and higher overall survival rates for breast cancer patients. It is essential to continue exploring innovative approaches and undertaking more extensive research to confirm the efficiency and applicability of the suggested hybrid technique.

## REFERENCES

- [1] The global challenge of cancer, Nature Cancer, vol. 1, January 2020, pp. 1-2.
- [2] Arun Upadhyay, "Cancer: An unknown territory; rethinking before going ahead," Genes & Diseases, vol. 8, Iss. 5, September 2021, pp. 655-661.
- [3] He, Z., Chen, Z., Tan, M., Elingarami, S., Liu, Y., Li, T., Deng, Y., He, N., Li, S., Fu, J., and Li, W., "A review on methods for diagnosis of breast cancer cells and tissues," Cell Proliferation, vol. 53(7), July 2020: e12822.
- [4] Zahoor Saliha , Lali Ullah Ikram , Khan Attique Muhammad, Javed Kashif and Mehmood Waqar , "Breast Cancer Detection and Classification using Traditional Computer Vision Techniques: A Comprehensive Review", Current Medical Imaging, vol. 16(10), 2020, pp. 1187 – 1200.
- [5] Champaign JL and Cederbom GJ., "Advances in breast cancer detection with screening mammography," The Ochsner journal, vol. 2(1), January 2000 , pp. 33-5, PMID: 21765659.
- [6] Heang-Ping Chan, Ravi K. Samala and Lubomir M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," The British Journal of Radiology, vol. 93(1108), 2020: 20190580
- [7] C. Kaushal, S. Bhat, D. Koundal and A. Singla, "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images," IRBM, vol. 40, Iss. 4, August 2019, pp. 211-227.



- [8] Dileep G and Gianchandani Gyani S G, "Artificial Intelligence in Breast Cancer Screening and Diagnosis," *Cureus*, vol. 14(10), October 2022: e30318.
- [9] Marshall J, Peshkin B, Yoshino T et al., "The Essentials of Multiomics," *The Oncologist*, vol. 27, Iss. 4, April 2022, pp. 272–284.
- [10] Sebastian AM and Peter D., "Artificial Intelligence in Cancer Research: Trends, Challenges and Future Directions," *Life*, vol. 12(12), November 2022 :1991.
- [11] Hamamoto R, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, Takahashi M, Jinnai S, Shimoyama R, Sakai A, et al., "Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine," *Cancers*, vol. 12(12), November 2020:3532.
- [12] Zi-Hang Chen, Li Lin, Chen-Fei Wu, Chao-Feng Li, Rui-Hua Xu and Ying Sun, "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine," *Cancer Communications*, vol. 41(11), November 2021, pp. 1100-1115.
- [13] Brindha Senthilkumar, Doris Zodinpuui, Lalawmpuii Pachuau, Saia Chenkual, John Zohmingthanga, Nachimuthu Senthil Kumar and Lal Hmingliana, "Ensemble Modelling for Early Breast Cancer Prediction from Diet and Lifestyle," *IFAC-PapersOnLine*, vol. 5(1), 2022, pp. 429-435.
- [14] Athraa H. Farhan and Mohammed Y. Kamil, "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 928, 2nd International Scientific Conference of Al-Ayen University (ISCAU-2020) 15-16 July 2020, Thi-Qar, Iraq, 072098.
- [15] M. B. A. Rasyid, Yunidar, F. Arnia and K. Munadi, "Histogram statistics and GLCM features of breast thermograms for early cancer detection," 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), Chiang Rai, Thailand, 2018, pp. 120-124.
- [16] Amin Alqudah and Ali Mohammad Alqudah, "Sliding Window Based Support Vector Machine System for Classification of Breast Cancer Using Histopathological Microscopic Images," *IETE Journal of Research*, vol. 68:1, 2022, pp. 59-67.
- [17] Gour M, Jain S and Sunil Kumar T, "Residual learning based CNN for breast cancer histopathological image classification," *International Journal of Imaging Systems and Technology*, vol.30(3), September 2020, pp. 621-635.
- [18] Ziba Gandomkar, Patrick C. Brennan and Claudia Mello-Thoms, "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks," *Artificial Intelligence in Medicine*, vol. 88, June 2018, pp. 14-24.
- [19] Beltran-Perez, C., Wei, HL. & Rubio-Solis, A., "Generalized Multiscale RBF Networks and the DCT for Breast Cancer Detection," *Int. J. Autom. Comput.*, vol.17, February 2020, pp. 55–70.
- [20] Abhinav Kumar, Sanjay Kumar Singh, Sonal Saxena, K. Lakshmanan, Arun Kumar Sangaiah, Himanshu Chauhan, Sameer Shrivastava and Raj Kumar Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, January 2020, pp. 405-421.
- [21] Li, X., Shen, X., Zhou, Y., Wang, X., and Li, Q., "Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)," *PLOS ONE*, vol.15(5), May 2020: e0232127.
- [22] Duc My Vo, Ngoc-Quang Nguyen and Sang-Woong Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, May 2019, pp. 123-138.
- [23] Saxena S, Shukla S and Gyanchandani M, "Breast cancer histopathology image classification using kernelized weighted extreme learning machine," *International Journal of Imaging Systems and Technology*, vol. 31, no.1, March 2021, pp. 168-179.
- [24] Alom, M.Z., Yakopcic, C., Nasrin, M.S., Taha T. M. and Asari V. K., "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," *J. Digit. Imaging*, vol. 32, August 2019, pp. 605–617.
- [25] Said Boumaraf, Xiabi Liu, Zhongshu Zheng, Xiaohong Ma and Chokri Ferkous, "A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images," *Biomedical Signal Processing and Control*, vol. 63, January 2021: 102192.
- [26] Burçak, K.C., Baykan, Ö.K. and Uğuz, H., "A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model," *J. Supercomput.*, vol. 77, January 2021, pp. 973–989.
- [27] Xie Juanying, Liu Ran, Luttrell Joseph and Zhang Chaoyang, "Deep Learning Based Analysis of Histopathological Images of Breast Cancer," *Frontiers in Genetics*, vol. 10, February 2019.
- [28] Jiang, Y., Chen, L., Zhang, H., and Xiao, X., "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PLOS ONE*, vol. 14(3), March 2019:e0214587.
- [29] Zhongyi Han, Benzhenq Wei, Yuanjie Zheng, Yilong Yin, Kejian Li and Shuo Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Scientific Reports*, vol. 7, June 2017:4172.
- [30] Rajesh Kumar, Rajeev Srivastava, and Subodh Srivastava, "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features," *Journal of Medical Engineering*, vol. 2015, August 2015: 457906.
- [31] Sheikh TS, Lee Y and Cho M., "Histopathological Classification of Breast Cancer Images Using a Multi-Scale Input and Multi-Feature Network," *Cancers*, vol. 12, no.8, July 2020:2031
- [32] Abdullah-Al Nahid , Mohamad Ali Mehrabi, and Yanan Kong, "Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering," *BioMed Research International*, vol. 2018, March 2018:2362108.
- [33] Zhu, C., Song, F., Wang, Y. et al. Breast cancer histopathology image classification through assembling multiple compact CNNs. *BMC Med Inform Decis Mak*, vol. 19, 198, October 2019.
- [34] L. Gong, J. Yang and X. Zhang, "Semi-Supervised Breast Histological Image Classification by Node-Attention Graph Transfer Network," in *IEEE Access*, vol. 8, August 2020, pp. 158335-158345.
- [35] Kalpana George, Shameer Faziludeen, Praveen Sankaran and Paul Joseph K, "Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion," *Computers in Biology and Medicine*, vol. 124, September 2020:103954.
- [36] D.C.R. Novitasari, A. Lubab, A. Sawiji, A.H. Asyhar "Application of Feature Extraction for Breast Cancer using One Order Statistic, GLCM, GLRLM, and GLDM", *Advances in Science, Technology and Engineering Systems Journal*, vol. 4, no. 4, 2019, pp. 115-120.
- [37] Hao, Y., Zhang, L., Qiao, S., Bai, Y., Cheng, R., Xue, H., Hou, Y., Zhang, W., and Zhang, G., "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix," *PLOS ONE*, 17(5), May 2022 :e0267955.
- [38] S. J. A. Sarosa, F. Utaminigrum and F. 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), Malang, Indonesia, 2018, pp. 54-59.
- [39] R. Touahri, N. AzizI, N. E. Hammami, M. Aldwairi and F. Benaïda, "Automated Breast Tumor Diagnosis Using Local Binary Patterns (LBP) Based on Deep Learning Classification," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-5.
- [40] Ahirrao, Sonal R. and Bormane, D. S., "A novel approach for Face Recognition using Local Binary Pattern," *International Journal of Image Processing and Vision Science*, vol. 1 : Iss. 1 , Article 6, July 2012.
- [41] Ü. Budak and A.B. Güzel, "Automatic Grading System for Diagnosis of Breast Cancer Exploiting Co-occurrence Shearlet Transform and Histogram Features," *IRBM*, vol. 41(2), April 2020, pp. 106-114.
- [42] F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, July 2016, pp. 1455-1462.
- [43] M. Adel, A. Kotb, O. Farag, M. S. Darweesh and H. Mostafa, "Breast Cancer Diagnosis Using Image Processing and Machine Learning for

- Elastography Images," 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCASST), Thessaloniki, Greece, 2019, pp. 1-4.
- [44] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider and R. Deters, "Breast Cancer Diagnosis with Transfer Learning and Global Pooling," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2019, pp. 519-524.
- [45] S. Punitha, A. Amuthan and K. Suresh Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," Future Computing and Informatics Journal, vol. 3(2), December 2018, pp. 348-358.
- [46] D. A. Zebari, H. Haron, D. M. Sulaiman, Y. Yusoff and M. N. Mohd Othman, "CNN-based Deep Transfer Learning Approach for Detecting Breast Cancer in Mammogram Images," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 256-261
- [47] H. -C. Lu, E. -W. Loh and S. -C. Huang, "The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2019, pp. 9-12
- [48] Cengiz, Enes, Kelek, Muhammed Mustafa, Oğuz, Yüksel and Yılmaz, Cemal, "Classification of breast cancer with deep learning from noisy images using wavelet transform," Biomedical Engineering / Biomedizinische Technik, vol. 67, no. 2, March 2022, pp. 143-150.
- [49] Murcia-Gómez D, Rojas-Valenzuela I and Valenzuela O, "Impact of Image Preprocessing Methods and Deep Learning Models for Classifying Histopathological Breast Cancer Images," Applied Sciences, vol. 12, no. 22, November 2022:11375.
- [50] Kode H and Barkana BD, "Deep Learning- and Expert Knowledge-Based Feature Extraction and Performance Evaluation in Breast Histopathology Images," Cancers, vol. 15, no.12, June 2023: 3075.
- [51] Assiri AS, Nazir S and Velastin SA, "Breast Tumor Classification Using an Ensemble Machine Learning Method," Journal of Imaging, vol.6, no.6, May 2020:39.
- [52] Aswathy, M.A. and Jagannath, M, "An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features," Med Biol Eng Comput, vol. 59, July 2021, pp. 1773–1783.
- [53] M. Samridha Majumdar, Payel Pramanik and Ram Sarkar, "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images," Expert Systems with Applications, vol. 213, Part B, March 2023:119022.