International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I12P230

Volume 70 Issue 12, 317-327, December 2022

Original Article

Local Attention-Based Descriptor Definition using Vision Transformer for Breast Cancer Identification

Anish Anurag [1], Aniket Das[2], Jaya H. Dewan[3], Rik Das[4], Govind Kumar Jha[5], Sudeep D. Thepade[6]

[1,5] *Vinoba Bhave University, Hazaribag, Jharkhand, India.*
[2]*Netaji Subhash Engineering College, Kolkata, West Bengal, India.*
[3,6] *Pimpri Chinchwad College of Engineering, Pune, Maharashtra India.*
[4]*Xavier Institute of Social Service, Ranchi, Jharkhand, India.*

[4]*Corresponding Author : rikdas78@gmail.com*

Received: 15 October 2022 Revised: 12 December 2022 Accepted: 20 December 2022 Published: 24 December 2022

Abstract - *Assorted morphological characteristics of breast cancer have categorized it as a heterogeneous life-threatening ailment for females. Researchers have proposed automated approaches for detecting malignant breast cancer from patient images for faster diagnosis with higher precision. Descriptor definition with local attention by dividing a given image into patches can result in a robust representation of the breast cancer histopathological images for identification of malignancy. This work has experimented with feature extraction with local attention-based vision transformers and has evaluated the classification results under multiple classification environments. The research paper has also compared the resilience of classical descriptors generated using handcrafted methods and automated pre-trained convolutional neural network (CNN)-based feature extraction. Three separate handcrafted feature extraction methods, namely Color Histogram, Local Binary Pattern (LBP), and Oriented FAST and Rotated BRIEF(ORB), are used in the process, along with pre-trained CNN-based feature extraction methods (InceptionNet-v1, EfficientNet-B7, and ResNet-50). The experimentation is performed using the BreakHis dataset, and the results have revealed the superior performance of vision transformer-based features as compared to all other individual features considered. Furthermore, early fusion-based descriptors with different combinations of handcrafted and deep-learning features are created to investigate any improvement in the generalization of descriptors. The results have indicated that the feature extracted with local attention-based vision transformer overfits with early fusion has the best performance when evaluated individually in three different classification environments.*

Keywords - *Breast cancer, Local attention, Vision transformer, Color histogram, Local Binary Pattern, Convolutional Neural Networks.*

1. Introduction

Histopathological images are the yardstick for automated breast cancer detection using computer-aided diagnosis. Breast cancer causes maximum mortality and morbidity among women worldwide, as reported by World Health Organization's (WHO) World Cancer Report. According to the survey on recent fatal encounters with breast cancer, morbidity is known to be the second leading cause of death, with a stated account of 14.7% [1]. The likelihood of surviving breast cancer can improve by 80% in the case of early detection [2]. The histopathological investigation has always outclassed all the recent advancements in molecular biology in diagnosing breast cancer recurrence [3].

Most of the abnormalities in the epithelial cells generating the benign lesion cannot give rise to breast cancer. However, cells exhibiting inappropriate division and irregular development are referred to as malignant or cancerous cells. Manual analyzing microscopic images is a highly complicated task based on the nature of their appearances [4][5]. However, developing an automated method for determining the disease's malignancy might prove helpful in ensuring early medical attention, which could be a gesture that saves lives [6] [7] [8]. In[9], various algorithms for segmenting nuclei to distinguish malignant and benign cases are tested on a dataset of 500 pictures. Cytological images used in [10] for the recognition of breast cancer are derived from fine-needle biopsies. Four different classifiers were trained as a result of the research. In [11], cytological scans are used for nuclei segmentation to predict breast cancer using diverse machine learning (ML) algorithms. The Cascade technique combined with the rejection option has reported a classification accuracy of 97% for breast cancer [12]. The method enables the resolution of straightforward situations at the first level while deferring the resolution of complex cases to the next level, capable of higher complexity pattern recognition. Due to the lack of datasets available to researchers, Spanhol et al. [13] proposed a dataset with 7909 breast histopathology pictures acquired from 82 patients. It was evaluated by obtaining textual features from six classes, categorizing the features with various classifiers, and reporting an accuracy of up to 85%.

Hence, the preceding discussion clearly establishes the significance of breast cancer research in the current context. This has led to exploring the representation learning technique of descriptor definition using deep neural networks [14]. Traditional CNNs require substantial computational resources, which is a bottleneck for real-time implementation [15]. The inductive bias is stronger in CNNs, which can exhibit lesser confidence in model regularization or data augmentation [16]. Hence for smaller datasets, CNNs are prone to overfitting. The authors identified gaps in the research described above, which can be addressed using self-attention-based vision transformers (ViT). They have relatively lesser inductive bias and have exhibited incredible performance on smaller datasets by dividing an image into patches and initiating positional embedding. Impressive content-based image classification (CBIC) results are unveiled by the self-attention-based vision transformers (ViT), which divide the image data into patches to extract rich feature vectors. The entire process of defining feature vectors with ViT is not only different from the existing pre-trained convolutional neural networks (CNNs) but also is expected to have improved feature quality with respect to the handcrafted feature extraction techniques. Hence, the objectives of this paper are:

- To investigate the robustness of feature vector extracted using vision transformer (ViT)
- To investigate any improvement in classification accuracy by fusing handcrafted features and deep representation learning features to feature vectors extracted using ViT.
- To investigate augmentation in feature generalization with early feature fusion.
- To investigate the performance of early feature fusion using representation learning-based features.

The above objectives are achieved by extracting feature vectors with three different handcrafted techniques followed by descriptor definition using pre-trained CNNs and ViT. A popular public dataset named “BreakHis" is a testbed to compute the classification results in this experimentation process [13]. All the generated features are evaluated using three different classifiers, namely, K Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Machine (SVM). The results have revealed the dominance of the local attention-based vision transformer (ViT) feature for higher classification accuracies in all three classification environments when the experimentation is carried out using a single descriptor definition. This study has investigated the possibility of enhanced feature generalization using the fusion of handcrafted features to vision transformer-based features for the first time.

Moreover, this paper has also proposed using a vision transformer for a representation learning-based approach. However, the scenario is observed to be different in the case of fusion-based approaches, and the classification results have highlighted the combination of EfficientNet-B7 with handcrafted features ensuing in higher classification results by denying the obvious superiority of participation of vision transformer features as a member of the fusion process. This has also indicated better feature generalization in the early fusion technique using conventional pre-trained CNNs compared to the vision transformer-based local attention approach.

2. Literature Review

A substantial volume of research work is carried out for the automated identification of breast cancer using computer-aided diagnosis [17]. The research is essential in the domain since breast cancer is identified as being one of the primary issues resulting in the fatality of the female population [18]. A transfer learning-based approach using CNN for early breast cancer identification is observed to be beneficial in controlling premature deaths [19]. The initiation of computer-based nuclear morphometry for cancer detection has been discussed in the literature for the last 40 years [20]. The widespread adoption of whole sliding images (WSI) and other digital pathology forms has produced several roadblocks, including the high execution cost and technology implementation, the technique's insufficient output when dealing with extensive clinical routines, implicit technical issues and pathologists' cultural resistance [21]. Publication of the "BreakHis" dataset and its unrestricted access to the entire world has enabled public access to the histopathological image dataset to the experts for breast cancer diagnosis [13]. Recent research has revealed promising classification accuracies using this dataset as a testbed for extracting both handcrafted and deep-learning features [22]. Classification of breast tumors from image data in [23] is carried out by using the sorted gray values’ mean derived from the test images as features. Extraction of morphological textured-based feature vectors is presented in [24] for the binary categorization of cancer images using SVM. A prominent application of color features is observed in [25] for feature vector representation applied in the classification of breast cancer. Global and local thresholding-based binarization for feature extraction is carried out in [26] for breast cancer detection from mammography images. The efficacy of binarization is also reported in [27] and [28] for detecting malignancy in breast cancer. Texture features are learned from ultrasound images for breast cancer identification [29]. Curvelet transform is used in [30] to extract potential features for the recognition of breast cancer.
Deep learning (DL) techniques have recently gained

popularity for successfully addressing image categorization issues. [31] [32]. With patch-based DL algorithms for breast tumor categorization mentioned in [25], a superior presentation of input data for tumor detection is witnessed. A class structure deep CNN was recently proposed in [56] and has demonstrated an average image-level 96.7% accuracy for the BreakHis dataset. A deep multichannel preserving autoencoder was used in another method [34], which reported an aggregate 99.36% accuracy. Additionally, in the different arguments [35], a CNN model is evaluated using cardiology, radiology, and gastroenterology as three distinct medical imaging applications. In [36], a two-level model is suggested, using the expectation maximization (EM) method in conjunction with a patch-level CNN and SVM or the multiclass regression technique. Neural codes, also known as Decaf features, are considered to act in some way as a trade-off between conventional handmade methods and CNN procedures that are task-specific [37] [38]. This hypothesis has produced robust descriptor creation by using a CNN that has already been trained as a feature extractor. Additionally, distinct classifiers developed with competitive classification accuracy are given the retrieved features as inputs [39] [40].

The efficiency of vision transformers using ultrasound images for breast cancer categorization is reported in [41]. Average classification accuracy of 93.02% is achieved for breast cancer categorization using the DeconvTransformer model [42]. Classification of histopathological images for revealing breast cancer is carried out in [43] using Transformer based self-supervised learning. Improvement in breast cancer identification is observed in unregistered Multiview mammograms with the use of transformers [44].

Hence, the application of diverse methods for breast cancer using content-based image classification is evident from the discussion, as mentioned earlier in the existing literature. But, primarily, the techniques are implemented separately and are not tested using an early ensemble approach. It is observed that none of the methodologies has attempted to examine the improvement of the feature generalization capacity of vision transformers by fusing the extracted features with the handcrafted ones. Moreover, it is also observed that the fusion of self-attention-based extracted descriptors from vision transformers is also not fused with pre-trained CNN-based representation learning features to understand the performance enhancement for classification and breast cancer analysis.

The authors have identified the gap and have designed diverse fusion-based descriptor definitions with assorted feature combinations to evaluate the classification performances in terms of achieved feature generalization. The results are encouraging for ensuring augmented precision over individual approaches and have identified the suitable combinations of features for better classification results.

3. Proposed Techniques

Descriptor definition has a fundamental contribution to ensuring high precision for CBIC. The authors in this work have intended to investigate the feature generalization capability of varied descriptor definition techniques for designing an automated computer-aided breast cancer diagnosis using histopathological image data.

Self-attention between image patches applied by vision transformer (ViT) based feature extraction has enabled it to learn global information superior to conventional CNNs having limited local receptive field. Therefore, representation learning-based feature generation primarily uses a ViT [55]. Furthermore, pre-trained CNN-based feature extraction is carried out, followed by the definition of handcrafted feature vectors to design the fusion-based approach and to evaluate the comparative generalization. A brief description of each of the techniques is mentioned in the subsections below.

3.1. Vision Transformer (ViT)

Two-dimensional image data is handled with a vision transformer by reshaping a given image $i \in \mathbb{R}^{H*W*C}$ to a series of flattened two-dimensional patches $i_p \in \mathbb{R}^{N*(P^2.C)}$. Here, the original image resolution is (H, W), and C represents the number of channels. (P, P) denotes each patch's resolution, and N denotes the total number of patches. Here, $N = HW/P^2$.

The Transformer uses a constant latent vector of dimension D across all the layers with which the flattened patches are mapped with a trainable linear projection as in equation 1.

$$z_0 = \left[i_{class};\ i_p^1 E;\ i_p^2 E; \dots;\ i_p^N E\right] + E_{pos},\quad E \in \mathbb{R}^{(P^2.C)*D}, E_{pos} \in \mathbb{R}^{(N+1)*D} \tag{1}$$

where $z_0^0 = i_{class}$ are a sequence of embedded patches

A learnable embedding is prepended to the series of embedded patches from which the image representation r is derived as the state at the output of the transformer device, as in equation 2. Positional information is retained in patch embedding by adding positional embeddings.

$$r = LN(z_L^0) \tag{2}$$

where z_L^0 is the Transformer encoder.

The transformer blocks present in the ViT model are responsible for creating a tensor to be processed by a classifier head with *softmax* to conclude the final class probabilities. The authors in this approach have extracted the feature vectors from the histopathological images by flattening the outputs of the final transformer block and using them as image representations to the classifier to evaluate classification accuracy. A block diagram of ViT is provided in Figure 1.

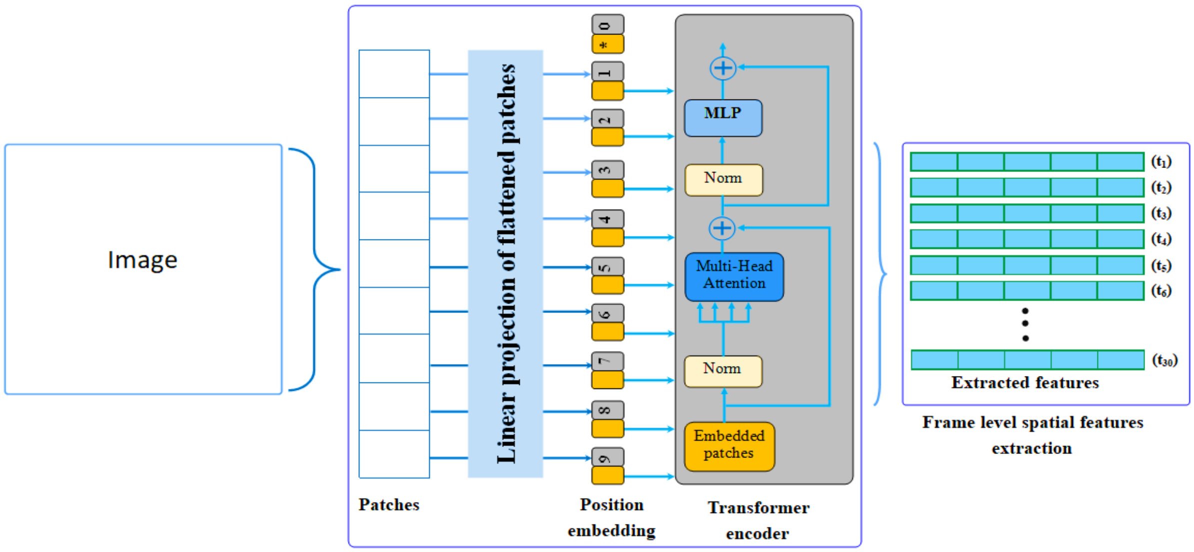


Fig. 1 Block Diagram of ViT-based Feature Extraction

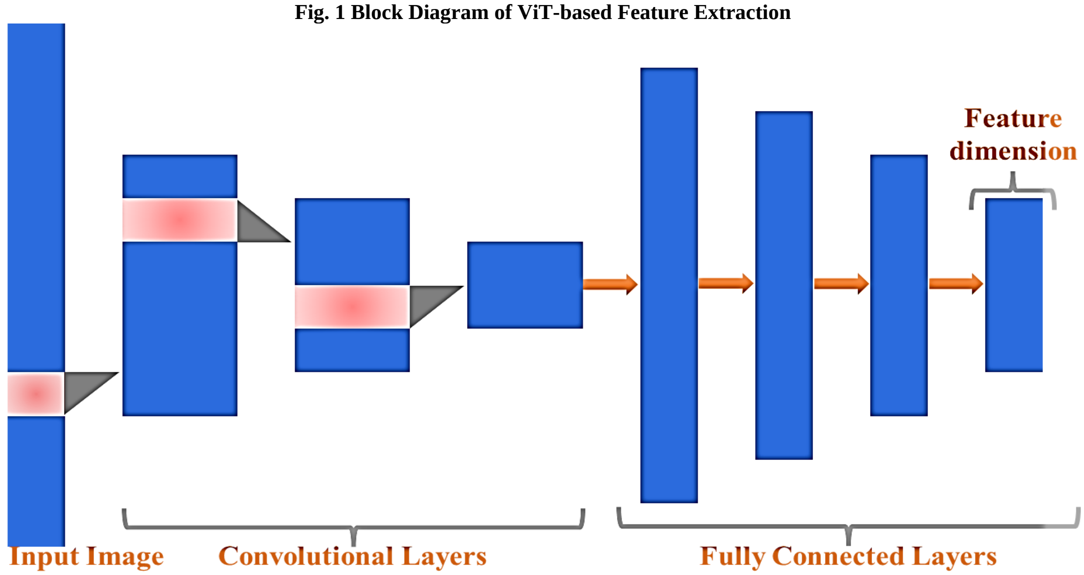


Fig. 2 Block Diagram of Pre-trained CNN-based feature extraction

3.2. Convolutional Neural Network-based Representation Learning

Feature extraction using DL techniques [46] is performed using pre-trained CNNs, namely, InceptionNet-v1 and EfficientNet-B7+ ResNet-50. The CNNs are trained using the imagenet dataset and are capable of categorizing 1000 unique classes. The last layer of the CNNs is considered a feature vector and is extracted from the architecture to provide input to the classifiers, as shown in Figure 2.

3.3. Color Histogram (CH)

One of the MPEG-7 feature sets is the CH, which is regarded as an effective image descriptor. These features can be adjusted for scale-invariant properties after image size normalization. This technique's invariance to image rotation and translation are also essential features[22].

A color histogram is created using the R, G, and B planes of an image. The image's histogram represents the probability dispersal of the color intensities. The global histogram is computed by taking the entire image into account. Color histograms are simple to create and are less affected by slight shifts in viewpoint. Color histograms can't give spatial information and are affected by lighting changes.

3.4. Local Binary Pattern (LBP)

The "local binary pattern (LBP) is a non-parametric operator that describes the local spatial structure of an image and achieves excellent texture classification results. The center pixel is matched to its surrounding pixels in a radius of 'P' pixels. If the surrounding pixels' intensity is higher than that of the center pixel, code '0' is assigned; otherwise, code '1'[47]. Concatenating these surrounding code bits yields the binary codeword for the center pixel. The decimal number is then generated from the codeword. The feature vector is formed by computing the histogram of the decimal codewords. The histogram has a total of 2d bins, where 'd' denotes the number of bits in the binary codeword. Equations 3 and 4 can thus be used to define the LBP descriptor for the center pixel(x_c,y_c).

$$LBP(x_c\, y_c) = \sum_{p=1}^{P} t(i_c - i_p)2^p \tag{3}$$

$$t(x) = \begin{cases} 1, & i_p \geq i_c \\ 0, & i_p < i_c \end{cases} \tag{4}$$

Where 'i_c' denotes the center pixel intensity, and 'i_p' denotes the surrounding pixel intensity. Also, t=0 if $i_c <= i_p$ else t=1.

3.5. Oriented FAST and Rotated BRIEF (ORB)

Oriented Features from Accelerated Segment Test (FAST) and Rotated Binary Robust Independent Elementary Features (BRIEF) combine to generate ORB. FAST and its variants, such as oFAST, are techniques for locating interesting areas in images[57]. The fact that these interest points are generally found near corners of an image suggests that the pixels with the highest gradients relative to their surrounding neighbors are typically the interest points.

The Oriented FAST (oFAST) was developed to overcome the issue of orientation for the corners. Centroid intensity is a straightforward and useful orientation measurement used by oFAST. The corner orientations can be determined using the standard moments as given in equation 5.

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{5}$$

Where I(x,y) is the point's intensity value (x,y), the standard moments are used to compute the centroid, as demonstrated in equation 6:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{6}$$

After locating the centroid, a vector $\overrightarrow{OC}$ is produced, where O denotes the corner's center and C denotes the centroid. Calculating the angle, as defined in equation 7, where *atan2* is a special case of the arc tangent function used with standard moments – gives the orientation.

$$\theta = atan2(m_{01}, m_{10}) \tag{7}$$

The interest points detected using oFAST are described using binary vectors using the rBRIEF algorithm.

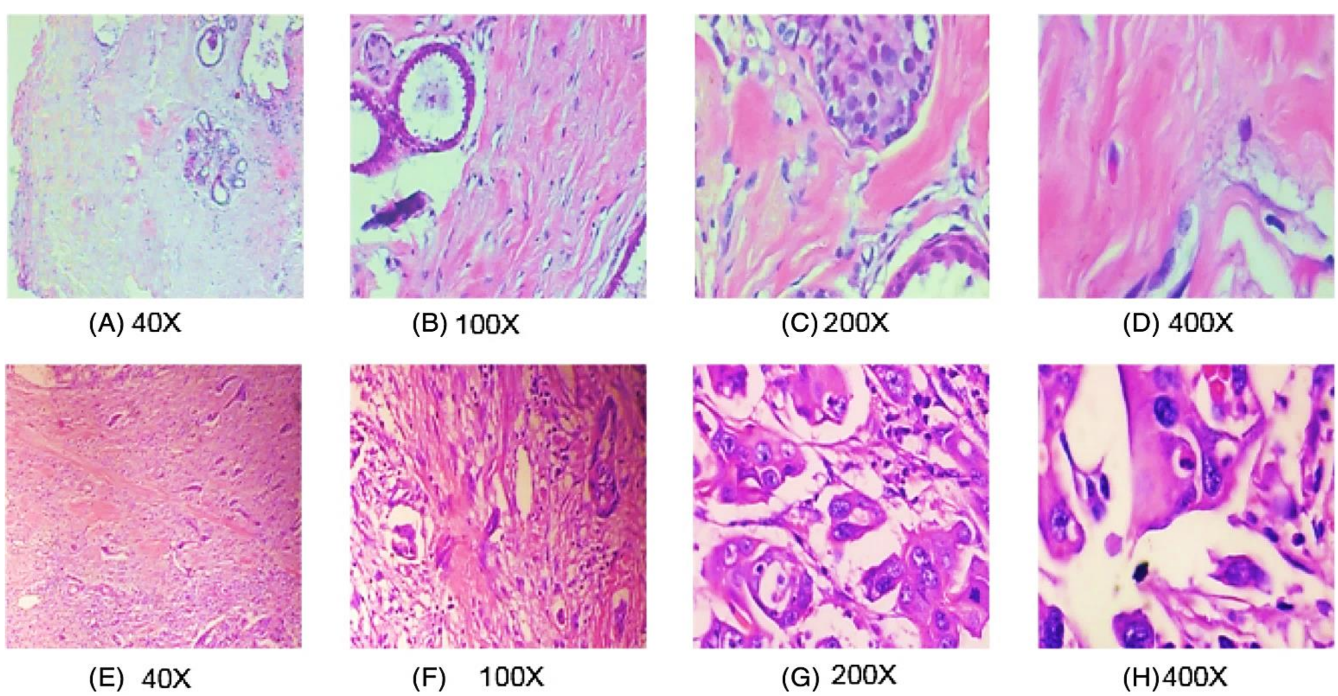


Fig. 3 Sample of BreakHis Dataset with Different Magnification Factors

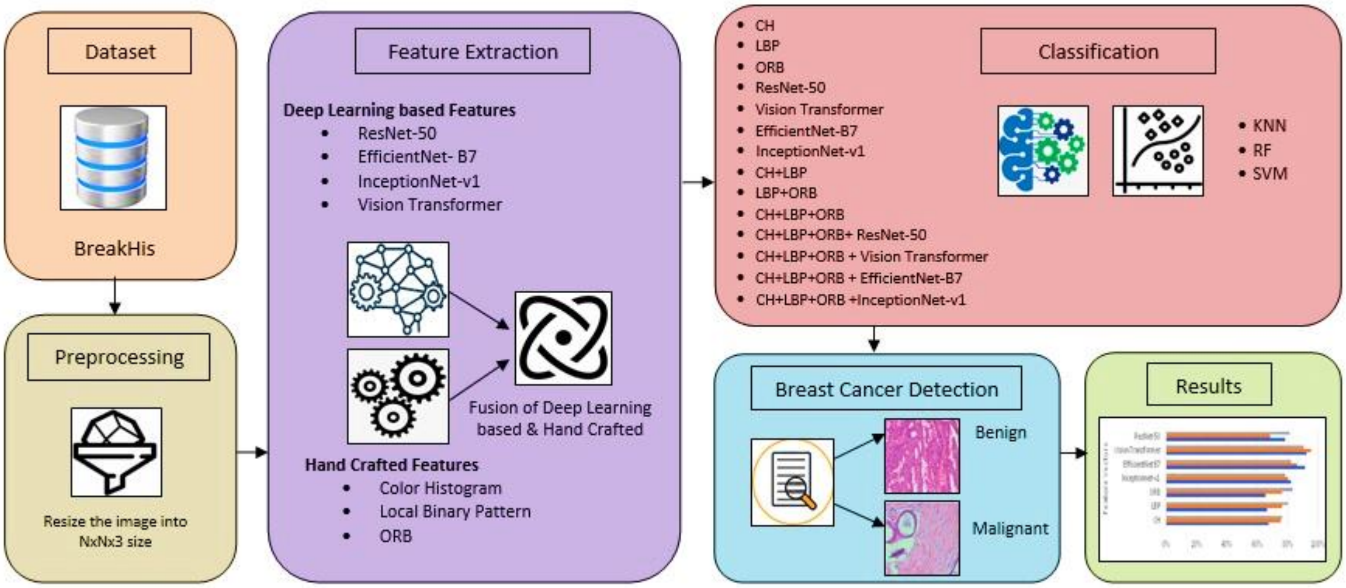


Fig. 4 Breast Image Classification using Proposed Technique

3.6. Experimental Setup

The experimental setup is based on a system comprising 2 CPU Cores assisted with 12 GB NvidiaK80 GPU and 12 GB RAM. The testbed is prepared with a widespread public dataset named BreakHis dataset having 7909 images categorized into eight different breast cancer sub-classes [13] [29]. These images are contributed by 82 anonymous patients of the Pathological Anatomy and Cytopathology (P&D) Lab Brazil. The images have four different magnification factors, namely, 40X, 100X, 200X, and 400X, respectively, and are divided into malignant and benign categories. The images are three-channeled in nature with a dimension of 700×460 and eight-bit per channel depth. The results are derived using five-fold cross-validation, and the number of training images is approximately 6327. An illustration of the dataset with sample images is shown in Figure 3.

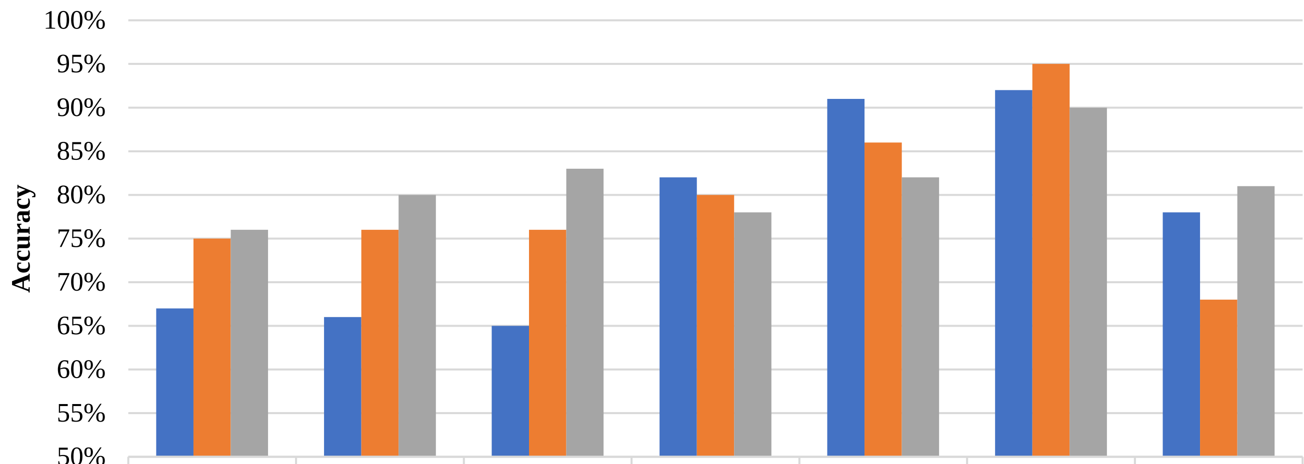
Figure 4 shows the block diagram of the proposed technique. The BreakHis dataset images are pre-processed. Handcrafted features, deep learning representations, and vision transformer-based features are extracted from the images to generate the feature vector of an image. All the feature vectors are normalized for creating an early fusion-based classification framework using diverse combinations. The process is envisioned to be useful for enhanced generalization of descriptors. The fused features and individual features are evaluated using three different classifiers, namely, support vector machine (SVM), random forest (RF), and K Nearest Neighbours (KNN). The evaluation metric is identified as the image level accuracy, which is the ratio of the number of correctly classified images (I_c) to the number of total images (I_t), as in equation 8.

$$Accuracy_{Image\_level} = \frac{I_c}{I_t} \quad (8)$$

4. Discussion and Analysis

Individual descriptor definition primarily uses handcrafted techniques and deep learning-based feature extraction. The feature vectors extracted using CH [49], LBP[50], and ORB[51] are evaluated for image-level accuracies using SVM, RF, and KNN classifiers. The accuracy recorded for the SVM classifier using CH features is 67% which is higher than the accuracies of LBP and ORB features, having subsequent accuracies of 66% and 65%. Although, in the case of RF and KNN classifiers, feature descriptors for the ORB technique have the highest accuracy compared to CH and LBP, respectively. However, the classification accuracy for features extracted using a vision transformer has exhibited the highest value of 95% using an RF classifier compared to any of the other handcrafted techniques and representation learning techniques of feature extraction. A comparative illustration of the classification accuracies with individual features is shown in Figure 5.

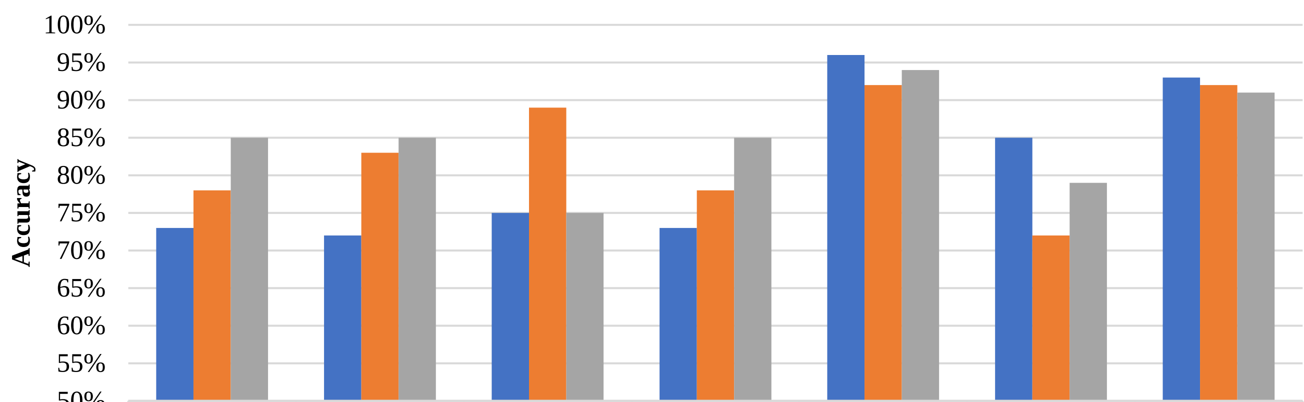
The illustration in Figure 5 has provided the image level accuracies achieved by different individual feature extraction techniques. Local attention-based features extracted using Vision Transformer have the highest accuracies of 92%, 95%, and 90% with SVM, RF, and KNN, respectively, among all the seven varieties of feature vectors extracted. EfficientNet-B7 [52] has recorded almost equivalent performance of 91% image-level classification accuracy to that of vision transformer (ViT) and much higher than InceptionNet-v1[53] (82% with SVM) and ResNet-50 [54] (70% with SVM) for classification using SVM.



	CH[49]	LBP[50]	ORB[51]	InceptionNet-v1[53]	EfficientNet-B7[52]	Vision Transformer	ResNet-50[54]
SVM	67%	66%	65%	82%	91%	92%	78%
RF	75%	76%	76%	80%	86%	95%	68%
KNN	76%	80%	83%	78%	82%	90%	81%

Feature Vectors

Fig. 5 Comparative Image Level Accuracies in Different Classification Environments for Individual Features



	CH + LBP	LBP + ORB	CH + LBP + ORB	CH + LBP + InceptionNet-v1	CH + LBP + ORB + EfficientNet-B7	CH + LBP + ORB + ResNet-50	CH + LBP + ORB + ViT
SVM	73%	72%	75%	73%	96%	85%	93%
RF	78%	83%	89%	78%	92%	72%	92%
KNN	85%	85%	75%	85%	94%	79%	91%

Feature Vector

Fig. 6 Comparative Image Level Accuracies in Different Classification Environments for Fused Features

Although, the remaining two classification results for EfficientNet-B7 are 86% and 82% for RF and KNN, respectively, which have outclassed the classification performances of the other two representation learning features extracted using InceptionNet-v1 (80% with RF and 82% with KNN) and ResNet-50 (68% using RF and 81% using KNN). Nevertheless, it is observed that features extracted using local attention with ViT have revealed the maximum image-level classification accuracies in all three classification environments (SVM, RF, and KNN).

Furthermore, different varieties of fused features are created using the following combinations:

- Color Histogram with LBP
- LBP with ORB
- CH with LBP and ORB
- CH, LBP, ORB and InceptionNet-v1
- CH, LBP, ORB and EfficientNet-B7
- CH, LBP, ORB and ResNet-50
- CH + LBP + ORB + Vision Transformer (ViT)

The comparative results for image-level classification accuracies for fused features are presented in Figure. 5.

The visualization in Figure 6 clearly reveals that the fusion of handcrafted features in different combinations, namely, CH with LBP, LBP with ORB, and CH, LBP with ORB, has improved the classification results in all the cases except for the fusion of three features evaluated using the KNN classifier. The accuracy is reduced to 75%, less than each handcrafted feature. Although, an increase in accuracies in most cases has indicated better feature generalization by introducing early feature fusion for handcrafted techniques.

Henceforth, feature fusion is carried out between representation learning features extracted using pre-trained CNNs and vision transformers to that of the handcrafted features. The results in Figure 6 demonstrate an improvement in the individual accuracies of handcrafted features with this fusion in the majority of the cases.

However, a drop in accuracy is observed in deep learning features due to fusion in some issues, namely, the classification of CH+LBP+ORB+ InceptionNet-v1 using an SVM classifier and for CH+LBP+ORB+ ViT using an RF classifier. Although, the performance of handcrafted techniques has improved even in these two cases.

An SVM classifier achieves maximum image level accuracy of 96% with feature fusion of CH + LBP + EfficientNet-B7. The classification results achieved by the fusion of handcrafted features to local attention-based ViT features are inferior to that of the EfficientNet-B7 for SVM and KNN and are equal for RF. However, the other two fused feature combinations, namely, CH + LBP + ORB + InceptionNet-v1 and CH + LBP + ORB + ResNet-50, are much lesser for all three classification environments.

Thus, the experimentation process has disclosed the following research insights:

- Individual descriptors defined using the local attention-based mechanism of vision transformer (ViT) possess superior feature generalization capacity compared to pre-trained CNNs and handcrafted features

- The classification performance of ViT features degrades with fusion compared to individual outcomes

- A careful combination needs to be done for the fusion-based approach using ViT features to avoid inferior classification outcomes due to overfitting.

- Fusion of handcrafted features with representation learning features in general extracted using pre-trained CNNs leads to higher accuracy due to robust descriptor definition.

Therefore, the research work has identified that features extracted using local attention-based vision transformers are prone to overfitting in the fusion-based scenario. The vision transformers can extract significantly rich feature vectors compared to pre-trained CNNs or handcrafted techniques when used as individual feature extractors. However, for the fusion-based scenario, a combination of handcrafted designs and features extracted using pre-trained CNNs results in superior generalization compared to the ViT model for enhanced classification accuracy.

5. Conclusion

The paper has conducted an extensive examination to identify robust feature vector extraction techniques for breast cancer classification using histopathological images.

The main contribution of this work is to examine the robustness of feature vectors extracted using a local attention-based vision transformer compared to traditional descriptors extracted using handcrafted techniques and deep learning-based automated techniques. In the process, three different handcrafted techniques, namely, CH, LBP, and ORB, are implemented along with pre-trained CNN-based (InceptionNet-v1, EfficientNet-B7, and ResNet-50) feature extraction techniques.

The research outcomes have clearly revealed the higher classification accuracies of local attention-based vision transformer features compared to the rest of the techniques for evaluating classification accuracies with single feature vector extraction. The process is followed by feature fusion of the different feature varieties to assess any enhancement in feature generalization capabilities to stimulate classification accuracies.

The results are affirmative and have exhibited enhanced image level accuracies compared to the individual descriptors. Hence, the feature fusion technique is observed to be efficient in improving the robustness of local attention-based features of vision transformers to ensure a high level of precision in designing computer-aided breast cancer diagnosis using histopathological images.

References

[1] Juan Carlos Martín-Sánchez et al., "Projections in Breast and Lung Cancer Mortality among Women: A Bayesian Analysis of 52 Countries Worldwide," *Cancer Research*, vol. 78, no. 15, pp. 4436–4442, 2018. *Crossref*, http://doi.org/10.1158/0008-5472.CAN-18-0187

[2] Ophira Ginsburg et al., "Breast Cancer Early Detection: A Phased Approach to Implementation," *Cancer*, vol. 126, no. S10, pp. 2379–2393, 2020. *Crossref*, http://doi.org/10.1002/cncr.32887

[3] Shweta Saxena, and Manasi Gyanchandani, "Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review," *Journal of Medical Imaging Radiation Sciences*, vol. 51, no. 1, pp. 182–193, 2020. *Crossref*, http://doi.org/10.1016/j.jmir.2019.11.001

[4] Mark C.Lloyd et al., "Using Image Analysis as a Tool for Assessment of Prognostic and Predictive Biomarkers for Breast Cancer: How Reliable is it?," *Journal of Pathology Informatics*, vol. 1, no. 1, p. 29, 2010. *Crossref*, http://doi.org/10.4103/2153-3539.74186

[5] Christian A. Falk Dahl StudPsychol et al., "A Study of Body Image in Long-Term Breast Cancer Survivors," *Cancer*, vol. 116, no. 15, pp. 3549–3557, 2010. *Crossref*, http://doi.org/10.1002/cncr.25251

[6] Icro Meattini et al., "Radiation Therapy for Young Women with Early Breast Cancer: Current State of the Art," *Critical Reviews in Oncology/Hematology*, vol. 137, pp. 143–153, 2019. *Crossref*, http://doi.org/10.1016/j.critrevonc.2019.02.014

[7] Farah J Nassar, Rihab Nasr, and Rabih Talhouk, "MicroRNAs as Biomarkers for Early Breast Cancer Diagnosis, Prognosis and Therapy Prediction," *Pharmacology & Therapeutics*, vol. 172, pp. 34–49, 2017. *Crossref*, http://doi.org/10.1016/j.pharmthera.2016.11.012

[8] Liaqat Ali et al., "A Novel Sample and Feature Dependent Ensemble Approach for Parkinson's Disease Detection," *Neural Computing and Applications*, 2022. *Crossref*, http://doi.org/10.1007/s00521-022-07046-2

[9] Marek Kowal et al., "Computer-Aided Diagnosis of Breast Cancer Based on Fine Needle Biopsy Microscopic Images," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1563–1572, 2013. *Crossref*, http://doi.org/10.1016/j.compbiomed.2013.08.003

[10] Paweł Filipczuk et al., "Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies," *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2169–2178, 2013. *Crossref*, http://doi.org/10.1109/TMI.2013.2275151

[11] Yasmeen Mourice George et al., "Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images," *IEEE Systems Journal*, vol. 8, no. 3, pp. 949–964, 2014. *Crossref*, http://doi.org/10.1109/JSYST.2013.2279415

[12] Bailing Zhang, "Breast Cancer Diagnosis from Biopsy Images by Serial Fusion of Random Subspace Ensembles," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, pp. 180–186, 2011. *Crossref*, http://doi.org/10.1109/BMEI.2011.6098229

[13] Fabio A. Spanhol et al., "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016. *Crossref*, http://doi.org/10.1109/TBME.2015.2496264

[14] Yann Le Cun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. *Crossref*, http://doi.org/10.1038/nature14539

[15] B. Suryakanth, and S A. Hari Prasad, "3D CNN-Residual Neural Network Based Multimodal Medical Image Classification," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 371–380, 2022. *Crossref*, http://doi.org/10.14445/22315381/IJETT-V70I10P236

[16] D. M. Bongulwar, V. P. Singh, and T. S N, "Evaluation of CNN based on Hyperparameters to Detect the Quality of Apples," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 232–246, 2022. *Crossref*, http://doi.org/10.14445/22315381/IJETT-V70I10P222

[17] Afsaneh Jalalian et al., "Computer-Aided Detection/Diagnosis of Breast Cancer in Mammography and Ultrasound: A Review," *Clinical Imaging*, vol. 37, no. 3, pp. 420–426, 2013. *Crossref*, http://doi.org/10.1016/j.clinimag.2012.09.024

[18] Sudha Prathyusha Jakkaladiki, and Filip Maly, "Potential Role of Artificial Intelligence in Breast Cancer Detection- A Review," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 130–139, 2022. *Crossref*, http://doi.org/10.14445/22315381/IJETT-V70I7P214

[19] Pratheep Kumar P, and V. Mary Amala Bai, "Breast Cancer Detection on Mammographic Images using Hyper Parameter Tuning & Optimization: A Convolutional Neural Network & Transfer Learning Approach," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 79–92, 2022. *Crossref*, http://doi.org/10.14445/22315381/IJETT-V70I9P208

[20] B. Stenkvist et al., "Computerized Nuclear Morphometry as an Objective Method for Characterizing Human Cancer Cell Populations," *Cancer Research*, vol. 38, no. 12, pp. 4688–4697, 1978.

[21] Andrew J.Evans et al., "2014 American Telemedicine Association Clinical Guidelines for Telepathology: Another Important Step in Support of Increased Adoption of Telepathology for Patient Care," *Journal of Pathology Informatics*, vol. 6, no. 1, p. 13, 2015. *Crossref*, http://doi.org/10.4103/2153-3539.153906

[22] Rik Das, Kanwalpreet Kaur, and Ekta Walia, "Feature Generalization for Breast Cancer Detection in Histopathological Images," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 14, no. 2, pp. 566–581, 2022. *Crossref*, http://doi.org/10.1007/s12539-022-00515-1

[23] Emmanuel Maggiori et al., "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, 2017. *Crossref*, http://doi.org/10.1109/TGRS.2016.2612821

[24] Mohammad Rastegari et al., "XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks," *European Conference on Computer Vision*, pp. 525-542, 2016. *Crossref*, http://doi.org/10.1007/978-3-319-46493-0_32

[25] Fabio Alexandre Spanhol et al., "Breast Cancer Histopathological Image Classification using Convolutional Neural Networks," *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567, 2016. *Crossref*, http://doi.org/10.1109/IJCNN.2016.7727519

[26] Smita Khairnar, Sudeep D.Thepade, and Shilpa Gite, "Effect of Image Binarization Thresholds on Breast Cancer Identification in Mammography Images using OTSU, Niblack, Burnsen, Thepade's SBTC," *Intelligent Systems with Applications*, vol. 10–11, p. 200046, 2021. *Crossref*, http://doi.org/10.1016/j.iswa.2021.200046

[27] Xiaobo Zhou, Xiaodong Wang, and Edward R Dougherty, "Binarization of Microarray Data on the Basis of a Mixture Model," *Molecular Cancer Therapeutics*, vol. 2, no. 7, pp. 679–84, 2003.

[28] Youssef Masmoudi, Metin Türkay, and Habib Chabchoub, "A Binarization Strategy for Modelling Mixed Data in Multigroup Classification," *2013 International Conference on Advanced Logistics and Transport*, pp. 347–353, 2013. *Crossref*, http://doi.org/10.1109/ICAdLT.2013.6568483

[29] Huiling Gong et al., "Ultrasound Image Texture Feature Learning-Based Breast Cancer Benign and Malignant Classification," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–8, 2021. *Crossref*, http://doi.org/10.1155/2021/6261032

[30] R. Karthiga, and K. Narasimhan, "Medical Imaging Technique Using Curvelet Transform and Machine Learning for the Automated Diagnosis of Breast Cancer From Thermal Image," *Pattern Analysis and Application*, vol. 24, no. 3, pp. 981–991, 2021. *Crossref*, http://doi.org/10.1007/s10044-021-00963-3

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. *Crossref*, http://doi.org/10.1145/3065386

[32] Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature*, vol. 542, pp. 115–118, 2017. *Crossref*, http://doi.org/10.1038/nature21056

[33] Syeda Sara Samreen, and Hakeem Aejaz Aslam, "Hyperspectral Image Classification using Deep Learning Techniques: A Review," *SSRG International Journal of Electronics and Communication Engineering*, vol. 9, no. 6, pp. 1-4, 2022. *Crossref*, https://doi.org/10.14445/23488549/IJECE-V9I6P101

[34] Yangqin Feng, Lei Zhang, and Juan Mo, "Deep Manifold Preserving Autoencoder for Classifying Breast Cancer Histopathological Images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 1, pp. 91–101, 2020. *Crossref*, https://doi.org/10.1109/TCBB.2018.2858763

[35] Nima Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016. *Crossref*, https://doi.org/10.1109/TMI.2016.2535302

[36] Le Hou et al., "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2424-2433, 2016. *Crossref*, https://doi.org/10.1109/CVPR.2016.266

[37] Jeff Donahue et al., "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," *Proceedings of the 31st International Conference on Machine Learning*, 2013.

[38] Artem Babenko et al., "Neural Codes for Image Retrieval," *European Conference on Computer Vision*, pp. 584-599, 2014. *Crossref*, https://doi.org/10.1007/978-3-319-10590-1_38

[39] Mircea Cimpoi, Subhransu Maji, and Andrea Vedaldi, "Deep Filter Banks for Texture Recognition and Segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3828–3836, 2015. *Crossref*, https://doi.org/10.1109/CVPR.2015.7299007

[40] Ali Sharif Razavian et al., "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 512–519, 2014. *Crossref*, https://doi.org/10.1109/CVPRW.2014.131

[41] Behnaz Gheflati, and Hassan Rivaz, "Vision Transformer for Classification of Breast Ultrasound Images," *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 480-483, 2021. *Crossref*, https://doi.org/10.1109/EMBC48229.2022.9871809

[42] Zhu He et al., "Deconv-Transformer (Dect): A Histopathological Image Classification Model for Breast Cancer Based on Color Deconvolution and Transformer Architecture," *Information Sciences*, vol. 608, pp. 1093–1112, 2022. *Crossref*, https://doi.org/10.1016/j.ins.2022.06.091

[43] Xiyue Wang et al., “TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 186–195, 2021. *Crossref,* https://doi.org/10.1007/978-3-030-87237-3_18

[44] Xuxin Chen et al., “Transformers Improve Breast Cancer Diagnosis from Unregistered Multi-View Mammograms,” *Diagnostics*, vol. 12, no. 7, p. 1549, 2022. *Crossref,* https://doi.org/10.3390/diagnostics12071549

[45] Nidhi Mongoriya, and Vinod Patel, "Review The Breast Cancer Detection Technique Using Hybrid Machine Learning," *SSRG International Journal of Computer Science and Engineering,* vol. 8, no. 6, pp. 5-8, 2021. *Crossref,* https://doi.org/10.14445/23488387/IJCSE-V8I6P102

[46] P. V. Rukmangadha, and R. Das, “Representation-Learning-Based Fusion Model for Scene Classification Using Convolutional Neural Network (CNN) and Pre-trained CNNs as Feature Extractors,” *Computational Intelligence in Pattern Recognition*, pp. 631–643, 2022. *Crossref,* https://doi.org/10.1007/978-981-16-2543-5_54

[47] T. Ojala, M. Pietikäinen, and D. Harwood, “A Comparative Study of Texture Measures with Classification Based on Featured Distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996. *Crossref,* https://doi.org/10.1016/0031-3203(95)00067-4

[48] Keerti Maithil, and Tasneem Bano Rehman, "Urban Remote Sensing Image Segmentation using Dense U-Net+," S*SRG International Journal of Computer Science and Engineering,* vol. 9, no. 3, pp. 21-28, 2022. *Crossref,* https://doi.org/10.14445/23488387/IJCSE-V9I3P103

[49] Bhagwati Charan Patel, and G.R.Sinha, “An Adaptive K-Means Clustering Algorithm for Breast Image Segmentation,” *International Journal of Computer Applications,* vol. 10, no. 4, pp. 35–38, 2010. *Crossref,* https://doi.org/10.5120/1467-1982

[50] Pavel Král, and Ladislav Lenc, “LBP Features for Breast Cancer Detection,” *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2643–2647, 2016. *Crossref,* https://doi.org/10.1109/ICIP.2016.7532838

[51] Soumya Deep Roy et al., “Computer Aided Breast Cancer Detection Using Ensembling of Texture and Statistical Image Features,” *Sensors,* vol. 21, no. 11, p. 3628, 2021. *Crossref,* https://doi.org/10.3390/s21113628

[52] Maheshvar Chandrasekar et al., “Breast Cancer Histopathological Image Classification using EfficientNet Architecture,” in *2020 IEEE International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET)*, pp. 1–5, 2020. *Crossref,* https://doi.org/10.1109/TEMSMET51618.2020.9557441

[53] Mohammed Abdulla Salim Al Husaini et al., “Thermal-Based Early Breast Cancer Detection Using Inception V3, Inception V4 and Modified Inception MV4,” *Neural Computing and Applications*, vol. 34, no. 1, pp. 333–348, 2022. *Crossref,* https://doi.org/10.1007/s00521-021-06372-1

[54] Arijit Das, and Srinibas Rana, “Exploring Residual Networks for Breast Cancer Detection from Ultrasound Images,” in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2021. *Crossref,* https://doi.org/10.1109/ICCCNT51525.2021.9580160

[55] Xiaoyu Yue et al., “Vision Transformer with Progressive Sampling,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 377-386, 2021. *Crossref,* https://doi.org/10.1109/ICCV48922.2021.00044

[56] Zhongyi Han et al., “Breast Cancer Multi-Classification from Histopathological Images with Structured Deep Learning Model,” *Scientific Reports,* vol. 7, no. 1, p. 4172, 2017. *Crossref,* https://doi.org/10.1038/s41598-017-04075-z

[57] Payal Chhabra, Naresh Kumar Garg, and Munish Kumar, “Content-based Image Retrieval System using ORB and SIFT Features,” *Neural Computing and Applications*, vol. 32, no. 7, pp. 2725–2733, 2020. *Crossref,* https://doi.org/10.1007/s00521-018-3677-9