

## A hybrid approach for classification Breast Cancer histopathology Images

Amr H. Abedhaliem<sup>1</sup>, Mohamed E. Wahed<sup>2</sup>, Mohamed S. Metwally<sup>1</sup>, Mohamed A. Atiea<sup>3</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, Suez University, Suez, Egypt

<sup>2</sup> Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

<sup>3</sup> Faculty of Computers and Informatics, Suez University, Suez, Egypt

### ARTICLE INFO

#### Article history:

Received 20 June 2021

Received in revised form 30 July 2021

Accepted 31 July 2021

Available online 19 August 2021

#### Keywords

Breast cancer;  
Convolutional neural network;  
Transfer learning;  
Classification handcrafted feature

### ABSTRACT

Breast cancer is a significant factor in female mortality. Automated identification and classification of breast histopathology image tissue characteristics using computer-aided diagnostic tools is an important step in disease identification and therapy. In this work, we propose an automated classification system which is based on mixing pre-trained deep Convolutional Neural Networks (CNN) as a feature extractor, and multilevel hand-crafted features. The pre-training models are used: ResNet18, Inception ResNet v2, ShuffleNet, and Xception. These hand-crafted features are extracted using Haralick textures, Rotation and Scale-invariant Hybrid image Descriptor (RSHD), Local Diagonal Extrema Pattern (LDEP), Speeded up robust features (SURF), Colored Histogram, and the Dense Invariant Feature Transform (DSIFT) set. All extracted features are reduced by the feature selection method (PCA) and used as a feature vector for the training of three classifiers: Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN). We evaluate the efficiency of the proposed methodology for public microscopy. The ICIAR-2018 dataset contains histopathology images of four classes: invasive carcinoma, in-situ carcinoma, benign tumors, and normal tissue. Experimental results show the accuracy of the proposed method at 96.97%.

### 1. Introduction

Cancer is a disease that splits aberrant cells uncontrollably and destroys other healthy tissues, which eventually leads to tumors. According to the World Health Organization (WHO), cancer is the second largest cause of worldwide death, and in 2020 it will cause over 6,9 million deaths [1,2]. Breast cancer is one of the main fatalities in women's disease. It is caused by delays and errors in diagnosis. As a result, a field dedicated to the early detection and treatment of breast cancer in women is required. The primary objective is to create systems that allow for early cancer detection and diagnosis, so that people can be treated and totally cured if cancer is detected early. A number of imaging technologies, including ultrasound, magnet resonance imaging, digital mammograms (DM) and histology, can be used to identify breast cancer. Digital mammograms are saved and processed easily and are cheap and compatible. One of the disadvantages is, however, digital mammography has a lower area and a larger pretreatment need [4,5]. An MRI provides more information and can be used for inside tissue biopsy. MRI is dangerous as pregnant women are not allowed to use it, and in certain individuals, the chemistry used to improve MRI can cause allergic reactions [6].

Ultrasound imaging is considered beneficial for pregnant women. Nevertheless, due to a lack of picture consistency, it is unable to distinguish accurately between malignant and benign tissue. Images in histopathology operate better than other kinds of pictures. It can indicate a number of cancers and may be stored for a long time to be examined. The diagnosis might differ depending on the experiences of scientists in the analysis of histological images [7]. Because WSIs are quite large, diagnosing disorder kinds by monitoring any patches is difficult for specialists. These challenges can be overcome by the use of computer-aided diagnostic systems. A significant need for improving diagnostics equipment with computer support, therefore, which would decrease the workloads of pathologists by providing them with quick and accurate diagnostics [8].

Robust automatic image classification inside a computer's field of view needs methods for extracting highly distinct combinations of features. As a result, the significant topic of computer vision analysis focuses on methods for extracting such descriptors. These strategies are also classified as a set of features that are created by handcrafted methods or learning-based methods.

Handcrafted descriptors square measure those which might be designed by researchers to extract specific characteristics of an image that involve finding the correct

\* Corresponding authors at: Suez University  
E-mail addresses: [amr\\_cs\\_2012@yahoo.com](mailto:amr_cs_2012@yahoo.com) (Amr H. Abedhaliem)

trade-off between accuracy and machine performance. Another category of feature generation is feature learning-based, supported by convolutional neural networks (CNN) [9,10], and mentioned as "deep methods" that play a supreme role in several medical tasks [9–11]. Thus, breast cancer classification and detection [12, 13]. CNN is a multilayered neural network, learning a set of convolutional filters at each layer [14, 15]. In contrast to manual characteristics, deep learning networks evaluate incoming images by evaluating a number of features immediately acquired from training images [16].

Where data availability is limited, traditional or handcrafted features outperform deep features in being able to explain the high contrast of features. On the one hand, deep learning needs broad datasets to generalize well on the learned weights, and the availability of this data remains an obstacle to medical data, which constitutes a bottleneck for learning convolutional neural networks. In such cases, transfer learning techniques [17] are the best option for medical clinical images for classification activities. Another advantage of handcrafted feature extraction is that it correlates well with the visual features that clinicians specify. These features can be easily obtained by object level, spatial level, and neutral field rather than deep learning methods. In recognition of the benefits inherited from both handcrafted features and deep learning (DL) features, a few recent studies [18, 19] have combined them with different methods to create a classification of medical images.

This work methodically conducted experimental work and led to the following contributions:

- The handcrafted approach to feature extraction from histopathological images is applied using six feature extraction techniques (Table 1) and analyzing classification performance using three classic classifiers: KNN, Random Forest (RF), and SVM to find the finest classifier in a multiple-classification BC with handcrafted features.

- The use of pre-trained Deep Neural Networks as feature extractors in breast cancer multiclass image-based classification utilising the ICIAR-2018 histopathological image dataset, including ResNet18, Inception ResNet v2, ShuffleNet, and Xception.

- Different performance assessment criteria, including accuracy, precision, sensitivity, and F1 score, are used to assess the performance of classifiers.

- This research compares the performance of handcrafted techniques with the technique of transfer learning and developed a computer-aided diagnosis tool based on ensembles of two approaches to the classification of breast cancer.

### Related Work

Image interpretation is measured using deep features differently from handcrafted features. Handcrafted algorithms are inspired by human perceptions of things like color, scale, surface, shape, texture, depth, and the ability to distinguish an object from the background. The DL approaches, on the other hand, are pixel-based

representations of the image features investigated via a wide range of iterations before the closest picture is converted. The previous majority of works for classification algorithms extracted features of nuclei and glands that have been segmented. In the experience of [20], which resulted in the object-level features extracted from nuclear structures for image-stained H & E having about 97 percent accuracy in detecting prostate cancer. Spatial features with regard to cell nuclei and glands have been openly utilized in previous literature [21-24]. Defect features and object-level spatial that include segmentation of the object before measuring features.

Numerous machine learning techniques have been used to analyse the ICIAR-2018 histopathology images. They employed a fine-tuned version of ALEXNET to image breast cancer and called it TK-ALEXNET, achieving an overall accuracy of 57%. By combining VGG-16 and VGG-19 networks, [46] achieves an average accuracy of approximately 92 percent. [48] presented an approach based on Inception ResNet V2 that obtained a transfer learning accuracy of approximately 90%. When considerable data augmentation is used, Inception-v4 outperforms the standard CNN baseline by a factor of 89 [49]. In [50], it was proposed to use a hybrid Convolutional Neural Network (CNN) architecture. For four classes, 87.5 percent classification accuracy was obtained. The authors of [51] developed a network architecture based on a pre-trained Xception model that achieves a classification accuracy of 92.50 percent on average. We attained a 96.1 percent classification accuracy on the test set using the hybrid deep convolutional neural network [52]. Recently, authors [53] The proposed automatic computer aided diagnosis (CADx) system divided the ICIAR-2018 data set into two classes. The benign group contains (normal and benign) images and the malignant group contains (situ carcinoma and invasive carcinoma) images and achieved accuracy 97.93 for binary classification.

In our study, we modeled the images directly without using segmentation. Many of the current raw image investigations have been used to extract features from medical databases, most of which use deep learning techniques.; As with pre-trained CNN, features for transfer learning may be extracted.

According to some research, they do not provide trained models already on independent sets of data quality features. Instead, they have been trained in advance on more massive data sets, but have been tuned to smaller and associated data sets, leading to improved output standards. For example, [25] was used by CNN to identify images of human epithelial-2, and discovered that the preset fine-tuned pre-trained outperformed those that had been trained from scratch on a smaller data set. In addition, it validated this technique in [26]. Many published research advantages of using deep learning classifiers, like classification of brain tumor histopathology images [27], breast cancer histopathology images [28-30].

There are also recent studies on the use of both HandCraft-Features and deep learning features. Use the

authors' [31] features HandCraft-along with deep learning features generalized to detect pathological images of breast cancer. They proposed a set of consecutive two-step strategies for the first classification mitosis by featuring HandCraft and depth of learning separately and then integrating them into the second stage to get the best results. However, their approach focused on the characteristics of the object level, such as morphology, density, and texture, which involves the segmentation of the candidates before the extraction features.

Similarly, authors in [32] stated that the combination of BoF and LBP descriptors with deep VGGf and Caffe-ref features with reduced dimensions improved classification accuracy. All these studies have paved the way for more open research possibilities of combining traditional HandCraft-Features with deep features where one approach inspires domain-specific responses; the other approach increases the generalizing capability. Through our efforts, we have thus sought to solve every single shortcoming of these literary works.

- Use raw images without segmentation directly.

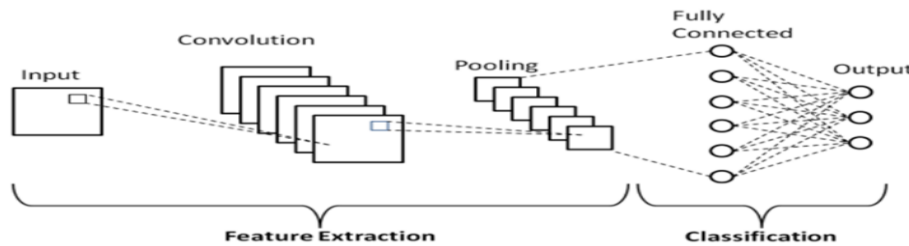
- highlight the need for large-scale trials in order to identify models for the optimal performance of the ensemble methodology approach.

- To validate our results using more appropriate performance measures.

## Methodology

### A. Deep Learning features extraction

Like regular neural networks, other than Convolution neural networks contain some additional layers and consist of three main blocks: the convolution layer, the pooling layer, and the fully connected layer, as illustrated in Figure 1. The layer of convolution utilizes convolutional windows that convolve with the input image and extract features. Then reduce the size of the extracted features by the pooling layer to make them faster in processing. There are several types of pooling, such as max-pooling, min-pooling, and average pooling. Then, after passing through a fully connected layer, the output is predicted into classes.



**Fig. 1:** Basic convolutional neural network (CNN) architecture

Transfer Learning (applying a model that has already been trained on a large labelled dataset for another job). The fine-tuning technique and the deep feature extraction technique are the two basic strategies in Transfer Learning. As a result, in our strategy, we apply a deep feature extraction strategy to extract deep features from images and then classify the retrieved features using a computer classifier. By removing the entirely linked layer from the top layer of the pre-trained model. Following the max-pooling layer, features are retrieved from the convolutional layer, which is then followed by a global average pooling layer. After that, the obtained features are merged to generate a vector of features. In this work, four pretrained ConvNet architectures were used: ResNet18 [33], Inception ResNet v2 [34], ShuffleNet [35], and Xception [36] with their default parameter settings with an average pooling implemented in the Keras [37] deep learning library.

### B. Handcrafted feature extraction

We used several handcrafted algorithms and their advantage is that they can turn raw images into feature descriptors without passing through the segmentation process. The handcrafted features are learned using descriptors such as Haralick Textures [38], Rotation and Scale-invariant Hybrid Image Descriptor (RSHD) [39],

Local Diagonal Extrema Pattern (LDEP) [40], speeded up robust feature (SURF) [41], Colored Histogram [42], and Dense Scale Invariant Feature Transform (DSIFT) [43].

### C. Proposed model

Figure 2 depicts the suggested model's general architecture, which was created to produce a Breast Cancer (BC) classifier utilising histopathology image. The research methodology's output is based on five basic steps: Step1 (Input Data set), Step2 (image pre-processing and augmentation) and Step3 (feature extraction by two techniques). Hand-crafted features and transfer learning-based model). In Step4 (dimension reduction technique), Step5 (classification technique and performance evaluation metrics).

#### Step 1 Input histopathology image dataset

The dataset was made public in the context of the great ICIAR-2018 challenge. It includes 400 images in four categories: normal carcinoma, brain carcinoma, on-site carcinoma, and invasive cancer. Figure 3(a–d) shows the above four classes when the tissue is viewed under a microscope following suitable staining.

#### Step 2 pre-processing (Normalization and augmentation)

Histological images show inter-image variability regarding contrast and color. Image normalization may improve the system robustness for the training stage as well as the prediction stage. Since staining, acquisition, and digitization processes are performed in different conditions and laboratories, histopathology may show heterogeneity. The idea behind normalization is to allow a robust classification framework and to avoid learning overfitting. We investigated the effectiveness of popular stain normalization techniques based on the Retinex theory developed by Land and Mc-Cann in [44].

The augmentation strategy is used as a preprocessing strategy to prevent problems with limited data size and unbalanced data. This pre-processing approach helps reduce overfitting in the case of the CAD system. Medical imaging uses data augmentation techniques to copy, flip, rotate, translate, and interpolate images. We used rotation and horizontal flipping for the data increase strategy here. Due to the fact that histology images are rotation and

reflection in nature, information might be lost if another strategy to increase is used. We rotated pictures by 90, 180, and 270 degrees to increase the size of the training data set.

### Step 3 Feature extraction

Feature extraction is the most essential stage in classification. Because each feature contains vital information. To extract features, two techniques were employed. Handcrafted features and a model based on transfer learning.

#### Hand-crafted features

In this, we have extracted features using Haralick textures techniques, Rotation and Scale-invariant Hybrid image Descriptor (RSHD), Local Diagonal Extrema Pattern (LDEP), Speeded up robust features (SURF), Colored Histogram, and Dense Invariant Feature Transform (DSIFT).

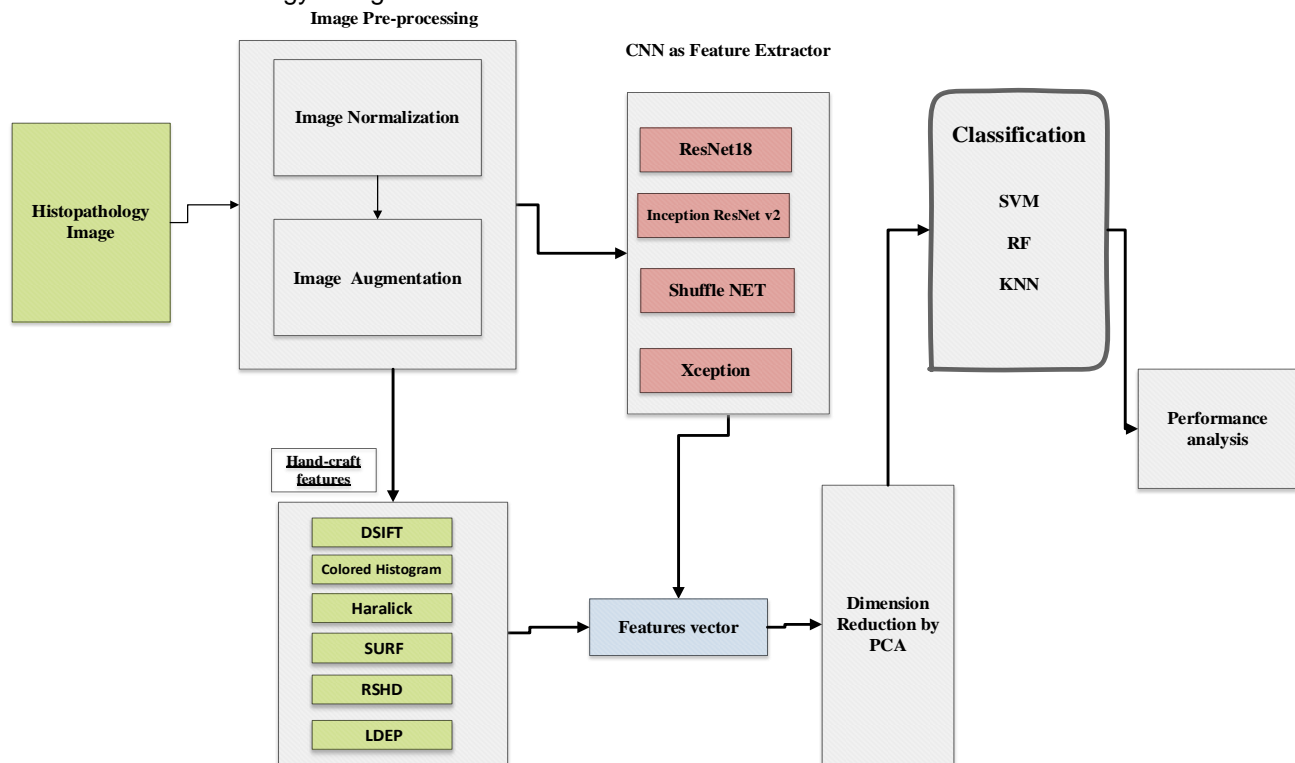


Fig. 2: Proposed model

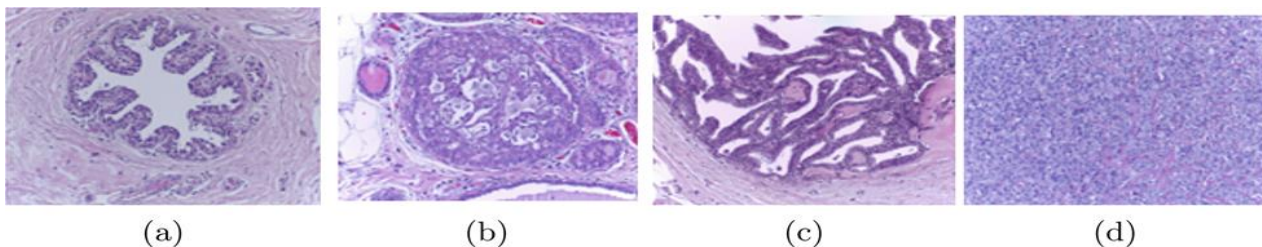


Fig. 3: Histology images. (a) Normal tissue (b) Benign lesion (c) in situ carcinoma (d) Invasive carcinoma

Pre-Trained Feature Here, features are extracted using various variations of the CNN model, ResNet18, Inception ResNet v2, ShuffleNet, and Xception (pre-trained) models.

#### Step 4 Dimensionality Reduction

To minimise feature dimension, PCA is applied to feature vectors. The features are harmonized before PCA is used. PCA was only trained on a training set, and both the training and test sets were projected into a lower-dimensional space. PCA is used with a 0.95 explained variance ratio.

#### Step 5 Classification

Three different classifiers are utilised (RF, SVM, and K-NN). The classifiers get the reduced images as feature vectors for classification. Every classifier is trained on a training set before being put to the test on a test set.

#### Performance measures

Several standard metrics are used to measure the success of the proposed diagnosis system: accuracy, recall, precision, F-measure

$$\text{Accuracy (AC)} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall (Sn)} = \frac{TP}{TP + FN}$$

$$\text{Precision (Pr)} = \frac{TP}{TP + FP}$$

$$F\text{-measure}(F1) = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Where TP (True Positives)-Correctly classified as having breast cancer, TN (True Negatives)-Correctly classified as not having breast cancer (False Positive)-Classified as having breast cancer but actually they don't have it. FN (False Negatives)-Classified as not having breast cancer but actually having cancer

## Results and Discussion

To evaluate the performance of the three classifiers along with all the handcrafted algorithms, a 10-fold cross validation approach was applied to the training set and then a test result was generated on the test set.

### A. Only handcrafted feature classification

We discovered that among the six approaches tested, color histogram and RSHD performed substantially better (Table 1). The reason for this is that the color histogram and RSHD encode colors and are rotation and scale-invariant. Because of the presence of proteins that respond to staining dye by generating a dark blue or purple hue for nuclei and a pink hue for cytoplasm in the cell, H & E stained histopathology images are extremely colour sensitive, and performing stain-color normalization as a preprocessing stage considerably enhances the colour features. However, all other handcrafted approaches employ the grey image as an input to further processing for color-sensitive histopathology images, resulting in the loss of critical information. The RF classifier yielded the highest classification results and dominated all other classifiers, followed by KNN and SVM respectively. The corresponding recalls and precision for different classifiers are shown in Figure 3.

**Table 1:** Comparative performance measures for handcrafted feature classification (without PCA).

Feature Extractors	Classifier	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)
Haralick Textures	SVM	44.44	43.48	45.45	44.44
	RF	45.45	43.48	47.62	45.46
	KNN	41.07	39.29	42.86	41.00
RSHD	SVM	77.4	79.75	74.63	77.11
	RF	80.65	81.08	80.25	80.66
	KNN	72.9	78.75	66.67	72.21
LDEP [10]	SVM	53.19	50.85	55.56	53.10
	RF	54.95	55.56	54.35	54.95
	KNN	52.63	55.56	50.01	52.64
SURF	SVM	47.93	47.37	48.54	47.95
	RF	45.41	43.55	47.62	45.49
	KNN	56.5	57.47	55.56	56.50
Color (Colored Histogram)	SVM	71.07	75	66.67	70.59
	RF	79.67	82.89	74.47	78.45
	KNN	75.84	79.75	71.43	75.36
DSIFT	SVM	62.78	67.74	57.47	62.18
	RF	67.26	71.59	62.5	66.74
	KNN	66.47	71.59	60.98	65.86

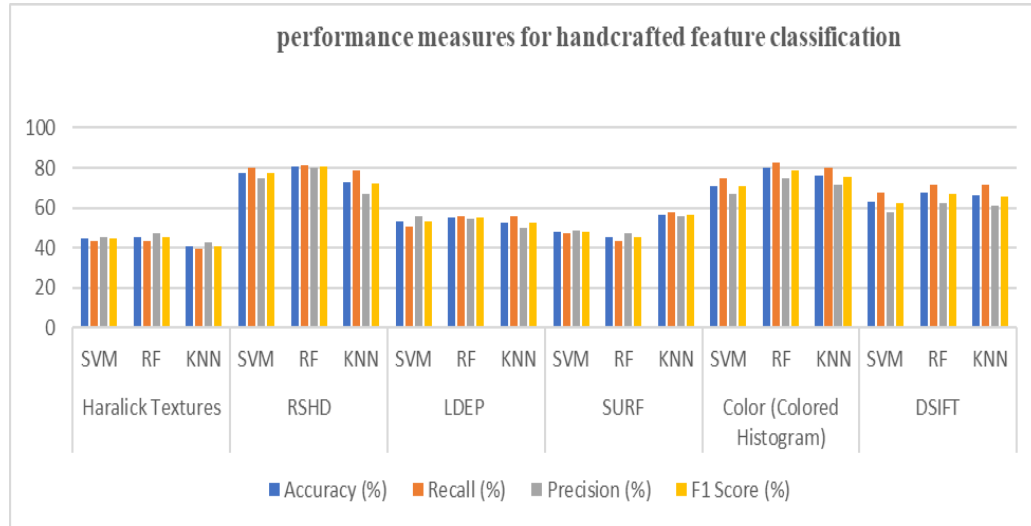


Fig. 3: Performance measures for handcrafted feature classification

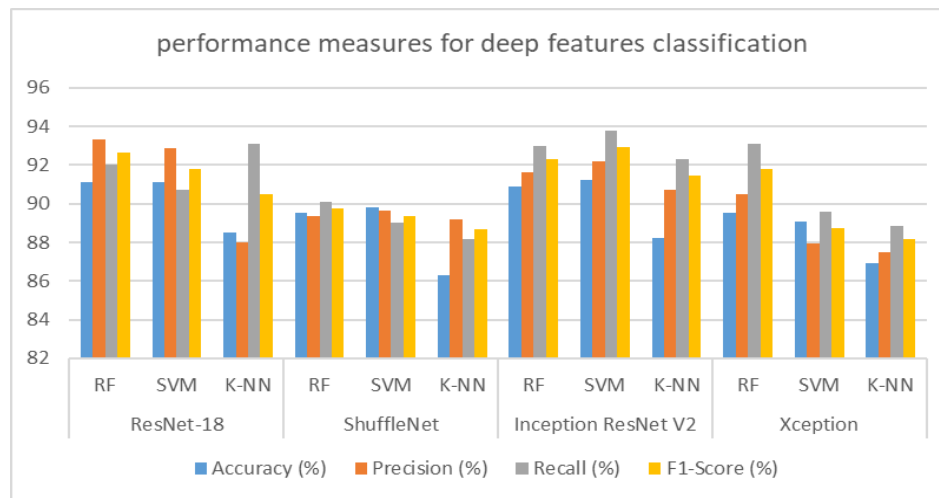


Fig. 4: Performance measures for deep feature classification

Table 2: Comparative performance measures for deep features classification.

Feature Extractor	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-18	RF	91.11	93.33	91.994	92.66
	SVM	91.11	92.86	90.701	91.77
	K-NN	88.52	88.00	93.094	90.48
ShuffleNet	RF	89.54	89.35	90.105	89.73
	SVM	89.79	89.66	89.005	89.33
	K-NN	86.28	89.21	88.195	88.70
Inception ResNet V2	RF	90.89	91.64	92.999	92.32
	SVM	91.23	92.2	93.8	92.94
	K-NN	88.25	90.69	92.28	91.48
Xception	RF	89.55	90.48	93.116	91.78
	SVM	89.07	87.93	89.586	88.75
	K-NN	86.91	87.50	88.826	88.16

### B. Compare deep learning features and different combinations of them.

On images using three different classifiers, RF, SVM, and KNN, we compared the performance of four pre-trained deep neural networks: ResNet18, Shuffle Net, Inception Resnet V2, and Inception-V3Net. With ResNet18, RF and SVM had the highest accuracy, but RF had the highest precision, recall, and f1score. The RF model is better than the SVM model at correctly classifying the positive class, which is always preferable in diagnostic histopathology practice. With Inception, ResNet V2 with SVM classifier had the best accuracy, precision, recall, and f1 score (Table 2, Figure 4), followed by ResNet 18 with RF classifier, which achieved the best performance. Table 3 shows different combinations of four pre-trained networks (ResNet18, Shuffle Net, Inception Resnet V2, and Inception-V3Net). The better accuracy was achieved by combining (Inception ResNet V2 + ResNet-18+ Xception) with the RF classifier by 94.61 Accuracy, and combining (Inception ResNet V2 + ResNet-18) with the RF classifier by 94.19 Accuracy.

### C. Ensemble of deep feature classification, Ensemble of handcrafted feature classification, and Ensemble of deep + handcrafted features

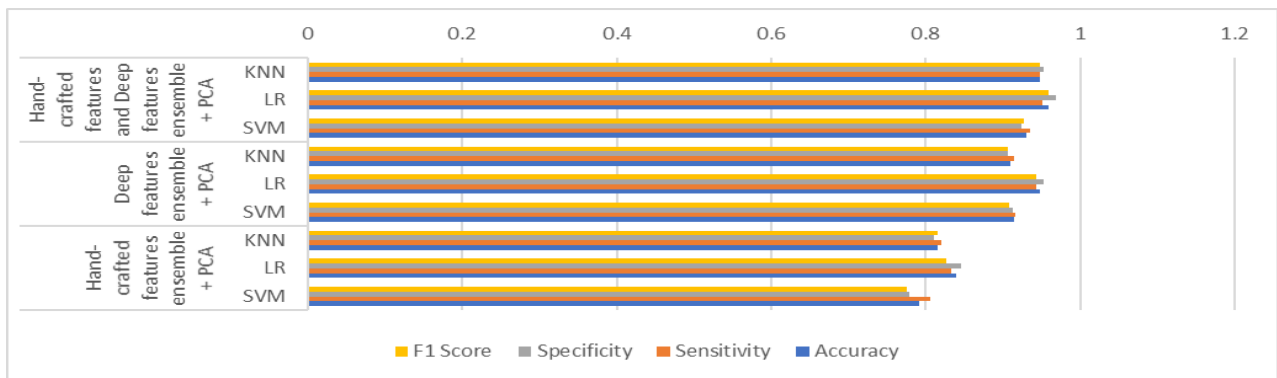
Table 4 using the ten-fold cross-validation procedure provides a comparison between the classification results achieved by the all-handcrafted features set, the all-deep-learned features set, and combining the deep-learned features feature set with the handcrafted features. The best classification performance is obtained by combining the deep learning features set with the handcrafted features and processing these combined features using feature selection (PCA) to choose a combination of features, as shown in (Table 4, Figure 5). In fact, the RF classifier was able to classify histopathology images with the greatest accuracy of 96.97 percent, the best recall of 95.88 percent, the highest precision of 97.88 percent, and the highest F1 of 95.87 percent. These results indicate the combined handcrafted features and deep learned features are able to achieve higher classification performance compared with the deep learned features alone and handcrafted features alone. The high classification sensitivity achieved by integrating the deep learning features feature set with the handcrafted features suggests that these combined features can detect positive breast cancers with high accuracy.

**Table 3:** Comparative performance measures for different combinations deep features classification.

Feature Extractors	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-18 + ShuffleNet + Inception ResNet V2	SVM	92.91	92.19	93.65	92.91
	RF	93.55	94.74	91.53	93.1
	KNN	90.91	88.89	93.02	90.91
ShuffleNet + Inception ResNet V2 + Xception	SVM	92.62	88.68	94	91.26
	RF	92.31	90.2	93.88	92
	KNN	91.3	84.8	92.17	88.33
Inception ResNet V2 + ResNet-18+ Xception	SVM	93.52	91.00	94.79	92.86
	RF	94.61	92.66	95.28	<b>93.95</b>
	KNN	93.67	94.81	92.41	93.59
ShuffleNet + ResNet-18+ Xception	SVM	94.32	95.4	93.26	94.32
	RF	93.87	88.62	95.61	91.98
	KNN	92.37	90.74	90.74	90.74
ResNet-18 + ShuffleNet	SVM	93.12	92.06	95.87	93.93
	RF	93.19	94.18	91.94	93.05
	KNN	90.85	89.58	94.51	91.98
Inception ResNet V2 + ResNet-18	SVM	93.86	91.82	95.28	93.52
	RF	94.19	92.80	95.87	94.31
	KNN	93.52	93.24	95.54	94.38
ResNet-18 + Xception	SVM	92.66	93.33	92.72	93.02
	RF	93.29	93.83	92.68	93.25
	KNN	91.21	91.74	93.46	92.59
ShuffleNet + Inception ResNet V2	SVM	93.14	94.00	94.00	94.00
	RF	93.85	94.32	93.26	93.79
	KNN	92.11	92.52	93.4	92.96
ShuffleNet + Xception	SVM	90.96	90.57	93.2	91.87
	RF	91.87	91.6	93.97	92.77
	KNN	91.63	91.38	93.81	92.58
Inception ResNet V2 + Xception	SVM	92.96	90.53	94.51	92.47
	RF	93.14	91.35	95.00	93.14
	KNN	92.91	93.01	93.66	93.33

**Table 4:** Comparative performance measures Ensemble of handcrafted feature classification, Ensemble of deep feature classification

Feature Extractors	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
All Hand-crafted features vector + PCA	SVM	81.48	81.97	81.08	81.52
	RF	84.97	84.33	84.51	84.42
	KNN	80.14	81.65	77.92	79.74
All Deep features vector + PCA	SVM	91.44	91.59	91.31	91.45
	RF	94.83	94.34	95.24	94.79
	KNN	90.98	91.38	90.63	91.00
All Hand-crafted features and all Deep features ensemble + PCA	SVM	94.97	94.55	95.42	94.98
	RF	96.97	97.88	96.83	97.35
	KNN	93.83	94.83	95.24	95.03

**Fig. 5.** Performance measures Ensemble of handcrafted feature classification, Ensemble of deep feature classification.

### Comparison with Similar Works

Table 5 shows the image-wise comparison between our model and the most recent approaches that used the ICIAR-2018 dataset. Using the ICIAR-2018 challenge dataset, our model outperformed the most recent approaches. With a combination of all hand-crafted features, all deep features, and an RF classifier, it achieved an accuracy of 96.97 percent.

**Table 5:** Comparative performance between our model and the most recent approaches that used the ICIAR-2018 dataset

Method	Image-Wise (%)
Nawaz W., et al. [45]	81.25
Wang, Z., et al. [46]	90
Ferreira C.A., et al. [47]	90
Vang Y.S., et al. [48]	87.5
Sarker M., I et al. [49]	89
Guo Y., et al. [50]	77
Kassani S. H., et al. [51]	92.5
Alzubaidi, et al. [52]	96.1
Proposal model	96.97

### Conclusion

In this study, we compared handcrafted features to a pre-trained model for multiclass classification of histopathological breast cancer images. The Deep Learning-based approach based on transfer learning as features extracted shows a remarkable performance superiority compared to handcrafted approaches. Among different combinations of classifiers, the better performance classification results achieved using handcraft approaches is 80.65% for four-class classification, while the best-obtained classification accuracy using the Deep Learning approach is 91.11% for four-class classification. The best-obtained classification accuracy using the all Deep Learning approach is 94.83%, and 94.61 % for combining (Inception ResNet V2 + ResNet-18+ Xception) for four-class classification. Additionally, data augmentation techniques are used, which help to improve classification accuracy even further. Integrating Deep Learning approaches may be shown to be a superior technique than combining handcrafted features, according to our findings. Hence, better performance classification results were achieved by combining handcrafted features and deep-learning features with 96.79% for four-class classification. We looked at the results in terms of precision and recall, which is important when categorizing medical images since we want to correctly distinguish the tumour image.



## References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*;68(6):394-424, 2018. <https://doi.org/10.3322/caac.21492> .
- [2] Organization WHO. Available from: available: <http://www.who.int/en/>.
- [3] M. C. Chun. Breast Cancer: Symptoms, Risk Factors, and Treatment, *Medical News Today* Available from: <https://www.medicalnewstoday.com/articles/37136.php>.
- [4] Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. 2009 Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 253(3):641-51. <https://doi.org/10.1148/radiol.2533082308> .
- [5] Sadaf A, Crystal P, Scaranelo A, Helbich T. 2011 Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. *Eur J Radiol*. 77(3):457–61. Available from: <http://dx.doi.org/10.1016/j.ejrad.2009.08.024>.
- [6] Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. 2017 Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Sci Rep*.7(1):1-10.
- [7] Robbins, P., Pinder, S., De Klerk, N., Dawkins, H., Harvey, J., Sterrett, G., Ellis, I. and Elston, C., 1995. Histological grading of breast carcinomas: a study of interobserver agreement. *Human pathology*, 26(8), pp.873-879.[https://doi.org/10.1016/0046-8177\(95\)90010-1](https://doi.org/10.1016/0046-8177(95)90010-1).
- [8] Rodenburg, B. (2016). Deep Learning in Histopathology Research Paper Business Analytics.
- [9] Nurmaini, S.; Darmawahyuni, A.; Sakti Mukti, A.N.; Rachmatullah, M.N.; Firdaus, F.; Tutuko, B. Deep Learning-Based Stacked Denoising and Autoencoder for ECG Heartbeat Classification. *Electronics* 2020, 9, 135. <https://doi.org/10.3390/electronics9010135>.
- [10] Yang, Z.; Leng, L.; Kim, B.-G. StoolNet for Color Classification of Stool Medical Images. *Electronics* 2019, 8, 1464. <https://doi.org/10.3390/electronics8121464>.
- [11] Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; Duan, Y. Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis. *Electronics* 2020, 9, 427. <https://doi.org/10.3390/electronics9030427>.
- [12] Abdel-Zaher, A.M.; Eldeib, A.M. Breast cancer classification using deep belief networks. *Expert Syst. Appl.* 2016, 46, 139-144. <https://doi.org/10.1016/j.eswa.2015.10.015>
- [13] Al-Zubaidi, L. Deep Learning Based Nuclei Detection for Quantitative Histopathology Image Analysis. Ph.D. Thesis, University of Missouri, Columbia, MO, USA, 2016.
- [14] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition." pp. 647-655.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks." pp. 1097-1105.
- [16] K. Bora, M. Chowdhury, L. B. Mahanta, M.K. Kundu, and A. K. Das, "Pap smear image classification using convolutional neural network," presented at the Tenth Indian Conference on Computer Vision, Graphics and Image Processing 2016 <https://doi.org/10.1145/3009977.3010068>
- [17] Torrey L, Shavlik J (2009) Transfer learning. *Handbook of research on machine learning applications*, vol 3.
- [18] Wang H, Roa AC, Basavanahally AN, Gilmore HL, Shih N, Feldman M, Tomaszewski J, Gonzalez F, Madabhushi A (2014) Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging* 1(3):034003. <https://doi.org/10.1117/1.JMI.1.3.034003> .
- [19] Zhang J, Xia Y, Xie Y, Fulham M, Feng DD (2018) Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. *IEEE J Biomed Health Inform* 22(5):1521-1530. <https://doi.org/10.1109/JBHI.2017.2775662>.
- [20] Jafari-Khouzani K, Soltanian-Zadeh H (2003) Multiwavelet grading of pathological images of prostate. *IEEE Trans Biomed Eng* 50(6):697-704.<https://doi.org/10.1109/TBME.2003.812194> .
- [21] Demir C, Yener B (2005) Automated cancer diagnosis based on histopathological images: a systematic survey. *Rensselaer Polytechnic Institute, Tech. Rep.*
- [22] Dhandra B, Hegadi R, Hangarge M, Malemath V (2006) Endoscopic image classification based on active contours without edges. In: 2006 1st international conference on digital information management. *IEEE*, pp 167-172. <https://doi.org/10.1109/ICDIM.2007.369348>
- [23] Keenan SJ, Diamond J, Glenn McCluggage W, Bharucha H, Thompson D, Bartels PH, Hamilton PW (2000) An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (cin). *J Pathology* 192(3):351-362. <https://doi.org/10.1002/1096-9896>.
- [24] Weyn B, Van De Wouwer G, Van Daele A, Scheunders P, Van Dyck D, Van Marck E, Jacob W (1998) Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Cytometry: The Journal of the International Society for Analytical Cytology* 33(1):32-40.[https://doi.org/10.1002/\(SICI\)1097-0320\(19980901\)33:1<32::AID-CYTO4>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0320(19980901)33:1<32::AID-CYTO4>3.0.CO;2-D)
- [25] Gao Z, Wang L, Zhou L, Zhang J (2016) Hep-2 cell image classification with deep convolutional neural networks. *IEEE J Biomed Health Inform* 21(2):416-428. <https://doi.org/10.1109/JBHI.2016.2526603>.
- [26] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Medical Imaging* 35(5):1299-1312. <https://doi.org/10.1109/TMI.2016.2535302>.
- [27] Xu Y, Jia Z, Ai Y, Zhang F, Lai M, Eric I, Chang C (2015) Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). *IEEE*, pp 947-951 <https://doi.org/10.1109/ICASSP.2015.7178109>.
- [28] Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermsen M, Manson QF, Balkenhol M et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* 318(22):2199-2210.
- [29] Litjens G, Bandi P, Ehteshami Bejnordi B, Geessink O, Balkenhol M, Bult P, Halilovic A, Hermsen M, van de Loo R, Vogels R et al (2018) 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *Gigascience* 7(6):giy065. <https://doi.org/10.1093/gigascience/giy065>.

- [30] Chen CL, Mahjoubfar A, Tai LC, Blaby IK, Huang A, Niazi KR, Jalali B (2016) Deep learning in label-free cell classification. *Scientific Reports* 6:21471. <https://doi.org/10.1038/srep21471>
- [31] Wang H, Roa AC, Basavanthally AN, Gilmore HL, Shih N, Feldman M, Tomaszewski J, Gonzalez F, Madabhushi A (2014) Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging* 1(3):034003. <https://doi.org/10.1117/1.JMI.1.3.034003>.
- [32] Zhang J, Xia Y, Xie Y, Fulham M, Feng DD (2018) Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. *IEEE J Biomed Health Inform* 22(5):1521-1530. <https://doi.org/10.1109/JBHI.2017.2775662>.
- [33] He K., Xiangyu Z., Shaoqing R., Jian S. (2016) "Deep residual learning for image recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770-778 <https://doi.org/10.1109/CVPR.2016.90>.
- [34] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, Arxiv.org. Available from: <https://arxiv.org/abs/1602.07261>.
- [35] Zhang X., Zhou X., Lin M., Sun J., "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pages 6848-6856.
- [36] Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, Arxiv.org. Available from : <https://arxiv.org/abs/1610.02357>.
- [37] Keras.io, Keras Documentation, Available from: <https://keras.io/>, 2018. (Accessed 10 June 2018)
- [38] Haralick RM, Shanmugam K: Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*:610-621, 1973 <https://doi.org/10.1109/TSMC.1973.4309314>.
- [39] Dubey SR, Singh S, Singh RK (2015) Rotation and scale invariant hybrid image descriptor and retrieval. *Comput Elect Eng* 46:288-302. <https://doi.org/10.1016/j.compeleceng.2015.04.011>.
- [40] Dubey SR, Singh S, Singh RK (2015) Local diagonal extrema pattern: a new and efficient feature descriptor for ct image retrieval. *IEEE Signal Processing Letters* 22(9):1215-1219. <https://doi.org/10.1109/LSP.2015.2392623>.
- [41] Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (surf). *Comput Vis Image Understand* 110(3):346-359 <https://doi.org/10.1016/j.cviu.2007.09.014>.
- [42] Jeong S: Histogram-based color image retrieval. Psych221/EE362 project report, 2001.
- [43] A. Bosch, A. Zisserman, and X. Munoz, "Image Classification using Random Forests and Ferns," 2007 IEEE 11th International Conference on Computer Vision, 2007. <https://doi.org/10.1109/ICCV.2007.4409066>.
- [44] E.H. Land, J.J. McCann, Lightness and retinex theory, *J. Opt. Soc.Am.* 61 (1971), 1-11 <https://doi.org/10.1364/JOSA.61.000001>.
- [45] Nawaz, W.; Ahmed, S.; Tahir, A.; Khan, H.A. Classification of breast cancer histology images using AlexNet. In International Conference Image Analysis and Recognition; Springer: Cham, Switzerland, 2018; pp. 869-876. Mahbod, A.; [https://doi.org/10.1007/978-3-319-93000-8\\_99](https://doi.org/10.1007/978-3-319-93000-8_99).
- [46] Wang, Z.; Dong, N.; Dai, W.; Rosario, S.D.; Xing, E.P. Classification of breast cancer histopathological images using convolutional neural networks with hierarchical loss and global pooling. In Proceedings of the International Conference Image Analysis and Recognition, Póvoa de Varzim, Portugal, 27-29 June 2018; pp. 745-753. [https://doi.org/10.1007/978-3-319-93000-8\\_84](https://doi.org/10.1007/978-3-319-93000-8_84).
- [47] Ferreira, C.A.; Melo, T.; Sousa, P.; Meyer, M.I.; Shakibapour, E.; Costa, P.; Campilho, A. Classification of Breast Cancer Histology Images through Transfer Learning Using a Pre-Trained Inception ResNet v2; Springer: Cham, Switzerland, 2018; pp. 763-770. [https://doi.org/10.1007/978-3-319-93000-8\\_86](https://doi.org/10.1007/978-3-319-93000-8_86).
- [48] Vang, Y.S.; Chen, Z.; Xie, X. Deep Learning Framework for Multi-Class Breast Cancer Histology Image Classification; Springer: Cham, Switzerland, 2018; pp. 914-922. [https://doi.org/10.1007/978-3-319-93000-8\\_104](https://doi.org/10.1007/978-3-319-93000-8_104).
- [49] Sarker, M.I.; Kim, H.; Tarasov, D.; Akhmetzanov, D. Inception Architecture and Residual Connections in Classification of Breast Cancer Histology Images. arXiv 2019, arXiv:1912.04619. Available online: <https://arxiv.org/abs/1912>.
- [50] Guo, Y.; Dong, H.; Song, F.; Zhu, C.; Liu, J. Breast Cancer Histology Image Classification Based on Deep Neural Networks; Springer: Cham, Switzerland, 2018; pp. 827-836. [https://doi.org/10.1007/978-3-319-93000-8\\_94](https://doi.org/10.1007/978-3-319-93000-8_94).
- [51] Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Breast cancer diagnosis with transfer learning and global pooling. arXiv 2019, arXiv:1909.11839. Available online: <https://doi.org/10.1109/ICTC46691.2019.8939878>.
- [52] Alzubaidi, L.; Al-Shamma, O.; Fadhel, M.A.; Farhan, L.; Zhang, J.; Duan, Y. Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model. *Electronics* 2020, 9, 445. <https://doi.org/10.3390/electronics9030445>
- [53] Attallah O, Anwar F, Ghanem NM, Ismail MA. 2021. HistocADx: duo cascaded fusion stages for breast cancer diagnosis from histopathological images. *PeerJ Computer Science*.