Refined Color Texture Classification using CNN and Local Binary Pattern

Khalid M. Hosny [1*], Taher Magdy [2], Nabil A. Lashin [1], Kyriakos Apostolidis[3] and George A. Papakostas[3*]

[1] Information Technology Department, Zagazig University, Zagazig 44519, Egypt.

[2] Computer Science Department, Sinai University, Arish, North Sinai, Egypt.

[3] HUMAIN-Lab, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece.

*Corresponding author: gpapak@cs.ihu.gr

Abstract

Representation and classification of color texture generating considerable interest within the field of computer vision. Texture classification is a difficult task that tends to assign unlabeled images or textures to the correct labeled class. Some of key factors such as scaling and viewpoint variations and illumination changes make this task challenging. In this paper, we present a new feature extraction technique for color texture classification and recognition. Our analysis shows that using LBP improves the classification task when compared to using CNN only. The presented approach aggregates the features extracted from Local Binary Patterns (LBP) and Convolution Neural Network (CNN) to provide discriminatory information, leading to better texture classification results. Almost all of the CNN model cases classify images based on global features that describe the image as a whole to generalize the entire object. LBP classifies images based on local features which describe the key points (image patches) in the image. We test the proposed approach experimentally over three challenging color image datasets (ALOT, CBT, Outex). Our results demonstrated that our approach achieved an improvement of up to 25% in the recognition accuracy over the traditional CNN models. We identify optimal combinations for each dataset and obtain high classification rates. The proposed approach is robust, stable, and discriminatory among the three datasets and noticed enhancement for classification and recognition compared to the state-of-the-art method.

Keyword: Color Textures; Classification; Convolution Neural Network; Local Binary pattern.

1. Introduction

The motivation of the usage of CNN models as features descriptor is the ability of the deep neural network to capture the high level features that can be key point for classification of the texture images. Texture plays a significant role in distinguishing objects in color images. Texture classification involves a-two phase process. The first phase is the extraction of the features, which provides a feature-based description for each texture type; this phase tends to select features unaffected by image transformation, such as scaling, translation, and rotation. The second phase tends to recognize the texture from the extracted features. Texture recognition is increasingly set a critical issue in computer vision that has many applications such as biomedical image processing [1-5], object detection [6-8], remote sensing [9-11] application fields.

The scale invariant feature transform (SIFT), proposed by Lowe [12], is a common sparse descriptor. The SIFT descriptor was changed by Mikolajczyk and Schmid by altering the gradient location orientation grid as well as the quantization parameters of the histograms [13]. Ghen et al [14]. proposed robust discriminative descriptor called Weber Local Descriptor (WLD) which based on human perception. Gabor wavelet descriptor is considered as one of the most widely used dense descriptor [15]. The Gabor wavelet has achieved extensive use in image recognition applications such as face recognition, scene analysis and motion tracking, and face recognition [16]. Ershad and Fekri [17] proposed a new approach which could analyze the texture based on its grain components and classify the texture from grain components histogram and statistical features. Local Binary Pattern (LBP) is indeed one of the texture descriptor methods used to describe and represent texture. Ojala et al. [18] presented it as an adequate gray-level image features descriptor. It was then applied to the color image by Mäenpää and Pietikäinen [19] and used in many color representation and recognition tasks such as facial expression recognition [20-21], gender classification [22-23], scene classification [24-25], medical image analysis [26-28] and 3D face recognition [29]. Ershad and Tajeripour [30] proposed a noise resistant and multi-resolution version of LBP called HCLBP and proposed an approach which adaptable to output images acquired using every kinds of digital camera. The convolution neural network (CNN) is a deep learning architecture that attracts widespread interest in real-world applications because it automatically extracts and recognizes features automatically, eliminating the need for manual feature extraction and selection techniques. Lately, Convolution Neural Networks (CNNs) are undergoing a revolution in color texture recognition applications [31-37]. Dabeer et al. [38] introduced a novel automatic technique for breast cancer detection based on CNN. Hosny et al. [39] proposed an automatic skin lesion classification approach based on transfer learning theory and a pre-trained deep neural network. Genovese et al. [40] introduced a new technique for the fusion of palmprint and inner finger texture using a single hand acquisition based on CNN. Umer et al. [41] presented a CNN technique that can identify COVID-19 patients from regular people from x-rays of the chest. Wu and Lee [42] used the CNN model for enhancing sound texture in acoustic scenes. H. Orii [43] et al. suggested a new recognition approach for tactile texture using CNN. Weskley et al. [44]

described a new approach based on CNN for classifying the degree of bread browning during baking. Hosny et al. [45] provided a precise improvement method on their approach for skin lesions classification centered on deep neural network AlexNet. They transferred learning which can accurately classify seven various types of lesions. Kassem et al. [46] proposed a novel technique using transfer learning and pre-trained deep neural network GoogleNet to classify eight various lesion types accurately. Hosny et al. [47] proposed a new approach for color texture recognition; the proposed approach has a significant benefit. It uses LBP and multi-channel moments to obtain local and global color texture features, leading to a high classification result. This success motivates us to propose a new technique for color texture classification based on the convolution neural network and local binary pattern to extract global and local features and combine them to obtain high color texture classification results.

The rest of the paper is arranged as follows. In section 2, Convolution neural network models and it’s architecture. In section 3, we introduce the local binary patterns descriptor. Section 4 presented the proposed method and simulation results shown in section 5.

2. Convolution neural network

The primary concept of why the use of CNN features is the capability of it to capture the high-level features that considered as the key point to separate the texture image. Convolution neural network (CNN) captures low detail features from the first convolution layers and then in last layers generates high detail features. Selecting the proper right set of features has always been the challenging task for machine learning algorithm. CNN has numerous extensive applications in different fields. Whereas image classification and recognition are challenging tasks for machine learning algorithms, CNN highly promising in terms of accuracy[48]. CNN generally consists of convolution, pooling, and fully connected layers, as shown in Figure 1.

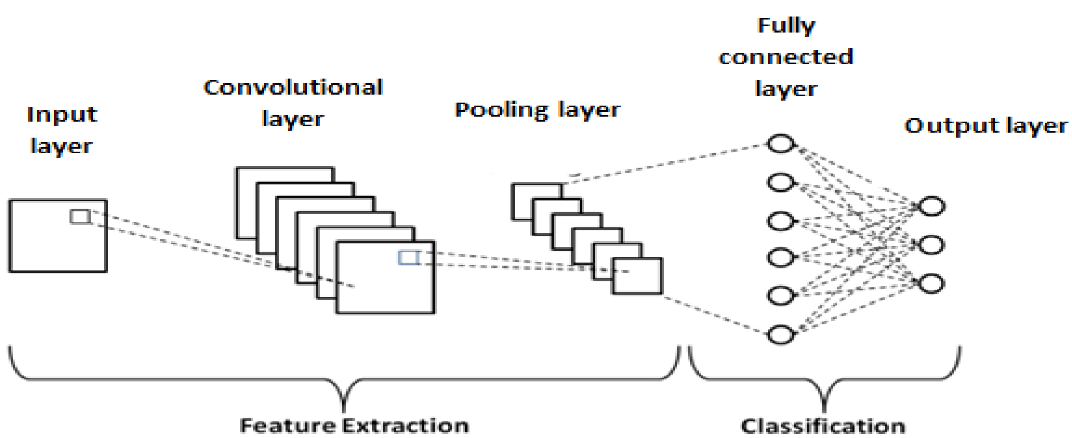

Figure 1. CNN model general architecture

A. Convolution layer

It is used to apply filters to the input feature map and produce convolved features. Two factors define the convolution layer:

- The size of the block extracted from the input feature map
- The depth of convolved features.

B. Pooling layer

Generally, applied after the convolution layer, it is a down-sampling (dimensional reduction) operation such as average pooling, min pooling, and max pooling.

C. Fully connected layer

It is typically found towards the end of the CNN architecture. It works on such a flattened input, in which each input is linked to all nodes.

Each model of CNN has its characteristics. Table 1 summarizes the description of different CNN architecture models, image input size, accuracy, and model parameters number.

Table 1: Architecture of different CNN models

CNN Model	Parameters Number	Model Description	Image Input Size
AlexNet [49]	60 M	5 Conv+3 fc layers	227 × 227
GoogleNet [50]	7 M	21 Conv+1 fc layers	224 × 224
Vgg16Net [51]	138 M	13 Conv+3 fc layers	224 × 224
Vgg19Net [52]	144 M	16 Conv+3 fc layers	224 × 224
SqueezeNet [53]	1 M	18 layers	224 × 224
ResNet-18 [54]	11 M	17 Conv+1 fc layers	224 × 224

ResNet-50 [55]	26 M	49 Conv+1 fc layers	224 × 224
ResNet-101	44 M	100 Conv+1 fc layers	224 × 224
MobileNet [56]	4 M	27 Conv+1 fc	224 × 224
Xception [57]	23 M	36 layers	229 x 229

3. Local Binary Patterns

The Local Binary Patterns [LBP] is considered as one of the most effective texture descriptors. It was introduced by Ojala et al. [18] for the first time, and it is used to represent the local features of an image which means the key points of an image. The classic LBP operator is described as a window of 3x3 pixels. The center pixel of this window is considered a threshold; if the neighboring pixel's value is less than the threshold value, the pixel value is labeled 0, else it is 1. This method will generate an 8-bit binary number converted to a decimal number, as shown in Figure 2.

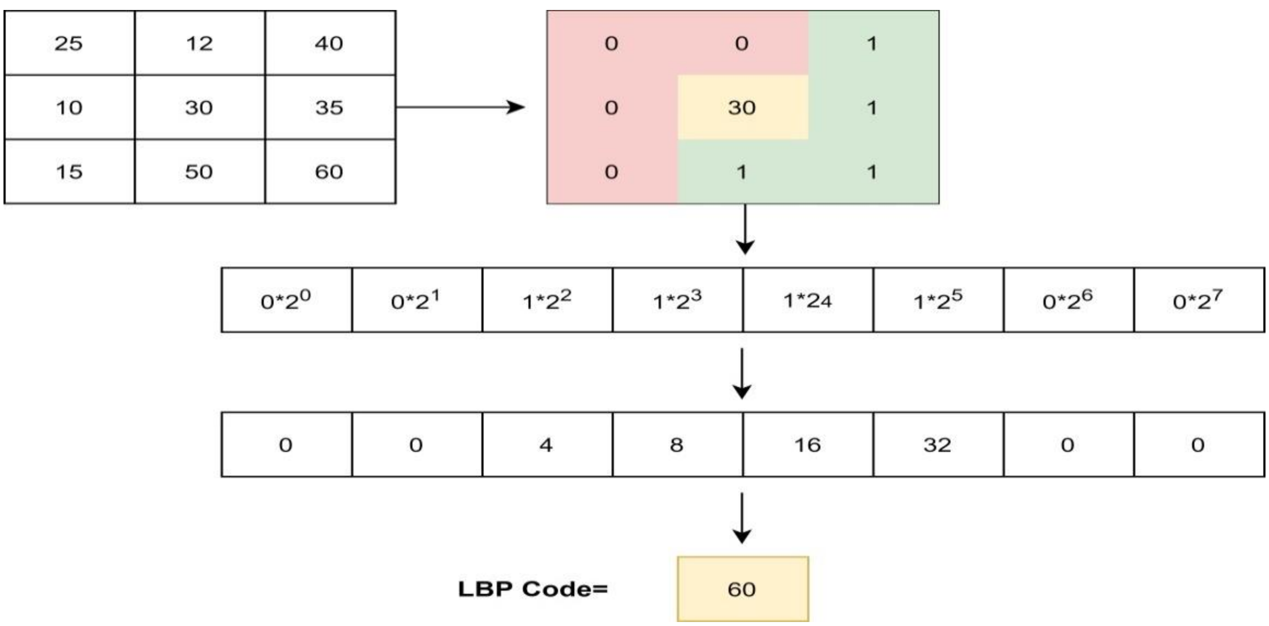

Figure 2: an example of LBP code calculation

The classic LBP descriptor can be expressed in the form below:

$$LBP_{I,J}(P_c) = \sum_{j=0}^{J-1} G(P_j - P_c)2^j, \quad (1)$$

$$\text{where} \quad G(m) = \begin{cases} 0, & \text{m} < 0 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

In Eq. (1), $P_c = P(m, n)$ is the central pixel at (m, n) location, and $P_j = P(m_j, n_j)$ is a neighboring pixel of the central pixel P_c, where

$$m_j = m + R\cos(2\pi \frac{j}{J}) \quad (3)$$

and

$$n_j = n - R\sin(2\pi \frac{j}{J}) \quad (4)$$

where J represents the neighbor pixels number, and R is the distance from P_c.

Ersahd and Tajeripour [58] proposed a new approach for color texture classification based on LBP and Kullback–Leibler divergence which used different sizes of neighborhood to achieve effective classification performance.

4. Proposed method

We proposed a novel feature extraction approach that aggregates the features extracted from Local Binary Patterns (LBP) and Convolution Neural Network (CNN) to obtain more accurate and consistent features for representation, leading to higher texture recognition accuracy results. The performance of our descriptors is assessed both individually and in combination using Local Binary Patterns (LBP). Firstly, We have fixed the image size at the pre-processing phase to be equal to the proposed image size of the pre-trained deep learning models (224×224 and 227×227), then we have made equalization of each image histogram to normalize under different illumination conditions, and then smooth it to reduce the noise introduced by the equalization. The feature extraction has been done by the CNN and LBP. The Local Binary Patterns are used to capture an image's local features representing a particular image portion. The radius of LBP neighborhoods was 1 and number of neighbors was 8. The dimensions of feature vector which extracted from CNN are [1, 59] Convolution Neural Network is used to capture global features that represent the image en block. The dimensions of feature vector which extracted from CNN are [1, 1000]. We aggregated the features of the two feature extraction approaches (CNN, LBP) in one vector. F_C= [w, x, y, z]; F_L= [l, m, n, o]; Agg Features= [F_C, F_L]. The dimensions of resultant feature vector are [1, 1059]. We find that the aggregation of features enhances efficiency and reliability over using CNN models individually. The flowchart of the proposed approach is shown in Figure 3.

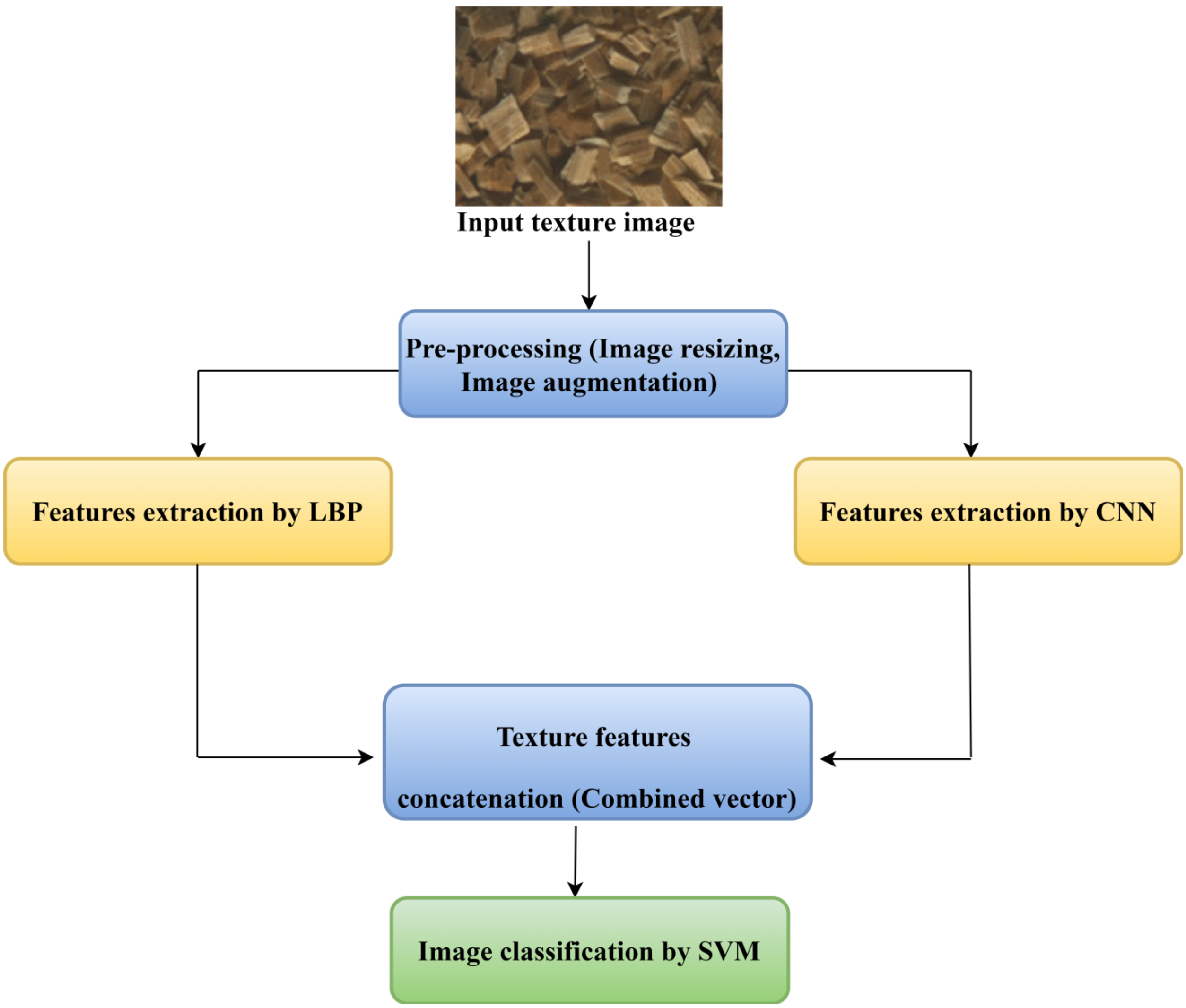

Figure 3: Proposed method flowchart

5. Experimentation and Result Discussion

5.1 Datasets

Numerical experiments carried out to validate and analyze the performance of the proposed descriptor. The experiments were executed on a Core i7 machine with 16 GB RAM. Three benchmark color textures dataset (ALOT, CBT, Outex) were used to perform the numerical experiments.

5.1.1 ALOT Dataset

ALOT dataset [59] is a series of color-image textures. The rotation angle, illumination condition is varied. More than 27.500 images are included in it. The original size of each image is 768 x 512. We centered on a fixed collection of 1500 texture images that varies from 15 classes of color textures. This collection has been proposed by [60]. Color texture samples of the ALOT database shown in Figure 4.

Figure 4: Texture samples of ALOT dataset.

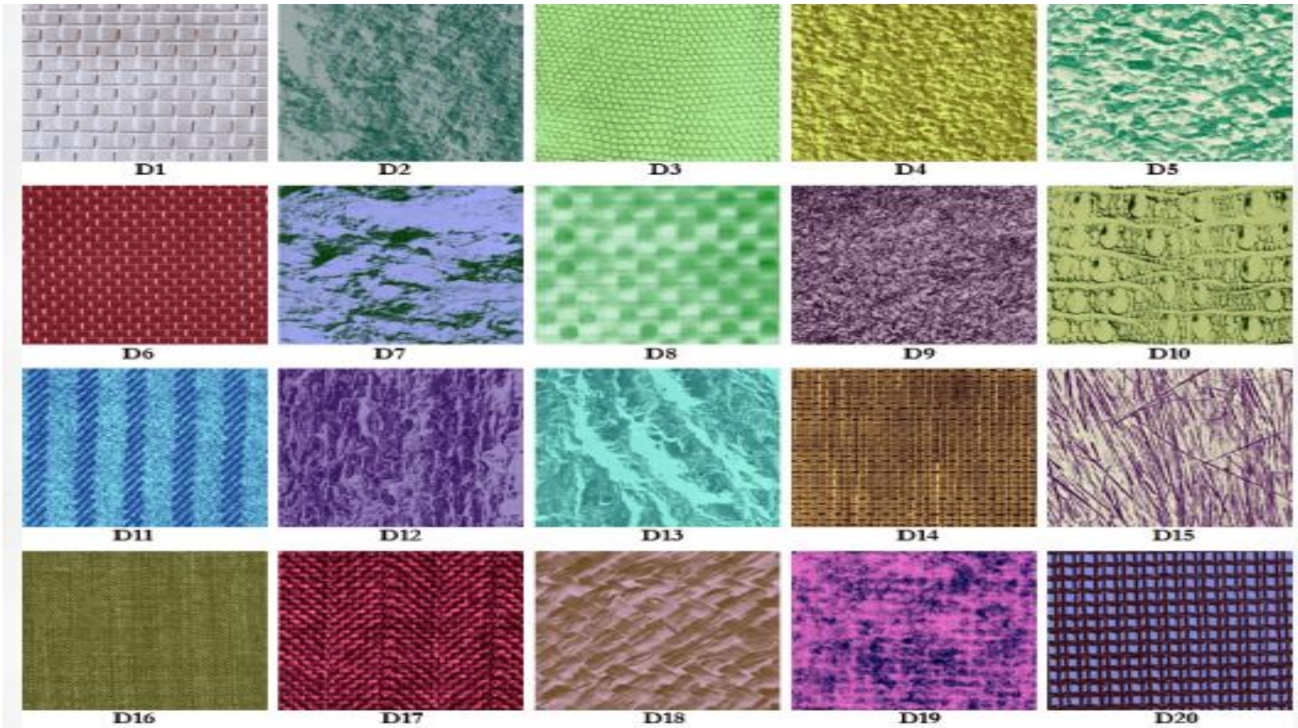

Figure 5: Texture samples of Colored Brodatz dataset (CBT).

5.1.2 Colored Brodatz Dataset

The second dataset, Colored Brodatz Texture (CBT) dataset [60], is a gray-scale-to-color extension of the original Brodatz dataset. This dataset has the advantage of keeping the original dataset's rich textural content while also providing a wide range of color content. It contains 112 color texture images that have various background intensities. The original size of each image is 512×512. We used a subset of 1500 color images from 15 classes. This subset has been proposed by Bianconi [61]. Texture samples of the CBT database shown in Figure 5.

5.1.3 Outex Dataset

The third dataset, the Outex dataset [62], is an extensive range of color textures; the range of textures is displayed in varied types of lighting, resolution, and rotated angles. It contains 320 surface color textures. The original size of each image is 746×538**.** A subgroup has been identified, containing 1500 images from 15 classes. This subgroup proposed by [60].Color texture samples of the Outex database shown in Figure 6.

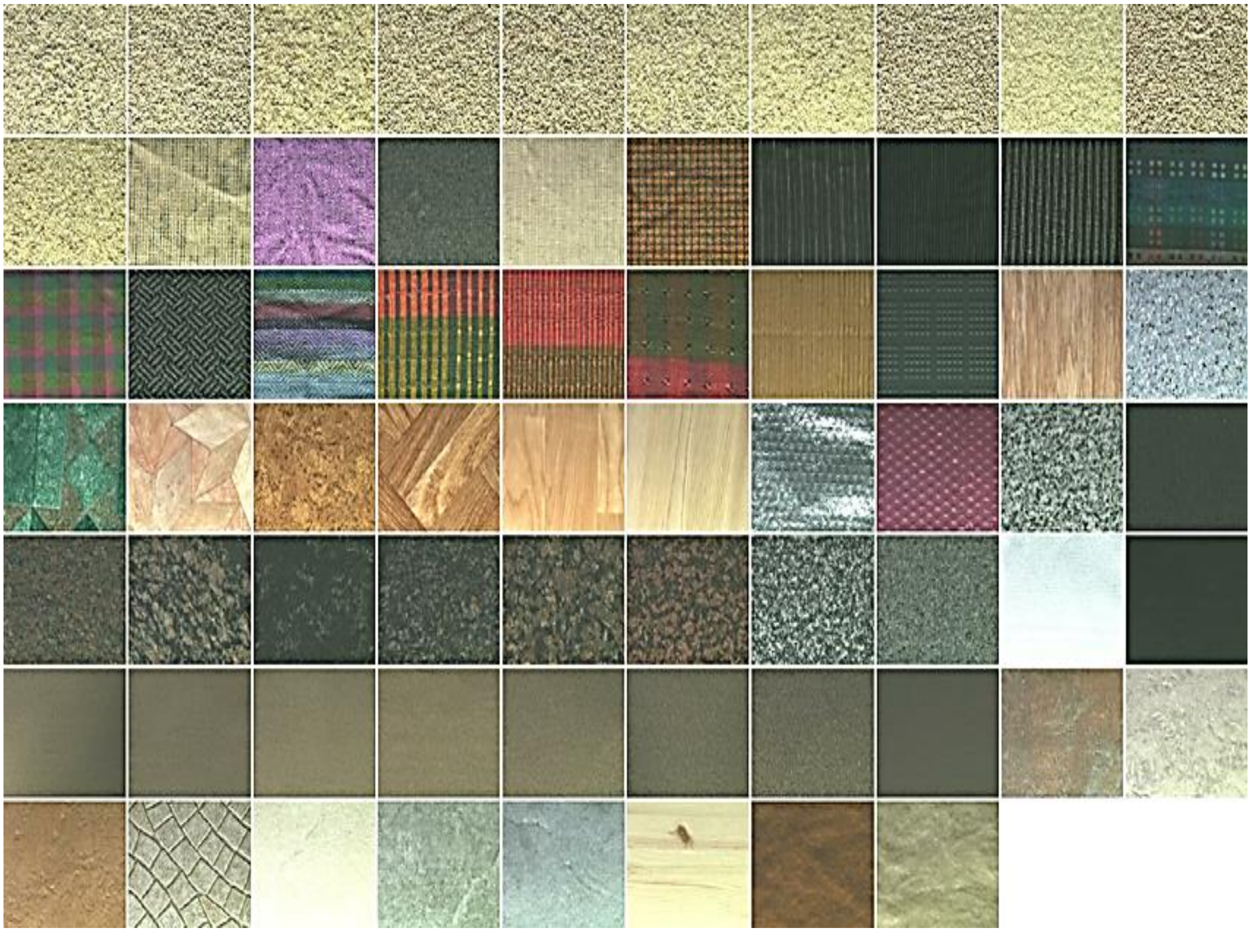

Figure 6: Texture samples of Outex dataset.

.

5.2 CNN models descriptors

To begin, we used CNN models to extract features for the grayscale and color image datasets. The number of features extracted from each image was 1000. The three color image datasets were split into training and testing sets at a ratio of 70:30, respectively in these simulations. The training set is used in classifier practice, while the testing set is used for testing, analysis and evaluation. Using Linear Support Vector Machine (SVM) classifier because of its efficiency to solve multi-class classification problems, we checked the performance of CNN models over Brodatz grayscale dataset. The results revealed that the performance of CNN has some limitations in terms of accuracy and need some enhancement. Table 2 details the results of CNN descriptor over grayscale Brodatz dataset.

Table 2: The accuracy of results obtained from CNN models and SVM Classifier over grayscale Brodatz dataset

Dataset → **Texture Descriptor**	**Brodatz dataset**

Alexnet	76.05%
Googlenet	79.25%
Vgg16Net	80.55 %
Vgg19Net	81.40%
SqueezeNet	79.95%
ResNet-18	90.12%
ResNet-50	92.50%
ResNet-101	77.20%
MobileNet	96.90%
Xception	80.00%

Then we checked the performance of CNN models over the three challenging datasets (ALOT, Outex, CBT). We have computed accuracy and confusion matrix for performing classification in all the given datasets .CNN models are trained once over these datasets. Our tests showed that their performance is good and fair but needs some way towards enhancing the accuracy results. The highest accuracy result of CNN models in the ALOT dataset was 90.89% by using Xception as a descriptor, in Outex dataset was 91.78% by using ResNet-101 as a descriptor and in CBT dataset was 86.45% by using AlexNet as a descriptor. Table 3 details the results (the best performed CNN in each dataset is indicated in bold) of CNN models using the SVM (Support Vector Machine) classifier for the three datasets (ALOT, CBT, Outex).

Table 3: The accuracy of results obtained from CNN models and SVM Classifier

Dataset ⟶ **Texture Descriptor**	**A LOT**	**CBT**	**Outex**
Alexnet	84.45%	**86.45 %**	87.70%
Googlenet	89.34%	84.23%	86.89%
Vgg16Net	88.67 %	82.67%	89.12%
Vgg19Net	88.23%	85.12 %	87.34%
SqueezeNet	78.89%	76.67%	89.78%
ResNet-18	81.12%	78.89%	80.00%
ResNet-50	90.45%	81.34%	80.45%
ResNet-101	89.56%	84.00%	**91.78%**
MobileNet	86.89%	85.56%	90.89%
Xception	**90.89%**	86.23%	90.23%

We have computed accuracy and confusion matrix for performing classification in all the given datasets. Figure 7 represents ROC curves of CNN models as descriptors using Support Vector Machine Classifier for ALOT dataset.

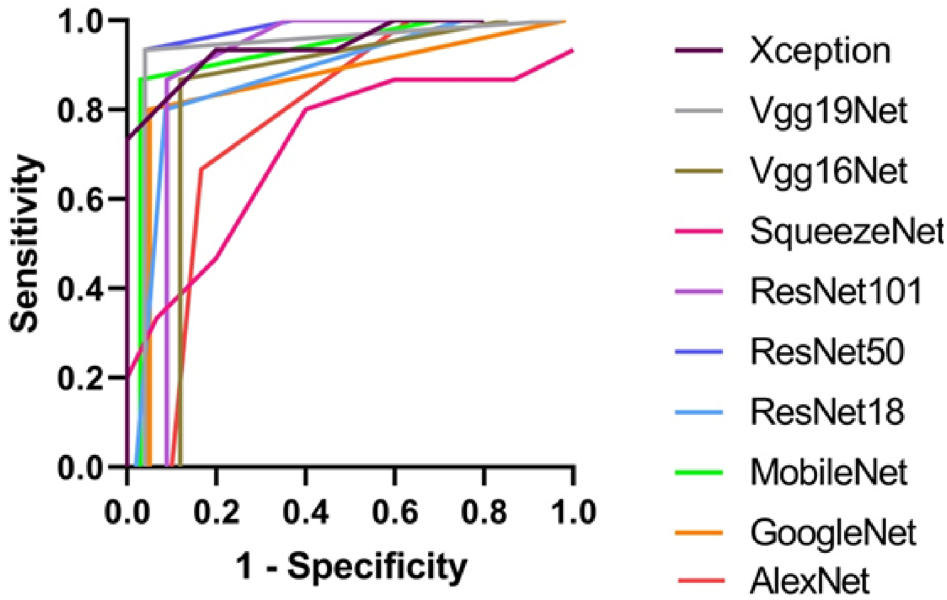

Figure 7: ROC curves of CNN models using SVM classifier for the ALOT dataset.

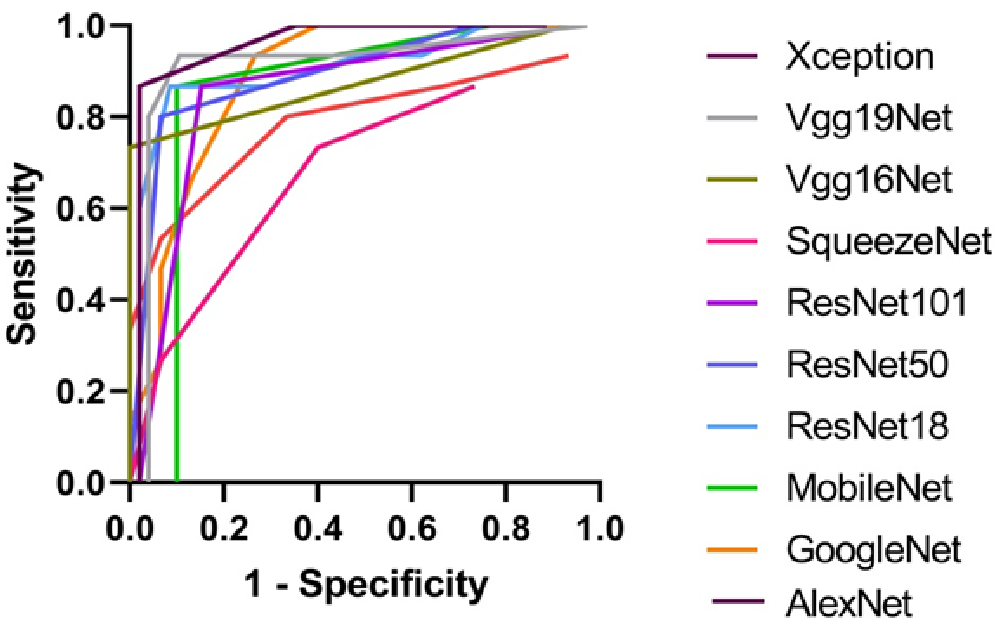

Figure 8.ROC curves of CNN models using SVM classifier for the CBT dataset.

Figure 8 and 9 represent ROC curves of CNN models as descriptors using Support Vector Machine Classifier for the CBT and Outex dataset, respectively. The ROC analysis revealed that the CNN model descriptors had a fair AUCs score (~0.8) which need to some improvement to achieve highest accuracy result. This led the authors to look forward to a local features extraction descriptor to use it with CNN model.

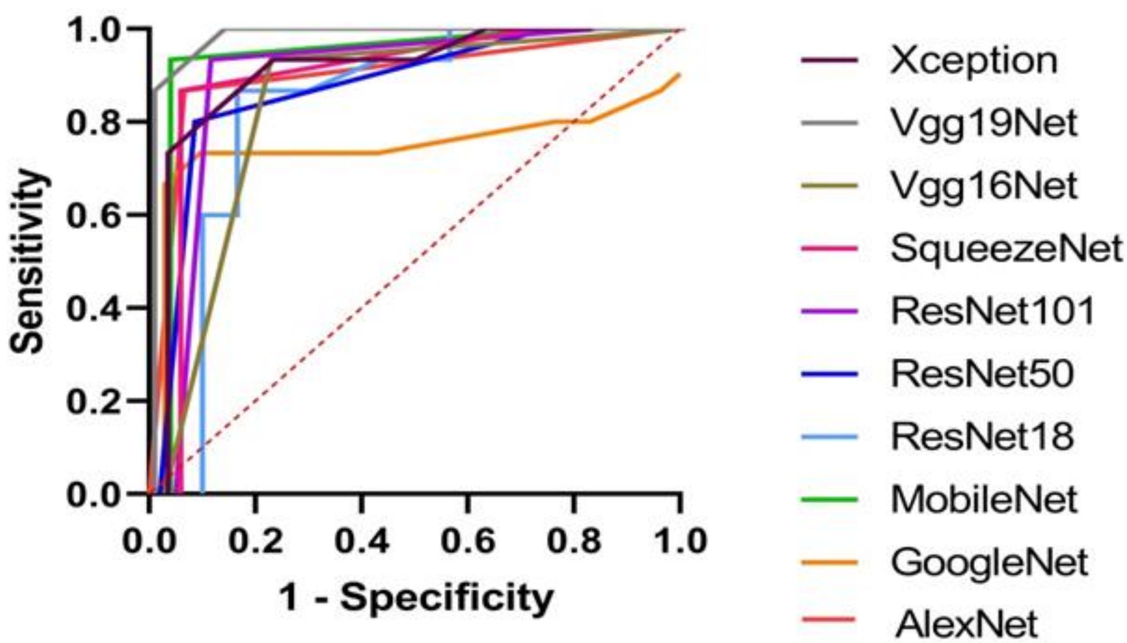

Figure 9.ROC curves of CNN models using SVM classifier for the Outex dataset.

5.2 LBP models descriptors

Secondly, we used traditional LBP descriptor to extract features for the color image datasets. The number of features extracted from each image was 59, using also the Support Vector Machine (SVM), we checked the consistency of CNN models over Brodatz dataset. The result showed that the performance needs some improvements. Table 4 shows the accuracy results of LBP using the SVM (Support Vector Machine) classifier for grayscale Brodatz dataset.

Table 4: Results of LBP using SVM classifier for the grayscale Brodatz dataset.

Dataset → **Texture Descriptor**	**Brodatz**
LBP Descriptor	78.20%

Then we checked the consistency of CNN models over the three challenging datasets (ALOT, Outex, CBT). CNN over the same datasets. Our analysis and tests showed that their performance is good and fair but needs some way towards enhancing the accuracy results. The accuracy result of LBP models in the ALOT dataset was 70.3%, in CBT dataset was 73.45% and in Outex dataset was 66.2%. Table 5 details the accuracy results of LBP using the SVM (Support Vector Machine) classifier for the three datasets (ALOT, CBT, Outex).

Table 5: Results of LBP using SVM classifier for the three dataset.

Dataset → **Texture Descriptor**	**A LOT**	**CBT**	**Outex**
LBP Descriptor	70.3%	73.45 %	66.2%

5.3 CNN+LBP descriptors

The first set of analysis confirmed that the accuracy result which obtained by using only CNN models as grayscale and color image feature descriptors have some limitation in classification accuracy and need for some improvement. Therefore, for the grayscale and color textures in the second experiment, we used an aggregated feature extraction descriptor (CNN+LBP). Convolution Neural Networks are used to capture global features that describe the image overall, while Local Binary Patterns are used to capture the local features of an image that describe a certain image part. We checked the performance of our proposed descriptor over Brodatz grayscale dataset and our result was remarkably and proved the effective performance of the descriptor over grayscale dataset. Table 6 details the proposed descriptors' results using the SVM classifier for grayscale Brodatz dataset.

Table 6: Summary results of the proposed descriptors using SVM classifier for grayscale Brodatz dataset.

Dataset → **Texture Descriptor**	**Brodatz dataset**
Alexnet+LBP	95.80%
Googlenet+LBP	95.25%
Vgg16Net+LBP	97.00 %
Vgg19Net+LBP	100.00%
SqueezeNet+LBP	98.20%
ResNet-18+LBP	94.32%
ResNet-50+LBP	99.20%
ResNet-101+LBP	94.30%
MobileNet+LBP	100.00%
Xception+LBP	97.00%

This performance encourages the authors to check descriptor performance over color datasets (ALOT, Outex, CBT). Firstly, the number of features extracted from each image was 2559 (2500 global features from the CNN model and 59 local features from LBP), we found that the increased number of features in the combinations would typically slow the machine learning stage significantly. So we have to reduce the number of features extracted from each image to 1059 (1000 global features from the CNN model and 59 local features from LBP). We have fixed the original image size at the pre-processing phase to be equal to the proposed image size of the pre-trained deep learning models (224×224 and 227×227) to be used in input layer. The three datasets are split between training and testing datasets at a ratio of 70:30, respectively. A routine test for detecting overfit of our descriptor by monitoring the loss and accuracy on the training and validation sets, our result demonstrated that our model has no overfit. In the three challenging datasets (ALOT, Outex, CBT), these tests confirm that the proposed descriptors show a clear advantage over the recognition accuracy, are effective, discriminant and outperform the existing state-of-the-art methods. Using Support Vector Machine Classifier (SVM), the highest accuracy result of our descriptors in the ALOT dataset was 100% by using Xception+LBP descriptor, in Outex dataset was 96.67% by using Xception+LBP and in CBT dataset was 100% by using ResNet-101+LBP descriptor. The lowest accuracy result of our descriptors in the ALOT dataset was 93.12% by using ResNet-18+LBP, in Outex dataset was 88.89% by using SqueezeNet+LBP and in CBT dataset was 94% by using ResNet-18+LBP.

Table 7 details the proposed descriptors' results using the SVM classifier for the A LOT dataset.

Table 7: Summary results of the proposed descriptors using SVM classifier for ALOT dataset.

Features descriptor	**RP Rate**	**WP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**accuracy**
Alexnet+LBP	0.953	0.004	0.953	0.951	0.953	0.949	0.955	0.952	95.30%
Googlenet+LBP	0.995	0.001	0.995	0.990	0.995	0.991	0.990	0.992	99.50%
Vgg16Net + LBP	0.975	0.003	0.997	0.994	0.997	0.993	1.000	0.997	97.50%
Vgg19Net + LBP	0.982	0.002	0.980	0.981	0.984	0.982	0.984	0.980	98.20%
SqueezeNet+LBP	0.948	0.005	0.949	0.945	0.946	0.948	0.948	0.946	94.89%
ResNet-18+LBP	0.931	0.006	0.932	0.930	0.931	0.932	0.929	0.928	93.12%
ResNet-50+LBP	0.998	0.002	0.996	0.996	0.998	0.997	0991	0.993	99.78%
ResNet-101+LBP	0.989	0.001	0.985	0.983	0.989	0.986	0.987	0.980	98.89%
MobileNet+LBP	0.969	0.002	0.968	0.968	0.968	0.960	1.000	0.969	96.89%
Xception+LBP	1.000	0.000	0.989	0.991	0.999	0.997	0.999	0.997	**100.00%**

Figure 10 represents ROC curves of proposed descriptors using SVM Classifier for ALOT dataset.

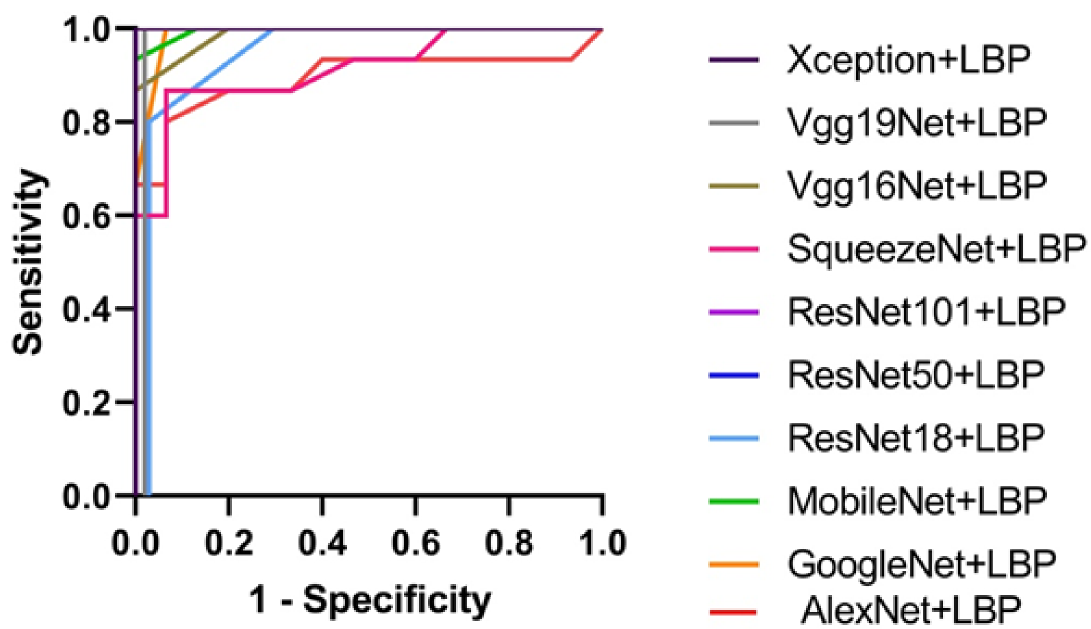

Figure 10: ROC curves of proposed descriptors using SVM classifier for the A LOT dataset.

Table 8 details precisely the results of the proposed descriptors using the SVM classifier for the CBT dataset.

Table 8: Summary results of proposed descriptors using SVM classifier for CBT dataset.

Features descriptor	RP Rate	WP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	accuracy
Alexnet+LBP	0.911	0.007	0.903	0.916	0.914	0.910	0.911	0.914	91.12%
Googlenet+LBP	0.922	0.004	0.920	0.919	0.920	0.922	0.922	0.923	92.23%
Vgg16Net + LBP	0.926	0.003	0.923	0.920	0.925	0.926	0.949	0.946	92.67%
Vgg19Net + LBP	0.933	0.004	0.945	0.932	0.933	0.931	0.996	0.945	93.34%
SqueezeNet+LBP	0.888	0.009	0.882	0.882	0.882	0.873	0.948	0.876	88.89%
ResNet-18+LBP	0.906	0.006	0.889	0.904	0.901	0.895	0.968	0.917	90.67%
ResNet-50+LBP	0.998	0.002	0.996	0.996	0.997	0.997	0991	0.993	92.45%
ResNet-101+LBP	0.945	0.004	0.942	0.945	0.943	0.939	0.997	0.945	94.45%
MobileNet+LBP	0.960	0.002	0.961	0.961	0.960	0.960	0.998	0.960	96.00%
Xception+LBP	0.966	0.001	0.965	0.960	0.962	0.962	0.961	0.966	**96.67%**

Figure 11 represents ROC curves of proposed descriptors using SVM classifier for CBT dataset.

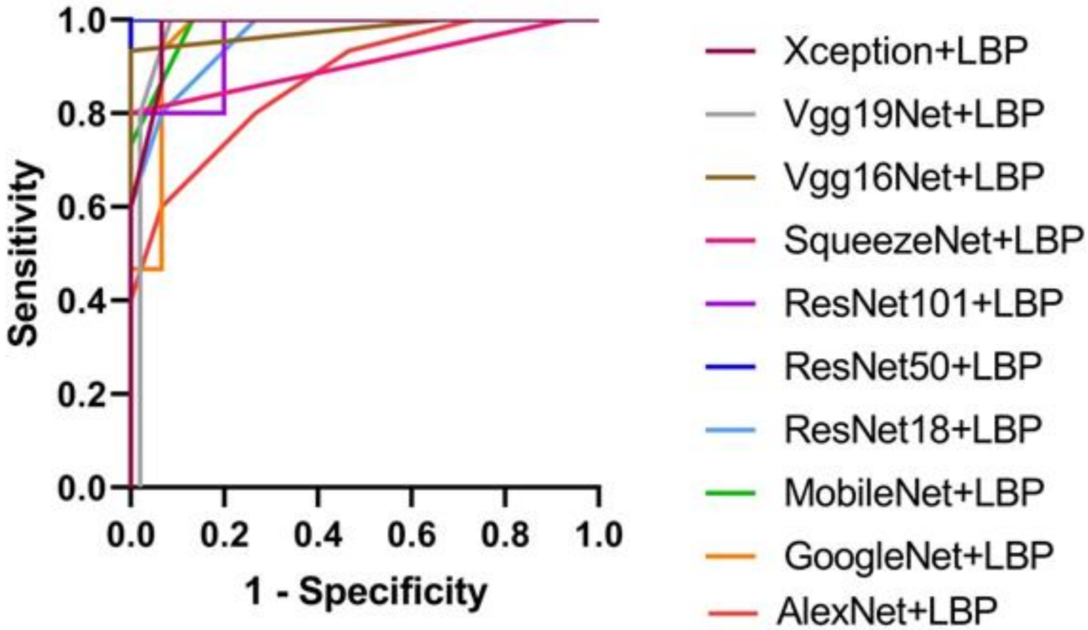

Figure 11: ROC curves of proposed descriptors using SVM classifier for the CBT dataset.

Table 9 details precisely the results of the proposed descriptors using the SVM classifier for the Outex dataset.

Table 9: Summary results of proposed descriptors using SVM classifier for Outex dataset.

Features descriptor	RP Rate	WP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	accuracy
Alexnet+LBP	0.977	0.001	0.975	0.977	0.975	0.972	.0979	0.973	97.76%
Googlenet+LBP	0.988	0.0003	0.985	0.988	0.985	0.981	0.983	0.956	98.80%
Vgg16Net+LBP	0.993	0.0001	0.990	0.993	0.990	0.989	0.990	0.985	99.34%
Vgg19Net+LBP	0.997	0.001	0.998	0.997	0.997	0.997	1.000	1.000	99.78%
SqueezeNet+LBP	0.973	0.002	0.972	0.973	0.974	0.968	0.968	0.966	97.34%
ResNet-18+LBP	0.940	0.005	0.932	0.930	0.933	0.935	0.930	0.928	94.00%
ResNet-50+LBP	0.977	0.001	0.976	0.978	0.976	0.974	0.975	0.973	97.78%
ResNet-101+LBP	1.000	0.000	0.998	0.993	0.999	0.996	1.00	0.998	**100.00%**
MobileNet+LBP	0.993	0.002	0.992	0.989	0.987	0.990	0.993	0.988	99.34%
Xception+LBP	0.997	0.001	0.995	0.991	0.999	0.997	0.999	0.997	99.78%

Figure 12 represents ROC curves of proposed descriptors using Support Vector Machine Classifier for Outex dataset.

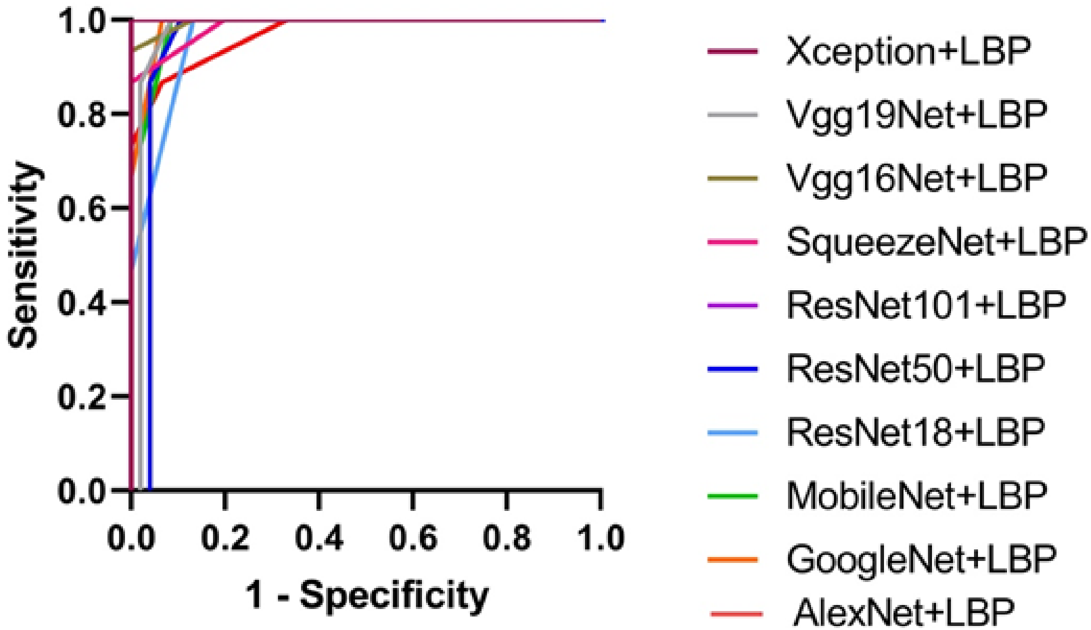

Figure 12: ROC curves of proposed descriptors using SVM classifier for the Outex dataset.

Table 10 is revealing the high accuracy result obtained by the proposed method for the three datasets.

Table 10: The accuracy of results obtained from proposed descriptors and SVM Classifier

Dataset → **Texture Descriptor**	**A LOT**	**CBT**	**Outex**
Alexnet + LBP	95.34%	91.12%	97.70%
Googlenet + LBP	99.56%	92.23%	98.80%
Vgg16Net+ LBP	97.56 %	92.67%	99.30%
Vgg19Net+ LBP	98.20%	93.34%	99.70%
SqueezeNet+LBP	94.89%	88.89%	97.34%
ResNet-18+LBP	93.12%	90.67%	94.00%
ResNet-50+LBP	99.78%	92.45%	97.78%
ResNet-101+LBP	98.89%	94.45%	**100.00%**
MobileNet+LBP	96.89%	96.00%	99.34%
Xception+LBP	**100.00%**	**96.67%**	99.78%

From the above given table 9, table 10 , it is clear that pretrained model Xception + LBP, Resnet + LBP and Xception+LBP stands out in the performance when tested with ALOT, Outex and CBT datasets, respectively. The result demonstrated that ResNet-18+LBP has the lowest accuracy rate in two datasets (A LOT and Outex) and also SqueezeNet+LBP in CBT dataset.

Figure 13 provided a comparison between our descriptor's promising accuracy result and the accuracy result obtained by CNN models for the ALOT dataset. Figure 14 provided a comparison between the promising accuracy result obtained by our descriptor and the accuracy result obtained by CNN models for the CBT dataset. Figure 15 provided a comparison between our descriptor's promising accuracy result and the accuracy result obtained by CNN models for the Outex dataset.

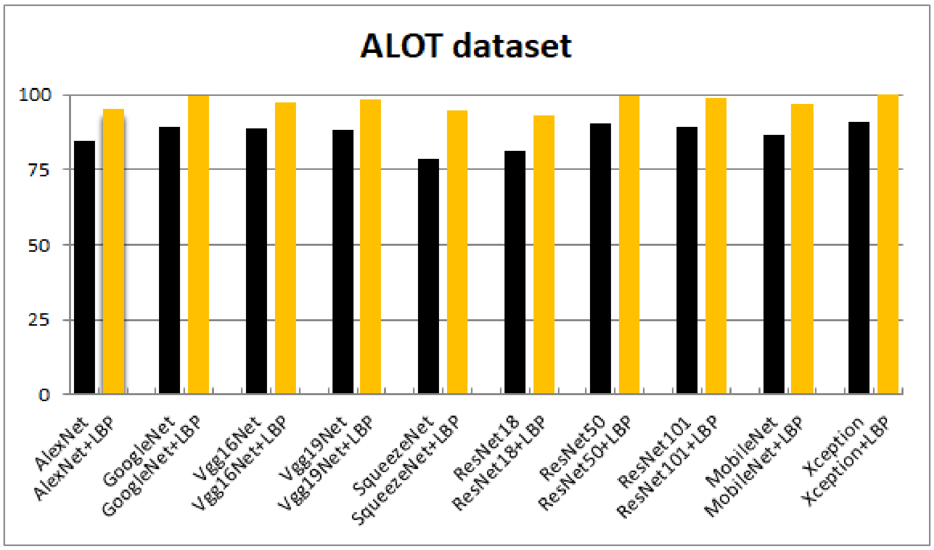

Figure 13: CNN vs CNN+LBP in terms of accuracy for the ALOT dataset.

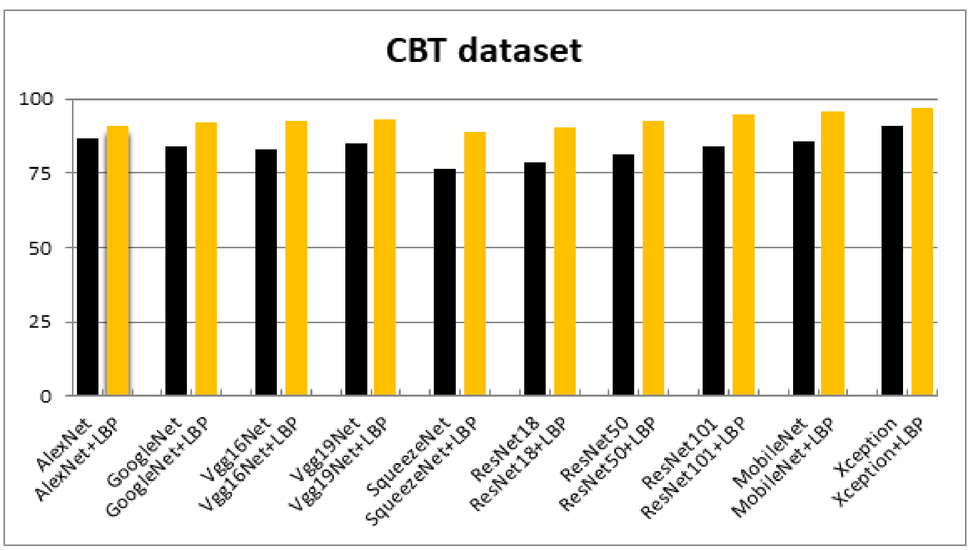

Figure 14: CNN vs CNN+LBP in terms of accuracy for the CBT dataset.

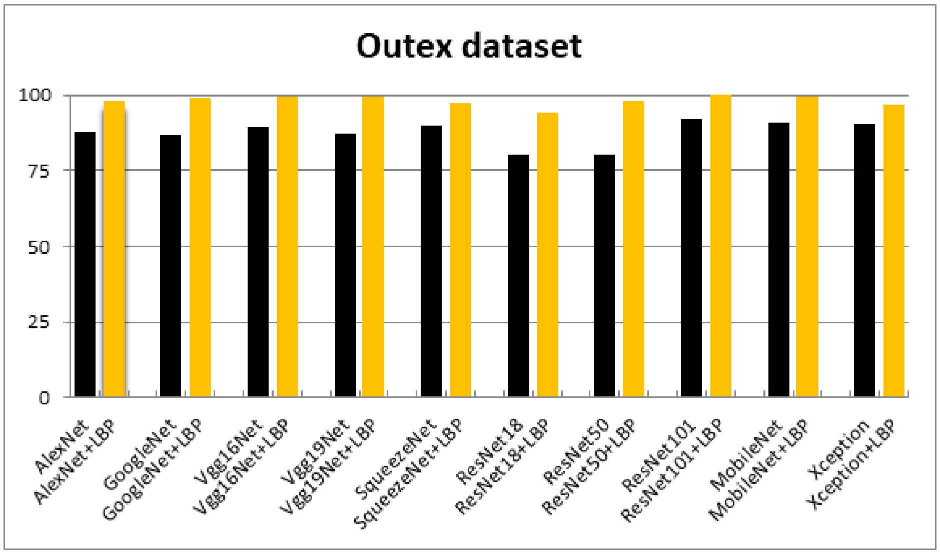

Figure 15: CNN vs CNN+LBP in terms of accuracy for the Outex dataset.

From previous results, we can observe that the accuracy increased when we used LBP+CNN as a descriptor for the three datasets, and to evaluate its performance, we presented a comparison between it and other states of the art methods that introduced by [63] and by [60]in Table 11.

Texture Descriptor	Datasets	
	Outex	A LOT
Legendre-c.f GLCM	83.2%	66.2%
Discr. Cheb. GLCM	84.9%	64.4%
LBP-RI	74.6%	70.6%
MLBP	93.6	71.4%
GWN	78.4%	88%
Alexnet + LBP	95.34%	97.70%
Googlenet + LBP	99.56%	98.80%
Vgg16Net+ LBP	97.56 %	99.30%
Vgg19Net+ LBP	98.20%	99.70%
SqueezeNet+LBP	94.89%	97.34%

ResNet-18+LBP	93.12%	94.00%
ResNet-50+LBP	99.78%	97.78%
ResNet-101+LBP	98.89%	**100.00%**
MobileNet+LBP	96.89%	99.34%
Xception+LBP	**100.00%**	99.78%

Table 11. Comparison between results of LBP+CNN and other descriptors for Outex and ALOT dataset.

Then we have evaluated our proposed descriptors performance when provided it with limited training set of texture, to detect if it's a robust approach or not. Table 12 is revealing the accuracy result obtained by the proposed method for the three datasets by providing only 20% for training set.

Dataset → Texture Descriptor	**ALOT**	**CBT**	**Outex**
Alexnet + LBP	65.12%	61.35%	70.70%
Googlenet + LBP	79.56%	72.27%	78.08%
Vgg16Net+ LBP	77.56 %	72.67%	71.30%
Vgg19Net+ LBP	78.00%	60.02%	66.70%
SqueezeNet+LBP	54.95%	68.82%	58.14%
ResNet-18+LBP	63.20%	65.25%	74.50%
ResNet-50+LBP	80.08%	72.25%	78.08%
ResNet-101+LBP	78.20%	74.00%	79.25%
MobileNet+LBP	66.42%	72.10%	73.48%
Xception+LBP	79.98%	76.70%	59.18%

Table 12. The accuracy of results obtained from proposed descriptors and SVM Classifier by providing 20% for training set

Table 13 is revealing the accuracy result obtained by the proposed method for the three datasets by providing only 50% for training set.

Dataset → Texture Descriptor	**ALOT**	**CBT**	**Outex**
Alexnet + LBP	74.34%	81.12%	80.20%
Googlenet + LBP	75.20%	71.43%	81.42%
Vgg16Net+ LBP	77.65 %	74.72%	79.50%

Vgg19Net+ LBP	80.20%	72.22%	83.60%
SqueezeNet+LBP	74.90%	69.48%	77.42%
ResNet-18+LBP	73.02%	71.70%	82.00%
ResNet-50+LBP	81.82%	62.22%	71.82%
ResNet-101+LBP	69.20%	76.50%	84.21%
MobileNet+LBP	70.99%	84.24%	71.25%
Xception+LBP	80.40%	85.40%	80.82%

Table 13. The accuracy of results obtained from proposed descriptors and SVM Classifier by providing 20% for training set

6. Conclusion

Overall, in this paper, a new feature extraction approach was presented for color texture recognition techniques that aggregate features extracted from Local Binary Pattern and Convolution neural network. features combinations provide more reliable, consistent performance. Our studies demonstrated that suggested descriptors have some good features such as stability and robustness against different orientations and remarkably accurate classification results in recognition task. The experiments were performed on three benchmark color datasets (ALOT, CBT, Outex) using Support Vector Machine as a classifier. We have obtained results demonstrating the efficiency, robustness, and high accuracy of the proposed classification technique against image transformation such as scaling, rotation, and translation. By analyzing classification and recognition accuracy, the findings are promising and encouraging.

Acknowledgment

This work was supported by the MPhil program "Advanced Technologies in Informatics and Computers", hosted by the Department of Computer Science, International Hellenic University, Kavala, Greece

References

[1] Jiaohua Qin, Wenyan Pan, Xuyu Xiang, Yun Tan, and Guimin Hou, "A biological image classification method based on improved CNN," Ecological Informatics, Volume 58, 2020.

[2] Z. Lai, H. Deng, "Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron," Comput Intell Neurosci, 2018.

[3] Pedro P. Rebouças Filho, Elizângela de S. Rebouças, Leandro B. Marinho, Róger M. Sarmento, João Manuel R.S.Tavares, Victor Hugo C. de Albuquerque, “Analysis of human tissue densities: A new approach to extract features from medical images” ,Pattern Recognition Letters, Volume 94,pp. 211-218, 2017.

[4] Doyle S., Feldman M., Tomaszewski J., and Madabhushi A., "A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," IEEE Trans. Biomed. Eng., vol. 59, no. 5, pp. 1205–1218, 2012.

[5] S. Hicks et al., "Dissecting Deep Neural Networks for Better Medical Image Classification and Classification Understanding," 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), pp. 363-368, 2018.

[6] Anil K. Jain, Nalini K. Ratha, Sridhar Lakshmanan, "Object detection using Gabor filters," Pattern Recognition, Vol 30, Issue 2, Pages 295-309, 1997.

[7] Qing Zhang, Jiajun Lin, Yanyun Tao, Wenju Li, Yanjiao Shi," Salient object detection via color and texture cues," Neurocomputing, Vol 243, pp. 35-48, 2017.

[8] P. Barmpoutis, K. Dimitropoulos, and N. Grammalidis, "Smoke detection using Spatio-temporal analysis, motion modeling and dynamic texture recognition," 2014 2nd European Signal Processing Conference (EUSIPCO), Lisbon, pp. 1078-1082, 2014.

[9] Li, Jiang et al. "Texture Analysis of Remote Sensing Imagery with Clustering and Bayesian Inference." International Journal of Image, Graphics and Signal Processing 7, 2015.

[10] Hemalatha, S., & Anouncia, S. M., "Unsupervised Segmentation of Remote Sensing Images using FD Based Texture Analysis Model and ISODATA." International Journal of Ambient Computing and Intelligence (IJACI), 8(3), 58-75, 2017.

[11] Frédéric Borne and Gaëlle Viennois, "Texture-based classification for characterizing regions on remote sensing images," Journal of Applied Remote Sensing, 11(3), 2017.

[12] D. Lowe, “Distinctive Image Features from Scale Invariant Key Points,” Int’l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[13] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[14] J. Chen et al., "WLD: A Robust Local Image Descriptor," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1705-1720, Sept. 2010.

[15] B. Manjunath and W. Ma, “Texture Features for Browsing and Retrieval of Image Data,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 8, pp. 837-842, Aug. 1996.

[16] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel NonStatistical Model for Face Representation and Recognition,” Proc. IEEE Int’l Conf. Computer Vision, 2005.

[17] Ershad, Shervan Fekri. “Color Texture Classification Approach Based on Combination of Primitive Pattern Units and Statistical Features.” *ArXiv* abs/1109.1133 (2011): n. pag.

[18] Timo Ojala, Matti Pietikäinen, David Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, Volume 29, Issue 1, pp. 51-59, 1996.

[19] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?" Pattern Recognition Letters, vol. 37, no. 8, pp. 1629–1640, 2004.

[20] G.A. Papakostas, D.E. Koulouriotis, E.G. Karakasis, V.D. Tourassis, "Moment-based local binary patterns: A novel descriptor for invariant pattern recognition applications," Neurocomputing, 99, pp. 358-371, 2013.

[21] Yuan LUO, Cai-ming WU, Yi ZHANG, "Facial expression feature extraction using hybrid PCA and LBP," The Journal of China Universities of Posts and Telecommunications, Vol 20, Issue 2, pp. 120-124, 2013.

[22] Yuchun Fang and Zhan Wang, "Improving LBP features for gender classification," 2008 International Conference on Wavelet Analysis and Pattern Recognition, pp.373-377, 2013.

[23] Caifeng Shan, "Learning local binary patterns for gender classification on real-world face images," Pattern Recognition Letters, Volume 33, Issue 4, pp. 431-437, 2012.

[24] S. Abidin, R. Togneri and F. Sohel, "Enhanced LBP texture features from time-frequency representations for acoustic scene classification," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.626-630, 2017.

[25] Giveki, D., Karami M., "Scene classification using a new radial basis function classifier and integrated SIFT–LBP features," Pattern Anal Applic 23, pp. 1071–1084, 2020.

[26] P. Král and L. Lenc, "LBP features for breast cancer detection," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, pp. 2643-2647,2016.

[27] E. Badeka, C. I. Papadopoulou and G. A. Papakostas, "Evaluation of LBP Variants in Retinal Blood Vessels Segmentation Using Machine Learning," 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020, pp. 1-7.

[28] Daniel O. Tambasco Bruno, Marcelo Z. do Nascimento, Rodrigo P. Ramos, Valério R. Batista, Leandro A. Neves, Alessandro S. Martins, "LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues," Expert Systems with Applications, Vol 55, pp. 329-340, 2016.

[29] Liangliang Shi, Xia Wang, Yongliang Shen, "Research on 3D face recognition method based on LBP and SVM", Optik, Vol 220, 2020.

[30] Ershad, Shervan Fekri and Farshad Tajeripour. "Color Texture Classification Based on Proposed Impulse-Noise Resistant Color Local Binary Patterns and Significant Points Selection Algorithm." *ArXiv* abs/1906.11010 (2019): n. pag.

[31] J. Chen et al., "WLD: A Robust Local Image Descriptor," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1705-1720, Sept. 2010.

[32] Bianco, Simone, Cusano, Claudio, Napoletano, Paolo, Schettini, Raimondo, "Improving CNN-Based Texture Classification by Color Balancing" J. Imaging, 3, pp. 3- 33, 2017.

[33] Cimpoi, M., Maji, S., Kokkinos, I. et al., "Deep Filter Banks for Texture Recognition, Description, and Segmentation," Int. J. Comput Vis 118, pp. 65–94, 2016.

[34] K. S. Loke, "Texture Recognition Using a Novel Input Layer for Deep Convolutional Neural Network," 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS), Singapore, Singapore, pp. 14-17, 2018.

[35] D. P. Van Hoai and V. T. Hoang, "Feeding Convolutional Neural Network by hand-crafted features based on Enhanced Neighbor-Center Different Image for color texture classification," 2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Ho Chi Minh City, Vietnam, pp. 1-6, 2019.

[36] Saikat Basu, Supratik Mukhopadhyay, Manohar Karki, Robert DiBiano, Sangram Ganguly, Ramakrishna Nemani, Shreekant Gayaka, "Deep neural networks for texture classification—A theoretical analysis," Neural Networks, Volume 97,2018.

[37] Wang TC., Zhu JY., Hiroaki E., Chandraker M., Efros A.A., Ramamoorthi R., "A 4D Light-Field Dataset and CNN Architectures for Material Recognition", Computer Vision – ECCV 2016, vol. 9907, 2016.

[38] S. Dabeer, M. M. Khan, S. Islam, "Cancer diagnosis in histopathological image: CNN based approach, Informatics in Medicine Unlock, vol 16, 2019.

[39] Hosny K.M., Kassem M.A., Fouad MM, "Classification of skin lesions using transfer learning and augmentation with Alex-net", PLOS ONE 14(5), 2019.

[40] Genovese, V. Piuri, F. Scotti and S. Vishwakarma, "Touchless Palmprint and Finger Texture Recognition: A Deep Learning Fusion Approach," 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 1-6, 2019.

[41] Umer, M., Ashraf, I., Ullah, S. *et al.*, "COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images", J Ambient Intell Human Comput, 2021.

[42] Y. Wu and T. Lee, "Enhancing Sound Texture in CNN-based Acoustic Scene Classification," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 815-819, 2019.

[43] H. Orii, S. Tsuji, T. Kouda and T. Kohama, "Tactile texture recognition using convolutional neural networks for time-series data of pressure and 6-axis acceleration sensor," 2017 IEEE International Conference on Industrial Technology (ICIT), pp. 1076-1080, 2017.

[44] Weskley da Silva Cotrim, Valéria Paula Rodrigues Minim, Leonardo Bonato Felix, Luis Antonio Minim, "Short convolutional neural networks applied to the recognition of the browning stages of bread crust," Journal of Food Engineering, Volume 277,2020.

[45] Hosny, K.M., Kassem, M.A. and Fouad, M.M., "Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet". *J Digit Imaging* 33,pp. 1325–1334, 2020.

[46] M. A. Kassem, K. M. Hosny and M. M. Fouad, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access*, vol. 8, pp. 114822-114832, 2020.

[47] Hosny, K.M., Magdy, T. & Lashin, N.A., "Improved color texture recognition using multi-channel orthogonal moments and local binary pattern", *Multimed Tools Appl*, 2021.

[48] Krizhevsky, Alex, Sutskever, Ilya and Hinton, Geoffrey E.. "ImageNet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems 25* , edited by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger , ,2012.

[49] R. Yoshihashi, R. Kawakami, M. Iida, T. Naemura, “Bird detection and species classification with time-lapse images around a wind farm: Dataset construction and evaluation,” Wind Energy, 20(12), 1983–1995, 2017.

[50] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[51] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[52] Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556.

[53] Iandola, Forrest & Han, Song & Moskewicz, Matthew & Ashraf, Khalid & Dally, William & Keutzer, Kurt., “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size”, 2016.

[54] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition”,2015.

[55] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[56] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, 2017.

[57] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, 2017

[58] S. ekri-Ershad and F. Tajeripour, “mpulse-Noise Resistant Color-Texture Classification Approach Using Hybrid Color Local Binary Patterns and Kullback–Leibler Divergence”, The Computer Journal, 2017.

[59] http://color.univ-lille.fr/datasets/alot.

[60] F. Bianconi, A.Fernandez, "Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform", Pattern Reognit. Lett. 48, 2014

[61] http://multibandtexture.recherche.usherbrooke.ca/colored%20_brodatz.html.

[62] https://computervisiononline.com/dataset/1105138685.

[63] C. Di Ruberto, L. Putzu and G. Rodriguez, “Fast and accurate computation of orthogonal moments for texture analysis”, Pattern Recognition, Volume 83, 2018.