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. Papakostas3

1Information Technology Department, Zagazig University, Zagazig 44519, Egypt
2Computer Science Department, Sinai University, North Sinai, Arish, Egypt
3MLV Research Group, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece

Correspondence should be addressed to Khalid M. Hosny; k_hosny@yahoo.com and George A. Papakostas; gpapak@teikav.edu.gr

Received 27 May 2021; Revised 20 October 2021; Accepted 23 October 2021; Published 16 November 2021

Academic Editor: Muhammad Haroon Yousaf

Representation and classification of color texture generate considerable interest within the field of computer vision. Texture classification is a difficult task that assigns unlabeled images or textures to the correct labeled class. Some 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. 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 that describe the image’s key points (image patches). Our analysis shows that using LBP improves the classification task when compared to using CNN only. We test the proposed approach experimentally over three challenging color image datasets (ALOT, CBT, and Outex). The results demonstrated that our approach improved up to 25% in the classification 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 has shown enhancement in classification and recognition compared to the state-of-the-art method.

1. Introduction

The motivation of the usage of CNN models as features descriptors is the ability of the deep neural network to capture the high-level features that can be a key point for the 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 extracting 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 as a critical issue in computer vision that has many applications such as biomedical image processing [1–5], object detection [6–8], and remote sensing [9–11] application fields.

The scale-invariant feature transform (SIFT), proposed by Lowe [12], is a common sparse descriptor. Mikolajczyk and Schmid changed the SIFT descriptor by altering the gradient location orientation grid and the quantization parameters of the histograms [13]. Chen et al. [14] proposed a robust discriminative descriptor called Weber local descriptor (WLD) based on human perception. Gabor wavelet descriptor is considered one of the most widely used dense descriptors [15]. The Gabor wavelet has extensively used image recognition applications such as face recognition, scene analysis, motion tracking, and face recognition [16]. Ershad and Fekri [17] proposed a new approach to analyze the texture based on its grain components and classify it from grain components.
histogram and statistical features. Local binary pattern (LBP) is 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 multiresolution version of LBP called HCLBP and proposed an approach that adapts to output images acquired using every kind 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 [14, 31–36]. Dabeer et al. [37] introduced a novel automatic technique for breast cancer detection based on CNN. Hosny et al. [38] proposed an automatic skin lesion classification approach based on transfer learning theory and a pretrained deep neural network. Genovese et al. [39] introduced a new technique for fusion palm print and inner finger texture using a single hand acquisition based on CNN. Umer et al. [40] presented a CNN technique that can identify COVID-19 patients from regular people from chest X-rays. Wu and Lee [41] used the CNN model for enhancing sound texture in acoustic scenes. Ori et al. [42] suggested a new recognition approach for tactile texture using CNN. Wesley et al. [43] described a new approach based on CNN for classifying the degree of bread browning during baking. Hosny et al. [44] provided a precise improvement method on their approach for skin lesions classification centered on deep neural network AlexNet. They used transfer learning which can accurately classify seven various types of lesions. Kassem et al. [45] proposed a novel technique using transfer learning and pretrained deep neural network, GoogLeNet, to accurately classify eight various lesion types. Hosny et al. [46] proposed a new approach for color texture recognition; the proposed approach has a significant benefit. It uses LBP and multichannel 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 its architecture are presented. In Section 3, we introduce the local binary patterns descriptor. Section 4 presents the proposed method and simulation results are shown in Section 5.

2. Convolution Neural Network

The primary concept of using CNN features is its capability to capture the high-level features that are considered the key point to separate the texture image. Convolution neural network (CNN) captures low detail features from the first convolution layers and then generates high detail features in the last layers. Selecting the proper right set of features has always been a challenging task for the machine learning algorithm. CNN has numerous extensive applications in different fields. Whereas image classification and recognition are challenging tasks for machine learning algorithms, CNN is highly promising for accuracy [47]. CNN generally consists of convolution, pooling, and fully connected layers, as shown in Figure 1.

2.1. Convolution Layer. It is used to apply filters to the input feature map and produce convolved features. The following two factors define the convolution layer.

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

2.2. Pooling Layer. Generally, applied after the convolution layer, it is a downsampling (dimensional reduction) operation such as average pooling, min pooling, and max pooling.

2.3. 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.

3. Local Binary Patterns

The local binary patterns (LBP) are considered 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 3 × 3 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.

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

\[
LBP_{I,j}(P_c) = \sum_{j=0}^{l-1} G(P_j - P_c)2^j, \tag{1}
\]

where

\[
G(m) = \begin{cases} 
0, & m < 0, \\
1, & \text{otherwise}. 
\end{cases} \tag{2}
\]

In (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
\[ m_j = m + R \cos \left( \frac{2\pi j}{J} \right), \]
\[ n_j = n - R \sin \left( \frac{2\pi j}{J} \right). \]

where \( J \) represents the neighbor pixels number and \( R \) is the distance from \( P_c \).

Ershad and Tajeripour [56] proposed a new approach for color texture classification based on LBP and Kullback–Leibler divergence, which used different neighborhood sizes to achieve effective classification performance.

4. The 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 with local binary patterns (LBP). Firstly, we have fixed the image size at the preprocessing phase to be equal to the proposed image size of the pretrained deep learning models (224 × 224 and 227 × 227). 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 CNN and LBP have done the feature extraction. 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 the number of neighbors was 8. The dimensions of the feature vector extracted from CNN are [1,57]. Convolution neural network is used to capture global features that represent the image en bloc. The dimensions of the feature vector 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 the 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.

5. Experimentation and Result Discussion

5.1. Datasets.

Numerical experiments were 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 datasets (ALOT, CBT, and Outex) were used to perform the numerical experiments.

5.1.1. ALOT Dataset. ALOT dataset [57] is a series of color image textures. The rotation angle and illumination conditions are varied. More than 27,500 images are included in it. The original size of each image is 768 × 512. We focused on a fixed collection of 1500 texture images that varies from 15 classes of color textures. This collection has been proposed by [58]. Color texture samples of the ALOT database are shown in Figure 4.
5.1.2. Colored Brodatz Dataset. The second dataset, colored Brodatz texture (CBT) dataset [58], is a grayscale-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 [59]. Texture samples of the CBT database are shown in Figure 5.

5.1.3. Outex Dataset. The third dataset, the Outex dataset [60], 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 $746 \times 538$. A subgroup has been identified, containing 1500 images from 15 classes. This subgroup was proposed by [58]. Color texture samples of the Outex database are shown in Figure 6.

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 the linear support vector machine classifier because of its efficiency in solving multiclass classification problems, we checked the performance of CNN models over the Brodatz grayscale dataset. The results revealed that the performance of CNN has some limitations.
in terms of accuracy and needs some enhancement. Table 2 details the results of the CNN descriptor over the grayscale Brodatz dataset.

Then we checked the performance of CNN models over the three challenging datasets (ALOT, Outex, and CBT). We have computed an 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 excellent and fair but needs to enhance the accuracy results. The highest accuracy result of CNN models in the ALOT dataset was 90.89%, using Xception as a descriptor. In the Outex dataset, it was 91.78% by using ResNet-101 as a descriptor. In CBT, the 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 classifier for the three datasets.

We have computed an 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. Figures 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 AUC score (~0.8), which needs improvement to achieve the highest accuracy result. This observation led the authors to look forward to a local features extraction descriptor with the CNN model.

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

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

5.4. CNN + LBP Descriptors. The first set of analyses confirmed that the accuracy results obtained using only CNN models as grayscale and color image feature descriptors have some limitations in classification accuracy and need some
improvement. Therefore, we used an aggregated feature extraction descriptor (CNN + LBP) for the grayscale and color textures in the second experiment. Convolution neural networks are used to capture global features that describe the image overall. In contrast, Local binary patterns capture the local features of an image that describe a particular image part. We checked the performance of our proposed descriptor over the Brodatz grayscale dataset, and our result was remarkable and proved the adequate performance of the descriptor over a grayscale dataset. Table 6 details the proposed descriptors’ results using the SVM classifier for the grayscale Brodatz dataset.

This performance encourages the authors to check descriptor performance over color datasets (ALOT, Outex, and 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 significantly slow the machine learning stage. 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 preprocessing phase to equal the proposed image size of the pretrained deep learning models (224 × 224 and 227 × 227) to be used in the input layer. The three datasets are split between training and testing datasets at a ratio of 70:30, respectively. Using 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, and CBT), these tests confirm that the proposed descriptors show a clear advantage over the recognition accuracy, are effective and discriminant, and outperform the existing state-of-the-art methods. Using SVM classifier, 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.

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

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

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

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

Figure 12 represents ROC curves of the proposed descriptors using support vector machine classifier for Outex dataset.

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

From Table 9 and Table 10, it is clear that the pretrained model Xception + LBP, ResNet + LBP, and Xception + LBP stand 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 (ALOT 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.

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 state-of-the-art methods introduced by [61] and by [58] in Table 11.

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

Table 13 reveals the accuracy result obtained by the proposed method for the three datasets by providing only 50% for the training set.
Table 4: Results of LBP using SVM classifier for the grayscale Brodatz dataset.

| Dataset texture descriptor | Brodatz (%) |
|----------------------------|-------------|
| LBP descriptor             | 78.20       |

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

| Dataset texture descriptor | ALOT (%) | CBT (%) | Outex (%) |
|----------------------------|----------|---------|-----------|
| LBP descriptor             | 70.3     | 73.45   | 66.2      |

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                |

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| 0.991    | 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 9: ROC curves of CNN models using SVM classifier for the Outex dataset.
Figure 10: ROC curves of the proposed descriptors using SVM classifier for the ALOT dataset.

Table 8: Summary results of the 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 | 0.968    | 0.991    | 96.00         |
| 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: ROC curves of the proposed descriptors using SVM classifier for the CBT dataset.
Table 9: Summary results of the 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  | 0.972     | 0.979 | 0.973    | 97.76    |
| GoogLeNet + LBP     | 0.988   | 0.0003  | 0.985     | 0.988  | 0.981     | 0.983 | 0.956    | 98.80    |
| Vgg16Net + LBP      | 0.993   | 0.0001  | 0.990     | 0.993  | 0.989     | 0.990 | 0.985    | 99.34    |
| Vgg19Net + LBP      | 0.997   | 0.001   | 0.998     | 0.997  | 0.997     | 1.000 | 1.000    | 99.78    |
| SqueezeNet + LBP    | 0.973   | 0.002   | 0.972     | 0.973  | 0.974     | 0.968 | 0.966    | 97.34    |
| ResNet-18 + LBP     | 0.940   | 0.005   | 0.932     | 0.930  | 0.933     | 0.935 | 0.930    | 94.00    |
| ResNet-50 + LBP     | 0.977   | 0.001   | 0.976     | 0.978  | 0.976     | 0.974 | 0.975    | 97.78    |
| ResNet-101 + LBP    | 1.000   | 0.000   | 0.998     | 0.993  | 0.999     | 0.996 | 1.00     | 100.00   |
| MobileNet + LBP     | 0.993   | 0.002   | 0.992     | 0.989  | 0.987     | 0.990 | 0.993    | 99.34    |
| Xception + LBP      | 0.997   | 0.001   | 0.995     | 0.991  | 0.999     | 0.997 | 0.997    | 99.78    |

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

Table 10: The accuracy of results obtained from the proposed descriptors and SVM classifier.

| Dataset texture descriptor | ALOT (%) | 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     |
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.
Table 11: Comparison between results of LBP + CNN and other descriptors for Outex and ALOT dataset.

| Texture descriptor          | Outex (%) | ALOT (%) |
|-----------------------------|-----------|----------|
| Legendre-c.f GLCM           | 83.2      | 66.2     |
| Discr. Cheb. GLCM           | 84.9      | 64.4     |
| LBP-RI                      | 74.6      | 70.3     |
| 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    |

6. Conclusion

Overall, this paper presents a new feature extraction approach 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 good features such as stability and robustness against different orientations and remarkably accurate classification results in recognition tasks. The experiments were performed on three benchmark color datasets (ALOT, CBT, and Outex) using support vector machine as a classifier. We have obtained results demonstrating the proposed classification technique’s efficiency, robustness, and high accuracy against image transformation such as scaling, rotation, and translation. By analyzing classification and recognition accuracy, the findings are promising and encouraging.

Data Availability

The datasets used in this study are from previously reported studies and datasets, which have been cited.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

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] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, “A biological image classification method based on improved CNN,” *Ecological Informatics*, vol. 58, Article ID 101093, 2020.

[2] Z. Lai and H. Deng, "Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 2061516, 2018.

[3] P. P. Rebouças Filho, E. S. Rebouças, L. B. Marinho, R. M. Sarmento, J. M. R. S. Tavares, and V. H. C. de Albuquerque, “Analysis of human tissues densities: a new approach to extract features from medical images,” *Pattern Recognition Letters*, vol. 94, pp. 211–218, 2017.

[4] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, “A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1205–1218, 2012.

[5] S. Hicks, M. Riegler, K. Pogorelov et al., “Dissecting deep neural networks for better medical image classification and classification understanding,” in *Proceedings of the 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 363–368, Karlstad, Sweden, June 2018.

[6] A. K. Jain, N. K. Ratha, and S. Lakshmanan, “Object detection using Gabor filters,” *Pattern Recognition*, vol. 30, no. 2, pp. 1205–1218, 1997.

[7] Q. Zhang, J. Lin, Y. Tao, W. Li, and Y. Shi, “Salient object detection via color and texture cues,” *Neurocomputing*, vol. 243, pp. 35–48, 2017.

[8] P. Barmoutsis, K. Dimitropoulos, and N. Grammalidis, “Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition,” in *Proceedings of the 2014 2nd European Signal Processing Conference (EUSIPCO)*, pp. 1078–1082, Lisbon, Portugal, September 2014.
[9] J. Li, R. William, and D. Buhl-Brown, “Texture analysis of remote sensing imagery with clustering and bayesian inference,” *International Journal of Image, Graphics and Signal Processing*, vol. 7, 2015.

[10] S. Hemalatha and S. M. Anouncia, “Unsupervised segmentation of remote sensing images using FD based texture analysis model and ISODATA,” *International Journal of Ambient Computing and Intelligence*, vol. 8, no. 3, pp. 58–75, 2017.

[11] F. Borne and G. Viennois, “Texture-based classification for characterizing regions on remote sensing images,” *Journal of Applied Remote Sensing*, vol. 11, no. 3, 2017.

[12] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[14] J. Shiguang Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, “WLD: a robust local image descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[15] B. S. Manjunath and W. Y. Ma, “Texture features for Browsing and Retrieval of image Data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.

[16] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local Gabor binary pattern histogram Sequence (LGBPHS): a novel non statistical model for face representation and recognition,” in *Proceedings of the IEEE International Conference on Computer Vision*, Beijing, China, October 2005.

[17] S. F. Ershad, “Color texture classification approach based on combination of Primitive pattern Units and statistical features,” *The International Journal of Multimedia & Its Applications*, vol. 3, no. 3, 2011.

[18] 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.

[19] T. Määnpää 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, and V. D. Tzourasis, “Moment-based local binary patterns: a novel descriptor for invariant pattern recognition applications,” *Neurocomputing*, vol. 99, pp. 358–371, 2013.

[21] L. U. O. Yuan, C. Wu, and Y. Zhang, “Facial expression feature extraction using hybrid PCA and LBP,” *The Journal of China Universities of Posts and Telecommunications*, vol. 20, no. 2, pp. 120–124, 2013.

[22] Y. Fang and Z. Wang, “Improving LBP features for gender classification,” in *Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 373–377, Hong Kong, China, August 2008.

[23] C. Shan, “Learning local binary patterns for gender classification on real-world face images,” *Pattern Recognition Letters*, vol. 33, no. 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,” in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 626–630, New Orleans, LA, USA, March 2017.

[25] D. Givens and M. Karami, “Scene classification using a new radial basis function classifier and integrated SIFT-LBP features,” *Pattern Analysis & Applications*, vol. 23, no. 3, pp. 1071–1084, 2020.

[26] P. Král and L. Lenc, “LBP features for breast cancer detection,” in *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2643–2647, Phoenix, AZ, USA, September 2016.

[27] E. Badeka, C. I. Papadopoulou, and G. A. Papakostas, “Evaluation of LBP Variants in Retinal Blood Vessels Segmentation using machine learning,” in *Proceedings of the 2020 International Conference on Intelligent Systems Computer Vision (ISCV)*, pp. 1–7, Fez, Morocco, June 2020.

[28] D. O. Tambasco Bruno, M. Z. do Nascimento, R. P. Ramos et al., “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] L. Shi, X. Wang, and Y. Shen, “Research on 3D face recognition method based on LBP and SVM,” *Optik*, vol. 220, 2020.

[30] S. Ershad Fekri and F. Tajeripour, “Color texture classification based on proposed Impulse-noise resistant color local binary patterns and significant points selection algorithm,” *Computer Vision and Pattern Recognition*, vol. 37, no. 1, Article ID 11010, 2019.

[31] S. Bianco, C. Casano, P. Napoletano, and R. Schettini, “Improving CNN-based texture classification by color Balancing,” *Journal of Imaging*, vol. 3, pp. 3–33, 2017.

[32] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi, “Deep filter Banks for texture recognition, description, and Segmentation,” *International Journal of Computer Vision*, vol. 118, no. 1, pp. 65–94, 2016.

[33] K. S. Loke, “Texture recognition using a novel input layer for deep convolutional neural network,” in *Proceedings of the 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS)*, pp. 14–17, Singapore, December 2018.

[34] 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,” in *Proceedings of the 2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pp. 1–6, Ho Chi Minh City, Vietnam, May 2019.

[35] S. Basu, S. Mukhopadhyay, M. Karki et al., “Deep neural networks for texture classification—a theoretical analysis,” *Neural Networks*, vol. 97, 2018.

[36] T. C. Wang, J. Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, “A 4D Light-field dataset and CNN architectures for material recognition,” in *Proceedings of the European Conference on Computer Vision ECCV’2016*, vol. 9907, Amsterdam, the Netherlands, October 2016.

[37] S. Dabeer, M. M. Khan, and S. Islam, “Cancer diagnosis in histopathological image: CNN based approach,” *Informatics in Medicine Unlocked*, vol. 16, 2019.

[38] K. M. Hosny, M. A. Kassem, and M. M. Fouad, “Classification of skin lesions using transfer learning and augmentation with Alex-net,” *PLOS ONE*, vol. 14, no. 5, Article ID e0217293, 2019.

[39] V. P. Genovese, F. Scotti, and S. Vishwakarma, “Touchless palmprint and finger texture recognition: a deep learning fusion approach,” in *Proceedings of the 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 1–6, Tianjin, China, June 2019.

[40] M. Umer, I. Ashraf, S. Ullah, A. Mehmood, and G. S. Choi, “COVInit: A convolutional neural network approach for
predicting covid-19 from chest x-ray images,” Journal of Ambient Intelligence and Humanized Computing, 2021.

[41] Y. Wu and T. Lee, “Enhancing sound texture in CNN-based acoustic scene classification,” in Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 815–819, Brighton, UK, May 2019.

[42] 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,” in Proceedings of the 2017 IEEE International Conference on Industrial Technology (ICIT), pp. 1076–1080, Toronto, ON, Canada, March 2017.

[43] W. d. S. Cotrim, V. P. Rodrigues Minim, L. Bonato Felix, and L. Antonio Minim, “Short convolutional neural networks applied to the recognition of the browning stages of bread crust,” Journal of Food Engineering, vol. 277, 2020.

[44] K. M. Hosny, M. A. Kassem, and M. M. Fouad, “Classification of skin lesions into seven classes using transfer learning with AlexNet,” Journal of Digital Imaging, vol. 33, no. 5, pp. 1325–1334, 2020.

[45] 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.

[46] K. M. Hosny, T. Magdy, and N. A. Lashin, “Improved color texture recognition using multi-channel orthogonal moments and local binary pattern,” Multimedia Tools and Applications, vol. 80, pp. 13179–13194, 2021.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Communications of the ACM Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 60, no. 6, pp. 84–90, 2012.

[48] R. Yoshihashi, R. Kawakami, M. Iida, and T. Naemura, “Bird detection and species classification with time-lapse images around a wind farm: dataset construction and evaluation,” Wind Energy, vol. 20, no. 12, pp. 1983–1995, 2017.

[49] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015.

[50] S. Liu and W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” in Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 730–734, Kuala Lumpur, Malaysia, November 2015.

[51] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Computer Vision and Pattern Recognition, Article ID 1556, 2014.

[52] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and < 0.5 MB Model Size,” Computer Vision and Pattern Recognition, 2016.

[53] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016.

[54] A. G. Howard, M. Zhu, B. Chen et al., “MobileNets: efficient convolutional neural networks for mobile vision applications,” Computer Vision and Pattern Recognition, 2017.

[55] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” Computer Vision and Pattern Recognition, 2017.

[56] S. Fekri-Ershad and F. Tajeripour, “Mpulse-noise resistant color-texture classification approach using hybrid color local binary patterns and Kullback–Leibler divergence,” The Computer Journal, vol. 60, no. 11, pp. 1633–1648, 2017.

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

[58] F. Bianconi and A. Fernandez, “Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform,” Pattern Recognition Letters, vol. 48, pp. 34–41, 2014.

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

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

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