

## Article

# A Dynamic Learning Method for the Classification of the HEp-2 Cell Images

Caleb Vununu <sup>1</sup>, Suk-Hwan Lee <sup>2</sup>, Oh-Jun Kwon <sup>3</sup> and Ki-Ryong Kwon <sup>1,\*</sup><sup>1</sup> Department of IT Convergence and Application Engineering, Pukyong National University, Busan 48513, Korea<sup>2</sup> Department of Information Security, Tongmyong University, Busan 48520, Korea<sup>3</sup> Department of Computer Software Engineering, Donggeui University, Busan 47340, Korea

\* Correspondence: krkwon@pknu.ac.kr; Tel.: +82-51-629-6257

Received: 28 June 2019; Accepted: 29 July 2019; Published: 31 July 2019



**Abstract:** The complete analysis of the images representing the human epithelial cells of type 2, commonly referred to as HEp-2 cells, is one of the most important tasks in the diagnosis procedure of various autoimmune diseases. The problem of the automatic classification of these images has been widely discussed since the unfolding of deep learning-based methods. Certain datasets of the HEp-2 cell images exhibit an extreme complexity due to their significant heterogeneity. We propose in this work a method that tackles specifically the problem related to this disparity. A dynamic learning process is conducted with different networks taking different input variations in parallel. In order to emphasize the localized changes in intensity, the discrete wavelet transform is used to produce different versions of the input image. The approximation and detail coefficients are fed to four different deep networks in a parallel learning paradigm in order to efficiently homogenize the features extracted from the images that have different intensity levels. The feature maps from these different networks are then concatenated and passed to the classification layers to produce the final type of the cellular image. The proposed method was tested on a public dataset that comprises images from two intensity levels. The significant heterogeneity of this dataset limits the discrimination results of some of the state-of-the-art deep learning-based methods. We have conducted a comparative study with these methods in order to demonstrate how the dynamic learning proposed in this work manages to significantly minimize this heterogeneity related problem, thus boosting the discrimination results.

**Keywords:** HEp-2 cell classification; HEp-2; deep learning; convolutional neural networks; discrete wavelet transform; artificial neural network; pattern recognition

## 1. Introduction

Computer-aided diagnostic (CAD) systems refer to all the techniques that aim to consolidate the automation of the disease diagnostic process. These systems have gained tremendous interest since the unfolding of various machine learning methods in the past decades because of their ability to perform any automatic task in a very reliable way. One of the most challenging tasks regarding those CAD systems is the complete analysis and understanding of the images that represent biological organisms. In the case of autoimmune diseases, indirect immunofluorescence (IIF) on human epithelial type 2 (HEp-2) cell patterns is the most recommended diagnosis methodology [1]. However, the manual analysis of the IIF images represents an arduous task that can take a substantial time. Moreover, the complexity of the images leaves an important part to the subjectivity of the pathologists, which can lead to some inconsistency in the diagnosis results [2]. That is the reason why CAD systems have gained critical attention in the past years. These systems can assist pathologists during the diagnostic process, mainly for the automatic classification of the different types of HEp-2 cells.

Different methods have been discussed in the literature and especially the methods presented during the different editions of the HEp-2 cell classification contest held by the International Conference on Pattern Recognition (ICPR) [3]. As a classical pattern recognition task, HEp-2 cell classification methods comprise a feature extraction or selection process that is followed by a classification one. Feature extraction remains the most important part of the procedure, because it consists of extracting the relevant information that can support an accurate discrimination of the different cell types. Commonly, we can distinguish two groups of methods in the literature: the conventional machine learning-based methods and the deep learning-based ones.

Conventional machine learning techniques entail hand-crafted features that are chosen for their capability of carrying relevant elements that are necessary for discrimination. Early efforts in that direction have been made by researchers such as Cataldo et al. [4], who proposed the gray level co-occurrence matrix and the discrete cosine transform (DCT) features, and Wiliem et al. [5], who adopted the codebooks generated from the DCT features and scale-invariant feature transform (SIFT) descriptors. Nosaka et al. [6] used the local binary patterns (LBP) as the features and given them as inputs to a linear support vector machine (SVM) for the classification step. Huang et al. [7] utilized the textural and statistical features in a hybrid fashion and fed them to a self-organizing map for the classification process.

Other statistical features, such as the gray-level size zone matrix, have been employed as the principal features in the work by Thibault et al. [8] and the nearest-neighbor classifier was adopted for the discrimination part. The same statistical features have been fed to an SVM in the work of Wiliem et al. [9], while a linear local distance coding method was used for extracting the features that were also utilized as the inputs of a linear SVM by Xu et al. [10].

Hybrid feature learning methods have also been utilized by the researchers in this field. In fact, Cataldo et al. [11] have proposed the use of a combination of different features such as morphological features, global texture descriptors like the rotation-invariant Gabor features [12], and also different kinds of LBP descriptors like the rotation-invariant uniform LBPs [13], the co-occurrence adjacent LBPs [14], the completed LBP [15], and also the rotation-invariant co-occurrence of adjacent LBPs, which are also adopted in [6]. Another hybrid feature extraction method can be found in the work of Theodorakopoulos et al. [16] where the authors proposed the combination of LBP and SIFT descriptors. Different other hand-crafted features can be seen in [17,18], and many others are listed in the quasi-exhaustive review made by Foggia et al. [3]. We can note that the performance of these methods exclusively relies on the discrimination potentiality afforded by the extracted features, leaving, again, an important part to the subjectivity of the user.

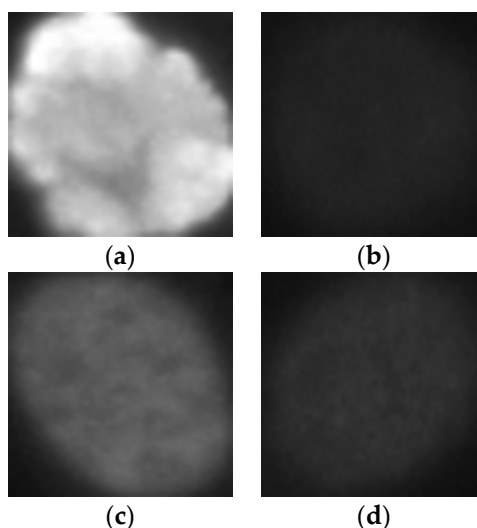
Automatic feature learning methods have been widely adopted since the unfolding of deep learning [19]. They have shown outstanding results in object recognition problems [20,21] and many researchers have adopted them as the principal tool for the HEp-2 cell classification. Unlike conventional methods, whose accuracy depends on the subjective choice of the features, deep learning methods, such as the now popular convolutional neural networks (CNNs), have the advantage of offering an automatic feature-learning process. In fact, many works have demonstrated the superiority of deep learning-based features over the hand-crafted ones for the HEp-2 cell classification task [2,22–30].

The first work to apply CNN to the HEp-2 cell classification problem was presented by Foggia et al. [2] during the 2012 edition of the ICPR HEp-2 cell classification contest. Although the results were outstanding, the datasets available at that time were not heterogeneous enough and needed a lot of improvements. Since then, many available datasets have been significantly diversified and the different proposed CNN models continue to push the limits in terms of the classification accuracy. Gao et al. [22] have presented a simple CNN architecture that was tested over different datasets. They were the first to test the data augmentation techniques, such as rotation in different angles, for the HEp-2 cell images. Li et al. [23] have adopted the deep residual inception model, the DRI, which combines two of the most popular CNN models, the ResNet [24] architecture and the “Inception” modules from the GoogleNet [25]. Phan et al. [26] have performed transfer learning, which

consists of using an already trained network in a new dataset, by using a model that was trained on the ImageNet dataset.

A complex transfer learning method has been proposed by Lei et al. [27] where they have used different proposed architectures of the pre-trained ResNet model and mixed it together in order to produce what they have named a cross-modal transfer learning approach. The results obtained via this method represent one of the state-of-the-art performances for the HEP-2 cell classification task. Another state-of-the-art performance was obtained in the work presented by Shen et al. [28] where the authors used the ResNet approach again, but with a deeper residual module, called the deep-cross residual (DCR) module, with a huge data augmentation. Other CNN based methods can be seen in [29,30].

The HEP-2 cell images datasets generally denote a significant heterogeneity. This is explained by the intra-class variations caused by the existence of different intensity levels. In Figure 1, we show some examples of the cellular images from one of the datasets used during the experiments. Both images in Figure 1a,b represent the fine speckled cell type. The first image comes from the positive intensity level and the second one from the negative intensity level. The intra-class variations can be clearly seen from these two images that are supposed to belong to the same class. We can imagine the complexity faced by the classifiers that attempt to output these two images as from the same class. The heterogeneity discussed here becomes much clearer when we look at the images depicted in Figure 1c,d. As for the previously mentioned type, both images belong to the same cellular type, the homogenous cells. Again, we can remark the intra-class variations. Moreover, the complexity becomes much noticeable when we compare the two negative intensity images shown in Figure 1b,d. The two images are really similar in two main points: the lack of perceptibility of the cellular shape and the weakness of the illumination. Note that the two images, while being quite indistinguishable, are supposed to belong to two different classes.



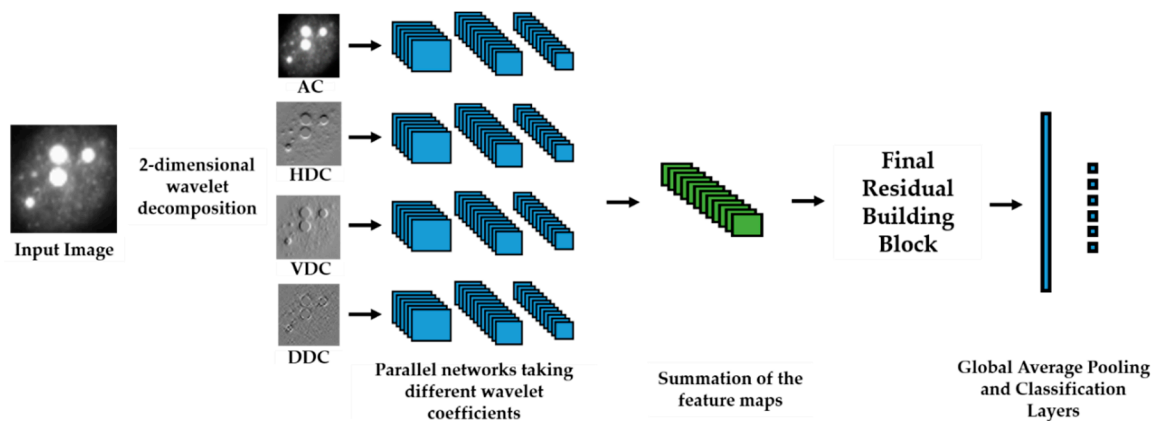
**Figure 1.** HEP-2 cellular images from the dataset. (a) The positive intensity fine speckled cell; (b) The negative intensity fine speckled cell; (c) The positive intensity homogeneous cell; (d) The negative intensity homogeneous cell.

This significant heterogeneity can cause certain problems in terms of the discrimination efficiency. Some of the actual state-of-the-art methods still encounter some trouble solving this problem, especially when the dataset exhibits a strong disparity. The method proposed in this work tries to specifically tackle this heterogeneity-related problem. We have chosen two datasets that show strong intra-class variations. Some methods, like the one proposed by Nigam et al. [31], try to solve this problem by performing a preliminary intensity-based separation that precedes the cell type classification itself. Even though this way of proceeding can lead to reasonably good final results, the fact of using two

steps in the classification part, after a burden future extraction process, clearly leads to elongating the global processing time. We propose a method that tries to minimize the heterogeneity in one step.

We propose a dynamic learning process with different networks taking different variations of the input image in parallel. The discrete wavelet transform (DWT) is used in order to produce the different variations. A two-dimensional DWT is performed over the images. The DWT coefficients are normalized and represented in an image form in order to feed the networks. Four different CNNs are used in parallel, each of them taking the different normalized wavelet coefficients as the inputs. The propagation of the error is done in parallel and, at a certain point of the networks, the high-level features from the feature maps of the four CNNs are concatenated and passed through a nonlinearity function in order mix up the different information learned by the four networks. Then, a final residual building block is formed using the output of the layer that performs the concatenation process. Finally, these nonlinearities are passed through a global average pooling layer and also a fully connected layer that will perform the final classification.

As explained in the next section, the different wavelet coefficients are expected to expose and significantly reinforce the localized changes in intensity, especially for the cellular types that exhibit strong differences in their levels of fluorescence illumination, thus, leading to a homogenization of not only the intensity, but also the shape, of the majority of the images of a given cell. The four different CNNs then learn dynamically these unveiled properties. The method was tested on the publicly available SNPHEp-2 [5] and ICPR datasets and the results show that the strong disparity of the datasets is really minimized, thus, boosting the finally obtained discrimination results. The schematic representation of the proposed method is shown in Figure 2.



**Figure 2.** The schematic representation of the proposed method. “AC”, “HDC”, “VDC”, and “DDC” stand for the approximation coefficients, horizontal detail coefficients, vertical detail coefficients and diagonal detail coefficients, respectively.

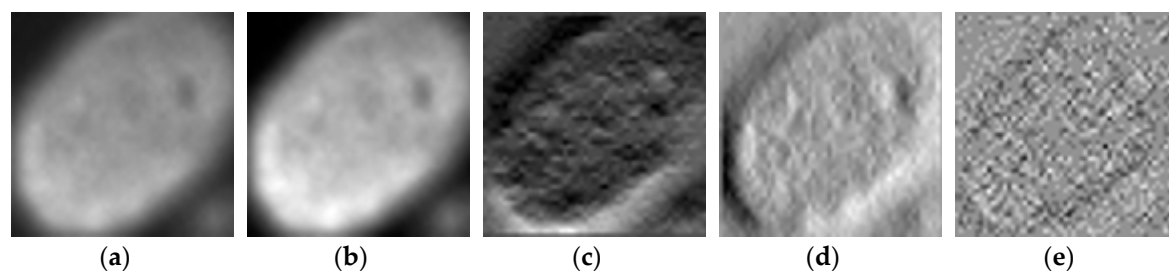
The remaining content of the paper is organized as follows. The next section presents in detail each step of the proposed scheme. Section 3 discusses about the obtained results and addresses a quasi-exhaustive comparative study using other methods.

## 2. Dynamic Learning for Cell Classification

### 2.1. Two-Dimensional Discrete Wavelet Transform Decomposition

Figure 2 illustrates the different steps of the proposed method. We have applied the DWT decomposition over the input image. The two-dimensional decomposition at the first level produces four different outputs. Following the high and low-frequency band decomposition afforded by the DWT, the approximation coefficients contain the low-frequency band information. The high-frequency related information can be mostly retrieved in the three remaining coefficients: the horizontal, the vertical, and the diagonal details. These four different coefficient matrices, after being represented

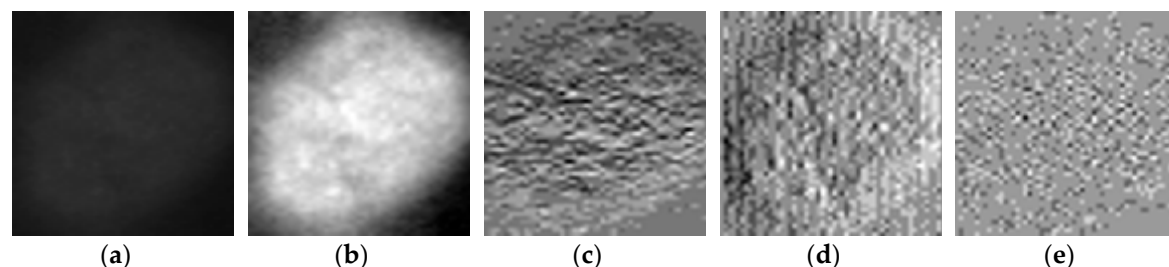
as images, will be considered as the inputs of the four different CNNs that will be used in parallel. In Figure 3, we show some wavelets coefficients of a positive intensity cellular image from the dataset.



**Figure 3.** Positive intensity Fine Speckled cell. (a) Original image; (b) the approximation; (c) the horizontal detail; (d) the vertical detail; and (e) the diagonal detail. The original image has a size of  $112 \times 112$ . The outputs of the DWT are all downsized by half ( $56 \times 56$ ). All the images in the figure were identically resized for the purpose of visualization.

In Figure 3, we have an example of the positive intensity fine speckled cell and the four different outputs of the DWT decomposition. In Figure 3a, we have the original image, with a strong illumination. Figure 3b represents the approximation's coefficients. They denote the result of the low-pass filtering process incorporated in the DWT decomposition. This low-pass filtering has an effect such that the regions that do not exhibit significant gray variations are reinforced. We can notice that in the resulting image shown in Figure 3b, the intensity values of the pixels are slightly reinforced. The slight reinforcement of the global illumination is clear by comparing the original image in Figure 3a and the approximation coefficients shown in Figure 3b. The images shown in Figure 3c–e represent the detail coefficients. These coefficients are representative of the high-frequency elements of the original image. As we can remark, all the three different detail coefficients describe the shape of the cell.

In Figure 4, we show an example of the negative intensity fine speckled cell and the four different outputs of the DWT decomposition. In Figure 4a, we have the original image, with a weak illumination. We can see how the negative intensity image does not provide any important cues regarding the shape of the cell. The image in Figure 4b portrays the approximation coefficients. We can see a noticeable reinforcement of the intensity level in the image. By comparing the images shown Figures 3b and 4b, we can clearly discern the homogenization that occurred in terms of the intensity. And this situation is precisely the result we were seeking for: a homogenization of the images from the two different fluorescence illuminations. Even though the original images shown in Figures 3a and 4a exhibit strong differences in terms of global illumination, their approximation coefficients shown respectively in Figures 3b and 4b denote a certain homogenization in their illumination.



**Figure 4.** Negative intensity Fine Speckled cell: (a) Original image; (b) the approximation; (c) the horizontal detail; (d) the vertical detail; (e) the diagonal detail. The original image has a size of  $112 \times 112$ . The outputs of the DWT are all downsized by half ( $56 \times 56$ ). All the images in the figure were identically resized for the purpose of visualization.

The images shown Figure 4c–e represent the three different detail coefficients. We have, respectively, the horizontal, vertical, and diagonal details. Even though the original image, in Figure 4a, does not



provide any noticeable information regarding the shape of the cell, we can remark how the detail coefficients really expose the geometric cues of the cells. Further, by comparing the detail coefficients of both the positive illumination image in Figure 3 and the negative illumination image in Figure 4, which englobe the images shown in Figures 3c–e and 4c–e, we can again discern the homogenization that occurred in terms of the perceptibility of the geometrical cues, which give us a comprehensive understanding of the cell's shape. Knowing that the geometrical patterns of the cells are really important for their discrimination, their reinforcement appears to be necessary. Moreover, the three detail coefficients act like the gradients at different directions.

Finally, the approximation and detail coefficients will minimize the divergences between the two kinds of illumination (intensity levels) and encourage a certain homogenization in terms of the intensity, with the approximation coefficients, and in terms of the geometrical patterns, with the three detail coefficients that reveal the cellular shape. These four different coefficients will be used as the inputs of four different networks that will learn in parallel how to dynamically encapsulate all the necessary information that will help in the discrimination of the different cellular types.

## 2.2. Dynamic Learning with Multiple Networks

The second part of the proposed method consists of feeding the four different coefficients to different networks. CNNs are used for that purpose. The networks will be utilized in a parallel paradigm. As shown in Figure 2, the first part of the proposed scheme has four networks that extract the features from the different coefficients. After this process, the features learned by the networks are concatenated in order to mix up the different information they contain and provide a final high-level feature vector. This vector is obtained after a global average pooling process and will be passed through a fully connected layer that will perform the final classification. The four CNNs have an identical structure that is depicted in Figure 5.

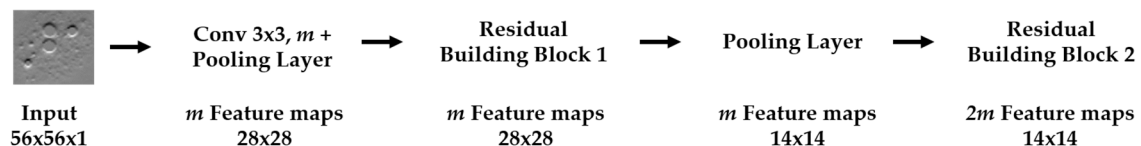


Figure 5. Structure of the four different networks.

The DWT decomposition process at the first level downsizes by half the original images. Since the original cellular images have a size of  $112 \times 112$ , the networks take the inputs of size  $56 \times 56$ , as denoted in Figure 5. For the totality of the convolution operations, we have used a receptive field of  $3 \times 3$ . The convolution operations are done in such a way that the output volumes keep the spatial extent of the input volume so that the down-sampling process is accomplished only by the pooling layers. This is in order to avoid the loss of the spatial information during the feature extraction process performed by the convolution operations. In Figure 5, “Conv  $3 \times 3$ ,  $m$ ” denotes a convolution operation with a filter size of  $3 \times 3$  and  $m$  different filters. The value of  $m$  is set to be 32.

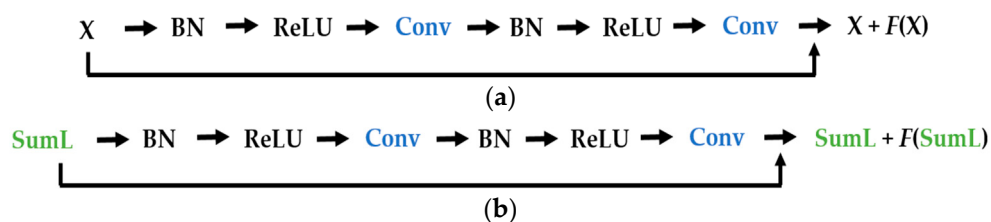
In Figure 5, we can see that the input coefficients are first given to a convolution layer, whose characteristics are described in the previous paragraph, followed by a pooling layer. This pooling layer has a stride and a padding of 2 in order to down-sample the input volume from the convolution layer by half. Thus, the output volume in this part of the network has a spatial extent of  $28 \times 28$  and contains 32 different maps. After that, a residual building block is constructed using the output of the previous layer. The output of the first residual building block (residual building block 1 in Figure 5) has the same characteristics as its input volume, which means that it still comprises 32 different feature maps of size of  $28 \times 28$ . A second pooling operation is conducted right before the second residual building block. The output volume of the residual building block 2 has 64 ( $2 \times m$  in Figure 5) features maps and a spatial extent of  $14 \times 14$ .

Residual building blocks comprise two convolution layers, two rectified linear unit (ReLU) activation layers, and two batch normalization (BN) layers. For a given variable  $x$ , the ReLU activation function is defined by:

$$f(x) = \max(x, 0). \quad (1)$$

BN [32] is a type of normalization method that is used between the layers in CNNs in order to solve the internal covariate shift problem. This problem refers to the change of the distribution of the activations of the layers during the training process. Knowing that the activations of a given layer are passed as inputs to the following layer, this problem has the consequence that every layer has to continuously adapt to the new distribution of its inputs. Ioffe et al. [32] have highlighted the fact that stabilizing the distribution of the activations of the different layers in very deep networks acts as a powerful regularizer that can reduce overfitting and speed up the training process. BN layers aim to find the parameters that minimize the distribution change during the training process by scaling and shifting (normalization process) the activations in a similar way iteration after iteration, thus stabilizing the network.

He et al. [33] claimed that using BN and ReLU before the convolution operation, the so-called full pre-activation method, can help to reduce overfitting in the network. Our residual building blocks are organized in that manner. As we can see in Figure 6, each convolution operation is preceded by the BN and ReLU layers. The two residual building blocks denoted in Figure 5 have the structure depicted in Figure 6a. The two convolution layers in each one of them do not alter the spatial extent of the entering volumes.



**Figure 6.** (a) Structure of the residual building blocks (1 and 2) in the four different networks; (b) Structure of the final residual building block after the summation layer (SumL).

The identity short-cut is defined as follows:

$$y = x + f(x, W), \quad (2)$$

where the term  $f(x, W)$  denotes the output of the final convolution operation,  $x$  being the input volume,  $W$  denoting the parameters to be learned during the training process, and  $y$  representing the final output volume of the residual building block. On the other hand, Figure 6b illustrates the identity short-cut as we will perform it after the summation of the feature maps from the four different networks. In that case, the variable  $x$  in Equation (2) represents the output of the summation layer. In Figure 7, we show the illustration of what we call the summation layer. After the second residual building blocks of the four networks, every final volume from the networks has 64 feature maps of size of  $14 \times 14$ , as discussed previously and shown in Figure 5. The first operation of the summation layer is to perform a concatenation of the feature maps from the different networks. The useful patterns learned by the different networks will be assembled in one volume. This single volume will have 256 different feature maps. In order to encapsulate all the information, and also to reduce the number of parameters for the following layers, a convolution operation with a filter size of  $1 \times 1$  is conducted over the volume inherited from the concatenation process. This operation is similar to a linear combination of the different features learned by the different networks. It is necessary to notice that the  $1 \times 1$  convolution operation does reduce the depth by combining the features but, also, it does preserve the spatial extent

of the input volume, so that we do not suffer any loss of the spatial information. The final volume after this summation process has 64 feature maps of the size of  $14 \times 14$ , as we can see in Figure 7.

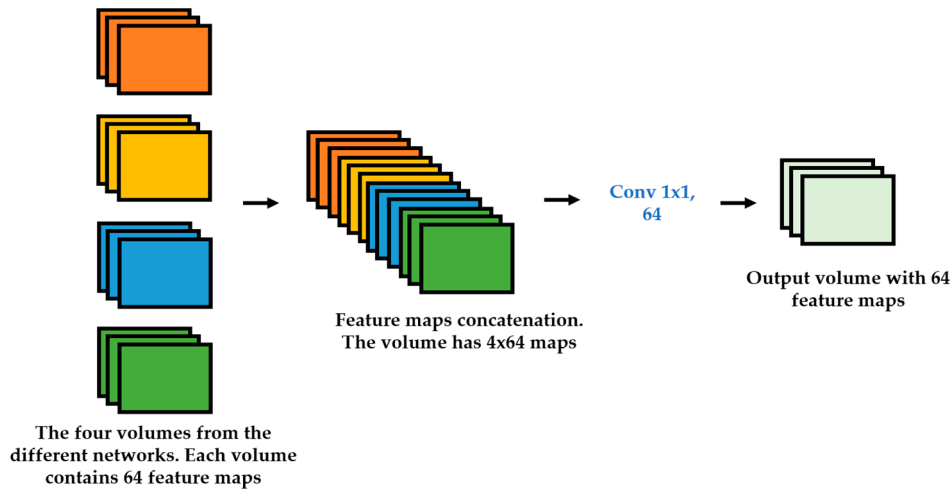


Figure 7. Illustration of the summation layer's structure.

This final volume will be used as the input of the final building block denoted in Figure 6b. This residual block contains also two convolution operations, as we can see in Figure 6b. Each one of these convolution operations utilize 128 different filters. Since the convolution operations preserve the spatial extent, the output volume of the residual building block 3 will be in the form of  $14 \times 14 \times 128$ . After a BN and a ReLU activation layer, finally, this volume will be used for the global average pooling. After applying the global average pooling, the final high-level feature vector will be in the form of  $1 \times 1 \times 128$ , which gives a one-dimensional vector containing 128 elements.

The high-level feature vector will be given to a fully connected layer that will perform the classification process. This layer uses the softmax function [34], which is defined as follows:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{i=1}^N e^{z_i}}, \text{ for } j = 1, \dots, N, \quad (3)$$

where  $N$  is the number of the classes, which means that it is the length of the input feature vector of the softmax function. The value  $z_j$  is the  $j^{\text{th}}$  element of that input vector, which means that it represents the feature associated with the  $j^{\text{th}}$  class in the input feature vector of the function. The values  $\sigma(z)_j$ , with  $j$  varying from 1 to the total number of classes  $N$ , represent the outputs of the softmax function, which can be interpreted as the normalized probabilities of every single class given a data. The networks learn by back-propagating the error [35] and using the cross-entropy error function [34] defined by:

$$E(y, z) = - \sum_{j=1}^N y_j \log(z_j), \quad (4)$$

where the values  $y_j$  denote the actual labels of the  $N$  classes for a given data. On the other hand, the values  $z_j$  represent the elements of the output vector  $\mathbf{z}$  of the softmax function computed using Equation (3).

Figure 8 shows the summary of the entire architecture of the proposed method. All the four networks being identical before the summation layer, in order to avoid redundancy, we have mentioned the sizes of the different volumes (the outputs of the layers) only for the first network. The same sizes apply for the remaining three networks in Figure 8. Note that all the inputs have also the same size, as mentioned before.



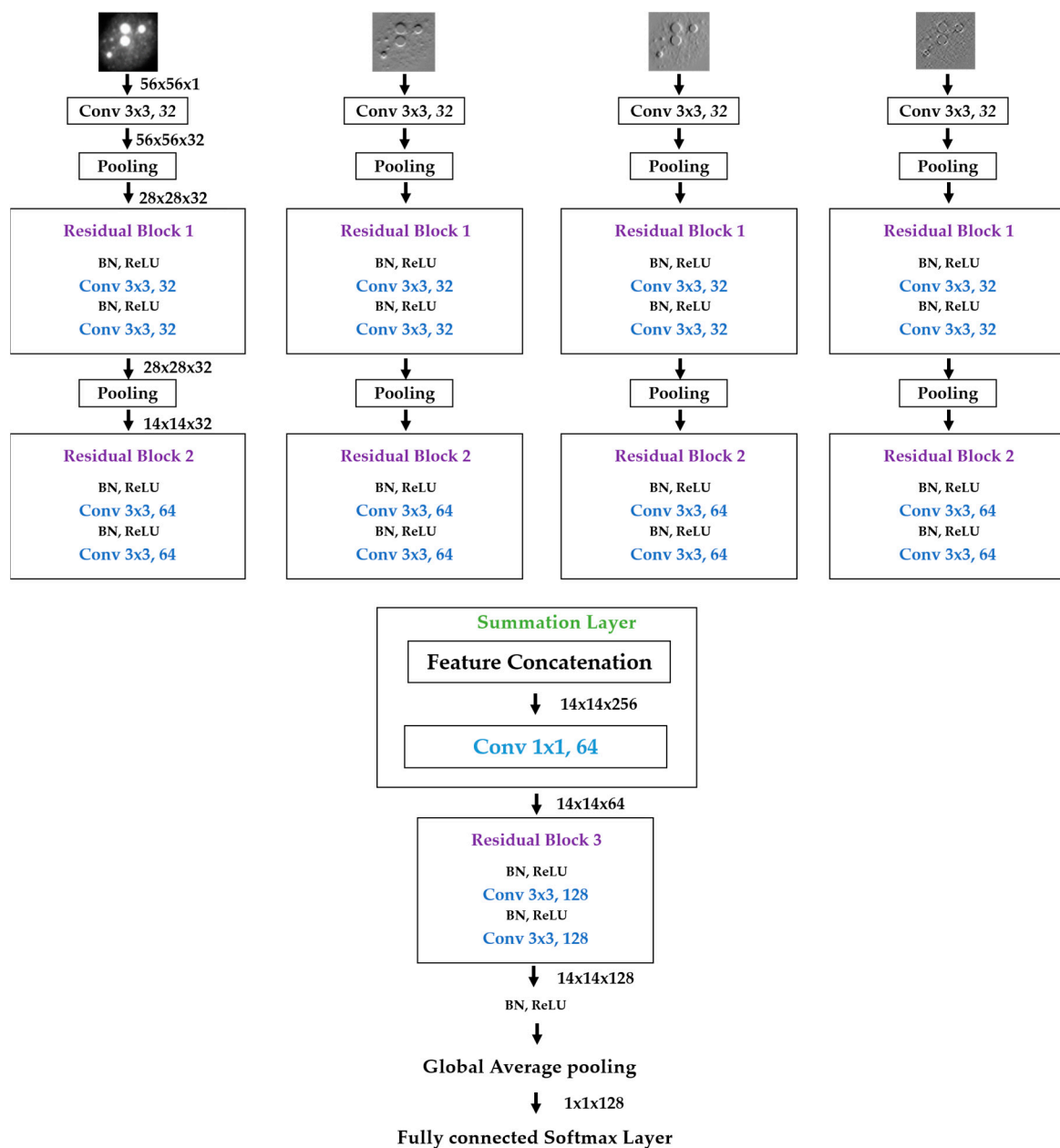


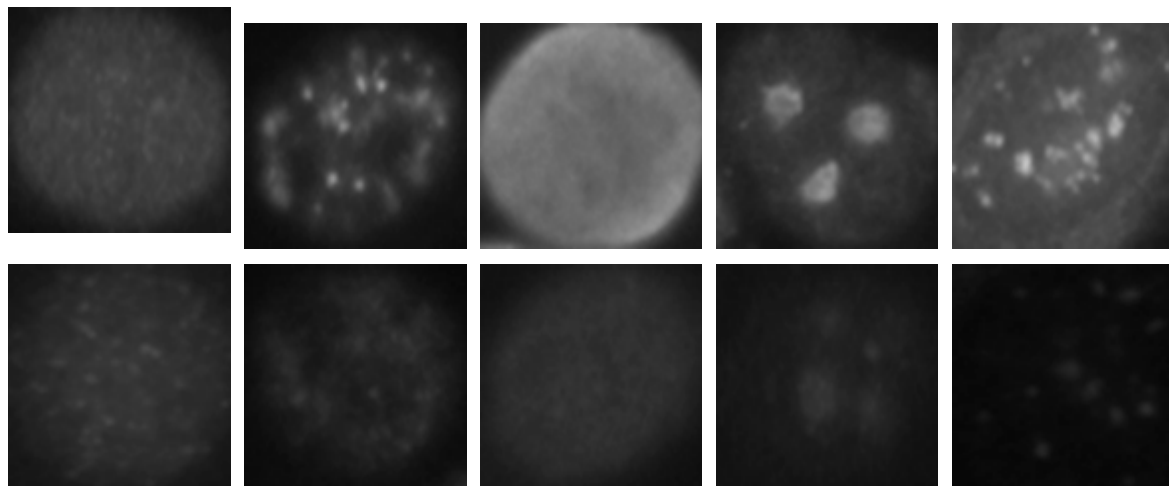
Figure 8. Summary of the architecture of the proposed scheme.

### 3. Results

The proposed method was mainly tested on the SNPHEp-2 dataset. This dataset was obtained between January and February 2012 at the Sullivan Nicolaides Pathology laboratory at Australia. It contains five patterns: the centromere, the coarse speckled, the fine speckled, the homogeneous and the nucleolar types. The images depicted in Figure 1, which represent the fine speckled and homogenous cells, were obtained from this dataset. It is composed of 40 different specimens and every single specimen image was captured using a monochrome camera, which was fitted on a microscope with a plan-Apochromat 20x/0.8 objective lenses and an LED illumination source. In order to automatically extract the image masks, which specifically delimit the cells' body, the DAPI image channel was utilized.

There are 1884 cellular images in the dataset, all of them are extracted from the 40 different cells' specimens. Different specimens were used for constructing the training and testing image sets, and

both sets were created in such a way that they cannot contain images from the same specimen. From the 40 specimens, 20 were used for the training sets and the remaining 20 were used for the testing sets. In total, there are 905 and 979 cell images for the training and testing sets, respectively. Each set (training and testing) contains five-fold validation splits of randomly selected images. In each set, the different splits are used for cross-validating the different models and hyperparameters, each split containing 450 images approximatively. The SNPHEp-2 dataset was presented by Wiliem et al. [5] and can be found at <http://staff.itee.uq.edu.au/lovell/snphep2/>. Every cellular type from this dataset is represented by one example image in Figure 9.



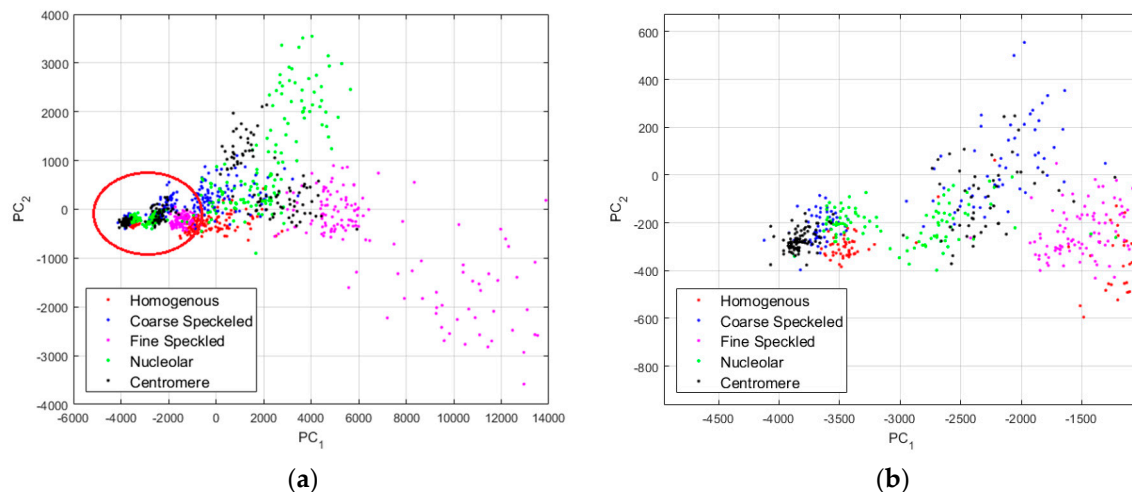
**Figure 9.** HEp-2 cell images from one of the SNPHEp-2 dataset. In the first row, we have the positive intensity images while the second row depicts the intermediate intensity images. For both rows we have, from left to right, the Homogeneous, Coarse Speckled, Fine Speckled, Nucleolar and Centromere cells.

The ICPR 2016 [36], on the other hand, dataset contains six different cellular types: the homogeneous (2494 images from 16 different specimens), the speckled (2831 images, 16 specimens), the nucleolar (2598 images, 16 specimens), the centromere (2741 images, 16 specimens), the nuclear membrane (2208 images, 15 specimens), and the Golgi (724 images, only 4 specimens). The dataset contains in total 13,596 images and can be downloaded at <http://mivia.unisa.it/datasets/biomedical-image-datasets/hep2-image-dataset/>.

The choice of mainly using the SNPHEp-2 dataset for the results shown in this work was justified by the fact that it exhibits a significant heterogeneity compared to the ICPR one, which makes it pose serious problems for the discrimination. The dataset contains only around 2000 images, but 40 different cells' specimens were used in order to produce these images. Besides the fact that the different sets (training and testing) comprise images from totally different specimens, the probability of finding many images that derive from the same specimen is significantly low. By comparison, the ICPR dataset, on which we have also tested the method and shown the results (only for the comparative study with the state-of-the-art methods), contains more than 10,000 images, but all of them were produced using only 83 specimens. There is a high probability to find in this dataset several images that derive from the same specimen. This fact reduces the heterogeneity of the dataset, making it susceptible to pose less complexity in terms of the discrimination compared to the SNPHEp-2. Note that, besides the fact that it contains more than 10,000 images, the ICPR dataset also exhibits strong complexity. We have just found that the complexity related to the variation of the fluorescence illumination, which is precisely the main problem that the proposed method aims to tackle, is more noticeable in the SNPHEp-2 dataset.

In Figure 10, we show the projections of the data from this dataset using the principal component analysis (PCA) [37].  $PC_1$  and  $PC_2$  denote the first and second principal component axis, respectively. All the projections are shown in Figure 10a, where we can clearly distinguish two main projection's sub-spaces. Many data points, the positive intensity ones, are spread all around the PCA space,

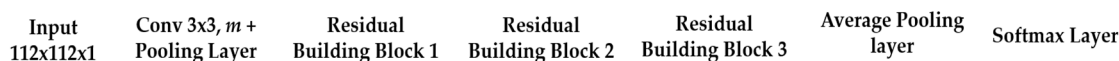
denoting their high heterogeneity. Specifically, the fine speckled cells, represented by the magenta dots, are scattered all over the PCA space. The projection sub-space that is surrounded by the red circle in Figure 10a represents the negative intensity data. We have zoomed this part and shown the result in Figure 10b. All the cellular types are present in that part of the PCA-space (see Figure 10b), and they represent the negative intensity subset. At the same time, all the cellular types are also present in the other part (outside of the red circle), and they represent the positive intensity subset. The purpose of the proposed method is to try to minimize this heterogeneity by homogenizing the images from the same class but belonging to the different fluorescence illumination's subsets.



**Figure 10.** (a) Projections of the data from the SNPHEP-2 dataset; (b) The zooming of the part surrounded by the red circle in (a).

Concerning the hyperparameters used during the training, different scenarios were used in order to optimize the discrimination results. These hyperparameters were selected via cross-validation using different splits, as clearly explained previously. All the following values concerning the hyperparameters are the ones that have provided the best results during the experiments. The initial learning rate was set to be 0.1 at the beginning of the training process. This value was then divided by a factor of 10 every 50 epochs. This schedule, as usually done in the literature with deep learning structures, aims to update the weights quickly at the beginning of the training process while the loss (error) is quite big, and then slowing down the update process, by decreasing the learning rate, when the loss decreases. The learning is terminated when the loss plateaus and does not change after decreasing the learning rate. The size of the mini-batch was set to be 128. The momentum was set to 0.9 and we have initialized the weights by following the process presented by He et al. [33]. All the experiments were conducted with MATLAB R2018b and performed on a computer with a Core i7 3.40 GHz processor and 8 GB of RAM. A GPU implementation was used with a NVIDIA GeForce GTX 1080 Ti with 11,264 MB of memory, which accelerates the training time.

We start by the results obtained using one single network that takes the original images as the input. The structure of this network follows the same structure of the proposed method, except the fact that there is no summation layer because there is only a single network. The network structure is depicted in Figure 11. The original images are given to a convolution layer that is followed by three residual building blocks. The residual blocks have the same structure as the ones described in Figure 6a. As previously mentioned, the original images have a size of  $112 \times 112$ , which is why the input, in this case, has that same size. Note that the input images of our parallel networks have a size of  $56 \times 56$  because of the down-sampling process performed by the DWT decomposition at the first level.



**Figure 11.** Structure of a single network used during the experiments.

The results are shown in the confusion matrix denoted in Figure 12. The same hyperparameters previously presented were used for training the single network. In all the confusion matrices shown in this work, “Homo”, “Coarse”, “Fine”, “Nucl”, and “Centro” denote, respectively, the homogeneous, coarse speckled, fine speckled, nucleolar and centromere cells.

Output Class	Target Class					
		Homo	Coarse	Fine	Nucl	Centro
	Homo	88.29	0	8.12	0	0
	Coarse	1.17	83.97	2.33	10.71	7.72
	Fine	6.42	0.05	88.54	1.39	0.99
	Nucl	1.83	6.59	0	79.30	3.18
	Centro	2.29	9.39	1.01	8.60	88.11

**Figure 12.** Confusion matrix of the results over the testing set by using the original cellular images as the inputs of one single network. The total accuracy of the network is 85.64%.

By analyzing the results shown in Figure 12, we retrieve the problems mostly faced while dealing with the HEP-2 cells’ discrimination: the confusion between the cells that share similarities in terms of the shape and intensity levels. The homogeneous cells are confounded mostly with the fine speckled cells. In Figure 8, by comparing the two cell types, we can remark that they share a lot of similarities in terms of their shape. If we take a look closely at the fine speckled cells’ classification in Figure 12, we can notice that the situation is bit worse than for the homogeneous cells. In fact, 8.12% of the fine speckled are classified as homogeneous while “only” 6.42% of the homogeneous are classified as fine speckled. This situation is explained by the fact that most of the negative intensity fine speckled cells, as we can see with the image shown in the second row and third column of Figure 8, do not show any distinguishable features when we compare them with the homogeneous (both positive and negative intensity) cells. Which means that the disparity related to the fluorescence illumination of the cells encourages the confusion between the cells. The results in Figure 12 have a total accuracy of 85.64%.

A second observation in Figure 12 is that the nucleolar cells are quasi-equivalently confounded with the coarse speckled and centromere cells. Even though we can notice the similarity between these three cell types in Figure 8, this remarkable confusion is also partly attributable to the lack of distinguishability between the negative intensity images of these three cellular types. The results shown in Figure 12 were obtained without any data augmentation process.

In order to improve the learning capability of the network, data augmentation was applied over the dataset. In every splitting set of the training data, the cells were rotated for 360° with the step of 18°, which is similar with that used in the experiments of Bayramoglu et al. [29]. Which means that the original training set was expanded by a factor of 20, a 360° quadrant containing 20 portions of 18°. We found that augmenting the training set really improves the accuracy over the testing set. The results shown in Figure 13 were obtained by using the same network as in Figure 12 but with data augmentation.

Output Class	Target Class					
		Homo	Coarse	Fine	Nucl	Centro
	Homo	90.41	0	6.81	0	0
	Coarse	0.37	87.53	2.17	8.98	6.70
	Fine	5.51	0	91.02	1.36	0.36
	Nucl	1.59	5.18	0	84.15	2.90
	Centro	2.12	7.29	0	5.51	90.04

**Figure 13.** Confusion matrix of the results over the testing set by using the original cellular images as the inputs of one single network with data augmentation. The total accuracy of the network is 88.63%.

The total accuracy obtained by using data augmentation is 88.63%. Compared to the previous case, where we did not augment the training set, the improvement is around 3%. But, the accuracy still remains low because, as we can notice by analyzing the results shown in Figure 13, the confusion between the cells that we have previously mentioned still persists. Even though augmenting the training set up to 20 times the original one improves the learning capability of the network, the problem related to the intra-class disparity is not solved at all. The different rotation processes applied over the cells do not decrease the intra-class heterogeneity inherited from the different levels of the fluorescence illumination, the intensity levels of the cells remaining the same even after the rotation.

As previously explained in Section 2, we have used the different DWT coefficients in order to attenuate the intra-class disparity by forcing an intra-class homogenization in terms of the intensity and also the shape. The results obtained by applying the four networks in parallel, as proposed in this work, are shown in the confusion matrix depicted in Figure 14. These results were obtained without any data augmentation process.

Output Class	Target Class					
		Homo	Coarse	Fine	Nucl	Centro
	Homo	96.53	0	4.27	0	0
	Coarse	0.11	95.95	0.98	3.96	2.58
	Fine	2.51	0	94.75	0	0
	Nucl	0	1.06	0	93.11	1.26
	Centro	0.85	2.99	0	2.93	96.16

**Figure 14.** Confusion matrix of the results over the testing set by using the proposed dynamic learning without data augmentation. The total accuracy is 95.3%.

By analyzing Figure 14, we see that the confusion between the cells has been drastically attenuated. The homogeneous cells are still confounded with the fine speckled cells, but only 2.51% of them were misclassified. On the other hand, the fine speckled cells still pose a complexity problem. But, globally,



compared to the previous case, the confusion is really minimized, only 4% of them being classified as homogeneous. This improvement demonstrates the minimization of the intra-class imbalances that has occurred specifically for the fine speckled cells. As explained before, many of the negative intensity fine speckled cells were confounded with both the negative and positive intensity homogeneous cells. However, now we can notice a remarkable homogenization process that allows an improvement of the fine speckled cells' discrimination.

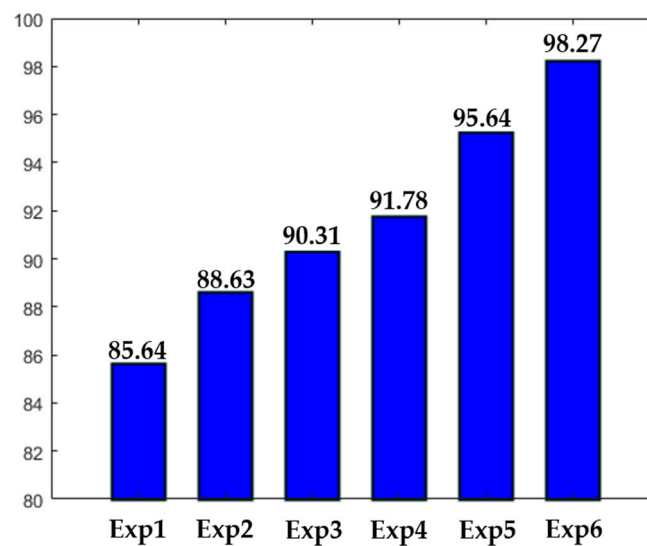
A second major observation is for the nucleolar cells. As for the previous case, these cells are still quasi-equivalently confounded with the coarse speckled and centromere cells. This cell type stills performs the least compared to the others but an impressive improvement can clearly be observed in the results. In Figure 14, the result of the nucleolar cells is marked in red. The centromere cells also remain confounded with the coarse speckled cells and also, relatively, with the nucleolar. The total accuracy of the network is 95.30%. Even though we can guess from here that the fact of increasing the learning capability of the networks by applying a huge data augmentation can correct the misclassification cases, we can already notice the remarkable intra-class homogenization that has been achieved by the parallel learning over the dataset. The results obtained when we apply data augmentation are shown in Figure 15.

Output Class	Target Class					
		Homo	Coarse	Fine	Nucl	Centro
	Homo	98.28	0	2.03	0	0
	Coarse	0	98.94	0	1.93	1.27
	Fine	1.72	0	97.97	0	0
	Nucl	0	0	0	97.48	0.04
	Centro	0	1.06	0	0.59	98.69

**Figure 15.** Confusion matrix of the results over the testing set by using the proposed dynamic learning with data augmentation. The total accuracy is 98.27%.

By applying data augmentation, the accuracy increases up to 98.27%. The confusion that we have discussed about previously still exists, but in a quite minimal manner. The nucleolar cells still perform the least but, more than 97% of them were well classified. The homogenization in terms of the intensity levels afforded the approximation coefficients, and in terms of the shape, with the different high-frequency components carried by the different details' coefficients, permit a dynamic learning that minimizes the disparity inherited from the different fluorescence illumination of the cells. The final result is that the global classification is boosted.

Different experiments were conducted in order to test different hypotheses. The results of the experiments are depicted in Figure 16. In Figure 16, "ExpN", for "Experiment N", refers to a given experimental scenario. Experiment 1 is the case discussed previously where a single network takes the original images without data augmentation. Experiment 2 is the same scenario with data augmentation applied over the training set, as previously mentioned.



**Figure 16.** Results of the different hypotheses during the experiments.

Experiment 3 refers to the case where three networks take only the detail coefficients. In this case, the accuracy is stacked at 90.31%, mainly because the intensity levels between the cells were not homogenized. Experiment 4 refers to the case where only a single network was used. But, in opposite with Experiments 1 and 2, the approximation coefficients were utilized as the inputs, rather than the original images. The accuracy is better than in Experiment 3, but the lack of homogenization in terms of the cellular shape, which is afforded by the detail coefficients, makes the accuracy low. It is necessary to mention that both Experiments 3 and 4 were conducted with data augmentation. Experiments 5 and 6 were discussed previously. The first denotes the case of the four parallel networks without data augmentation, and the latter with data augmentation.

In Table 1, we show the results of the different methods in the literature. We separate the hand-crafted features-based methods with the ones that utilize deep learning. The texture features [31], the DCT and SIFT descriptors [5], and also the LPB descriptors [6] achieve, respectively, 80.90%, 82.50%, and 85.71%. For all these methods, including the deep learning-based ones discussed in the next paragraph, we have used cross-validation in the same way in order to minimize the biases related to the hyperparameters and the training-testing split.

**Table 1.** Comparative study for the SNPHEP-2 dataset.

Method	Authors	Description	Accuracy
Hand-crafted features	Nigam et al. [31]	Texture features + SVM	80.90%
	Wiliem et al. [5]	DCT features + SIFT + SVM	82.50%
	Nosaka et al. [6]	LPB + SVM	85.71%
Deep Learning	This work	Experiment 1	85.64%
	Gao et al. [22]	5 layers CNN	86.20%
	Bayramoglu et al. [29]	4 layers CNN	88.37%
	This work	Experiment 2	88.63%
	This work	Experiment 5	95.3%
	Li et al. [23]	Deep Residual Inception Model	95.61%
	Lei et al. [27]	Cross-modal transfer learning	95.99%
	Shen et al. [28]	Use of a Deep-Cross Residual Module	96.26%
	This work	Experiment 6	98.27%

For the deep learning-based methods, we have also used cross-validation in the same way. Except for our results, where different scenarios (with or without data augmentation) were experimented, all the results shown here for the deep learning-based methods were obtained after applying data augmentation. The state-of-the-art methods proposed in [23,27,28] accomplish 95.61%, 95.99%, and

96.26, respectively, for this highly heterogeneous dataset. Our method clearly boosts the discrimination result on this particular dataset.

We have also tested our method on the most popular ICPR 2016 dataset [32]. The description of this dataset was discussed previously where we have compared it with the SHPHEP-2 dataset. The hand-crafted features-based methods perform quite poorly on this dataset. It is necessary to mention that for all the deep learning-based methods (only for this dataset, not for the previous one), we just show the results as they were reported by the authors in their works. We have experimented only the hand-crafted features-based methods because all of them were proposed at the time when this dataset was not available. The results are shown in Table 2.

**Table 2.** Comparative study for the ICPR 2016 dataset.

Method	Authors	Description	Accuracy
Hand-crafted features	Nigam et al. [31]	Texture features + SVM	71.63%
	Wiliem et al. [5]	DCT features + SIFT + SVM	74.91%
	Nosaka et al. [6]	LPB + SVM	79.44%
Deep Learning	Gao et al. [22]	5 layers CNN	96.76%
	Xi et al. [29]	VGG-like network	98.26%
	Li et al. [23]	Deep Residual Inception Model	98.37%
	Lei et al. [27]	Cross-modal transfer learning	98.42%
	Shen et al. [28]	Use of a Deep-Cross Residual Module	98.62%
	This work	Experiment 6	98.89%

One thing to notice is that the difference in accuracy between the state-of-the-art methods and our proposed one for the ICPR dataset is not significant. They perform quite similarly. But, as we can see in Table 1, the difference is a bit more noticeable on the much more heterogeneous and complex dataset, the SNPHEP-2. Also, all of the deep learning-based methods have been proposed especially for the ICPR dataset, which is why all of them provide outstanding results for it.

#### 4. Conclusions

HEP-2 cell classification is one of the most important steps for the automated diagnosis of autoimmune diseases. We have proposed a classification method that is specifically designed in order to tackle the problem of the intra-class imbalances inherited from the different fluorescence illuminations of the cells. A dynamic learning method was proposed in order to minimize the intra-class disparity by encouraging a certain homogenization in terms of the intensity levels and the shape of the cells. DWT decomposition was applied over the original images and the approximation, horizontal, vertical and diagonal details' coefficients were utilized as the inputs of four different networks that perform the learning in parallel. Finally, the features learned by the different networks were mixed up via a summation layer in order to encapsulate in one single high-level feature vector all the different characteristics that were previously learned by the different networks.

The proposed method was mainly tested on the SNPHEP-2 dataset, which exhibits a much greater heterogeneity compared to the others. The results show that the proposed dynamic learning method manages to minimize this heterogeneity by forcing the cellular images to expose some similarities in terms of their intensity and shape, which significantly boosts the discrimination results compared to the other methods. The proposed scheme was also tested on the popular ICPR dataset and it performs similarly with the actual state-of-the-art methods on that dataset.

**Author Contributions:** Conceptualization, C.V.; Data curation, C.V.; Funding acquisition, S.-H.L. and K.-R.K.; Investigation, C.V.; Methodology, C.V.; Project administration, S.-H.L. and K.-R.K.; Software, C.V.; Supervision, S.-H.L. and K.-R.K.; Validation, S.-H.L., O.-J.K. and K.-R.K.; Writing – original draft, C.V.; Writing – review & editing, C.V. and O.-J.K.

**Funding:** This research received no external funding.

**Acknowledgments:** This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2019-2016-0-00318) supervised by the IITP (Institute for Information & communications Technology Promotion), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2016R1D1A3B03931003, No. 2017R1A2B2012456), and Ministry of Trade, Industry and Energy for its financial support of the project titled “the establishment of advanced marine industry open laboratory and development of realistic convergence content”.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rigon, A.; Soda, P.; Zennaro, D.; Iannello, G.; Afeltra, A. Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose. *Cytom. B Clin. Cytom.* **2007**, *72*, 472–477. [[CrossRef](#)] [[PubMed](#)]
2. Foggia, P.; Percannella, G.; Soda, P.; Vento, M. Benchmarking hep-2 cells classification methods. *IEEE Trans. Med. Imag.* **2013**, *32*, 1878–1889. [[CrossRef](#)] [[PubMed](#)]
3. Foggia, P.; Percannella, G.; Saggese, A.; Vento, M. Pattern recognition in stained hep-2 cells: Where are we now? *Pattern Recognit.* **2014**, *47*, 2305–2314. [[CrossRef](#)]
4. Cataldo, S.D.; Bottino, A.; Ficarra, E.; Macii, E. Applying textural features to the classification of HEp-2 cell patterns in IIF images. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 689–694.
5. Wiliem, A.; Wong, Y.; Sanderson, C.; Hobson, P.; Chen, S.; Lovell, B.C. Classification of human epithelial type 2 cell indirect immunofluorescence images via codebook based descriptors. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Tampa, FL, USA, 15–17 January 2013; pp. 95–102.
6. Nosaka, R.; Fukui, K. Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns. *Pattern Recognit.* **2014**, *47*, 2428–2436. [[CrossRef](#)]
7. Huang, Y.C.; Hsieh, T.Y.; Chang, C.Y.; Cheng, W.T.; Lin, Y.C.; Huang, Y.L. HEp-2 cell images classification based on textural and statistic features using self-organizing map. In Proceedings of the 4th Asian Conference on Intelligent Information and Database Systems, Part II, Kaohsiung, Taiwan, 19–21 March 2012; pp. 529–538.
8. Thibault, G.; Angulo, J.; Meyer, F. Advanced statistical matrices for texture characterization: Application to cell classification. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 630–637. [[CrossRef](#)] [[PubMed](#)]
9. Wiliem, A.; Sanderson, C.; Wong, Y.; Hobson, P.; Minchin, R.F.; Lovell, B.C. Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching. *Pattern Recognit.* **2014**, *47*, 2315–2324. [[CrossRef](#)]
10. Xu, X.; Lin, F.; Ng, C.; Leong, K.P. Automated classification for HEp-2 cells based on linear local distance coding framework. *J. Image Video Proc.* **2015**, *2015*, 1–13. [[CrossRef](#)]
11. Cataldo, S.D.; Bottino, A.; Islam, I.U.; Vieira, T.F.; Ficarra, E. Subclass discriminant analysis of morphological and textural features for hep-2 staining pattern classification. *Pattern Recognit.* **2014**, *47*, 2389–2399. [[CrossRef](#)]
12. Bianconi, F.; Fernández, A.; Mancini, A. Assessment of rotation-invariant texture classification through Gabor filters and discrete Fourier transform. In Proceedings of the 20th International Congress on Graphical Engineering (XX INGEGRA), Valencia, Spain, 4–6 June 2008.
13. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
14. Nosaka, R.; Ohkawa, Y.; Fukui, K. Feature extraction based on co-occurrence of adjacent local binary patterns. In Proceedings of the 5th Pacific Rim Symposium on Advances in Image and Video Technology, Part II, Gwangju, South Korea, 20–23 November 2012; pp. 82–91.
15. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663.
16. Theodorakopoulos, I.; Kastaniotis, D.; Economou, G.; Fotopoulos, S. Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space. *Pattern Recognit.* **2014**, *47*, 2367–2378. [[CrossRef](#)]
17. Ponomarev, G.V.; Arlazarov, V.L.; Gelfand, M.S.; Kazanov, M.D. ANA hep-2 cells image classification using number, size, shape and localization of targeted cell regions. *Pattern Recognit.* **2014**, *47*, 2360–2366. [[CrossRef](#)]

18. Shen, L.; Lin, J.; Wu, S.; Yu, S. Hep-2 image classification using intensity order pooling based features and bag of words. *Pattern Recognit.* **2014**, *47*, 2419–2427. [[CrossRef](#)]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
20. LeCun, Y.; Huang, F.J.; Bottou, L. Learning methods for generic object recognition with invariance to pose and lighting. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, USA, 27 June–2 July 2004.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
22. Gao, Z.; Wang, L.; Zhou, L.; Zhang, J. Hep-2 cell image classification with deep convolutional neural networks. *IEEE J. Biomed. Health Inf.* **2017**, *21*, 416–428. [[CrossRef](#)]
23. Li, Y.; Shen, L. A deep residual inception network for HEp-2 cell classification. In Proceedings of the Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Québec City, QC, Canada, 14 September 2017.
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
26. Phan, H.T.H.; Kumar, A.; Kim, J.; Feng, D. Transfer learning of a convolutional neural network for HEp-2 cell image classification. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 16 June 2016; pp. 1208–1211.
27. Lei, H.; Han, T.; Zhou, F.; Yu, Z.; Qin, J.; Elazab, A.; Lei, B. A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning. *Pattern Recognit.* **2018**, *79*, 290–302. [[CrossRef](#)]
28. Shen, L.; Jia, X.; Li, Y. Deep cross residual network for HEp-2 cell staining pattern classification. *Pattern Recognit.* **2018**, *82*, 68–78. [[CrossRef](#)]
29. Bayramoglu, N.; Kannala, J.; Heikkilä, J. Human epithelial type 2 cell classification with convolutional neural networks. In Proceedings of the IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE), Belgrade, Serbia, 2–4 November 2015; pp. 1–6.
30. Xi, J.; Linlin, S.; Xiande, Z.; Shiqi, Y. Deep convolutional neural network based HEp-2 cell classification. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 77–80.
31. Nigam, I.; Agrawal, S.; Singh, R.; Vatsa, M. Revisiting HEp-2 cell classification. *IEEE Access* **2015**, *3*, 3102–3113. [[CrossRef](#)]
32. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 630–645.
34. Bengio, Y. Learning deep architecture for AI. *Foundat. Trends Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
36. Lovell, B.C.; Percannella, G.; Saggese, A.; Vento, M.; Wiliem, A. International contest on pattern recognition techniques for indirect immunofluorescence images analysis. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 74–76.
37. Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journ. Educ. Psychol.* **1933**, *24*, 417–441. [[CrossRef](#)]

