



Cystoscopy Image Classification Using Deep Convolutional Neural Networks

Seyyed Mohammadreza Hashemi^a, Hamid Hassanpour^a, Ehsan Kozegar^{b,*}, Tao Tan^c

^aFaculty of Computer Engineering and IT, Shahrood University of Technology, Shahrood, Iran

^bUniversity of Guilan, Guilan, Iran

^cEindhoven University of Technology, Eindhoven, the Netherlands

(Communicated by J. Vahidi)

Abstract

In the past three decades, the use of smart methods in medical diagnostic systems has attracted the attention of many researchers. However, no smart activity has been provided in the field of medical image processing for diagnosis of bladder cancer through cystoscopy images despite the high prevalence in the world. In this paper, two well-known convolutional neural networks (CNNs) and a multilayer neural network was applied to classify bladder cystoscopy images. One of the most important issues in training phase of neural networks is determining the learning rate because selecting too small or large learning rate leads to slow convergence, volatility and divergence, respectively. Therefore, an algorithm is required to dynamically change the convergence rate. In this respect, an adaptive method was presented for determining the learning rate so that the multilayer neural network could be improved. In this method, the learning rate is determined using a coefficient based on the difference between the accuracy of training and validation according to the output error. In addition, the rate of changes is updated according to the level of weight changes and output error. The proposed method was evaluated on 720 bladder cystoscopy images in four classes of blood in urine, benign and malignant masses. Based on the simulated results, the second proposed method (CNNs) achieved at least 17% decrease in error and increased the convergence speed of the proposed method in the classification of cystoscopy images, compared to the other competing methods.

Keywords: Cystoscopy Images, Medical Image Classification, MLP Neural Network, Adaptive Learning Rate, CNNs.

2010 MSC: 68T10

*Corresponding author

Email addresses: Smr.hashemi@shahroodut.ac.ir (Seyyed Mohammadreza Hashemi), h.hassanpour@shahroodut.ac.ir (Hamid Hassanpour), kozegar@guilan.ac.ir (Ehsan Kozegar), t.tan1@tue.nl (Tao Tan)

1. Introduction

The increasing occurrence of diseases in various countries has doubled the requirement for novel techniques for accurate diagnosis and early detection of diseases. The timely diagnosis of dangerous diseases such as cancer can save thousands of lives annually. For decades, different types of diagnostic imaging modalities have emerged to assist physicians. Since the evaluation of medical images by humans may be error-prone, complicated and time-consuming in many cases, introducing new intelligent techniques for medical image classification significantly contributes to a more precise diagnosis. Recently, the rate of bladder cancer has considerably increased in different countries. In addition, bladder cancer has been ranked ninth in the world.

The most important way to deal with cancer is using medical imaging, which allows specialists to detect the disease at early stages, thereby preventing further complications of the disease. Cystoscopy is one of the most common methods of bladder imaging. In this technique, a narrow camera called a cystoscope is inserted into the urethra and bladder so that the physician could accurately evaluate these areas in terms of possible diseases, such as the causes of bleeding and stricture, enlargement of the prostate, and bladder tumor. It is notable that the camera is linked to a monitor and the lesion can be simultaneously shown to the patients. 2

Despite the effectiveness of cystoscopy, specialists have faced challenges in interpreting its images. For example, the quality of cystoscopy images is not identical due to different imaging conditions and bladder environment. Moreover, these types of images are taken at different angles, from different areas of the bladder and in various sizes. Furthermore, poor image quality and the lack of uniform formatting of images complicate their interpretation. All of these factors lead to low diagnostic accuracy of interpreting cystoscopy images by specialists. Given the mentioned challenges, designing a computer aided diagnosis system as a second interpreter will be significantly helpful in classifying bladder cystoscopy images. The existence of such a system can enhance the ability of specialists to identify lesions.

To the best of our knowledge, no computer-aided diagnostic system has been developed in cystoscopy images. In this paper, a computer-aided system was presented for the diagnosis of blood in urine, benign, and malignant masses in cystoscopy images for the first time. The proposed classifier is based on the neural network. It is worth noting that neural networks are an array of artificial neurons, and information available is introduced and processed with a mathematical model. The main advantage of neural networks is the construction of a model using existing data. The function and accuracy of the neural networks depend on the network structure and the number of inputs [1]. The method of training is one of the most important issues in the training process of a neural network. In [2], researchers analyzed the error rates of a multi layer perceptron (MLP) network. In the event that the shape of the error levels is far from the second-degree error levels, then they will include flat and sloping regions. In this case, the backpropagation (BP) algorithm will have a low return rate at a constant rate.

Therefore, a small learning rate should be selected in order to avoid this situation. As a result, the calculation of the optimal weight vector is carried out slowly due to the small gradient in flat areas. As such, there is a need for a dynamic algorithm. Research on the dynamic change in the learning rate of the BP algorithm has been reported in [2]. Basically, all of these methods dynamically increase or decrease the learning rate with a constant coefficient based on the observations of error signal. While the positive impacts of these methods have been approved in many cases, they sometimes cause divergence. Other well-known methods to accelerate the learning phase are the second-order techniques such as Newton's method, BROYDEN-GOLD FARB-SHANNO method and a group of methods based on learning automata (LA). In some studies, researchers used variable-

structure learning automata (VSLR) as well as fixed-structure learning automata (FSLA) to find the appropriate values for the parameters of the BP learning algorithm.

In [2], LA-based methods for matching the parameters of the VLR learning algorithm for training the MLP network and also various methods for changing the learning rate dynamics were examined. In addition to the aforementioned methods, some researchers presented dynamic synapses based on chemical processes in which nonlinear differential equations are used to determine the amount of weights at a moment [3-5]. However, in the model presented in [6], the dynamic synapse weight function was estimated with a first-order differential equation. In the training phase, the parameters can be adjusted using several gradient based methods in which the gradient descent is one of the most traditional techniques to modify the parameters in the neural network. In this method, the network parameters are changed in the opposite direction to the error gradient [6]. In the conjugate gradient method, the search is performed between conjugate directions, which results in faster convergence, compared to the maximum reduction method [7]. In the Levenberg–Marquardt (LM) algorithm, both the first derivative (i.e. gradient) and the second derivative (i.e. Hessian) are used to correct the parameters. The most important advantage of this method is that the learning rate is not constant, and the algorithm can change the learning rate adaptively [8].

Rules of training with an observer called SpikeProp state that the internal status of neuron increases linearly for a small sample area that activates the neuron based on the back propagation error assuming a linear increase [9]. Moreover, it has been mathematically proven that SpikePropis also correct without the help of the linear hypothesis [10]. To improve the convergence of 5

neural networks, more training schemes can be utilized such as applying a MOMENTUM term [11] or using a combination of QUICKPROP with resilient propagation [12]. Delshad et al. used a self-adaptive learning rate in training phase but their method suffers from high computational complexity [13].

In this paper, a novel accurate approach based on MLP is introduced for cystoscopic image classification. As mentioned above, it is highly important to determine the learning rate for the sustainability of the learning process. Assigning a large learning rate can make the system unstable. Therefore, the learning rate must be minimized to some extent to avoid divergence. It should be noted that too small learning rate could prolong the training time [14]. To address this problem, an adaptive approach is proposed in the present study to select the appropriate learning rate. In this paper, two well-known CNN architectures including VGG16, ResNet50, are trained via transfer learning strategy to implement an automatic classification system for Cystoscopy Image.

The rest of this paper is organized as follows: in section 2 and 3 proposed classification methods are explained in details. The experimental results are described in section 4 and conclusions and discussions are presented in section 5.

2. PROPOSED METHOD IN MLP

In this section, the proposed method to classify the bladder cystoscopic images is described in details. The main steps of the proposed method include feature extraction, dimension reduction, and improving the neural network with an adaptive learning rate, which are described below. In this paper, two well-known CNN architectures including VGG16, ResNet50, are trained via transfer learning strategy to implement an automatic classification system for Cystoscopy Image.

2.1. FEATURE EXTRACTION AND DIMENSION REDUCTION

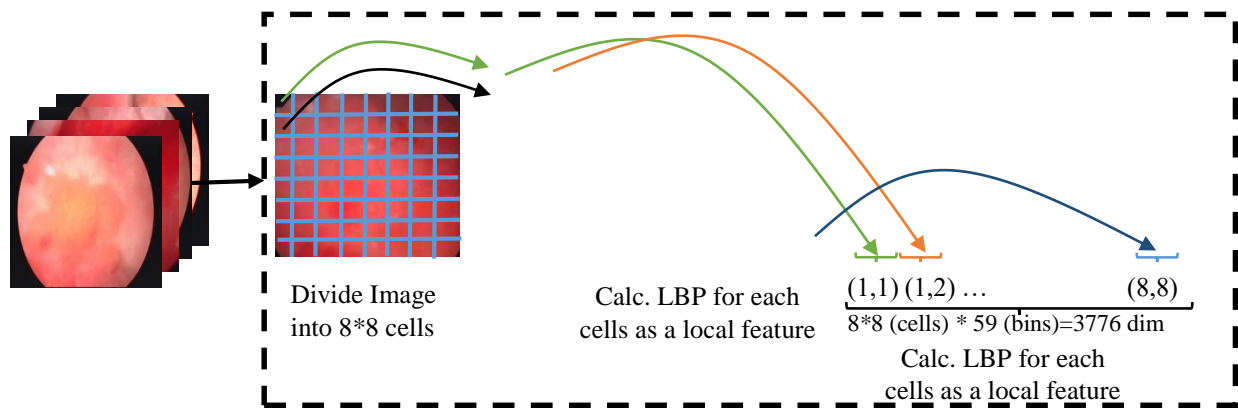


Figure 1: Feature extraction steps in the proposed technique by the LBP method

One of the most important steps in the diagnosis of diseases is feature extraction from medical images. In this respect, robust and powerful features must be exploited. In this paper, Local Binary Patterns (LBP) is used to extract such discriminative features. LBP algorithm is one of the strongest texture descriptors in machine vision. The most important advantage of this method is its robustness to rotation and high intensity changes. Moreover, the LBP operator has extremely simple computations. With regard to the high speed and simplicity of the calculations, this feature can be easily used in medical applications. Due to aforementioned reasons, this method is utilized for feature extraction in the present study.

As shown in Figure 1 the input images were divided into 8x8 cells with 32x32 dimensions, and LBP was calculated for each cell independently. Afterwards, these features are concatenated to form a single feature vector. The number of extracted features in this method is extremely high. Due to the small number of medical images which results in curse of dimensionality issue, the principal component analysis (PCA) is exploited for dimension reduction.

PCA is one of the most traditional transformations originally introduced by Pearson and independently developed by Hotelling in 1933 to analyze the structures of variance-covariance matrices and correlation coefficients. This method has shown proper performance in applications such as finding linear combinations with large or small relative variability, reducing size of data and interpreting data. The primary purpose of this analysis is decreasing the amount of data, which include a large number of variables with internal correlations, in such a way that the maximum amount of information is preserved in the data. This is through the transformation of the data (variables) into new variables called the principal components. They are independent and are prioritized based on their corresponding Eigen values.

In this paper, due to the high number of extracted features from medical images by LBP, this method was used for feature reduction. As shown in Figure 2, after computing the Eigen vectors, the new features mapped in the PCA space were calculated based on their value and importance in Eigen values.

2.2. MULTILAYER PERCEPTRON (MLP) NEURAL NETWORK

MLP is a supervised learning algorithm that essentially consists of two main paths. The forth path, where the input vector is fed to the neural network and the output of the network is calculated. In this path, the network parameters remain constant. The backpath: after production of output in

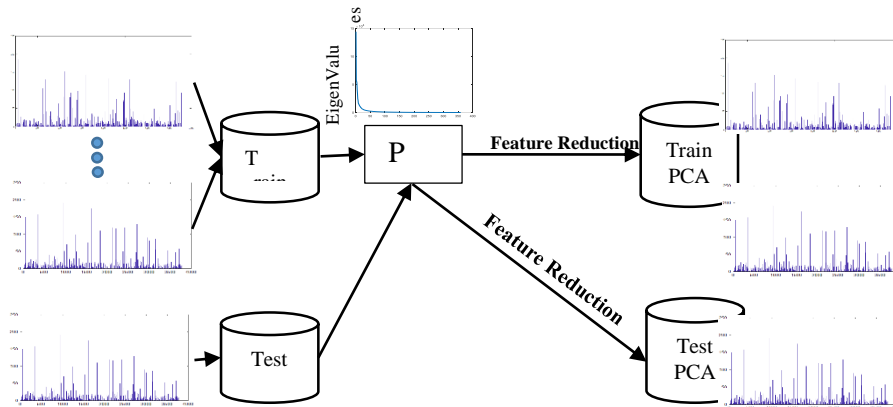


Figure 2: Feature Reduction in Proposed Method

the previous phase, the difference between the desirable output (observed) and output calculated by the network is determined. The error signals in the backpath of the output layer are backpropagated throughout the network and the network parameters are modified. The mentioned dual process is repeated until the network reaches the desirable output. The learning process stops when the error obtained is lower than a predefined threshold. The search hypothesis space in this method is the large space defined by all possible values for weights. The gradient descent method attempts to reach a suitable hypothesis by minimizing the error. However, there is no guarantee that optimization algorithm will reach an absolute minimum. The BP algorithm is usually repeated thousands of times with the same training data. Different conditions can be used to terminate the algorithm, such as stopping after repeating for a certain number of times, stopping when the error is lower than a specified value, and stopping when an error in a validation set follows a particular rule.

2.3. ADAPTIVE LEARNING RATE

After calculating the attribute from the images of the training and test set, a method should be used to classify the images. As mentioned in the previous sections, the MLP algorithm is applied for this purpose. The goal is to design a multilayer neural network, in which the learning rate is defined as a variable based on the performance of classification of cystoscopy images. In this regard, the training error is compared with validation error. If the difference is high, it means that the system's generalizability is low and the network is probably encountered with the problem of overfitting due to the high learning rate on the data. Therefore, this rate should be reduced by a coefficient. Nonetheless, if this gap (the difference between the training and validation error) is too low, the network might not be properly trained (i.e. it is underfitted) due to a low learning rate. Therefore, the learning rate should be elevated by a quantitative coefficient. In addition, the learning rate must be adjusted based on the difference in education accuracy and validation. In this article, the learning rate was evaluated based on the idea of the above mentioned idea on neural networks. In this method, the rate of changes is updated according to the number of weight changes. In this regard, more updating will be carried out if there is a high amount of changes and vice versa. Typically, the rate of learning changes varies in layers, in a way that weight changes are greater in the first iterations, and the changes gradually decrease after that. This process is performed here using the n-interval definition to estimate the error and determine the coefficient rate of learning.

According to the error created, the learning rate coefficient is updated for n stages. The process is such that if the error is greater than alpha and less than beta, the learning rate will change in the

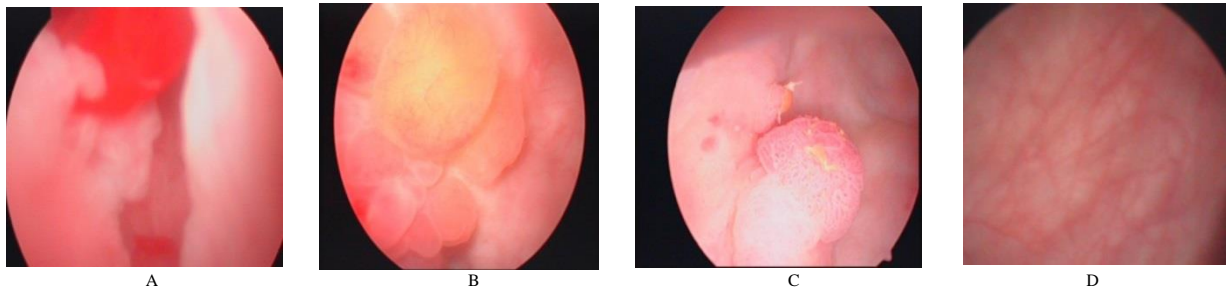


Figure 3: Examples of database images. A) An example of a malignant bladder tumor, B) An example of a benign bladder tumor, C) An example of blood bladder image, d) An example of normal bladder image

next step. Similarly, it is applied for the values of resulting errors, where the learning rate increases based on the level of error increase in each alpha. In addition, the learning rate decreased based on the amount of error decrease in each alpha, and if the resulting error is above a value of π , the learning rate value of 0.1 will be applied.

2.4. STRUCTURE OF PROPOSED METHOD IN MLP

As explained in the proposed method, the method presented in this paper consists of three important steps, including the extraction of features from cystoscopy images and feature reduction, and adaptive learning rates. In this section, the structure used for this purpose is assessed. As mentioned in the feature extraction section, the LBP method was applied to extract the feature in bladder cystoscopy images. To extract this feature, input images were divided into 8×8 cells with 32×32 dimensions. Reducing the cell size leads to the overlooking of several features. In fact, very small areas do not show a proper fit. Increased size of cells causes the improper combination of characteristics of the various regions. LBP is calculated for each area in the next step. In addition, the number of output bins was considered to be 59. In fact, $64 \times 59 = 3776$ features were extracted per each image after applying this method.

In the next step, the PCA method is employed for feature reduction with regard to the high number of features. In addition, an Eigen value chart is used to determine the number of features. As shown in Figure 4, after the first 15 valuable features in PCA, the value of other features is extremely low. Hence, only this number of features was applied for the next stage.

Specifications and parameters related to the network learning process are shown in Table 1 using the gradient descent method.

It should be noted that the number of hidden layer neurons is calculated according to the validation set. In fact, the MLP network is trained and evaluated with the number of neurons in the middle layer, and the best performance was shown by the network with eight neurons in the middle layer in the validation set. Figure 5 shows the accuracy of the network for different neurons.

Obtaining the range of changes in learning rate with regard to the training set is extremely crucial. To this end, the range of changes in the learning rate regarding to the validation set error are shown in Table 2.

3. CNNs

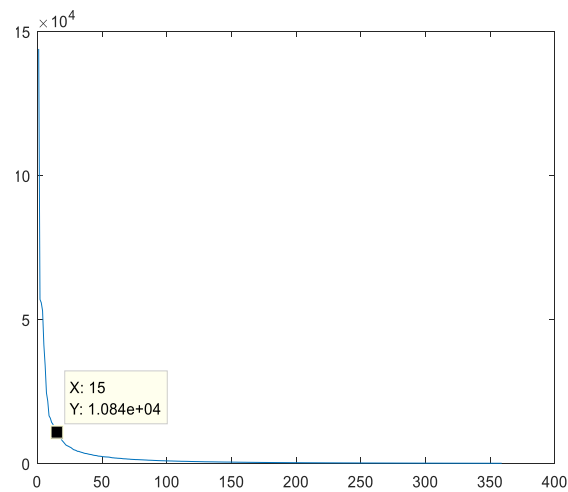


Figure 4: Details Extraction Values after Applying PCA in Cystoscopy Images

Table 1: Parameters of MLP Network for Disease Diagnosis in Bladder Cystoscopy Images

Parameters	Values
Number of latent layers	1
Number of neurons in the input layer	15
number of neurons in the latent layer	8
number of neurons in the output layer	3
maximum number of repetitions	200
The threshold for stopping learning	0.1

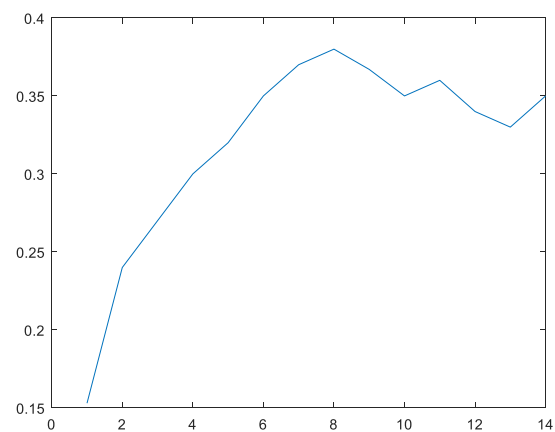


Figure 5: The Accuracy of the Network with Different Number of Neurons in the Validation Set

Table 2: Parameters of MLP Network for Disease Diagnosis in Bladder Cystoscopy Images

Range (validation error)	Learning rate
<0.5	0.01
>0.5 and <0.6	0.03
>0.6 and <0.7	0.06
Else	0.1

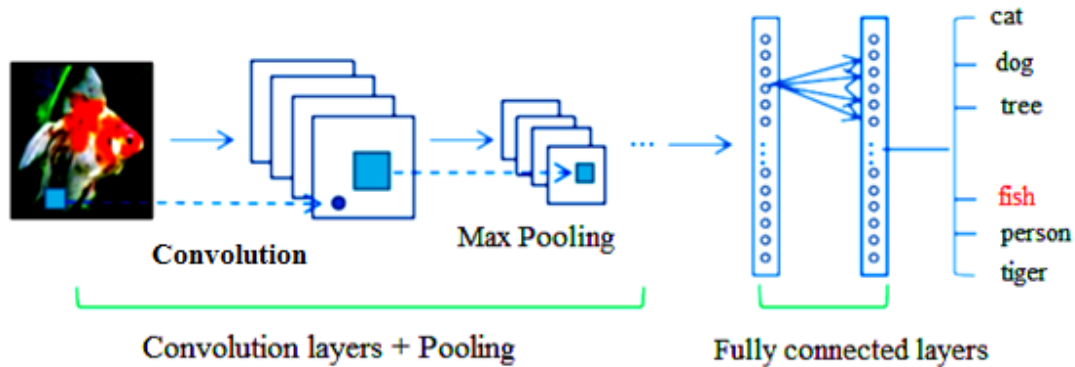


Figure 6: The overview of CNNs

In this paper, two well-known CNN architectures including VGG16, ResNet50, are trained via transfer learning strategy to implement an automatic classification system for Cystoscopy Image.

Artificial neural networks are computational models which follow the general principles of brain functions with their neurons and their binding synapses. The main feature of these networks is the ability to direct the supervised learning process. During this process, neural networks are prepared to learn the behavior of a system, using labeled data which contains pairs of system inputs and outputs.

The CNNs[15] are the evolved form of artificial neural networks, the main focus of which is on applications with repetitive patterns in different areas of model space, especially for image recognition. Using the methodology applied in their layering, these networks are able to extract features in a way that is called unsupervised feature learning in which the discriminative features are extracted using convolution filters whose weights are adjusted during training phase. For image recognition applications, several basic architectures of CNNs have been successfully designed and used for complex visual image tasks.

Two well-known architectures of the CNNs, which are evaluated in the present study are explained in the following:

3.1. VGG16

Simonian et al. developed VGG architecture at ILSVRC 2014 competition and reached the error 7.3%, with the view that deeper architecture improves the accuracy[16]. The VGG 16 network, as illustrated in, contains 16 convolution layers or 16 parametric layers. Compared to AlexNet, VGG uses much more convolutional layers with smaller filter size within each layer. Moreover, it consists of consequent convolutional layers which increase the field of view of filters. For example, convolving two 3×3 filters results in a 5×5 filter having wider field of view. However, VGG has too many

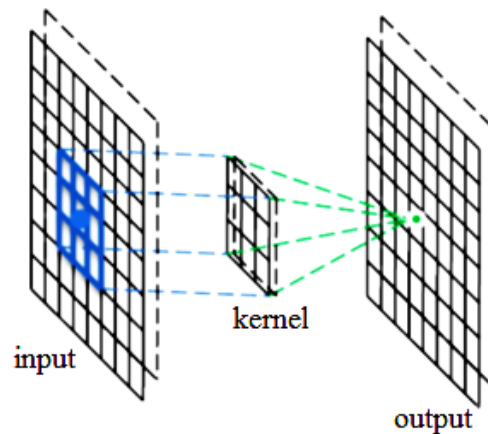


Figure 7: convolutional operation

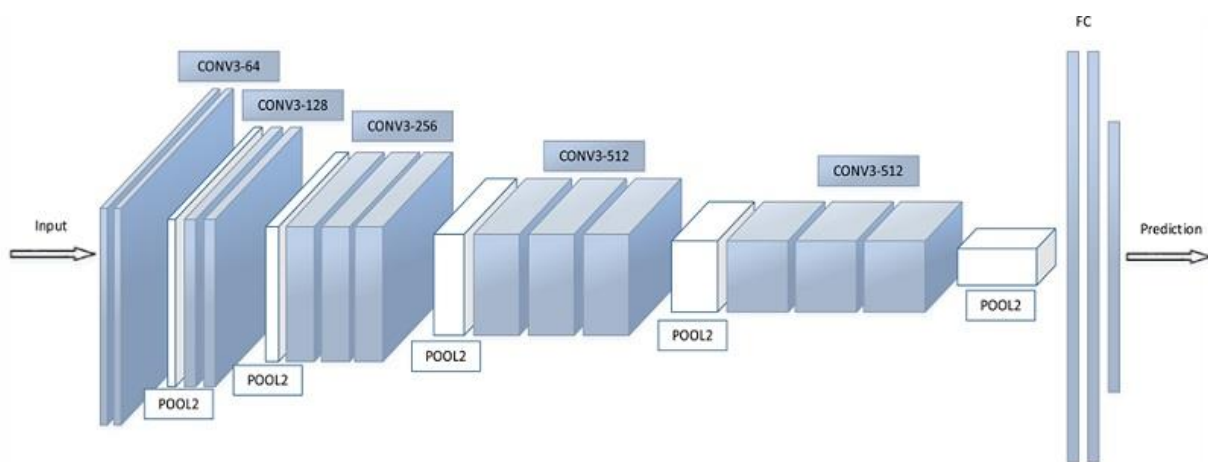


Figure 8: VGG Architecture [16]

parameters to be trained that probably encounter the network with over-fitting problem especially for small data sets.

3.2. ResNet 50

A common approach to develop CNN architectures is that increasing the depth of the CNN (i.e. increasing the number of convolution layers) without increasing the number of parameters to be trained leads to extracting more semantic and discriminative features. Therefore, making CNNs deeper generally improves the classification accuracy. ResNet is regarded as one of the greatest architectures introduced by He et al.[17]and was able to succeed in many competitions. This architecture with a depth of 152 layers is known as the deepest architecture until that year, and is still one of the deepest architecture by now, indicating that how deep a deep learning architecture can be. ResNet (which stands for Residual Networks) includes several residual modules stacked on each other, which actually form the main building of the ResNet architecture.[17] demonstrates an overview of the residual module. The main novelty of ResNet is applying specific pointers called skip pointers which directly connects convolution layers to their top layers. They proved that skip pointers improves the optimization of weights of the network. Simply, the residual module has two ways and can perform a

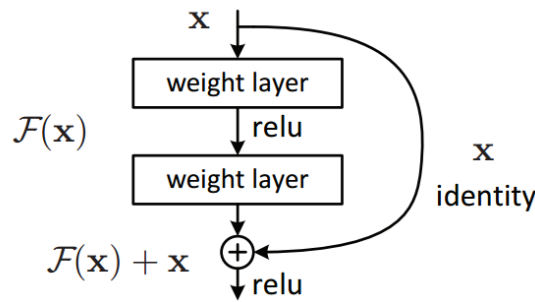


Figure 9: ResNet block [17]

series of operations on the input or reject all these steps. These residual modules are stacked together to form a complete network.

4. EXPERIMENTS AND RESULTS

In this section, the parameters and values used for the proposed method as well as the database of cystoscopy images of the bladder are introduced. In the next step, the parameters related to the training and test are explained in details. It is notable that all implementations were performed on a computer with the specification of Core (TM) i7 M620 CPU, 6GB of memory with a Python programming language. In addition, it is worth noting that all results of the presented work are obtained after 20 runs. In fact, the mean of the results obtained is reported from 20 times of random selection and different implementation.

4.1. DATABASE

The database presented in this study was prepared from medical cystoscopy images of a medical center in the Netherlands, which contains classes of normal, blood in urine, benign masses and malignant masses. Database images are from different angles and various areas of the bladder taken by the cystoscope. The size and quality of the images vary due to different situations. Totally, each class consists of 180 images of size 476x540 pixels. From the database images, one-third was considered for test and two-thirds was regarded for training and validation. Due to the limited number of samples, the strategy of k-fold cross-validation (kFCV) is exploited. In this paper, k is equal to 10. Figure 3 illustrates the sample images of this database.

4.2. METHODS COMPARISON

In order to determine the effect of change in the learning rate, we first assume the learning rate to be constant. Here, the learning rate was considered 0.1. The error of the fixed method for the set of training is shown in Figure 10. As shown, the best error rate was 56.67 in the training stage.

In the next stage, the effect of adaptively changing the learning rate on the cystoscopy images was evaluated. Figure 11 shows the error rate of the neural network with a variable and adaptive learning rate to the error rate of the validation. As observed in the figure, the proposed method achieved an error rate of 49.44 which outperforms the fixed method in training phase. Another effective parameter in network training is the use of a powerful method for the initial weighing of the network.

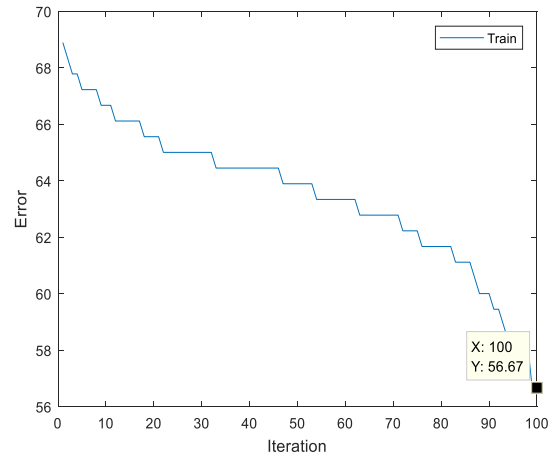


Figure 10: Error Rate of MLP Network with Constant Learning Rate

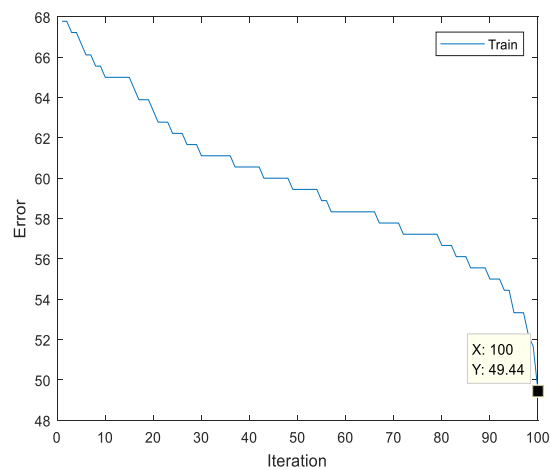


Figure 11: Error Rate of MLP Network with Variable Learning Rate

Table 3: Comparison of Methods for Classification of Bladder Cystoscopy Images (level of accuracy) after 20 repetitions

Criteria	Vgg16	Resnet	MLP Method	
			Randomized weighing and constant learning rate	Randomized weighing and variable learning rate
Mean	58.68	54.37	38.18	41.57
Standard deviation	0.52	0.67	0.59	2.28

Figure 11 shows the results of the proposed method with the variable learning rate. As shown in the figure, this method had a better performance in the training set, compared to the previous methods. The proposed method's error rate was 41.44 in the training stage.

The proposed method was then evaluated on the test set. Table 3 shows the results of comparing the proposed method and competing methods.

As shown in Table 3, the proposed method had a better performance in classifying the cystoscopy images, compared to the other techniques.

5. CONCLUSIONS AND DISCUSSIONS

In this paper, a dynamic and powerful method was presented based on smart methods for classification of cystoscopy images. LBP was applied in the first proposed method followed by principal component analysis for feature extraction and reduction, respectively. Afterwards, the MLP neural network was exploited for training and assessing the classifier of bladder cystoscopy images. In this paper, the second proposed method to classify the bladder cystoscopic images in two well-known CNN architectures including VGG16, ResNet50, are trained via transfer learning strategy to implement an automatic classification system for Cystoscopy Image. The results obtained were assessed on a database of 720 images. According to the results, the second proposed method outperforms other competing methods in classification of cystoscopy images.

REFERENCES

- [1] Krishnan, G.S., K. Sivanarulselman, and P. Betty, SURVEY ON BRAIN TUMOUR DETECTION AND CLASSIFICATION USING IMAGE PROCESSING.
- [2] M.R.Meybodi," Results on a Strongly absolutely expedient learning automata, " Proc. Of OU Inference Conf. 86, ed. D. R. Mootes and R. Butrick, athens, Ohio: Ohio university prees. Pp.197-204.
- [3] J. S. Liaw, and T. W. Berger, "Dynamic Synapse: A New Concept of Neural Representation and Computation", Hippocampus, Vol. 6, pp. 591-600, 1996.
- [4] H. H. Namarvar, J. S. Liaw, and T. W. Berger, "A New Dynamic Synapse Neural Network for Speech Recognition", Proc. IEEE International Joint Conference on Neural Networks (IJCNN), Vol. 4, pp. 2985-2990, 2001.
- [5] H. Markram, D. Pikus, A. Gupta, and M. Tsodyks, "Potentialfor Multiple Mechanisms, Phenomena and Algorithms for Synaptic Plasticity at Single Synapses", Neuropharmacology, Vol. 37, pp.

489-500, 1998.

- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation", D. Rumelhart and J. McClelland: editors. *Parallel Data Processing*, Vol. 1, Chapter 8, Cambridge, MA, MIT Press, pp. 318-362, 1986. W10466
- M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, Boston, MA, PWS Publishing, 1996
- [7] M. T. Hagan, and M. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm", *IEEE Trans. Neural Networks*, Vol. 5, No. 6, pp. 989-993, 1994.
- [8] Bohte, S.M., Kok, J.N., La Poutré, H., "Error-backpropagation in temporally encoded networks of spiking neurons", *Neurocomputing*, 48, 17-37 (2002).
- [9] Yang, J., Yang, W., Wu, W., "A remark on the error-backpropagation learning algorithm for spiking neural networks", *Applied Mathematics Letters*, 25 (??), 1118-1120 (2012).
- [10] Xin, J., Embrechts, M., "Supervised learning with spiking neural networks", In: *Proceedings of International Joint Conference on Neural Networks*, IEEE, Piscataway, USA, 1772-1777 (2001).
- [11] McKennoch, S., Liu, D., Bushnell, L.G., "Fast modifications of the SpikeProp algorithm", In: *Proceedings of the International Joint Conference on Neural Networks*, IEEE, Piscataway, USA, 3970-3977(2006).
- [12] Delshad, E., Moallem, P., Monadjemi, S.A.H., "Spiking neural network learning algorithms: Using learning rates adaptation of gradient and momentum steps", In: *5th International Symposium on Telecommunications (IST)*, IEEE, 944-949 (2010).
- [13] Sha, D., Bajic, V.B., "An Optimized Recursive Learning Algorithm For Three-Layer Feedforward Neural Networks For MIMO Nonlinear System Identifications", *Intelligent Automation and Soft Computing*, Vol. 19, No.11, pp. 1-15, 2011.
- [14] LeCun, Y., & Bengio, Y. (1998). Convolutional networks for images, speech, and time series. In *The handbook of brain theory and neural networks* (pp. 255–258). MA, USA: MIT Press.
- [15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Retrieved from <https://arxiv.org/pdf/1409.1556.pdf>
- [16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Computer Vision and Pattern Recognition* (pp. 770–778). IEEE. W10466