Al-Qadisiyah Journal of Pure Science

Volume 26 | Number 1 Article 22

1-7-2021

Eye Diseases Classification Using Back Propagation with Welch Estimation Based-Learning Rate

Hanaa M. Ahmed
University of Technology, 52 St., Baghdad, Iraq, 110113@uotechnology.edu.iq

Shrooq R. Hameed
University of Technology, 52 St., Baghdad, Iraq, 0110607@student.uotechnology.edu.iq

Al-Qadisiyah Journal of Pure Science Vol.(26) Issue (1) (2021) pp. Comp. 22–30

Al-Qadisiyah Journal of Pure Science

ISSN(Printed): 1997-2490 ISSN(Online): 2411-3514

DOI: /10.29350/jops.

http://qu.edu.iq/journalsc/index.php/JOPS

Eye Diseases Classification Using Back Propagation with Welch Estimation Based-Learning Rate

Authors Names

a. Hanaa M. Ahmed

b. Shrooq R. Hameed

Article History

Received on: 22/10 /2020
Revised on: 24 /12/2020
Accepted on: 10 /1/ 2021

Keywords:

Eye Diseases Classification
Back Propagation
Cyclic Learning Rate
Welch based Power Spectral Estimation.

DOI: https://doi.org/ 10.29350/jops.2021.26. 1.1231

ABSTRACT

Human eye disease classification adapted by many pieces of research in the last decade due to the importance of eye organ for humans and the evolution in classification techniques. External eye diseases classified in this paper using back propagation with non-linear cyclic learning rate based on welch estimation as a step for improving the performance of back propagation. As a result, Classification accuracy achieved in this paper is (93.22%).

1. Introduction

The human eye is the most important organ among the other four senses, so, early detection of external eye diseases becomes an important issue. Pattern Recognition has a wide area of applications. One of these applications is a medical application.

Several pieces of research conducted to classify one or more eye diseases, and different performance accuracy achieved according to different preprocessing and classification techniques used. Some of the recent works in this field are: (Patwari, Arif et al. 2011) detected ocular cataract disease in which four digital images of that disease with three images of normal eye utilized. The images of the ocular cataract converted to grayscale, in turn, binary format of that image obtained to detect intensity variation. Lastly, edge detection and circularity detection was utilized and compared with normal images that act as ground truth. The accuracy achieved was 94.96%. while, (Gunay, Goceri et al. 2015) used 18 healthy eye images with 12 eye images infected by conjunctivitis disease, these images were segmented by GrapCut algorithm to extract the sclera region. After that, features extracted including RGB thresholding, Hessian matrix to extract blood vessels, and Gray-Level Co-occurrence Matrix fed to classifier model to get accuracy as 96%. (Grassmann, Mengelkamp et al.

[a]University of Technology, 52 St., Baghdad, Iraq, E-Mail: 110113@uotechnology.edu.iq

[b]University of Technology, 52 St., Baghdad, Iraq, E-Mail: 0110607@student.uotechnology.edu.iq

2018) classified 13 classes of Age-related Macular Degeneration (AMD) disease classified from color fundus images which preprocessed then trained with different Convolutional Neural Networks (CNN), finally, random forest built based on the result of CNNs. The model implemented on different datasets with definite signs of late and early AMD and the accuracy achieved is 84.2%. whereas, (Malik, Kanwal et al. 2019) classified eye diseases depend on symptoms recorded and categorized hierarchically by medical experts. Various classification techniques used like Decision Tree (DT), Random Forest (RF), Naive Bayesian (NB), and Artificial Neural Network (ANN) algorithms. The researchers concluded that tree-based methods give better accuracy than ANN does and the performance positively correlated with the amount of available data. The first level diagnosis achieved higher accuracy followed by second and third levels diagnosis respectively. (Akram and Debnath 2020) extracted eye region from a face image then classified with two approaches CNN and Support Vector Machine (SVM). The purpose is to recognize seven eye diseases including cataracts, trachoma, conjunctivitis, corneal ulcer, ectropion, periorbital cellulitis, and Bitot's spot of vitamin A deficiency. The classification accuracy achieved using CNN is 98.79% while classification accuracy achieved using SVM is 96.13%. (Ahmad and Hameed 2020) proposed Hierarchal Multi-label Classification (HMC) to classify seven eye diseases (Cataract, Conjunctivitis, Corneal-Ulcer, Stye, Dilated-Pupil, Sub-Conjunctival, and Pterygium) in a Hierarchal manner. The seven eye diseases sub-divided into four classes according to the part of the eye where the disease causes a problem. The four classes further sub-divided into a number of sub-classes equal to the number of eye diseases belong to that class. Vector of Color and texture features extracted from eye diseases dataset fed to the first level of HMC and its prediction, in turn, fed to second level of HMC and the result provides final classification prediction. The accuracy achieved is 75.7142%.

Machine Learning proof that it can produce very good accuracy in diagnosing different medical conditions like chest X-ray, breast cancer, and many others. (Ting, Pasquale et al. 2019) Back Propagation (BP) is one of the most popular supervised learning techniques in which data trained to reduce the error between predicted output and actual output. For the first time, the input pattern with the connection weights and biases fed to BP. Small random values assigned to connection weights. Then, the activation of each neuron in each layer is the weighted sum of the input pattern and its connection weights and bias computed at each neuron in each layer. The neuron output of the output layer compared to the actual output of that neuron to estimate the error. Finally, this error used to adjust connection weights and biases. The adjustment process at the output layer uses different parameters, which are learning rate (LR), momentum, activation function derivative, error term, and the input pattern, while error adjustment at the hidden layer uses the error of the output layers in addition to these parameters. This process repeated iteratively until the error reaches a minimum. (Hossain, Rahman et al. 2013) instead of using fixed LR that tuned experimentally to achieve the best BP performance, LR values changed gradually between upper and lower limits during BP training in a cyclic manner. (Smith 2017) many functions used for developing LR include Sine, Cosine, Exponential, Polynomial, and Triangular functions. (Wu, Liu et al. 2019) all of these functions change the LR value linearly from upper bound to lower bound or visa verse. Contribution of this paper is that welch based Power Spectral Density (PSD) was conducted to generate LR values that decreases or increases between upper and lower bounds non-linearly.

This paper classifies seven different external eye diseases (Cataract, Conjunctivitis, Corneal-Ulcer, Dilated-Pupil, Stye, Sub-Conjunctival, and Pterygium) in addition to the normal eye as seen in خطأ! لم

يتم العثور على مصدر المرجع. Color and texture features extracted and normalized to feed into BP. The extracted features trained by BP.

This paper organized as follows: In section two, the related work discussed. In section three, the learning rate, and its development discussed in detail. While the extracted features explained in section four. Meanwhile, Results & discussions presented in section five. Finally, conclusions illustrated in section six.

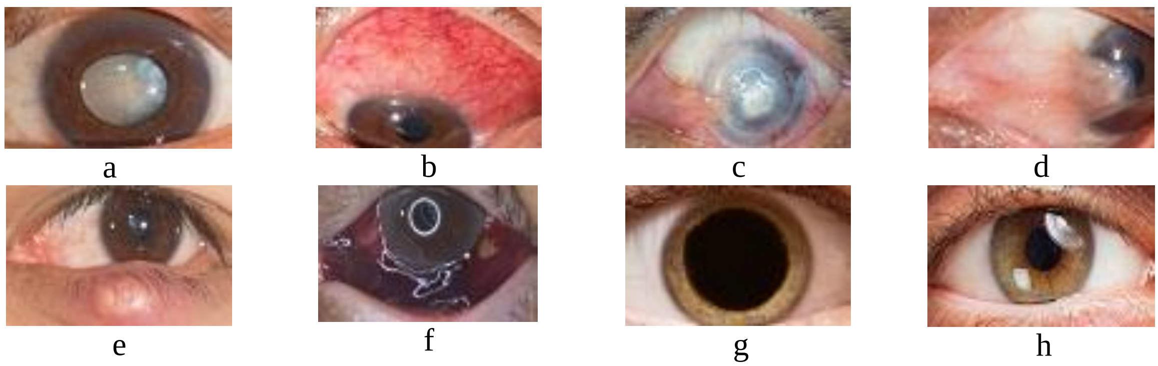

Fig. 1- (a) Cataract Disease; (b) Conjunctivitis Disease; (c) Corneal-Ulcer Disease; (d) Pterygium Disease; (e) Stye Disease; (f) Sub-Conjunctival Disease; (g) Dilated Pupil Disease; (h) Normal Eye.

2.Related Work

Cyclical Learning Rate (CLR) is a new mechanism that is used to improve the performance of BP. Smith (Smith 2015) proposed a method to tune LR within a range during the BP training phase. Smith estimated the range of LR experimentally that was used in turn to train BP with two linear CLR (triangular and exp_range) policies. These policies implemented on different datasets like CIFAR-10, AlexNet, GoogleNet, and the results compared. While in (Loshchilov and Hutter 2016) Loshchilov proposed a restart mechanism for generating learning rate values that improved the convergence rate in gradient-based optimization methods. Restart mechanism implemented using the Cosine function with four different instantiations on CIFAR-10 and CIFAR-100. After that Bo-Yang Hsueh in (Hsueh, Li et al. 2019) proposed a new LR scheduler Hyperbolic Tangent Decay and compared its result with the Cosine scheduler. He also analyzed the performance of both step decay and exponential decay separately and compared the results. Yanzhao in (Wu, Liu et al. 2019) provided a comparative study of 13 LR scheduler (fixed, fixed step size, variable step size, inverse time, polynomial, triangular, triangular_exp, Sine, Sine_exp, Cosine) through finding optimal LR values, optimal LR ranges, and optimal epoch numbers for a schedule update. He also proposed three evaluation metrics in terms of cost, utility, and robustness in comparing the result of 13 LR policies. Finally, (Hameed and Ahmed 2020) classified the eye diseases mentioned in خطأ! لم يتم العثور على مصدر المرجع. utilizing from parabola scheduler and obtain classification accuracy equal to 89.83%.

3. Learning Rate

LR is one of BP hyper-parameters that determines the step size BP should take to move downhill towards local minima where the difference between predicted output and actual output as minimum as possible or zero. (Smith 2017, Wu, Liu et al. 2019) to achieve the best BP performance LR tuned appropriately. It is not easy to Tune the parameter of LR.

If LR set too small BP process train slowly, on the other hand, if LR is too high BP process oscillates widely prevents the training model from improving. (Wilson and Martinez 2001, Wu, Liu et al. 2019) so adaptive LR is considered in which separate LR is computed for each neuron or layer based on the gradient of a neuron. (Smith 2015) the main disadvantage of adaptive LR is that its computation cost-effective. (Smith 2017) instead, CLR comes to overcome the previous problems in which LR value changes within a range (upper bound and lower bound). (Smith 2015, Smith 2017) There are different CLR policies, like triangular policy in which LR increases linearly from lower bound to upper bound. While, exp_range policy by which LR decline from upper bound to lower bound exponentially. However, welch based-PSD policy uses parabola function to generate LR values that decline from upper bound to lower bound non-linearly, and Hann policy depends on cosine function to generate LR values increases from lower bound to upper bound.(Smith 2015). In this paper welch policy was conducted to generate LR values.

3.1 Welch Based-PSD Learning Rate

Welch method is a non-parametric periodogram-based PSD. It based on three main steps. In the first step, the signal is broken into overlapping segments. In the second step, the modified periodogram of these segments computed. While in the third step, the mean of these periodograms is calculated. The result is spectral estimate is obtained.(THOMAS, John et al. 2015) welch estimation method illustrated as follows: let $\boldsymbol{X(i), i = 0, \dots, N-1}$ is a sample of length $\boldsymbol{N}$ segmented into $\boldsymbol{K}$ blocks of length $\boldsymbol{L}$ with $\boldsymbol{D}$ overlapping samples such that the $\boldsymbol{K}$ block or segment is defined as (Welch 1967):

$$\mathbf{X_k = X(i + (k-1)D)} \tag{3.1}$$

Where $\mathrm{i} = 0, \dots, \mathrm{L} - 1$.

To calculate welch estimation, once, the signal samples divided into overlapped blocks as shown in Fig. 2. Each block convoluted with a data window. Data window has a length equal to the length of blocks (Welch 1967)

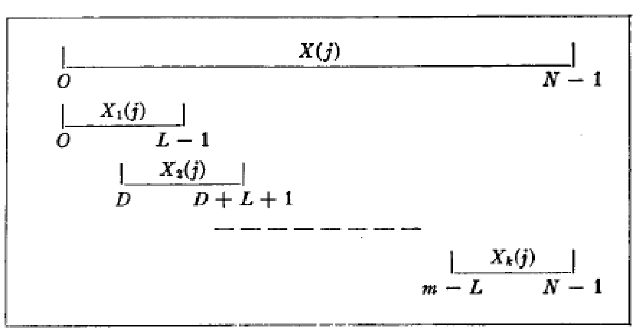

Fig. 2. signal sample segmentation.(Welch 1967)

Fourthly, window function also known as the tapering function utilized to shape and cut the segment of the signal of length N for spectral analysis. There are many windowing functions such as rectangular, Hann, Hamming, Blackman, Kaiser, and Bartlet. Hann's window selected in this paper. It also called the raised-cosine window since it depends on cosine function; it has a cosine-liked shape as shown in

. and its Equation as (Jwo, Wu et al. 2019)

$$win(m) = 0.5\left(1 - \cos\left(2\pi\frac{m}{M+1}\right)\right)R_N(m) \tag{3.2}$$

Where $m = 0,1,\dots,M$.

Its advantage is that it minimizes side lobes (Jwo, Wu et al. 2019). Fifthly, Fast Fourier Transform (FFT) is computed for each windowed block since the length of the sampling signal is not multiple of the length of signal period, so, applying FFT on it make many undesired components and spectral leakage in the discrete spectrum. Therefore, FFT is applied to each windowed block separately to produce a frequency spectrum with no spectral leakage (Jwo, Wu et al. 2019) as follows (Welch 1967):

$$f_k(n) = \frac{1}{L}\sum_{j=0}^{L-1} X_k(j) \times win(j) \times e^{-2kijn/L} \tag{3.3}$$

Where $i = \sqrt{-1}$. After applying Eq.3.4 k modified periodograms computed from Eq.3.5.

$$p_k(B_n) = \frac{L}{U}|f_k(n)|^2 \tag{3.4}$$

Where $k = 0,1,\dots,K$, and

$$B_n = n/L \tag{3.5}$$

Where $n = 0,1,\dots,L/2$. And

$$U = \frac{1}{L}\sum_{j=0}^{L-1} win(j)^2 \tag{3.6}$$

Lastly the spectral estimation of welch is the mean of modified periodograms.

$$\hat{s}(B_n) = \frac{1}{K}\sum_{k=1}^{K} p_k(B_n) \tag{3.7}$$

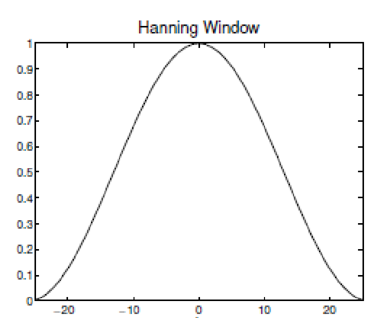

Fig. 3. The shape of the Hanning window. (Stoica and Moses 2005)

4. Feature Extraction

Features extracted from the data set of 590 samples taken from https://www.shutterstock.com/ consist of seven eye diseases in addition to a normal eye. Two kinds of features are utilized in this paper color histogram features and texture features since each kind of eye disease are distinguished by its color and texture as seen in Fig. 1. Each image sample converted from RGB color space to HMMD color space and image color quantized into 32-bin based on the MPEG-7 standard to decrease the size of memory requirement. Moreover, Law's texture features also utilized since they detect particular texture features like spots, waves, ripples, intensity levels, edges by multiplying five vectors with each other to form 25-2D masks convoluted with the intensity channel of the image. Each vector responsible for enhancing one of the features mentioned previously. As a result, 18 texture features extracted. Texture features concatenated with color features and normalized to form input patterns to BP.

5. Results & Discussions

BP has LR and momentum as hyper-parameters. To improve the performance of BP, one of these hyper-parameters can be adapted for this purpose, so BP with PSD-based welch LR applied on a dataset of 590 samples contains seven eye diseases plus normal eye as shown in Fig. 1. Instead of the fixed value assigned to LR, 32 LR values are generated by using PSD-based welch as explained in sub-section 3.1 the LR values decay shown in Fig. 4.

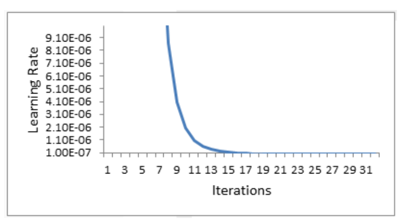

Fig. 4. PSD based Welch-based LR for one cycle.

The training phase starts with an upper limit of LR and during the training process, the value of LR decreases non-linearly until it reaches the lower limit this represents a complete cycle then a new cycle started by reassigning the upper limit to the LR value and the process was continuing until training phase end. Fig. 5 shows the declination of training error during the training phase. The weights of BP are random numbers generated with Gaussian distribution and normalized within the interval $\left[\frac{-1}{\sqrt{N}}, \frac{1}{\sqrt{N}}\right]$ where N equal to the number of input layer neurons. The obtained classification accuracy is promising as the confusion matrix illustrated in Table 1. The classification accuracy is 93.22%. While f-score=94.47763%, sensitivity is 93.10875%, and specificity is 98.9438%.

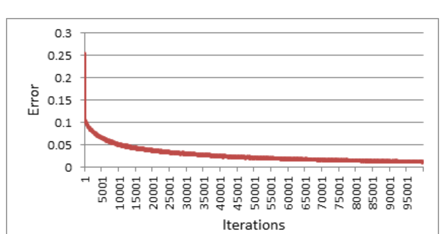

Fig. 5. The error of Training BP with LR based welch estimation.

		Predicted							
	Eye Disease	**Cataract**	**Conjunctival**	**Corneal Ulcer**	**Stye**	**Pterygium**	**Sub Conjunctival**	**Dilated Pupil**	**Normal**
Actual	**Cataract**	8	0	0	0	0	0	0	0
	Conjunctival	0	22	0	0	1	1	0	0
	Corneal Ulcer	0	1	8	0	0	0	0	0
	Stye	0	2	0	6	0	0	0	0
	Pterygium	0	3	0	0	25	0	0	0
	Sub Conjunctival	0	0	0	0	0	12	0	0
	Dilated Pupil	2	0	0	0	0	0	9	0
	Normal	0	0	0	0	0	0	0	20

Table 1. Confusion Matrix for Classification result of BP with LR based Welch Estimation.

6. Conclusions

In this, paper seven external eye diseases in addition to normal eye classified in presence of welch-based LR. Since LR values generated by PSD-based welch are non-linear. the accuracy achieved is very good. The comparative study present in

Authors	**Method used**	**Accuracy (%)**
(Patwari, Arif et al. 2011)	Image Matching	94.96
(Gunay, Goceri et al. 2015)	*	96
(Grassmann, Mengelkamp et al. 2018)	CNN	84.2
(Malik, Kanwal et al. 2019)	DT	85.81
(Malik, Kanwal et al. 2019)	NB	81.53
(Malik, Kanwal et al. 2019)	RF	86.63
(Malik, Kanwal et al. 2019)	NN	85.98
(Akram and Debnath 2020)	CNN	98.79
(Akram and Debnath 2020)	SVM	96.13
(Ahmad and Hameed 2020)	HMC	75.7142
(Hameed and Ahmed 2020)	BP with Parabola LR	89.83
Our	BP with Welch-based LR	93.22

Table 2.

Authors	**Method used**	**Accuracy (%)**

(Patwari, Arif et al. 2011)	Image Matching	94.96
(Gunay, Goceri et al. 2015)	*	96
(Grassmann, Mengelkamp et al. 2018)	CNN	84.2
(Malik, Kanwal et al. 2019)	DT	85.81
(Malik, Kanwal et al. 2019)	NB	81.53
(Malik, Kanwal et al. 2019)	RF	86.63
(Malik, Kanwal et al. 2019)	NN	85.98
(Akram and Debnath 2020)	CNN	98.79
(Akram and Debnath 2020)	SVM	96.13
(Ahmad and Hameed 2020)	HMC	75.7142
(Hameed and Ahmed 2020)	BP with Parabola LR	89.83
Our	BP with Welch-based LR	93.22

Table 2. Comparative Study.

References

1. Ahmad, H. M. and S. R. Hameed (2020). Eye Diseases Classification Using Hierarchical MultiLabel Artificial Neural Network. 2020 1st. Information Technology To Enhance e-learning and Other Application (IT-ELA, IEEE.

2. Akram, A. and R. Debnath (2020). "An automated eye disease recognition system from visual content of facial images using machine learning techniques." Turkish Journal of Electrical Engineering & Computer Sciences **28**(2).

3. Grassmann, F., et al. (2018). "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography." Ophthalmology **125**(9): 1410-1420.

4. Gunay, M., et al. (2015). Automated detection of adenoviral conjunctivitis disease from facial images using machine learning. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), IEEE.

5. Hameed, S. R. and H. M. Ahmed (2020). "Eye Diseases Classification using Back Propagation with Parabola Learning Rate." Al-Qadisiyah Journal Of Pure Science **26**(1): 1–9-1–9.

6. Hossain, M., et al. (2013). "Implementation of back-propagation neural network for isolated Bangla speech recognition." arXiv preprint arXiv:1308.3785.

7. Hsueh, B.-Y., et al. (2019). Stochastic gradient descent with hyperbolic-tangent decay on classification. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE.

8. Jwo, D.-J., et al. (2019). Windowing Design and Performance Assessment for Mitigation of Spectrum Leakage. E3S Web of Conferences, EDP Sciences.

9. Loshchilov, I. and F. Hutter (2016). "SGDR: stochastic gradient descent with restarts. CoRR abs/1608.03983 (2016)." arXiv preprint arXiv:1608.03983.

10. Malik, S., et al. (2019). "Data Driven Approach for Eye Disease Classification with Machine Learning." Applied Sciences **9**(14): 2789.

11. Patwari, M. A. U., et al. (2011). Detection, categorization, and assessment of eye cataracts using digital image processing. The First International Conference on Interdisciplinary Research and Development, Thailand.

12. Smith, L. N. (2015). "No more pesky learning rate guessing games." CoRR, abs/1506.01186 **5**.

13. Smith, L. N. (2017). Cyclical learning rates for training neural networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE.

14. Stoica, P. and R. L. Moses (2005). "Spectral analysis of signals."

15. THOMAS, E., et al. (2015). "Power Spectral Density Computation using Modified Welch Method." IJSTE-International Journal of Science Technology & Engineering **2**(4).

16. Ting, D. S. W., et al. (2019). "Artificial intelligence and deep learning in ophthalmology." British Journal of Ophthalmology **103**(2): 167-175.

17. Welch, P. (1967). "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms." IEEE Transactions on audio and electroacoustics **15**(2): 70-73.

18. Wilson, D. R. and T. R. Martinez (2001). The need for small learning rates on large problems. IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222), IEEE.

19. Wu, Y., et al. (2019). Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks. 2019 IEEE International Conference on Big Data (Big Data), IEEE.