Eye Diseases Classification Using Back Propagation With Parabola 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

QJPS

Al-Qadisiyah Journal of Pure Science

Eye Diseases Classification using Back Propagation with Parabola Learning Rate

Authors Names	ABSTRACT
a. Hanaa M. Ahmed b. Shrooq R. Hameed **Article History** Received on: 19/10 /2020 Revised on: 21/11 /2020 Accepted on: 1/12/ 2020 **Keywords:** *Eye Diseases Classification* *Back Propagation* *Cyclic Learning Rate* *Parabola Function.* **DOI:**https://doi.org/ 10.29350/qjps2021.26.1.1220	Human eye disease classification adapted by many pieces of research in the last decade due to the importance of (eye) optical organ for humans and the evolution in classification techniques. Seven external eye diseases plus normal eye classified in this paper using back propagation with linear cyclic learning rate-based parabola function. The classification accuracy achieved in this paper is (89.83%).

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.

ªUniversity 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 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) The novelty of this paper comes from using the Parabola function for generating LR values utilizing from parabola function.

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 Fig. 1. Color and texture features extracted and normalized to feed into BP. The extracted features trained by BP. The contribution of this paper is that the parabola function proposed to be a new cyclic function utilized in this paper as a learning rate.

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, proposed method as well as block diagram explained in fifth Section, Results & discussions presented in section six. Finally, conclusions illustrated in section seven.
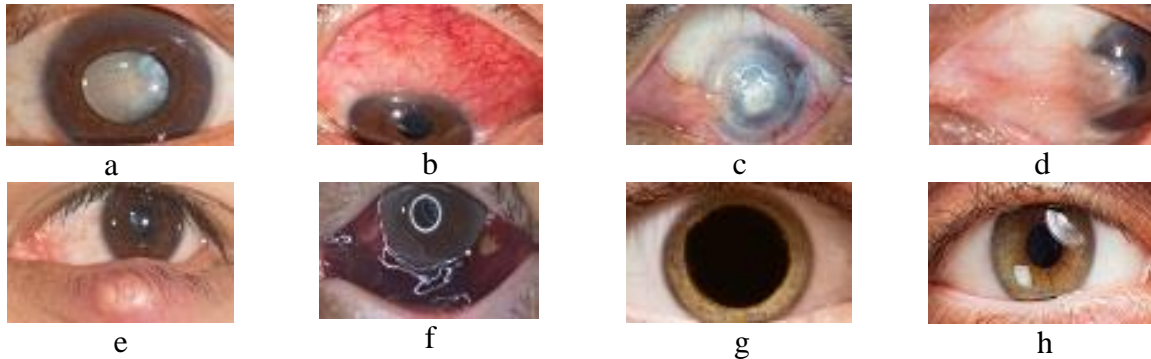

Fig. 1- (a) Cataract Disease; (b) Conjunctivitis Disease; (c) Corneal-Ulcer Disease; (d) Pterygium Disease; (e) Stye Disease; (f) Sub-Conjunctival Disease; (g) Dilated Pupil Disease; (h) Normal Eye.

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 (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. Finally, 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.

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 range (upper bound and lower bound). (Smith 2015, Smith 2017) CLR uses a worm restart mechanism to reassign the initial value of LR after several iterations or epochs are completed. This number of iterations or epochs forms a cycle. (Loshchilov and Hutter 2016, Loshchilov and Hutter 2016, Hsueh, Li et al. 2019) LR uses scheduling annealing as an LR policy as follows:(Wu, Liu et al. 2019)

$$g(t) = |u - l| \times f(t) + min(u, l) \tag{3.1}$$

t, Indicates current iteration while u refers to upper bound of LR and l refers to lower bound of LR. Finally,$f(t)$ the function used to change the value of LR smoothly between upper and lower bounds. Many functions used to replace $f(t)$ such as exponential function, cosine function, sine function, triangular function, and hyperbolic-tangent decay function. In this paper, the parabola function is utilized that explained in the following sub-section.

3.1 Parabola Learning Rate

Parabola is a set of points on the plane that have the same distance from the line called directrix and a fixed point called focus which does not lie on directrix as shown in Fig. 2. (Adams and Essex 1999)
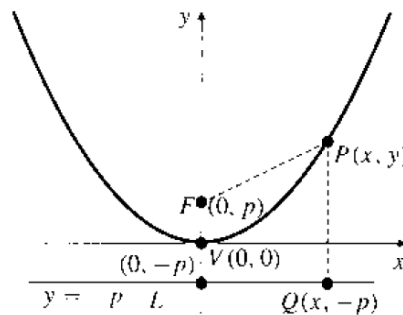

Fig. 2- The parabola $4py = x^2$ with focus $f(0, p)$and directrix $y = -p$.

The parabola function shifted either horizontally or vertically by c units. To shift the parabola equation horizontally by c units adds c units to x-axis. If $c < 0$ parabola function shifted to right while if $c > 0$ parabola function shifted to left. On the other hand, to shift the parabola function vertically by c units, add c units to y-axis. If $c < 0$, the parabola function shifted downward, while if $c > 0$ parabola function shifted upward as explained in
Fig. 3.(Adams and Essex 1999)
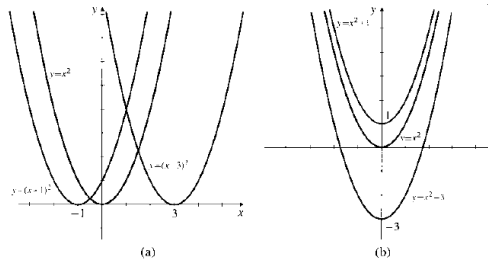
Fig. 3

Fig. 3.(Adams and Essex 1999)

Fig. 3- (a) parabola function $y = x^2$ shifted horizontally; (b) parabola function $y = x^2$ shifted vertically.

The parabola function utilized by this paper shifted horizontally to right by one unit to become parabola function and its equation and shape as:
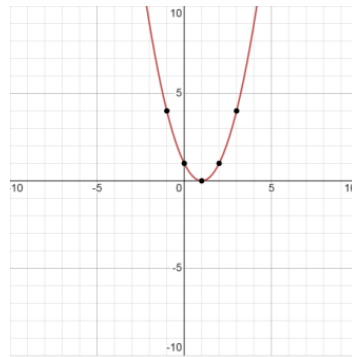
$$f(t) = (t - 1)^2 \qquad (3.2)$$

Fig. 4- Parabola Function.

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. Proposed Method

BP with parabola LR applied on a dataset of 590 samples contains seven eye diseases plus normal eye as shown in Fig 5 partitioned into two subsets according to the 80:20 ratios. The first subset is the training set and the second is the testing set. The feature vectors achieved in section four normalized and classified using BP with a parabola LR schedule. The weights of BP generated randomly 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. Fig. 5 illustrates the steps that both training and testing phases passed through. The weights obtained from the training phase utilized to evaluate the testing set.
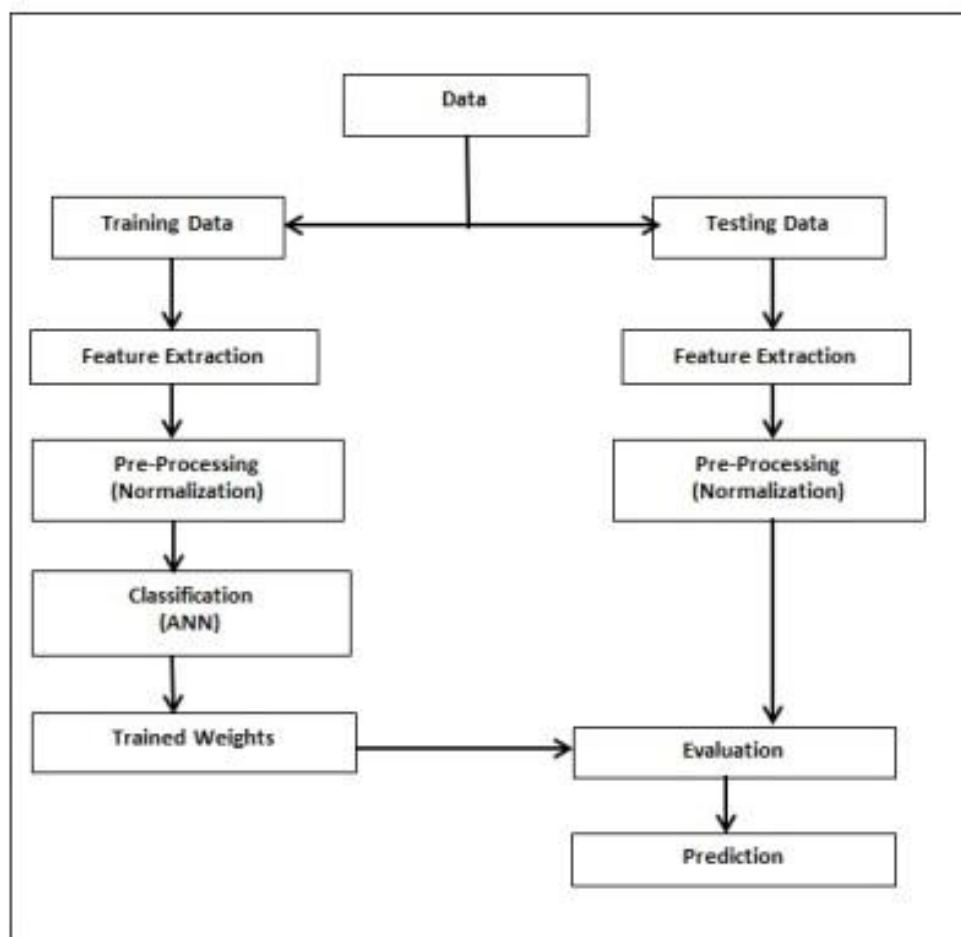

Fig. 5- Block Diagram of the Proposed Method.

6. 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 parabola 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, the parabola LR scheduler as explained in Eq.(1) follows the parabola

equation from maximum to minimum limits causes learning rate values to change smoothly and linearly during the training phase as shown in Fig. 5Instead of the fixed value assigned to LR, the parabola LR scheduler as explained in Eq. (1) follows the parabola equation from maximum to minimum limits causes learning rate values to change smoothly and linearly during the training phase as shown in

. The learning rate starts with a value of 0.05 and decreases at each iteration until it reaches 0.006 after 32 iterations are completed. At this point, one cycle is finished, and then a new cycle re-begins. This process continued until BP stopping criteria met.
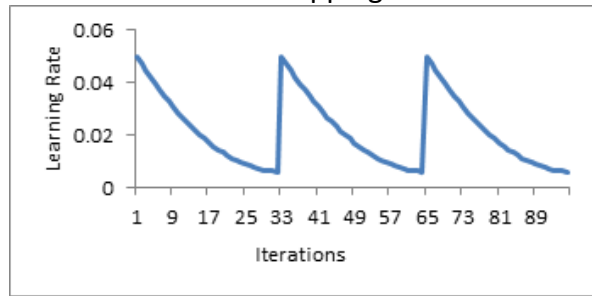

Fig. 6- Shape of Parabola LR for Three Cycles.

At the beginning of the training phase, the training error declined dramatically, and then the training error decreased smoothly as shown in Fig. 7. The error obtained at the first iteration of BP training was 0.25 then, the error decreased dramatically after a few iterations to reaches 0.1. After that, the error declines smoothly to reach 0.0202 after 100000 iterations. The normalized feature vectors classified using BP with parabola LR schedule. The upper limit of LR is 0.05 while the lower limit is 0.006. The training phase starts with the upper limit of LR and during the training process, the value of LR gradually decreases until it reaches the lower limit at this time a single cycle is complete. The length of a cycle is 32 iterations. Then the upper limit reassigned to the LR according to Eq. 1 and the process was continuing until the training phase end.
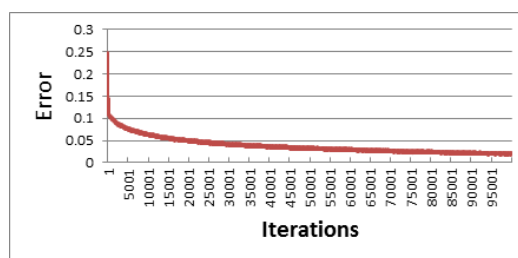

Fig. 7- Error of Training BP with Parabola LR.

The classification accuracy is 89.83% while f-score=88.7767% and sensitivity are 98.564%, specificity is 88.754%, and its Area Under Curve (AUC) is 94.1973%. Table 1 explains the confusion matrix of the classification result of this method. As seen in Table 1. Normal eye classified 100% correctly, while classification accuracy of cataract, corneal ulcer, and sub-conjunctival over 90%. On the other hand, stye, dilated pupil, and pterygium classification accuracy are greater than 80%, finally, conjunctivitis classification accuracy is more than 70%. The last four eye diseases affect the overall accuracy negatively, in particular, the last one.

	Predicted							
Eye disease	**cataract**	**conjunctivitis**	**Corneal ulcer**	**stye**	**pterygium**	**Sub conjunctival**	**Dilated pupil**	**Normal**
Cataract	17	0	1	0	0	0	0	0
conjunctivitis	0	10	1	0	1	1	0	0
Corenal -ulcer	0	0	10	0	0	1	0	0
stye	0	1	0	8	0	0	0	0
petrygium	0	2	0	1	21	0	0	0
Sub-conjunctivial	0	1	0	0	0	9	0	0
Dilated-pupil	2	0	0	0	0	0	9	0
Normal	0	0	0	0	0	0	0	22

(Leftmost label: **Actual**)

Table 1. Confusion Matrix for Classification result of BP with Parabola LR.

7. Conclusion

In this paper, tuning LR with parabola function presented. Parabola function is simpler than any other function like sine, cosine, or hyperbola-tangent functions. Parabola function produces good results and enhances BP performance significantly. The comparative study present in Table 2

Table 2. Comparative Study.

Authors	Method used	Accuracy (%)
[5]	Image Matching	94.96
[6]	CNN	84.2
[7]	DT	85.81
[7]	NB	81.53
[7]	RF	86.63
[7]	NN	85.98
our	BP with parabola LR	89.83

References

[1] Adams, R. A., and C. Essex (1999). Calculus: a complete course, Addison-Wesley Boston.

[2] 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).

[3] 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).

[4] 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.

[5] 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.

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

[9] Loshchilov, I. and F. Hutter (2016). "Sgdr: Stochastic gradient descent with warm restarts." 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

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

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

[15] 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.

[16] 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.