COMPARATIVE ANALYSIS OF DEEP LEARNING OPTIMIZATION METHODS FOR IMAGE CLASSIFICATION

[1]Amita Khatana, [2]V.K Narang, [3]Vikas Thada
[1]Research Scholar, [2]Director, [3]Associate Professor
[1]Department of Computer Science and Engineering,
[1]Amity University, Gurgaon, India

Abstract: In the field of computer engineering, image classification is emerging very rapidly for classifying images into different categories. A allot of approaches have been made to enhance the performance of image classification itself. Artificial neural network, support vector machine, decision tree are most popular approaches used for image classification. This paper focuses on one of the methods of image classification that is the deep artificial neural network. The deep artificial neural network consists of two or more hidden layers to extract the features of the image. The performance of artificial neural network depends upon its internal parameters such as weights, bias value, learning rate, batch size, number of neurons, hidden layers etc. To improve and updates the internal parameters of neural network, different optimization methods are used. The paper shows the comparative analysis of different deep learning optimization methods such as stochastic gradient descent, ADAM, RMSProp, Adamax, Adadelta.

IndexTerms - Deep learning, ADAM, stochastic gradient descent, Adamax, Adadelta, image classification, MNIST.

I. INTRODUCTION

Image classification is one of the techniques of image processing in which an image is a process to classify into a pre-defined category. Different methods are used for the classification purpose such as artificial neural network, SVM, K-Means etc. The artificial neural network is one of the most commonly used methods for the classification. The artificial neural network is a computational model that inspired from human brain. In the human brain, different neurons are connected to each other through 'axons' and transmit electrical signals from one neuron to another neuron. Similarly, the artificial neural network consists of a computation unit 'neuron' that performs all the computations and transmits useful information from one layer to another layer. Each neuron has an activation function to calculate the output of that neuron. The most popular types of activation functions are sigmoid, ReLU, leaky ReLU, tanh (Hyperbolic tangent), softmax etc.

Artificial neural networks are broadly categorized into two categories i.e. single layer feed forward network and multi-layer feed forward network. In single layer feed forward network only one computational layer exists. Input layer is directly connected to the output layer of the model. In multi-layer feed forward network, one or more hidden layers exist. The neurons of hidden layers are connected with neurons of forward layer. There are different models of multilayer feed forward networks used for image classification such as deep artificial neural network, recurrent neural network, CNN, DBN etc. Deep artificial neural network is one of the famous models of deep learning. This model consists of one or more hidden layers to perform computations. In this research work we have implemented multilayer feed forward deep neural network and applied supervised learning for classification.

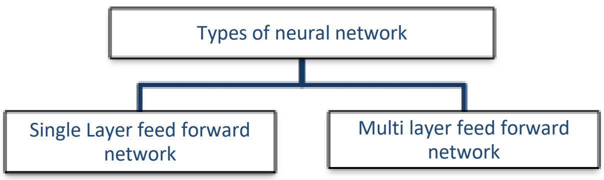


Fig 1: Types of neural network

The paper is organized into 6 sections as follows: section 1 covers introduction of artificial neural network, section 2 discussed about the proposed system and dataset used for the classification, section 3 explains design and implementation steps of deep neural network, section 4 shows comparative analysis of different parameters like optimization algorithm, neurons, time etc and section 5 ends up with the conclusion and future scope of research work.

II. Proposed System and Dataset

A) Proposed System

Deep Learning is an emerging technique of machine learning. The main aim of the machine learning algorithms is to trained models so that they can be generalized the problem. Nowadays, deep learning is broadly used to solve the image classification problems. In this paper, a deep neural network model is proposed to perform digit classification. The MNIST dataset was used to perform handwritten digit classification. A deep neural network was designed to compute the performance of different optimization algorithm such as SGD, ADAM, Adagrad, Adadelta, Adamax, Rmsprop on different parameters. This research work carried on keras framework with theano backend. Keras provided facility of predefined optimizers. A comparative analysis was shown on different parameters like a number of neurons, execution time, epochs, dataset etc.

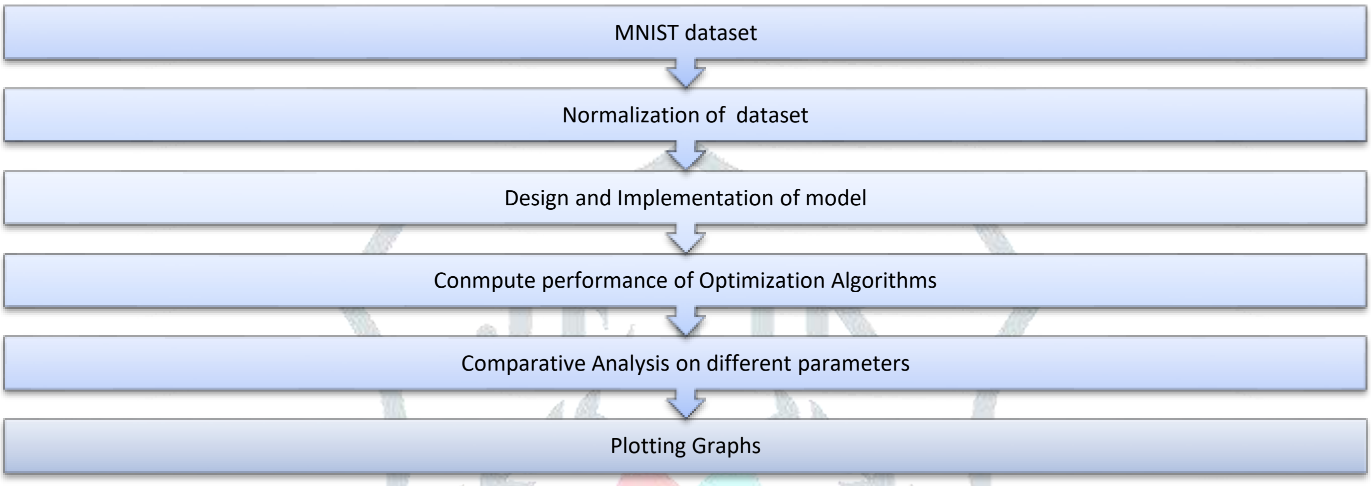


Fig 2: A block diagram of proposed system

B) Dataset

The MNIST dataset is used for the classification purpose; this was published by Yann LeCun at New York University. The dataset was downloaded from their official website http://yann.lecun.com/exdb/mnist/. The dataset divided into two sets: a training set and testing set. The training set contains 60,000 images of ten classes (0-9). The testing set contains 10,000 labeled test set. Table 1 shows the distribution of the dataset. The pre-processing was done on the dataset that includes normalization of data. In normalization process, a one hot encoding scheme applied on the dataset. Fig 3 shows the sample of MNIST dataset [31].

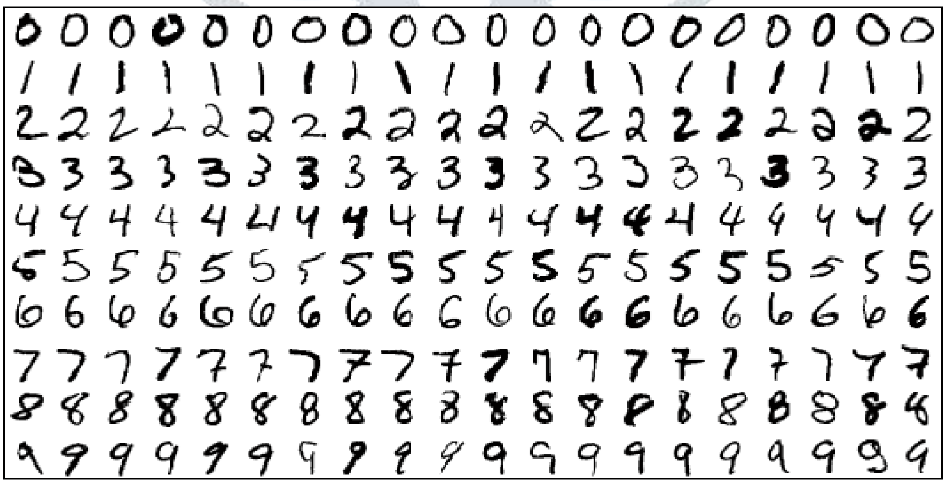


Fig 3: MNIST images [31]

The One-Hot Encoding (OHE) scheme is used for the multiclass classification. According to this scheme it shows a Boolean representation of integer digit. In which each pixel value was divided by 255 that is a maximum value of pixel for the sample image [31]. This type of representation is called one-hot encoding (OHE). If value of original label is 0.0 then its equivalent to OHE: [1.0.0.0.0.0.0.0.0.0.]

Table 1: Distribution of MNIST Dataset

Digit	Number of Training Samples	Number of Testing Samples
0	5923	980
1	6742	1135
2	5958	1032
3	6131	1010
4	5842	982
5	5421	892
6	5918	958
7	6265	1028
8	5851	974
9	5949	1009
Total	60,000	10,000

III. DESIGN AND IMPLEMENTATION OF MODEL

First we import the library needed to create a model. After that we define the layers and number of neurons of model in Keras. The following network consists of two hidden layers with 512 neurons. The input of network is 784-dimensional arrays. In this research work number of neurons and number of hidden layer and optimization methods are varied in neural network. A comparative analysis is performed after varying different parameters of neural network.

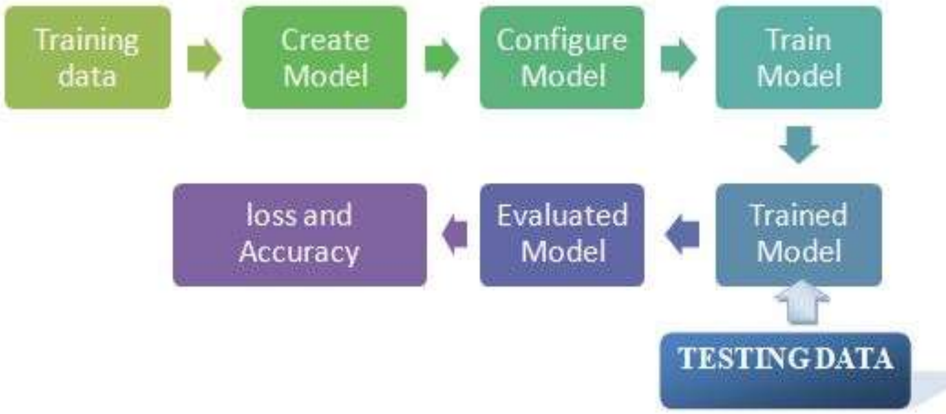


Fig 4: Workflow diagram for implementing neural network

B) Model Configuration

In this section, we define the optimization method such as stochastic gradient descent, RMSProp, ADAM, etc and specify the loss type. There are two loss types one is binary cross entropy which is used for binary classification and another is categorical cross entropy which is used for multiclass classification. After that we specify the metric, there are three matrices that we can define: Accuracy, Precision, and Recall. For this model we specify accuracy metric.

- **Accuracy**: This is the proportion of correct predictions with respect to the targets [2].
- **Precision:** This denotes how many selected items are relevant for a multiclass classification [2].
- **Recall:** This denotes how many selected items are relevant for a multiclass classification [2].

Objective function for binary classification

= -t log (p) – (1-t) log (1-p)

Objective function for Multiclass classification

$$= -\sum_{i-1}^{n} ti \log(pi)$$

Where,

t = target output, p = predicted output

C) Model Training

The configuration model is trained using fit () function in Keras. The fit () function specify different parameters used for the training of network. This function specifies following the parameters such as number of epochs, dataset, and batch size used for the training.

Epochs: Epochs is the number of times the model processed the training dataset. The Each iteration gives some loss and accuracy value. The used optimizer updates the weights value in each iteration so that the objective function and loss can be are minimized.

Dataset: Dataset contains training data and one-hot-encoding training labels, testing data and one-hot encoded testing labels

Batch size: The size of training instance is called batch size.

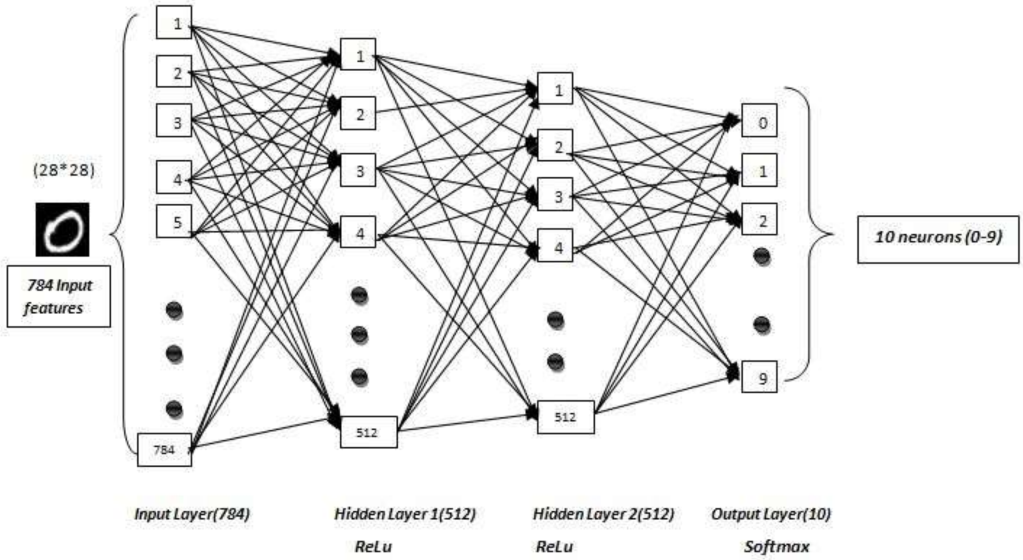


Fig 5: Deep neural network

D) Model Evaluation

Now the trained model, is get ready for processing testing data. The evaluated model checks the model's performance on the whole testing data. It show the cost and accuracy value for testing dataset. The parameters and hyperparameters affect the performance of the network.

Regularization

Regularization comes with a dropout technique that helps to overcome the over fitting problem. In over fitting problem network become biased to the training dataset. So, in dropout technique we randomly drop some neuron units from hidden layers. This dropout technique improves the performance of the network

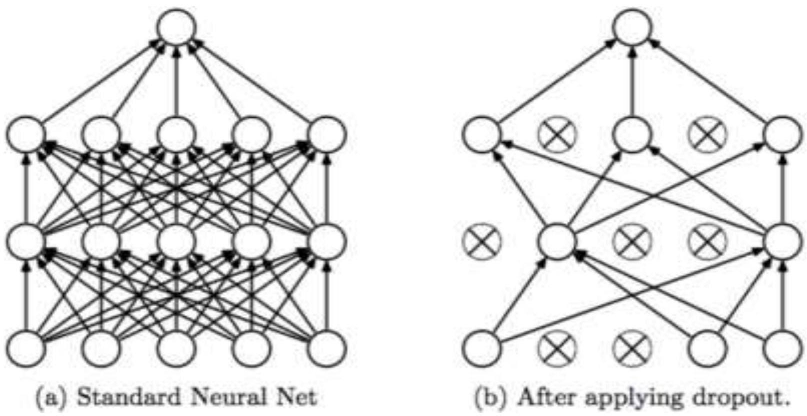


Fig 6: Dropout Technique [2]

IV. COMPARATIVE ASALYSIS

This section, discusses a comparison of all the optimization algorithms performances seen so far. This is done through tables and graph.

1) The accuracy of Optimization methods

This section shows performance of different optimization methods on each epoch. The given figure shows, how the testing accuracy of neural network varies with number of epochs. The testing accuracy archived from ADAM and RMSProp is nearly equal, but and different from stochastic gradient descent. Learning is not about training longer but use of smart optimization techniques that increases the accuracy of learning in terms of prediction and classification.

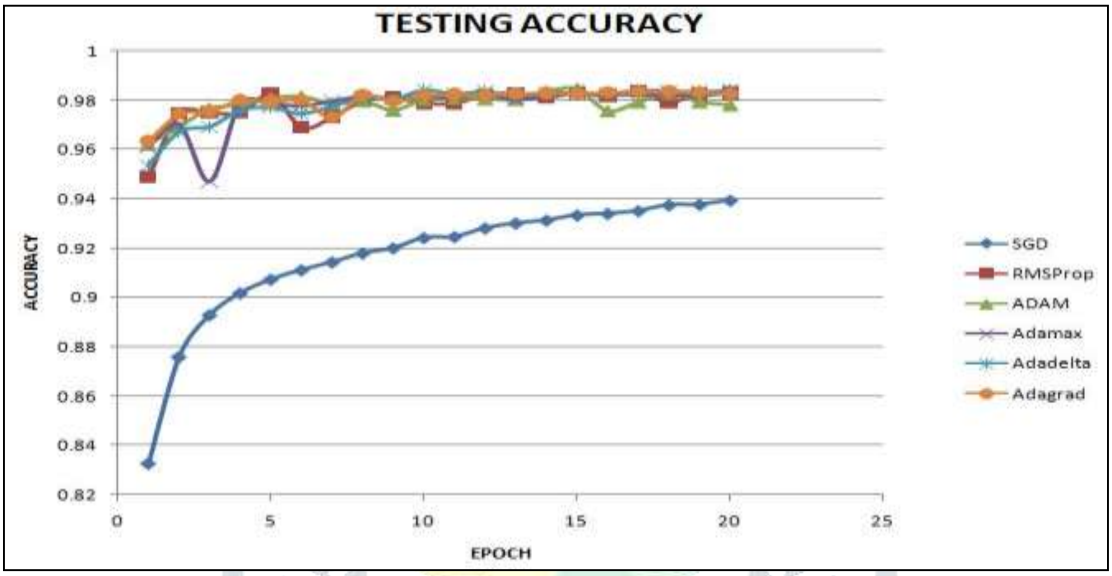
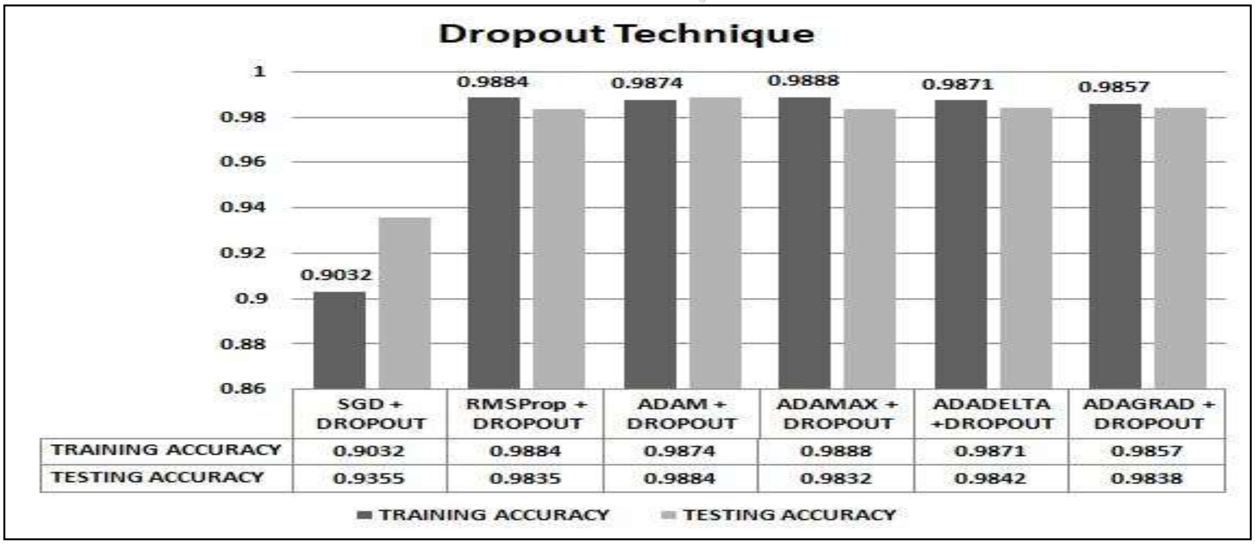


Fig 7: Comparative analysis of Testing Accuracy

2) Comparative analysis of dropout technique

This section compares the performance of optimization methods with and without dropout technique. The figure8 and figure 8 shows a clear view, how dropout technique affects the accuracy of the deep neural network. According to the graph dropout technique increase the testing accuracy except for SGD optimizer for MNIST dataset.



	SGD + DROPOUT	RMSProp + DROPOUT	ADAM + DROPOUT	ADAMAX + DROPOUT	ADADELTA +DROPOUT	ADAGRAD + DROPOUT
TRAINING ACCURACY	0.9032	0.9884	0.9874	0.9888	0.9871	0.9857
TESTING ACCURACY	0.9355	0.9835	0.9884	0.9832	0.9842	0.9838

Fig 8: Performance analysis with dropout technique

Table 2: Performance of dropout technique

OPTIMIZAION	Testing Accuracy (without dropout)	Testing Accuracy (with dropout)	Remark (+% & -%)
SGD	93.95%	93.55%	Decrease (-0.40%)
RMSProp	98.28%	98.28%	No change
Adam	97.82%	98.41%	Increase (+1.41%)
Adamax	98.28%	98.32%	Increase (+0.04%)
Adadelta	98.41%	98.42%	Increase (+0.01%)
Adagrad	98.27%	98.32%	Increase (+0.05%)

Table 2 shows the affect of dropout technique on the performance of different optimization algorithm and figure 9 shows SGD testing accuracy on different dropout values. The curve indicates that how the accuracy values vary with dropout values. At dropout=0.4 SGD optimizer give best result but at 0.5 it shows lower accuracy.

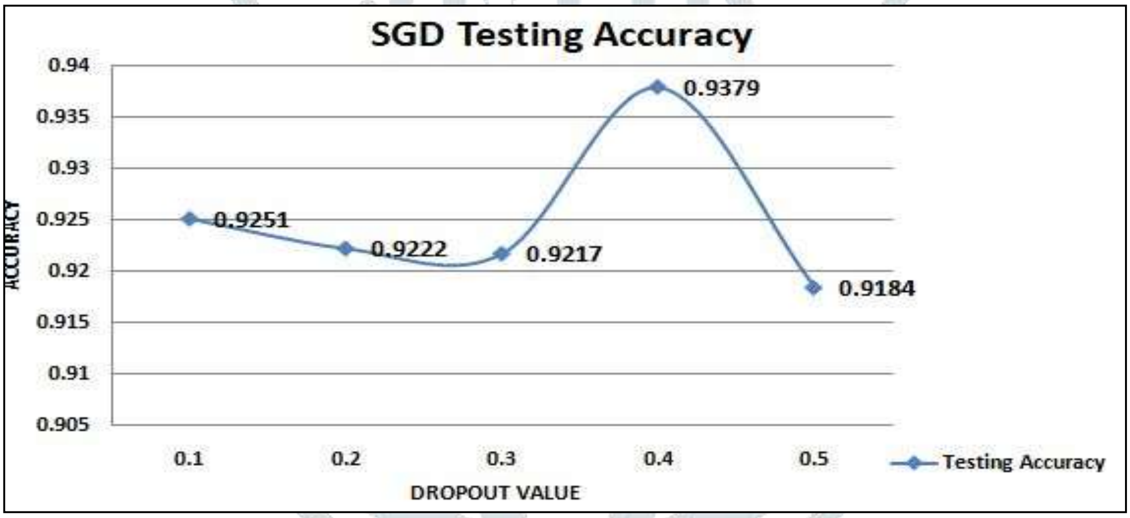


Figure 9: SGD testing accuracy on different dropout value

3) Increasing the number of neurons

The following graph shows number of neurons affects the performance of the deep neural network. As shown in the figure, when we increase the number of neuron of hidden layer, it increases the performance of the system. The following graph shows the performance of the ADAM optimization algorithm of number of neurons v/s run time. It is clear that the run time is directly prepositional to the model complexity. As the model complexity increases the number of parameter to optimize also increases and so the run time.

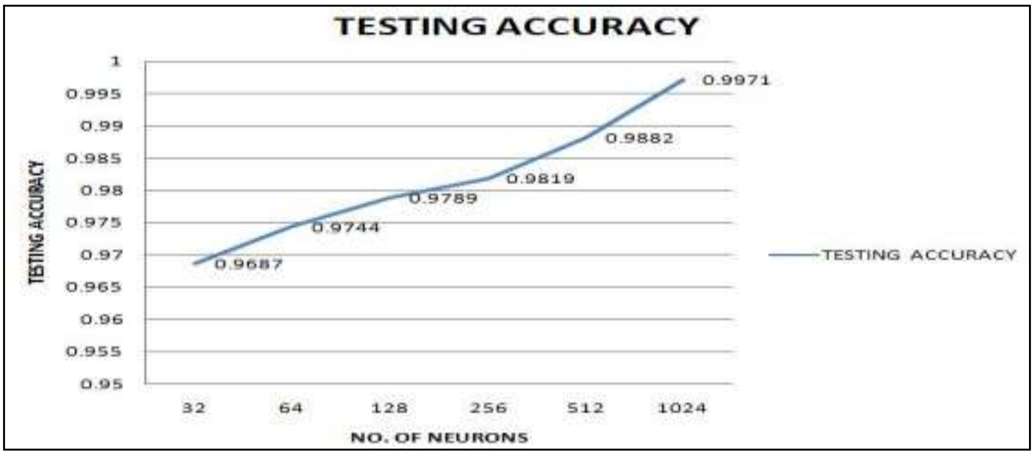


Fig 10: ADAM testing accuracy with number of neurons

4) Execution Time

In the above section we discussed how accuracy of deep neural network is vary by increases the number of neurons. But on the other hand, execution time is increases with number of neurons. The following graph shows time taken by ADAM algorithm when number of neurons increases and epochs =20.

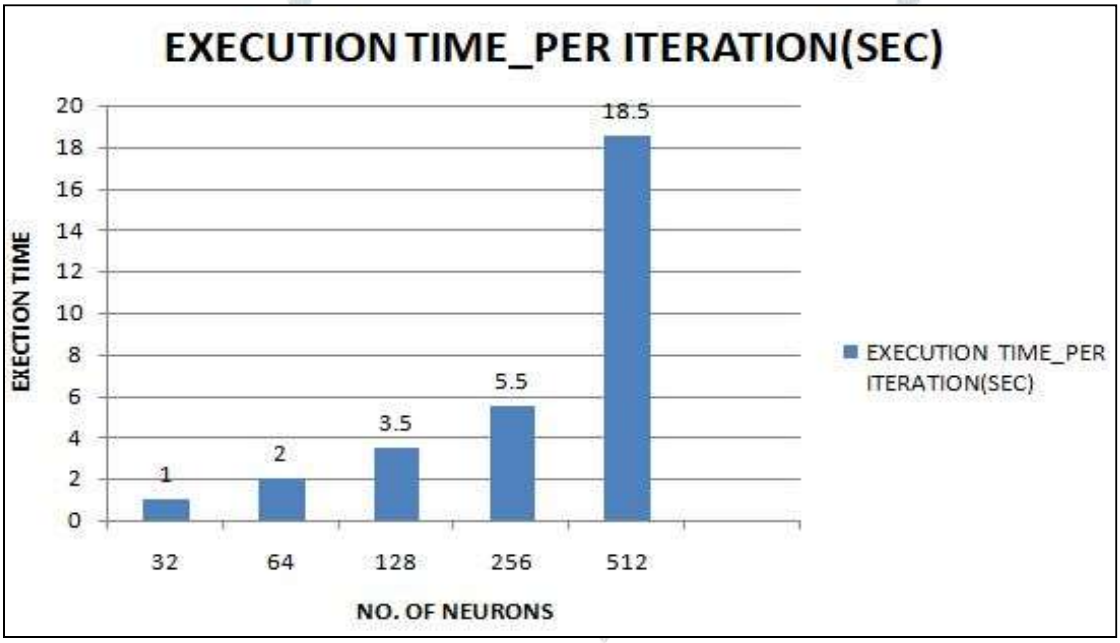


Fig 10: Execution time per iteration

6) Performance Summary

The performance of different optimization methods such as stochastic gradient descent, ADAM, RMSProp, Adamax, Adagrad, Adadelta on deep neural network are shown in the tabular format. According to the table 3 Adadelta gives best result on MNIST dataset. The performance of adagrad ,adamax, and RMSProp is also good on MNIST dataset. The dropout technique increases the testing accuracy for ADAM, Adamax, Adagrad, Adadelta. But in case of SGD optimizer dropout technique decreases the testing accuracy.

TABLE 3: Performance Summary

OPTIMIZATION METHOD	TRAINING ACCURACY	TESTING ACCURACY

Stochastic Gradient Descent	93.77%	93.95%
ADAM	98.85%	97.82%
RMSProp	98.87%	98.28%
Adamax	100%	98.29%
Adadelta	100%	98.41%
Adagrad	100%	98.27%
SGD+Dropout	90.32%	93.55%
ADAM+ Dropout	98.74%	98.41%
RMSProp+ Dropout	98.58%	98.28%
Adamax+ Dropout	98.88%	98.32%
Adadelta+Dropout	98.11%	98.42%
Adagrad + Dropout	98.57%	98.38%

CONCLUSION

In this research work, a deep neural network was designed to classify handwritten digits with the help of MNIST dataset. The aim was to get maximum accuracy on digit classification using different optimization methods. The accuracy obtained by different models was compared so that we get the best result. The neural network parameters and hyperparameters affect the performance of deep neural network so this research work examined the accuracy of network on varying different parameters like number of neurons, number of hidden layers, dataset etc. The performance of various optimization algorithms stochastic gradient descent, ADAM, RMSProp, Adamax, Adadelta optimization methods are analyzed on purposed deep neural architecture. The Adadelta optimizer gave best result on MNIST dataset. The performances of other optimizers were also good.

The future scope lies in trying different image datasets of varying sizes and checking the performance of the different optimizers using the proposed architectures. Further in this research only feed forward deep neural network have been used. The performance of image classification task gives very good accuracy when convolutional neural networks are used. The future work lies in selecting the best architecture of CNN and getting top notch accuracy. Concept of image augmentation can also be applied when image datasets are of small size.

References

[1] A.G. Ivakhnenko, “Polynomial theory of complex systems” IEEE Transaction on System, Man and Cybernetic vol. 1,no 4, pp. 364-378, 1971.

[2] Xuedam Du,Yinghao Cai, Wang, and Leijie Zhang “ Overview of Deep Learning” 31st Youth Academic Annual Conference of Chinese Association of Automation Wuhan China November 11-13-2016.

[3] Siddhartha Sankar Nath, Janynyaseni Kar, Girish Mishra, Sayan Chakraborty , Nilanjan Dey “ A Survey of Image Classification Methods and Techniques ” ICCCICCT 2014.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. Neural Information Processing Systems, Nevada, 2012

[5] Henrik Petersson, David Gustafsson and David Bergstroom “ Hyperspectral Image Analysis using Deep Learning- a Review” IEEE 2016.

[6] Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez Ramos and Pascual Campoy “A Review of Deep Learning Methods and Applications for unmanned Aerial Vehicles ” Hindawi Journal of Sensors 2017.

[7] H. Kamitomo and C. Lu, “3-d face recognition method based on optimum 3-d image measurement technology,” Artificial Life and Robotics, vol. 16, no. 4, pp. 551–554, 2012.

[8] Walaa Hussein Ibrrahim, Ahmed AbdelRhman Ahmed Osman, Yusra Ibrahim Mohamad” MRI Image Classification Using Neural Network” ICCEEE, 2013.

[9] S.Kim, B.Park, B.S Song, and S.Yang, “ Deep belief network based statistical feature learning for fingerprint liveness detection, ” Pattern Recog. Lett., vol 77, ,pp. 58-65,2016.

[10] Gang Liu, Liang Xiao, Caiquan Xiong “ Image Classification with deep belief network and improved gradient descent” IEEE 2017.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradientbased learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] Travis Williams, Robert Li “ Advanced Image Classification using Wavelets and Convolution Neural Network” IEEE 2016.

[13] Narek Abroyan , “ Convolutional and Recurrent Neural Network for real time data classification” The Seventh International Conference on Innovative Computing Technology (INTECH 2017).

[14] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, “Deep Learning”, Book in preparation for MIT Press, 2016, on-line version available at http://www.deeplearningbook.org.

[15] Michael A.Nielsen, “Neural Networks and Deep Learning”, Determination Press, 2015.

[16] Marek Dabrowski, Justyna Gromada, Tomasz Michalik Orange Centrum “A Practical study of neural network –based image classification model trained with transfer learning method” FedCSIS 2016.

[17] C. Lu and X. Tang, “Surpassing human-level face verification performance on lfw with gaussianface,” arXiv preprint arXiv:1404.3840, 2014.

[18] L. Deng and D. Yu, “Deep learning: methods and applications,” Foundations and Trends in Signal Processing, vol. 7, no. 3-4, pp. 197–387, 2013.

[19] J. A. Hertz,” Introduction to the theory of neural computation.“ Boulder, USA: Westview Press, 1991.

[20] J. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” Neural Processing Letters, vol. 9, no. 3, pp. 293–300, 1999.

[21] Y. Xiong and R.Zuo, “ Recognization of geochemical anomalies using a deep autoencoder network,”Computer Geosci-UK, vol.86, pp. 75-82, 2016.

[22] Zejian Shi, Minyong Shi, Chunfang Li,” The prediction of character based on recurrent neural network,” IEEE computer society, Wuhan China, 2017

[23] Taro Ishitakl, Ryolchlro Obukata, Tetsuya Oda, Leonard Baroll, “Application of deep recurrent neural network for prediction of user behavoiur in Tor Network,” 31st International Conference on Advanced Information Networking and Application Workshops, 2017.

[24]Marek Daabrowski, J. Gromada, T. Michalik,” P practical study of neural network-based image classification model trained with transfer learning method,” Federated Conference on Computer Science and Information Systems pp. 49-56, 2016.

[25] B. Wang, K. Yager, D.Yu, Minh Hoai,” X- ray Scattering image classification using deep learning,” IEEE Winter Conference on Application of Computer Science,2017

[26] Nur Anis Mohmon and Norsuzila ya acob,” A review on classification of satellite image using artificial neural network (ANN),” IEEE 5th Control and System Graduate Research Colloquium,2014.

[27] R.Jyothi, Y.K. SundaraKrishna, V. Srinivasa Rao,” Paper Currency recognition for color images based on Artificial Neural Network,” International Conference on Electrical , Electronics and Optimization Techniques (ICEEOT), 2016.

[28]M.Abadi, A. Agarwal, p. Barham, E. Brevdo, Z.f. Chen, C. C itro,et at.,” Tensorflow : Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv: 1603.04467,2016.

[29]R. Collobert, S.Bengio, and J.Mariethoz, “ Torch: a modular machine learning software library,” Idiap, No.EPFL-REPORT-82802, 2002.

[30]R. AI-Rfou, G.Alain, A.Almahairi el at.,” Theano ; a python framework for fast computation of mathematics expression,” arXiv preprint arXiv : 1605.02688,2016.

[31] LeCun, Y., C. Cortes, and C.J. Burges, MNIST handwritten digit database. AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist, 2010.