An empirical comparison and research on Neural Network model and its variants for TensorFlow framework using Mixed National Institute of Standards and Technology (MNIST) Database

Srishti Bhalla
Computer Science and Engineering,
Bharati Vidyapeeth’s College of Engineering
New Delhi,India

Palak Dhingra
Computer Science and Engineering
Bharati Vidyapeeth’s College of Engineering
New Delhi, India

Narina Thakur
Computer Science and Engineering,
Bharati Vidyapeeth’s College of Engineering
New Delhi,India

Saumya Tiwari
Computer Science and Engineering,
Bharati Vidyapeeth’s College of Engineering
New Delhi,India

Abstract: Neural Network is the technological breakthrough in the field of Deep Learning and is driving some of the most ingenious inventions of today’s century. It is a simple processing unit which is massively parallel, capable to store knowledge and apply this stored knowledge for Prediction. This paper summarizes Neural Network model, its variants and Learning techniques. These models have been implemented on TensorFlow framework using Python language. Experiments are conducted on MNIST dataset and accuracy is used as performance measure, its value shows the effectiveness of correct labeling. The results and analysis showed that Recurrent Neural Network performs better with small dataset whereas Convolutional approach is preferred in large-enough dataset for image labeling.

Keywords— Neural, Network, Deep learning, TensorFlow, Recurrent, Convolutional, Image labeling, MNIST

I. INTRODUCTION

Neural network, a subset of Artificial intelligence was originated in 1943 by McCulloch and Pitts, the pioneers of neural networks, in the hope to get human-like learning ability. Neural Network (NN) also called Artificial Neural Network (ANN) is named after its artificial representation of human nervous system. Its major components includes -

1. Dendrites- It takes input from other neuron sin form of an electrical impulse.
2. Cell Body- It generates inferences from those inputs and decides what action to take.
3. Axon Terminal- It transmits outputs in form of electrical impulse.

To replicate the nervous system ability to learn, the general structure of neural network consists of –

1. Input Layer- The inputs i.e. training examples are fed through this layer.
2. Hidden Layer- It’s an intermediate layer between input and output layers which helps the Neural Network to learn the complex relationships between data.
3. Output Layer- This layer yields the response to the inputs.

With advancement of technology the processing of inputs i.e. data through these neurons have taken a leap and variants of models have been developed to process the raw information. These Neural Network Models are used in various fields like information retrieval, machine translation, image recognition software, self-driving cars etc. Neural network offers a number of advantages, the ability to learn being the main one [1]. Other advantages include their ability to handle noisy data, no requirement of any prior assumptions about the distribution of the data is needed, they can map and approximate any complex function or nonlinearity [2], multiple training algorithms are also available [3]. Due to these advantages variants of neural network have found their applications in different fields. The paper is divided into 6 sections. Section 2 discusses the concept of Neural Network and its different architectures. Section 3 discusses its learning techniques. In section 4 literature survey of applications of different neural network models is done. In section 5 the experimental setup details for training deep, recurrent and convolutional neural networks are provided and Section 6 compares and discusses the results for the classification performance of these models.

II. NEURAL NETWORKS

Neural Network is an information processing model consisting of an inter-connection of a number of neurons/nodes with weights assigned to each connection. Neural Networks are also known as connectionist models, parallel distributed processing systems and neuromorphic models [4]. Neural Networks are specified using three parameters [5]:

Architecture- It maintains the topological relationships of the nodes. They also have an activation level specified for them.

Activity rule- The way in which the activation level of the neurons is updated in response to each other.

Learning rule- The way in which the weights of the neurons are updated as a function of time.

A. Different Architectures of Neural Network:

1) Single layer perceptron- This is the basic model of Neural Network containing an input and an output layer interconnected through weights. It uses supervised learning technique, usually the delta rule which compares the desired output with the calculated output. The computed output is compared with a threshold value and if it's greater the neuron is activated and it produces 1(generally) as output. The single layer perceptron can only learn linear separable problems.

2) Multi-layer feed-forward perceptron- This architecture contains an input layer, two or more hidden layers and an output layer interconnected in a feed forward way as shown in figure 1. It uses a non-linear function for example sigmoid or logistic as an activation function to introduce non-linearity in the computation. Generally, back-propagation (supervised technique) is used to train this network. It can classify data that are not linear.

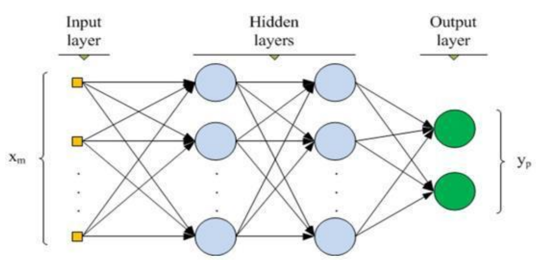

Fig.1. Multi- layer perceptron [6]

3) Convolutional Neural Network (CNN) - CNN is inspired from the visual cortex in the brain. The characteristic of specialized components looking for certain features forms the basis for this network. The architecture consists of Convolutional, non-linear, pooling, fully connected layers as shown in figure 2. The Convolutional layer consists of a set of filters also known as kernels containing a small receptive field, characterized by weights [7]. This filter is convolved across the height and the width of the input to produce activation or a feature map. The filter enables the network to learn to activate itself in the presence of a special feature observed at some spatial arrangement in the input. More the filters more information is gathered about the input volume. The output of the first Convolutional layer acts as an input for the second layer and so on. The non-linear layers between the Convolutional layers introduce non-linearity in the output of the previous layer. RELU (Rectified linear units) layers, have more computational efficiency than the sigmoid or Tan-h activation functions [8]. The fully connected layers receive the activation map after it has been passed through the non-linear layer and outputs the probability as to which class the input belongs to. It acts as a trained classifier. Pooling layer- It is inserted between successive Convolutional layers. The most common non-linear function to apply pooling is max-pooling [9]. The relative location of the features is important rather than the exact location of a feature which has been observed. It reduces the spatial size to a great extent which reduces the computational costs and also controls over fitting. The hyper-parameters stride (how the filter convolves around the input volume) and padding (input vector is padded with zeroes to preserve the dimensions of the output) control the size of the output.

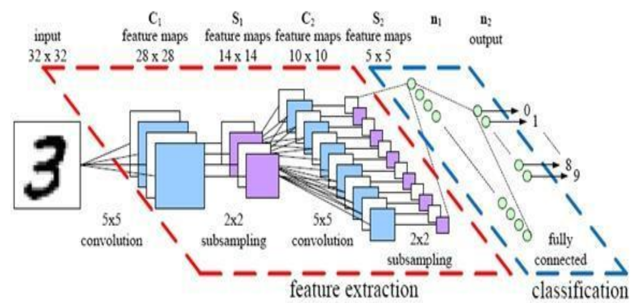

Fig.2. Basic structure of Convolutional Neural Network [10]

4) Recurrent Neural Network (RNN) **-** Since in NN it's assumed that the inputs and outputs are not dependent on each other but considering this in real life scenarios may be a bad choice. For example, if in a sentence, the next word is to be predicted, then the preceding words should be known. RNN uses sequential information and is called Recurrent because for every element in a sequence the same task is performed and thus the parameters are shared at every step, with output being dependent on the previous computation [11] as shown in figure 3. Due to the sharing of parameters the number of parameters learnt is reduced [12]. Calculations done so far are captured by a 'memory' in the network. Practically RNN can make use of information of only a few steps back. The key feature of RNN is its hidden state, which captures some information about a sequence. There are various applications of RNN in language modeling and generating text, machine translation, speech recognition and image description. Most commonly used RNNs are LSTMs (Long Short Term Memory). Training in RNN is similar to traditional Neural Network. Back propagation algorithm is used for training this network. The gradient at each output depends on the previous time steps along with the current step due to the sharing of parameters by all time steps. For example, to calculate the gradient at t = 4 we would need to propagate through three steps and sum up the gradients. This is called Back propagation through time (BPTT).Various RNN models are:

a) Bidirectional RNNs- They are based on the idea that the output at time t not only depends on the previous elements in a sequence, but also on future elements. They can be thought of as two RNNs placed on top of each other.

b) Deep RNNs- They are similar to Bidirectional RNN, but they have multiple layers per time step. This enables their learning capacity to be higher but it also requires more training data.

c) LSTM Networks- The difference between these networks and RNN is in the function used by them to calculate hidden state but the architecture is the same. The cells (memories) in LSTMs take previous state along with the current input as input and decide the contents of the memory.

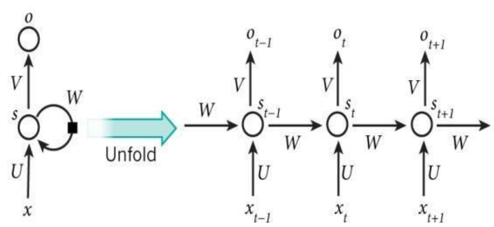

Fig.3. Basic structure of Recurrent Neural network [12]

III. Learning techniques for Neural networks:

A. Unsupervised learning- It differs from supervised learning in the sense that it doesn't need desired output or as we can say, a teacher to learn. It studies the inputs provided to it, and maps them based on the similarities of their features. Because of this property it can be used in clustering data. Basically its task is to infer an underlying data structure of the random inputs provided to it.

1) *Self-organizing map-* A self-organizing map (SOM) introduced by Kohonen [13] is an example of this. SOM maps a multidimensional feature vector to a lower dimensional grid for the ease of view. Each of the feature vector nodes are fully connected to the output grid of nodes with weights assigned to each of the connection. Similar inputs are clustered together in the map.

B. *Supervised learning technique***-** In this technique, the learning and adjustments of weights is done under the supervision of a teacher in the form of desired output presented for every input. Its aim is to minimize the mean squared error function. Basically the difference or the error between the target and the actual output is minimized. The most common supervised technique used to train NN is back propagation. Back propagation is a key algorithm for training Neural Network. It is used along with gradient descent algorithm. It calculates gradient of a cost function (generally mean squared error) with respect to all the weights in the network. The weights are then updated with the help of the calculated gradients which are fed into the optimized method, to minimize the error function. The back-propagation algorithm is divided into two phases-

Forward pass-During the forward pass, the outputs are calculated and compared with the targets. The mean squared error is calculated for each neuron in the output layer.

Backward pass- During the backward pass, the weights of each neuron is updated such that it becomes closer to the target, thereby minimizing the error.

Back propagation, in different forms and name, has its applications in many other areas, for example weather forecasting or analyzing numerical stability. Its general application independent name is "reverse-mode differentiation [14]."

IV. LITERATURE SURVEY:

A) Applications of NN-

Over the decade Neural Network has found its applications in many areas.

Neural Network was used in the implementation of Relevance feedback model in IR and a comparison was made between this model and Probabilistic relevance feedback model [15]. The test results showed that Probabilistic model is better than NN. NN has been applied for environmental causes for example to model hourly NOx pollutants concentrations [16] where it was observed that they perform better than regression based models, or to predict and forecast groundwater levels [17]. It has also been used to classify cancer into diagnostic classes based on gene expression [18] and to study the presence of disease conditions with the help of liver ultrasonic images [19]. [20] Implemented a controller for a mobile robot using the learning ability of NN for its design.

This model can be used for fault diagnosis. [21] Demonstrated a chemical process to show the ability of NN to detect and diagnose faults and [22] proposed a procedure for fault diagnosis of rolling element bearings and It has also found its application in the field of finance like forecasting economic time series data [23] or predicting stock price movement [24]

.In [24] NN model was compared with SVM and was found to be significantly better.

B) Applications of CNN-

These networks have their applications mainly in the field of image [25] and video classification [26] and natural language processing. CNN has been used for human face recognition [27, 28] and along with a rule based algorithm it was used to detect facial expressions [29]. CNN was used to classify 1.2 million images in imagenet-2010 contest [30]. In this method, Dropout [31] was used to reduce over-fitting and the resultant architecture was similar to columnar CNN [32], which resulted in top-1 error rate of 37.5% and top-5 error rate of 17% on the test data. The multilayer perceptron (MLP) and the Convolutional Neural network were compared [33] for handwriting recognition. It was proved that CNN is better than MLP for visual input classification. An experiment was conducted for sentence level classification using CNN trained on top of pre-trained word vectors [34]. Several variants of CNN were used and their performance was compared. Dynamic CNN was proposed for semantic modeling of the sentences [35] and was tested on 4 experiments giving an excellent performance. A CNN based model was proposed for speech recognition [36] and its performance was found to be better than the regular NN with the same parameters.

C) Applications of RNN

A model was developed that addressed sentence embedding's using RNN with LSTM cells [37]. On a web search task, the LSTM-RNN embedding performed significantly better than several existing methods. RNN has successfully been used for speech recognition task [38, 39, 40] and machine translation task [41]. Similarly in [42] LSTM was used for English to French translation task, where its performance was improved when the order of the words in source sequences were reversed. Multidimensional RNN along with connectionist temporal classification was used for handwriting recognition [43]. Deep RNN's were used for opinion extraction formulated as a sequence labeling task[44].It was observed that deep narrow RNN performs better than shallow wide RNN with the same parameters. In [45] RNN were applied to stock pattern recognition. A method was developed for evaluating RNN with context transition performances which provided fruitful results in trading analysis and prediction. RNN along with CNN has been used for image description [46,47]and scene labeling[48].

V. EXPERIMENTAL SETUP

The classification performance of different Neural Network models has been compared on Mixed National Institute of Standards and Technology (MNIST) dataset using TensorFlow using Python. MNIST dataset consists of a database of handwritten digits divided into testing and training examples according to table I and II. Since the dataset contains labels, variants of Neural Network are trained and their accuracy is calculated by finding the number of times the correct label is given to the sample.

Data- The data set consist of 60000 examples of which 55000 examples are used for training and rest 5000 examples are used for validation. Every MNIST data example has two parts: an image of a handwritten digit and a corresponding label (the range of label values is 0-9). Each greyscale image is 28 pixels by 28 pixels (the range of pixel values is 0-255) and so the output tensor for the training set images is of shape [55000, 784].

TABLE I. TRAINING AND TESTING DATASET

	Training Set	Testing Set
No. of images	60000	10000
No. of rows	28	28
No.of columns	28	28

TABLE II. TRAINING AND TESTING LABELS

	Training Set Label	Test Set Label
No. of items	60000	10000

Activation unit -We employ the standard softmax activation for the output layer. For the hidden layers we use the rectifier linear activation: f(x) = max{0, x}.

Network Training-We use the standard multiclass cross-entropy as the objective function when training the variants of Neural networks. We use stochastic gradient descent with a default learning rate(0.001). We update weights after minibatches of 100 images for Deep Neural Network and 128 minibatches for Recurrent and Convolution Neural Network.

Since our objective is to compare the NN models on how they correctly and efficiently recognize and label the images to one of the 10 classes of dataset. We define Deep Neural Network, RNN and CNN and pass the MNIST dataset to each one of them. These inputs characterized by weights are applied to number of hidden layers containing around 500 nodes which in turn is fed to the activation unit after which the output is compared to intended output with cost function and then optimization function is employed to minimize the cost, this whole process is repeated for backward propagation, that is training is done for 10 epochs for all the variants of Neural Network.

VI. RESULTS AND CONCLUSION

The experimental results show that Recurrent Neural Network outperforms both Deep Neural Network and Convolution

Neural Network as shown in table III. The reason RNN performs better than CNN is the limited or not enough dataset. It is also seen that RNN performs better than CNNs with less data.

TABLE III. ACCURACY COMPARISON BETWEEN DNN, RNN AND CNN

MODEL	ACCURACY (%)
DNN	**95.13**
RNN	**98.27**
CNN	**97.64**

REFERENCES

[1] Antsaklis, Panos J. "Neural networks for control systems." *IEEE Transactions on Neural Networks* 1.2 (1990): 242-244.

[2] Vellido, Alfredo, Paulo JG Lisboa, and J. Vaughan. "Neural networks in business: a survey of applications (1992–1998)." *Expert systems with applications* 17.1 (1999): 51-70.

[3] Tu, Jack V. "Advantages and disadvantages of using artificial Neural networks versus logistic regression for predicting medical outcomes." *Journal of clinical epidemiology* 49.11 (1996): 1225-1231.

[4] Lippmann, Richard. "An introduction to computing with Neural nets." *IEEE Assp magazine* 4.2 (1987): 4-22

[5] MacKay,David JC. *Information theory, inference and learning algorithms*.Cambridge university press, 2003.

[6] http://www.cs.us.es/~fsancho/?e=135

[7] Rutkowski, Leszek, and Jörg Siekmann. *Artificial Intelligence and Soft Computing: ICAISC 2004, 7th International Conference, Zakopane, Poland, June 7 - 11, 2004 ; Proceedings*. Berlin: Springer, 2004.

[8] Goldberg, Yoav. "A primer on Neural network models for natural language processing." *arXiv preprint arXiv:1510.00726* (2015).

[9] Boureau, Y-Lan, Jean Ponce, and Yann LeCun. "A theoretical analysis of feature pooling in visual recognition." *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.

[10] http://parse.ele.tue.nl/education/cluster2

[11] Stojanovski, Dario, et al. "Finki at SemEval-2016 Task 4: Deep learning architecture for Twitter sentiment analysis." *Proceedings of SemEval* (2016): 149-154.

[12] http://www.wildml.com/2015/09/Recurrent-Neural-networks-tutorial-part-1-introduction-to-rnns/

[13] Kohonen, Teuvo. "Self-organized formation of topologically correct feature maps." *Biological cybernetics* 43.1 (1982): 59-69.

[14] Griewank, Andreas. "Who Invented the Reverse Mode of Differentiation?." *Optimization Stories, Documenta Matematica, Extra Volume ISMP (2012)* (2012): 389-400.

[15] Crestani, Fabio. "Comparing Neural and probabilistic relevance feedback in an interactive information retrieval system." *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. Vol. 5. IEEE, 1994.

[16] Gardner, M. W., and S. R. Dorling. "Neural network modelling and prediction of hourly NO x and NO 2 concentrations in urban air in London." Atmospheric Environment 33.5 (1999): 709-719.

[17] Taormina, Riccardo, Kwok-Wing Chau, and Rajandrea Sethi. "Artificial Neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon." Engineering Applications of Artificial Intelligence 25.8 (2012): 1670-1676.

[18] Khan, Javed, et al. "Classification and diagnostic prediction of cancers using gene expression profiling and artificial Neural networks." *Nature medicine* 7.6 (2001): 673-679.

[19] Kalyan, Karthik, et al. "Artificial Neural network application in the diagnosis of disease conditions with liver ultrasound images." Advances in bioinformatics 2014 (2014).

[20] Mohareri, Omid, Rached Dhaouadi, and Ahmad B. Rad. "Indirect adaptive tracking control of a nonholonomic mobile robot via Neural networks." *Neurocomputing* 88 (2012): 54-66.

[21] Hoskins, Josiah Collier, and D. M. Himmelblau. "Artificial Neural network models of knowledge representation in chemical engineering." *Computers & Chemical Engineering* 12.9 (1988): 881-890.

[22] Samanta, B., and K. R. Al-Balushi. "Artificial Neural network based fault diagnostics of rolling element bearings using time-domain features." *Mechanical systems and signal processing* 17.2 (2003): 317-328.

[23] Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. "Forecasting with artificial Neural networks:: The state of the art." *International journal of forecasting* 14.1 (1998): 35-62.

[24] Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial Neural networks and support vector machines: The sample of the Istanbul Stock Exchange." *Expert systems with Applications* 38.5 (2011): 5311-5319.

[25] Ciresan, Dan C., et al. "Flexible, high performance Convolutional Neural networks for image classification." *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. Vol. 22. No. 1. 2011.

[26] Karpathy, Andrej, et al. "Large-scale video classification with Convolutional Neural networks." *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.

[27] Lawrence, Steve, et al. "Face recognition: A Convolutional Neural-network approach." *IEEE transactions on Neural networks* 8.1 (1997): 98-113.

[28] Nebauer, Claus. "Evaluation of Convolutional Neural networks for visual recognition." *IEEE Transactions on Neural Networks* 9.4 (1998): 685-696.

[29] Matsugu, Masakazu, et al. "Subject independent facial expression recognition with robust face detection using a Convolutional Neural network." *Neural Networks* 16.5 (2003): 555-559.

[30] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep Convolutional Neural networks." *Advances in Neural information processing systems*. 2012.

[31] Hinton, Geoffrey E., et al. "Improving Neural networks by preventing co-adaptation of feature detectors." *arXiv preprint arXiv:1207.0580* (2012).

[32] D.C. Cire¸san, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. High-performance Neural networks for visual object classification. Arxiv preprint arXiv:1102.0183, 2011.

[33] Simard, Patrice Y., David Steinkraus, and John C. Platt. "Best practices for Convolutional Neural networks applied to visual document analysis." *ICDAR*. Vol. 3. 2003.

[34] Kim, Yoon. "Convolutional Neural networks for sentence classification." *arXiv preprint arXiv:1408.5882* (2014).

[35] Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A Convolutional Neural network for modelling sentences." *arXiv preprint arXiv:1404.2188* (2014).

[36] Sander Dieleman, Benjamin Schrauwen, "End-to-end learning for music audio", *Acoustics Speech and Signal Processing (ICASSP) 2014 IEEE International Conference on*, pp. 6964-6968, 2014.

[37] Palangi, Hamid, et al. "Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.4 (2016): 694-707.

[38] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep Recurrent Neural networks." *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013.

[39] Waibel, Alex. "Modular construction of time-delay Neural networks for speech recognition." *Neural computation* 1.1 (1989): 39-46.

[40] Robinson, Tony, Mike Hochberg, and Steve Renals. "The use of Recurrent Neural networks in continuous speech recognition." *Automatic speech and speaker recognition*. Springer US, 1996. 233-258.

[41] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078* (2014).

[42] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with Neural networks." *Advances in Neural information processing systems*. 2014.

[43] Graves, Alex, and Jürgen Schmidhuber. "Offline handwriting recognition with multidimensional Recurrent Neural networks." *Advances in Neural information processing systems*. 2009.

[44] Irsoy, Ozan, and Claire Cardie. "Opinion Mining with Deep Recurrent Neural Networks." *EMNLP*. 2014.

[45] Kamijo, Ken-ichi, and Tetsuji Tanigawa. "Stock price pattern recognition-a Recurrent Neural network approach." *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*. IEEE, 1990.

[46] Vinyals, Oriol, et al. "Show and tell: A Neural image caption generator." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[47] Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[48] Pinheiro, Pedro HO, and Ronan Collobert. "Recurrent Convolutional Neural Networks for Scene Labeling." *ICML*2014.