Performance Analysis of ANN For Apparel Classification Application of Computer Vision

Miss. Pooja Tupe[1], Mr. P. M. Vibhute[2], Mr. M. A. Sayyad[3]
[1]Dept of Electronics &Telecomm.
[2, 3]Asst. Professor
[1, 2, 3] Sanjivani College of Engg. (SCOE), Kopargaon

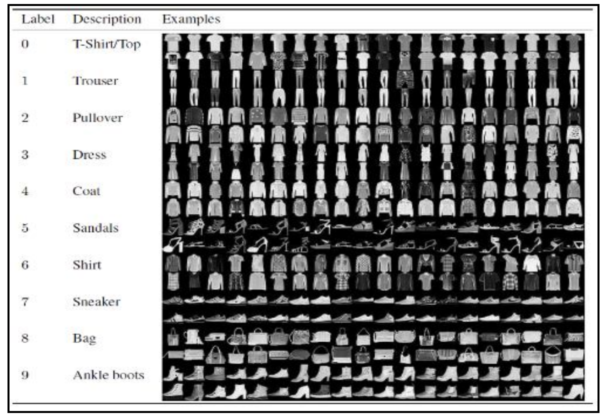
Abstract*- Apparel Classification is an approach of classifying apparel from a given image using Artificial Neural Networks (ANNs) which has been proven to perform remarkably well in field of image recognition. Proposed model shows impressive results on the benchmark dataset of Fashion-MNIST (Modified National Institute of Standards and Technology database). The dataset used to train simple 3, 6 and 12-layer neural networks, then compared the results with different epochs and finally, visualized the predictions.*

Keywords- Artificial Neural Network (ANN), Epochs, Fashion-MNIST, Image Classification.

I. INTRODUCTION

Computer vision is one of the rapidly advancing area and enabling new applications[2]. Computer vision methods are Image classification, object detection, and Neural style transfer. It involves extracting, analysing, and understanding of useful information from a single image or multiple images in a sequence.In this project, the computer vision task used is image classification. The classification[1] of data is a data mining or machine learning technique in which the algorithm tries to establish the relationship between input feature vectors and output variables, generally categorical, thereby adapt (learn) or build a model for prediction. This paper is focused on supervised learning and the aim is to increase the classification accuracy of the model by analyzing the factors which make a sufficient impact on the accuracy of the results obtained through ANNs.

Neural Network (NN) models are well suited to domains where large labeled datasets are available, since their capacity can easily be increased by adding more layers or more units in each layer. However, big networks with millions or billions of parameters can easily overfit even the largest of datasets. Correspondingly, a wide range of techniques for regularizing NNs have been developed[6].Fashion-MNIST is a dataset of Zalando's fashion article images that contains the 60,000 trained images of 10 different classes[2] as shown in Fig.1 and 10,000 testing images and the original size of the picture is 51*37 which will be resized to 28*28 grayscale.



Class names and example images in Fashion- MNIST Dataset.

II. PROBLEM DEFINITIION

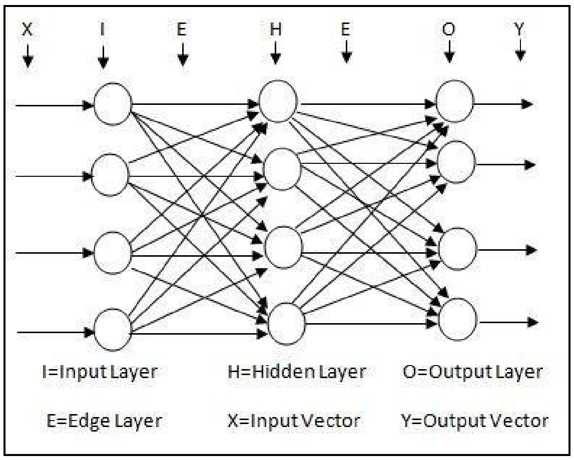
Image classification is a fundamental topic in the field of machine learning [2][5][9]. With the development of Neural Network architecture, researchers have achieved remarkable performance on this task comparing to classic models.Deep Neural Networks have been applied to a multitude of problems to achieve very good performances.

This work explores the idea of classifying Fashion MNIST images with variants of ANNs. Deep Learning concept likeANN used to build an image classification model which will learn to distinguish 10 different fashion apparel images into their respective categories. The types of fashion apparel include a T-shirt, Shirts, Bag, Coat, Trousers, etc[2]. This field has applications in social media, e-commerce and in the law sectors used to identify suspects with their clothes.

The problem of identifying a visual entity from an image is a very trivial problem for a human-being to perform, it is very challenging for a computer algorithm to do the same with human level accuracy[9]. The algorithm must be invariant to a number of variations in order to successfully identify and classify the images. For example, different illumination conditions, different scale and viewpoint variations, deformations, occlusion may influence the algorithm to wrongly predict the image class.

III. ARTIFICIAL NEURAL NETWORK

The Artificial Neural Networks are input-output processing systems inspired from human brain that is they consist of neurons which is fundamental unit of the human brain [1][4]. In the given Fig.2 ANN contains multiple layers and these layers contain multiple neurons. Each neuron is associated with a transfer function. Layers are connected with each other's through the edges. Each edge is associated with a numeric value called as weight. The ANNs are configured properly before applying the dataset to it. The working of ANN is based on back-propagation algorithm whose general working is as follows:



Artificial Neural Network (ANN)

- ***Step1:*** *Each edge layer is assigned the weights according to randomized technique or function.*
- ***Step2:*** *Inputs are applied to the input layer.*
- ***Step3:****The weighted sum of the inputs is determined for each neuron. i.e. (w1x1+w2x2+…wnxn) where wi= weights of corresponding edges xi=corresponding inputs.*
- ***Step4:*** *The weighted sum acts as an input for the transfer function.*
- ***Step5:*** *The output of each neuron is calculated up to the output layer by applying transfer function.*
- ***Step6:*** *At the output layerthe errorwhich is the difference between computed output and given outputis determined. The weights are then readjusted to minimize the error according to the back-propagation training algorithm.*

The step 2 to step 6 above have been iterated with all instances of given data set containing input-output vector pairs and each iteration is called an epoch[1]. Training is stopped when the network gives the optimized performance. The stopping criteria's of training are:

The maximum number of epochs is reached.
The minimum gradient is reached.
Best validation performance is achieved.
Mean Square Error(MSE) is minimized or zero. (The goal is reached).

IV. RELATED WORK

Lukas Bossard et. al [7] worked on the same dataset as ours to classify the images into clothing categories. Their focus was mainly to use Random Forests, Support Vector Machine (SVM) and Transfer Forests for the task. Their SVM baseline has an accuracy of 35.07% and the best transfer forest model obtained an accuracy of 41.3 %. Fashion-MNIST, a fashion product images dataset intended to be a drop-in replacement of MNIST and whilst providing a more challenging alternative for benchmarking machine learning algorithm [4].The images in Fashion-MNIST are converted to a format that matches that of the MNIST dataset, making it immediately compatible with any machine learning package capable of working with the original MNIST dataset. Khoi Hoang et. al [2] compares the performance of different models on the Fashion-MNIST and CIFAR-10 dataset. The paper also examines different features extraction techniques such as Principle Component Analysis (PCA), Autoencoder to boost the performance of the model. This [10] paper describes a well-defined and improved system for the humanassistance, based on wearable computer vision using Android ICS operating system.

Zhang, et. al [8] summarizes the some of the most important developments in neural network classification research. Specifically, the issues of posterior probability estimation, the link between neural and conventional classifiers, learning and generalization trade off in classification. Tripathi k. et. al [1] explores a novel approach for classification of data on four benchmark datasets from the perspective of ANNs and its intricacies. The proposed approach is successful in overcoming the drawback of over-fitting of data exists in the classification domain. Fischetti M. [3] discuss the peculiarity of 0-1 Mixed Integer Linear Program (MILP) models, and describe an effective bound-tightening technique intended to ease its solution.

V. PROPOSED METHODOLOGY

Classification is an example of Supervised Learning. Known class labels help indicate whether the system is performing correctly or not. This information can be used to indicate a desired response, validate the accuracy of the system, or be used to help the system learn to behave correctly. The known class labels can be thought of as supervising the learning process.The proposed methodology based on design of the NN architecture by deciding the number of layers and activation functions. Model Implemented with a simple 3-layer Neural Network. Next, system compared the classification accuracy between three depths, a 3-layer Neural Networks (NN-3), a 6-layer Neural Network (NN-6) and a 12-layer Neural Network (NN-12) with 5 epochs, to see if more layers mean higher accuracies. To improve accuracy, the model re-trained with 50 epochs.

A. Architecture of the Networks

The images are of size 28 x 28. So, image matrix is converted to an array, rescaled between 0 and 1, reshaped so that it's of size 28 x 28 x 1, and fed as an input to the network. Pixel location and its value at particular pixel is acting as a Features. In ANN, the first layer 'flatten' used to flattens the data, so that a (28x28) shape flattens to 784. The second layer is a dense layer with a ReLU activation function and has 128 neurons. The last layer is a dense layer with a SoftMax activation function that classifies the 10 categories of the data and has 10 neurons. For 6-layer network, 3 and for 12 -layer network 9 more hidden layers added keeping the same activation functions, shapes and settings, so the only difference is the depth of the network.

B. Compiling the model

The sparse categorical cross entropy function is used as error estimate or loss function. Sparse categorical cross entropy is defined as,

$$H_{p(y)} = -\sum_{i} y_i * \log pi \qquad (1)$$

where pi is the predicted probability distribution of class i and yi is the true probability distribution of class i. As there are ten classes in the dataset the summation is computed over 10 terms. After every minibatch iteration of the training processthe total cross entropy error computed[9]. The 'Adam' optimizer used for optimization of the loss function. An architectural summary of all the models has been shown in Fig.3.

C. Training and Evaluating Model

The set of data is divided into 3 sets.

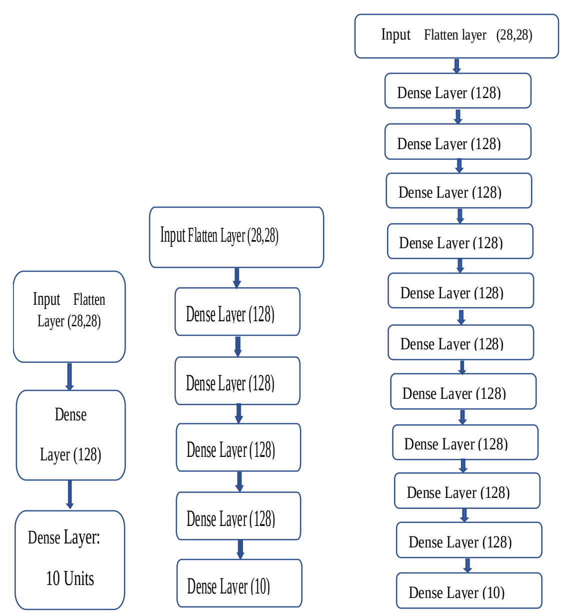
1.Training Set: This is the data that used to adjust the weights of the network by the training algorithm.

2.Validation set(Development set): This data is used during training to assess how well the network is currently performing -the performance of the network on this data may be used to guide the training in some way (e.g. controlling the learning rate, deciding when to stop training, choosing between several trained networks).

3.Test set: This is the genuine test data and, ideally, should be used once only, after training is complete[11].

The NumPy array target for training data, the target for training label data, verbose, batch size and the number of epochs. The first 2 arguments are simply training data and the labels, and the verbose argument determines how to visually view the progress of training, with the value 1 giving us an animated progress bar. The batch size determines the number of samples for the training process to go through until the gradient gets updated with new data. If not determined, batch size defaults at 32.

In proposed system for epochs value set at 5, but to improve accuracy, need much more than 5 epochs, as the model was retrained with 50 epochs. The trained network tested against the wholetest dataset and the loss and accuracy printed out. The loss should be as close as possible to the loss of the last epoch of the training process. This would mean that we didn't overfit or underfit the data, since it reacts in the same way to the test data as it did to the training data.



(a) (b) (c)

Neural Network Layers Architecture (a) 3_layer NN (b) 6_layer NN (c) 12_layer NN.

VI. RESULTS AND DISCUSSIONS

A. Accuracy and Loss

The proposed model successfully implemented ANN to perform the task of Apparel classification. Training shows results per epoch, with each epoch, the loss decreases and the accuracy increases, meaning the model is improving. The metric to compare the model, is the accuracy which shows how much the network classifies images correctly. Because the different model has different deptharchitecture and size it is not fair to compare the accuracy with same number of epochs and same learning rate. These hyperparameters should be adjusted independently to maximize accuracy foe each model on the validation dataset.

From Table 1 for the 3-layer Neural Network the test loss is 36.34% and accuracy is 87.09%, which is pretty close to the training metrics at the 5th epoch. With the NN-6, test loss is 36.22 % and the test accuracy is 87.07%. With the NN-12, test loss increased to 40.40% and accuracy is slightly lower at 84.88%. The loss should be small, and the accuracy should be as large as possible, so this network performed a little worse than above.

It can be seen that as the number of layers increased, training accuracy reduced as well as training loss increased. The overall trend with the increasing layer size seems to be that the loss function is increasing, and the accuracy is slightly decreasing.

Table 1. Effects of increasing no. of layers on accuracy %(5 epochs)

Parameters	*Layer size*		
	3-NN	*6-NN*	*12-NN*
Training Accuracy	89.03	88.78	87.67
Training Loss	30.02	30.12	34.22
Validation Accuracy	87.77	87.72	85.47
Validation Loss	32.88	33.79	38.72
Test Accuracy	**87.09**	**87.07**	**84.88**
Test loss	36.34	36.22	40.50

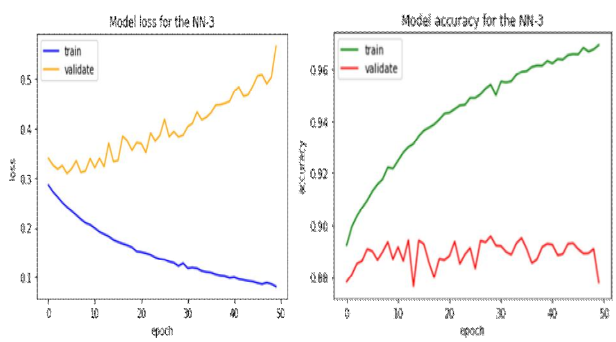
Table 2. Effects of Increasing No. of Epochs on Accuracy %(50 epochs)

Parameters	**Layer size**		
	3- NN	***6- NN***	***12NN***
Training Accuracy	96.94	96.25	94.53
Training Loss	8.12	9.60	16.30
Validation Accuracy	87.81	88.31	89.39
Validation Loss	56.67	66.84	47.39
Test Accuracy	**87.52**	**87.91**	**88.96**
Test loss	61.40	73.58	49.05

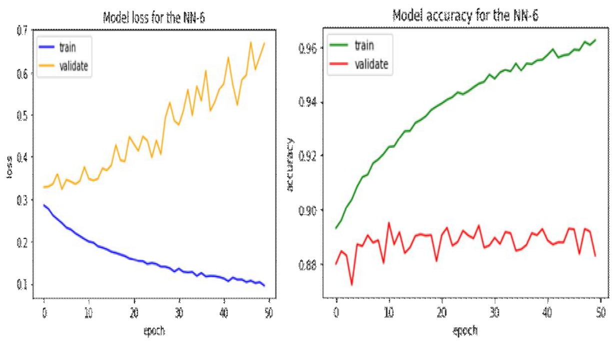
From Table 2 it can be seen that increasing number of epochs causes increase in Test Loss function but at the same time Accuracy improved slightly. Training loss is much lower than the previous results.

For all three models, from the Fig.4, it is observed as the epochs increase with the training data set is that the loss is decreasing down to 0, and the accuracy is increasing up to 1, both representing the ‘perfect score’. This is a sign of over-fitting, which is the motivation behind validating the model. The validation shows that the loss function increases past 0.50 for NN-3 and NN-6, and slope stabilizes at around 0.40 for the NN-12. The accuracy slope stabilizes between 0.88 and 0.90 for all 3 models.

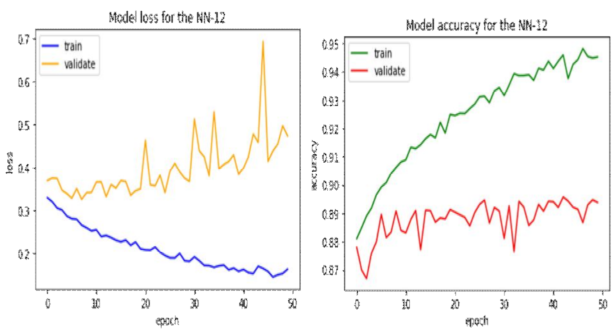
So, it looks like the 12-layer NN isperforming better on the validation set. Thus, these visualize the importance of the validation step.



(a)3-Layer NN



(b)6- Layer NN



(c)12-Layer NN

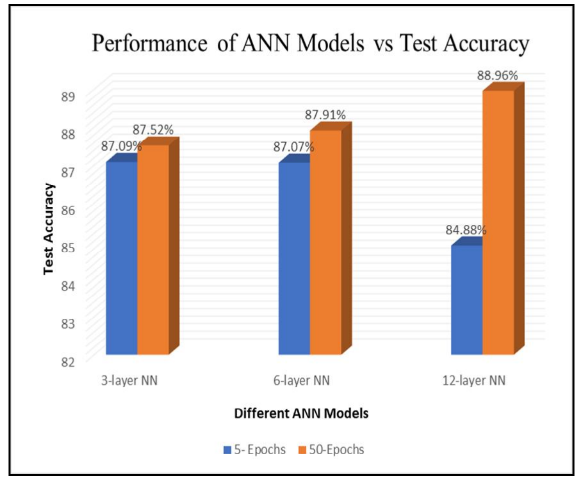
Model Accuracy and Loss Plots.

From the plots of 12-layer NN, it can be seen that the validation accuracy almost became stagnant after 4-5 epochs and rarely increased at certain epochs. In the beginning, the validation accuracy was linearly increasing with loss, but then it did not increase much. The validation loss shows that this is the sign of overfitting, similar to validation accuracy it linearly decreased but after 4-5 epochs, it started to increase. This means that the model tried to memorize the data andsucceeded. Table 3 shows the performance of ANN classifiers with their test accuracies.

Table 3 Performance of ANN Models vs TestAccuracies.

NN Models	Classifiers Accuracy	
	5 Epochs	*50 Epochs*
3-layer NN	87.09 %	87.52 %
6-layer NN	87.07 %	87.91 %
12-layer NN	84.88 %	**88.96 %**

Graph 1. Performance of ANN Models vs TestAccuracies



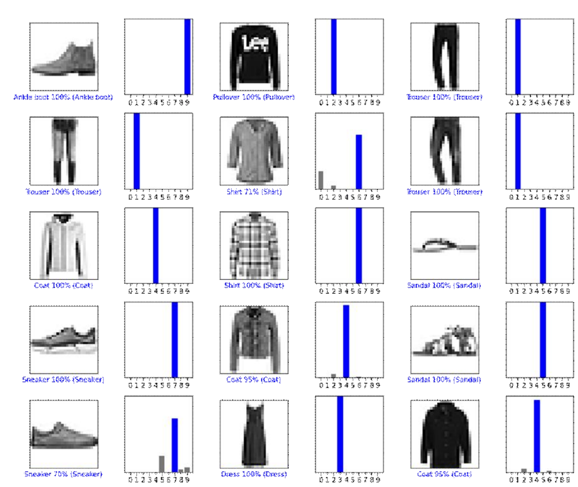
B. Visualized Predictions

Fig.5 shows 15 visualized data points, with the labelled images, and the probability graph beside them. If the label is red, that means the prediction did not match the true label; otherwise, it's blue. It looks like NN-12 model got 2 images wrong and 13 images correct.

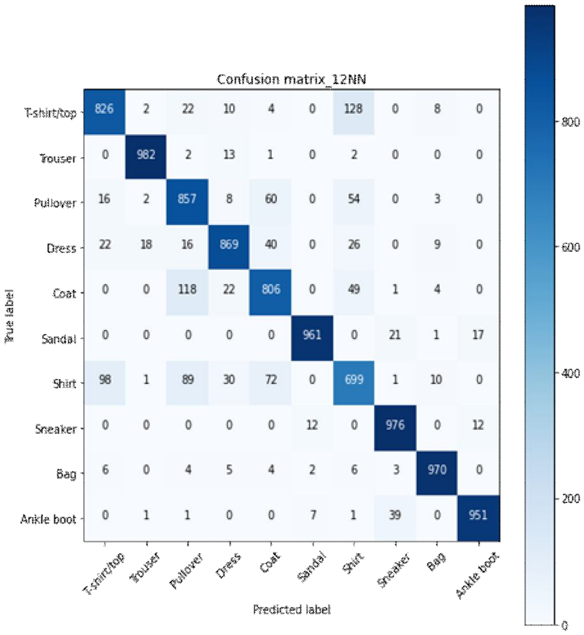
C. Confusion Matrix

In the field of machine learning and specifically the problem of statistical classification, a Confusion Matrix, also known as an Error matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

From Fig.6 we can see that the misclassification is usually between similar types of objects, such as t-shirts and shirts [12]. A row represents an instance of the actual class (i.e. an actual surgical step), whereas a columnrepresents an instance of the predicted class (i.e. the predicted surgical step).



Predictions made by the 12-Layer NN model



Confusion Matrix

D. Classification Report

Fig.7shows the classification report. Classification report helps in identifying the misclassified classes in more detail. One can observe for which class the model performed bad out of the given ten classes.

The classifier is underperforming for class 6 (shirts) regarding both precision and recall. For class 0 and class 2, the classifier is lacking precision. Also, for class 4, the classifier is slightly lacking both precision and recall.

	precision	recall	f1-score	support
0	0.85	0.83	0.84	1000
1	0.98	0.98	0.98	1000
2	0.77	0.86	0.81	1000
3	0.91	0.87	0.89	1000
4	0.82	0.81	0.81	1000
5	0.98	0.96	0.97	1000
6	0.72	0.70	0.71	1000
7	0.94	0.98	0.96	1000
8	0.97	0.97	0.97	1000
9	0.97	0.95	0.96	1000
accuracy			0.89	10000
macro avg	0.89	0.89	0.89	10000
weighted avg	0.89	0.89	0.89	10000

Classification report of 12- Layer NN.

VII. CONCLUSION

Three different Neural Networks, which are 3- layers NN, 6-layers NN, and 12-layers NN are implemented in Keras and trained on the test and validation dataset of Fashion-MNIST dataset. All the networks are trained and are evaluated in the accuracy metric of the test dataset of Fashion-MNIST. This process helps to understand how different architectures are designed and tested for image classification application.

From the comparison of different architectures 12-layers NN has the highest accuracy (88.96%) with low loss (49.05%). Compared to the different methods in Kaggle these networks are well performed. This accuracy could be increased using deeper network. For the future work, Convolutional Neural Network would be interesting to apply on this dataset.

REFERENCES

[1] Tripathi K, R. G. Vyas and A. K. Gupta, "The classification of data: A novel artificial neural network (ANN) approach through exhaustive validation and weight initialization", International Journal of Computer Sciences and Engineering, Vol. 6, No. 5, pp.241-254, 2018.

[2] Khoi Hoang, "Image Classification with Fashion-MNIST and CIFAR-10",California State University, Sacramento,2018,.

[3] Fischetti, M. and Jo, J., "Deep neural networks and mixed integer linear optimization", Constraints, 23(3):296–309, 2018.

[4] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms", Zalando Research Mühlenstraße 25, 10243 Berlin, arXiv:1708.07747 [cs.LG], 2017 .

[5] Rohit Patki, Suhas Suresha, "Apparel Classification using CNNs", Stanford University, 2016.

[6] Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of neural networks using dropconnect", In Proceedings of the 30th international conference on machine learning (ICML13), pages 1058–1066, 2013.

[7] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. "Apparel classification with style",In Asian Conference on Computer Vision (ACCV), pages 321–335. 2012.

[8] Zhang, Guoqiang Peter. "Neural networks for classification: a survey", IEEE Trans. Systems, Man, and Cybernetics, Part C 30 (2000): 451-462.

[9] Shobit Bhatnagar, Deepanway Ghosal, and Maheshkumar H. Kolekar, "Classification of Fashion Article Images using Convolutional Neural Networks", Fourth International Conference on Image Information Processing (ICIIP)2017.

[10] Vibhute Pritish M. and Sayyad M A., "Wearable Computer Vision based Human Assistance System using Android 4.0.3 ICS Operating System", IJCA Proceedings on International Conference on Recent Trends in Engineering and Technology 2013 ICRTET(4):4-7, May 2013.
[11] Keshav Singh, Artificial Neural Networks 101, [Internet], 2017, Available at: https://analyticsindiamag.com/artificial-neural-networks-101.
[12] Tensorflow, "Train your first neural network: basic classification," 2017. [Internet]. Available: https://www.tensorflow.org/tutorials/keras/basic_classification.