LEVERAGING TENSOR FLOW FOR SPEECH RECOGNITION AND IMAGE CLASSIFICATION LINKED TO CONVOLUTED NEURAL NETWORKS (CNN) FOR EFFICACIOUS DEEP LEARNING

Hardik Chaudhary, Vipul Goyal

ABSTRACT

Deep learning these days is playing a very major role in natural signal and information processing,like speech recognition and image classification. Deep learning technologyis immersed on the basis of the human brain. In deep learning, artificial neurons network automaticallytrained itself by large datasets that discovers connected patterns without the help of a human.Deep learning detects a pattern in unstructured data like image, sound, video, and text.For Image classification, CNN in deep learning is very popular. In many patterns, CNN performs bettercompared to human in a large dataset like image. In our research, we have used python with Kerasfor binary image classification. In this, we are using an animal's dataset, namely cat and dog, for image classification. Four different parameters with four different combinations have been appliedin CNN for comparison. It is shown that for Binary image classification combination of sigmoid classifier and Relu activation function gives higher classification accuracy than any other combination of classifier and activation function.

I. INTRODUCTION

Deep learning is a technology inspired by the functioning of the human brain. In deep learning, networks of artificial neurons analyse large datasets to automatically discover underlying patterns, without human intervention. [9] In deep learning, a computer learns to classify images, text, and sound. The computer is trained with large image datasets, and then it changes the pixel value of the picture to an internal representation, where the classifier can detect patterns on the input image. [4] Deep learning for image classification is becoming essential use of a machine learning method. To increase performance, the application of neural networks to learning tasks contains more than one hidden layer. Deep learning is part of a broader family of machine learning methods based on learning data representation, as opposed to hard code machine algorithms. [7] One of the most as often as possible utilized Deep learning techniques for picture characterization is the convolutional neural system (CNN). CNN gains legitimately from the picture information, consequently dispensing with manual component extraction. [4] Common issue in picture grouping utilizing Deeplearning is low execution on account of overfitting. To build execution and forestalling overfitting enormous datasets and models utilized. CNN has less association and hyperparameters that make CNN model simple to prepare and perform somewhat more regrettable than different models. [7] In this paper, a Deeplearning convolutional neural system dependent on Keras and Tensor Floware conveyed utilizing python for parallel picture

arrangement. In this examination, 10000 unique pictures, which contain two sorts of creatures, specifically feline and canine, are utilized for the order. Fig. 1.1 shows the case of the dataset.

Fig. 1 Dataset Sample

In this paper, four different structures of CNN are compared on the CPU system, with a combination of different classifiers and activation functions, namely softmax, sigmoid classifiers, and Relu, Tanh activation functions. For computation and processing, we are using Tensorflow and Keras framework. Tensorflow is one of the libraries utilized for picture arrangement in Deeplearning. Tensorflow is an open-source programming library created by Google in 2015 for numerical calculation. Keras is an open-source neural system library written in python, and it is equipped for running on MxNet, Deep learning, Tensorflow, and Theano. It intended to empower quick experimentation with Deepneural systems. The principal segment of this paper contains a general presentation about Deeplearning, TensorFlow, Keras, and dataset. The second section contains the basic theory about CNN, classifiers, and activation functions. The third section of this paper contains a literature review, research methodology, and the final section contains the experimental setup and results.

II. BASIC THEORY

A. Neural Network: Neural Network receives input and passes it through a number of hidden layers. Each hidden layer has a set of neurons, where each neuron is fully connected to all neurons in the previous layer. Each layer in a single layer functions independently. The last layer in the neural network is called „output layer," which represents the class to which input belongs.

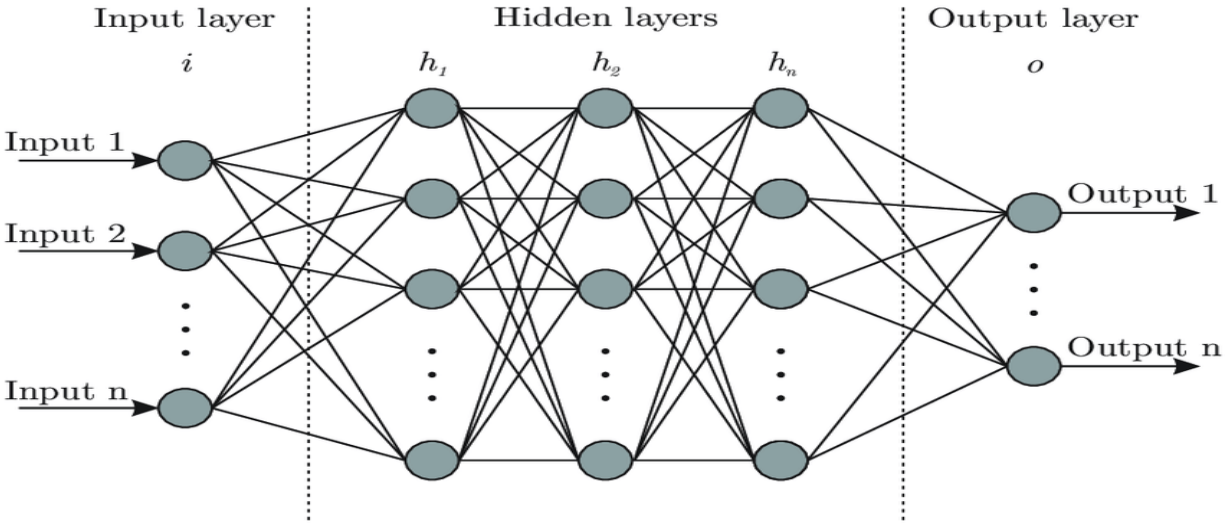


Fig 2.Neural Network Architecture

B. Convolutional Neural Network (CNN): Convolutional Neural Network is a special type of feed-forward artificial neural network, which inspired by the visual cortex. In CNN, the neuron in a layer is just associated with a little area of the layer before it, rather than all the neurons in a completely associated way, so CNN handle-less measures of loads and furthermore less number of neurons.

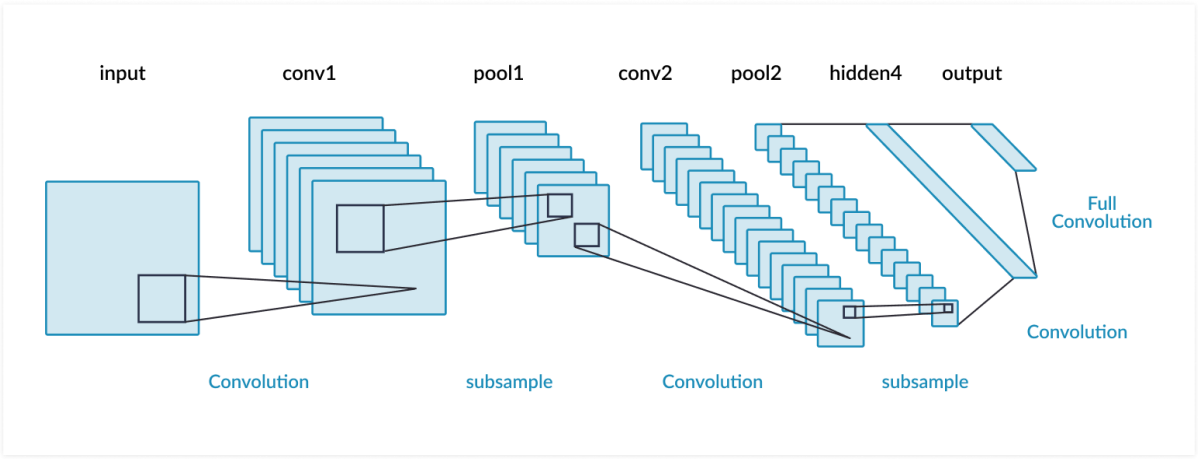


Fig 3 Convolutional Neural Network

Relu Activation Function: Relu F(x) = max (x, 0), is mostly used deep learning activation function, for hidden layers. A corrected direct unit has yielded „0‟ if the info is under „0‟, and crude yield „otherwise.‟Relu is the least difficult non-direct enactment work. Research has indicated that the real result is a lot quicker for enormous systems preparing. Most structures like TensorFlow, make it easy to utilize depend on concealed layers.

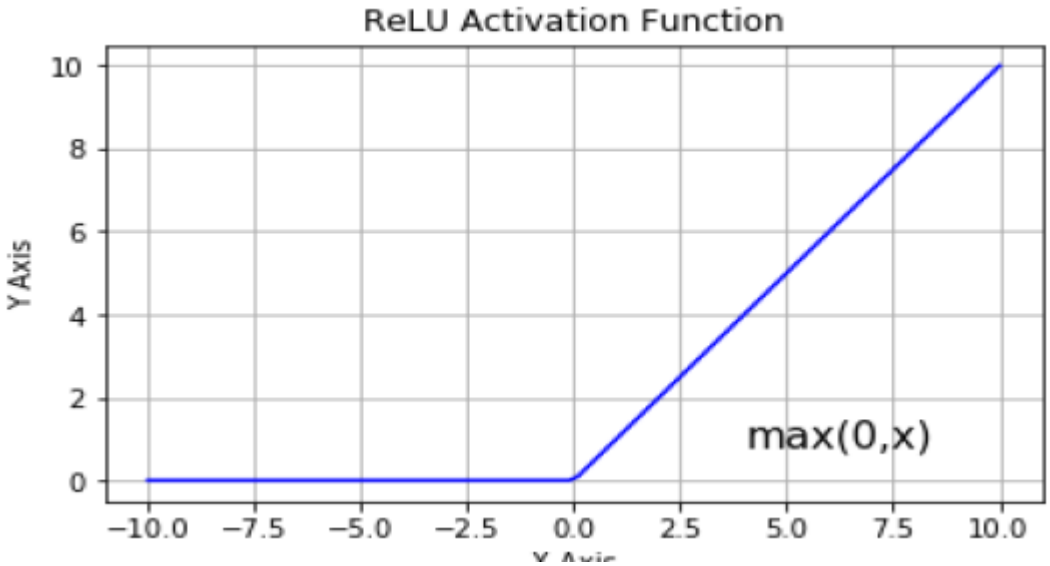


Fig. 4Relu activation function

C. Tanh activation function: Tanh function [tanh(x) = (ex - e-x) / (ex + e-x)] produces output in range of -1 to +1. It is continuous function, which produces output for every „x" value.

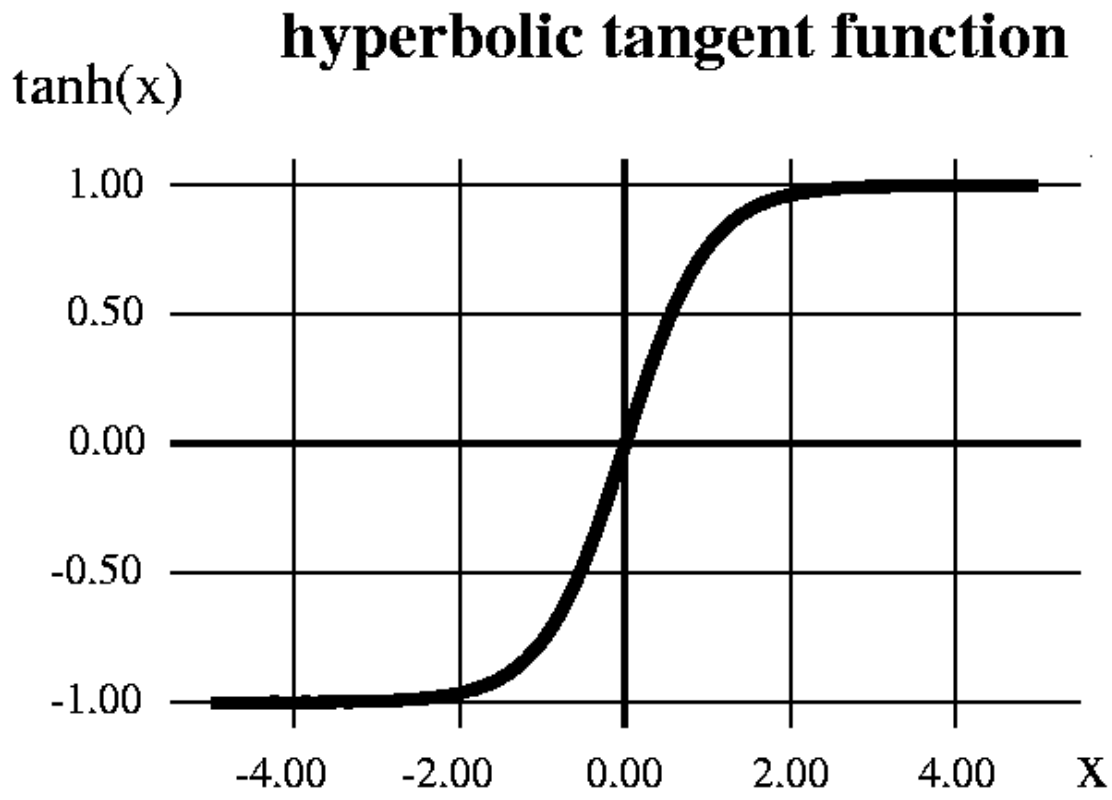


Fig. 5 Tanh activation function

D. Sigmoid classifier: Sigmoid classifier [F(x) = 1/ (1 + e-x)] takes any range of real numbers and returns the output value, which falls in the range of „0"to „1". It produces the curve in the „S" shape. The sigmoid classifier is mainly used for binary data classification.

E. Softmax classifier: The softmax classifier [F(x) = exi / (Σkj=0 exj)] squashes the outputs of each unit to be between 0 and 1, just like a sigmoid classifier. But it also divides each output such that the total sum of the outputs is equal to 1. The output of the softmax classifier is equivalent to a categorical probability distribution, and it tells you the probability that any of the classes are true. Softmax classifier is used for multiple data classification.

III. LITERATURE REVIEW

Hasbi Ash Shiddieqy, FarkhadIhsanHariadi, Trio Adiono "Implementation of Deep-Learning based Image Classification on Single Board Computer," In this paper, a deep-learning algorithm based on convolutional neural-network is implemented using python and learning for image classification, in which two different structures of CNN are used, namely with two, and five layers and It conclude that the CNN with higher layer performs classification process with much higher accuracy. Rui Wang, Wei Li, Runnan Qin and JinZhong Wu "Blur Image Classification based on Deep Learning," In this paper, a convolution neural network (CNN) of Simplified-Fast-Alexnet (SFA) based on the learning features is proposed for handling the classification issue of defocus blur, Gaussian blur, haze blur, and motion blur four blur type images. The experiment results demonstrate that the performance of classification accuracy of SFA, which is 96.99% for simulated blur dataset and 92.75% for natural blur dataset, is equivalent to Alexnet and superior to other classification methods. Sameer Khan and Suet-Peng Yong, "A Deep Learning Architecture for Classifying Medical Image of Anatomy Object," In this paper, a modified CNN architecture that combines multiple convolutions and pooling layers for higher-level feature learning is proposed. In this, medical image anatomy classification has been carried out, and it shows that the proposed CNN feature representation outperforms the three baseline architectures for classifying medical image anatomies. Ye Tao, Ming Zhang, Mark Parsons "Deep Learning in Photovoltaic Penetration Classification," this paper proposed a deep learning-based algorithm to differentiate photovoltaic events from other grid events, and it concludes that a deep convolutional neural network can achieve higher classification accuracy than a fully connected model.

IV. RESEARCH METHODOLOGY

The flow diagram of the proposed methodology is shown in fig. 4.1. Each square of the proposed stream outline is plainly named and speaks to preparing steps. Utilizing this technique, we look at four changed structures of CNN, with four unique blends of classifiers and actuation capacities.

In the initial step picture, the dataset is readied, there are four records in the dataset, which contains 10000 pictures of pooches and felines, were 8000 pictures utilized for preparing and 2000 pictures utilized for testing purposes. In the subsequent advance, characterize parameters for picture arrangement to python. In the third step, make CNN with two convolutional layers. At that point, we select various blends of initiation capacities and classifiers for examination purposes. In

the subsequent stages, we fit the made CNN to picture dataset and Train, Test the framework with preparing and test datasets individually. At last, we acquire the exactness for various CNN structures and think about this correctness's for execution estimation, and afterward get the resultant CNN structure.

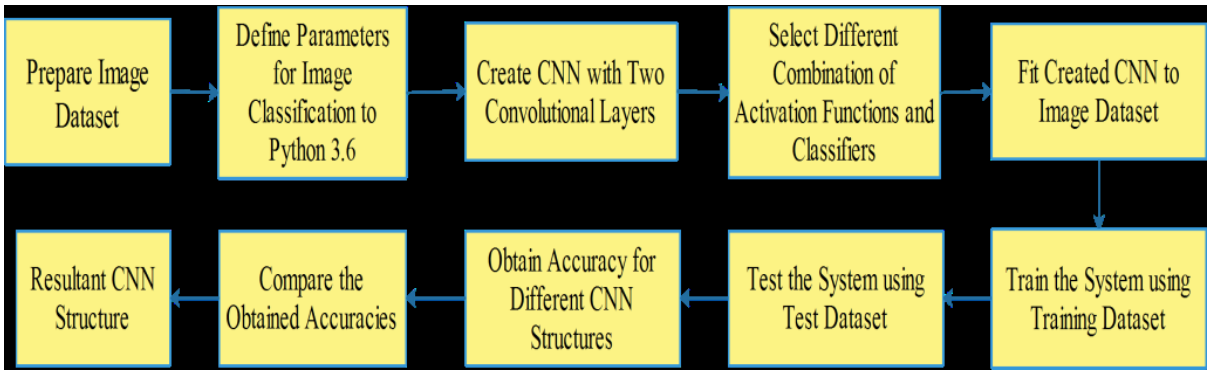


Fig. 6. Proposed Methodology

V. EXPERIMENTAL SETUP

In this paper, we perform probes windows 10 in python 3.6 on the CPU framework and make the CNN model dependent on Keras and TensorFlow libraries. The CNN model utilized for tests appears in fig 5.1. This model, for the most part, comprises four layers, including, convolutional, pooling, leveling, and completely associated layers.

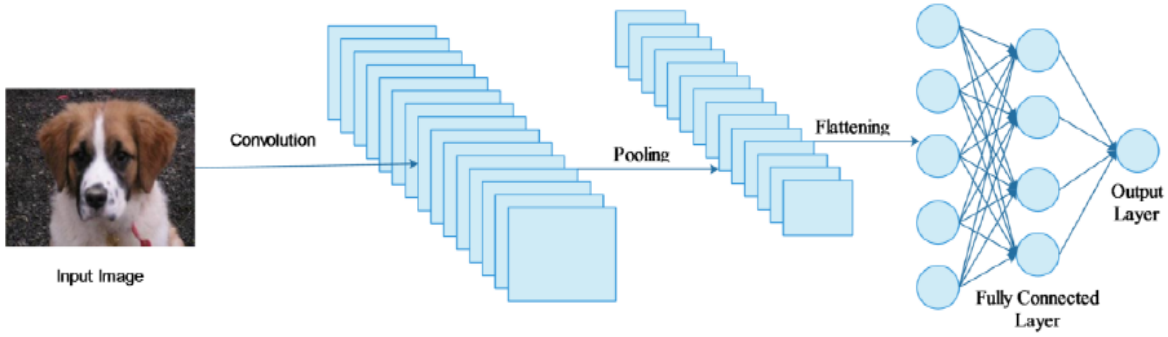


Fig. 7 Convolutional neural network model

For the convolutional layer, the size of the information picture is set to 64*64 pixels with three channels (RGB). To extricate the highlights from the picture, we utilize 32 channels of size 3*3 pixels. For the pooling layer, we utilize a window of size 2*2 pixels, which used to pack the first picture size for additional preparation. For execution estimation, we utilize two initiation works in particular, Relu (Rectified direct unit), Tanh (Hyperbolic digression), and two classifiers, specifically Softmax, Sigmoid. In the examination, we utilize a blend of these enactment capacities and classifiers and dissect what mix gives better arrangement exactness for paired picture grouping.

Table 5.1 Combinations of activation function and classifier

Serial number	Activation Function	Classifier
1.	Softmax	Relu
2.	Sigmoid	Relu
3.	Softmax	Tanh
4.	Sigmoid	Tanh

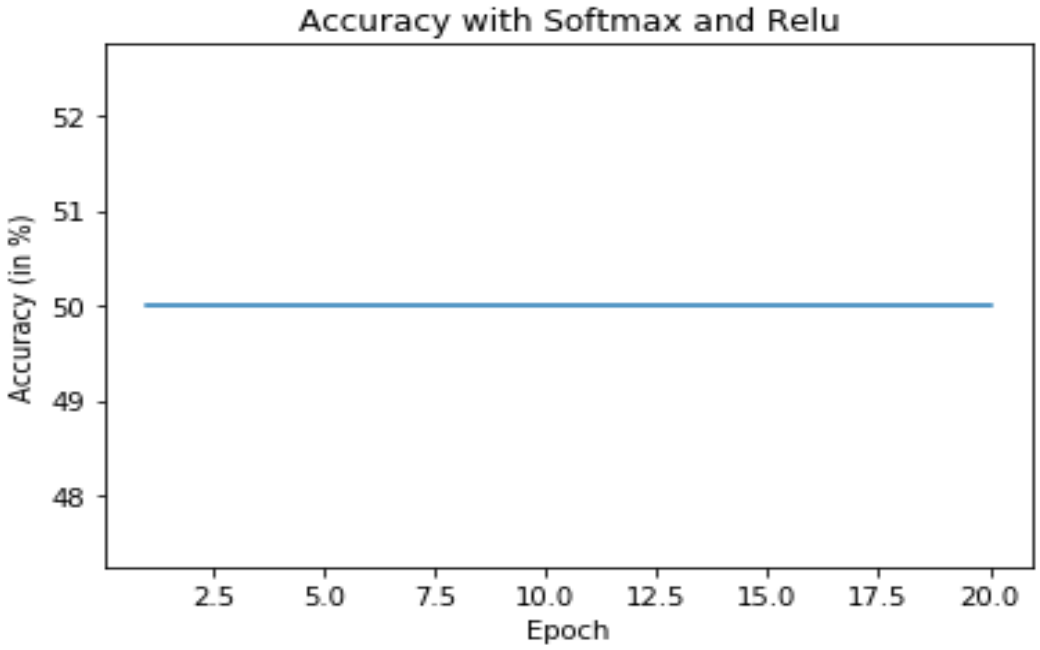
Subsequent to actualizing all the above parameters in python, we train and test the CNN model utilizing preparing and test datasets, and afterward acquire precision for various CNN structures. After then, we think about the acquired correctness and discover a CNN structure with higher exactness.

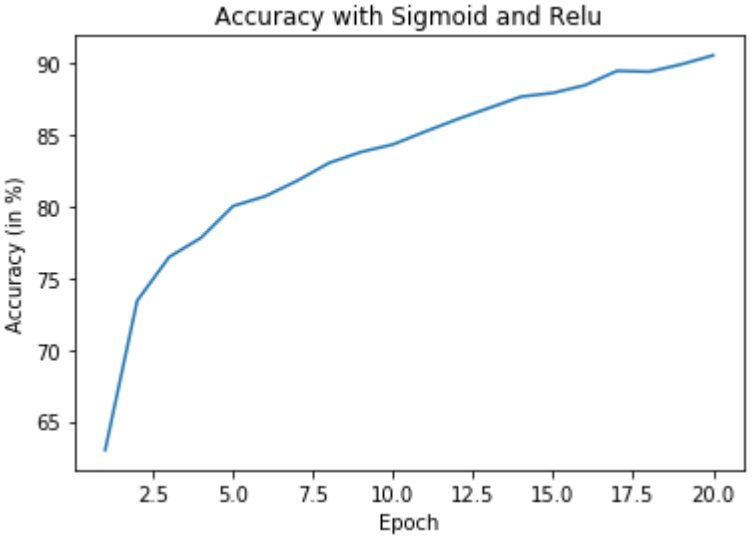
VI. RESULT

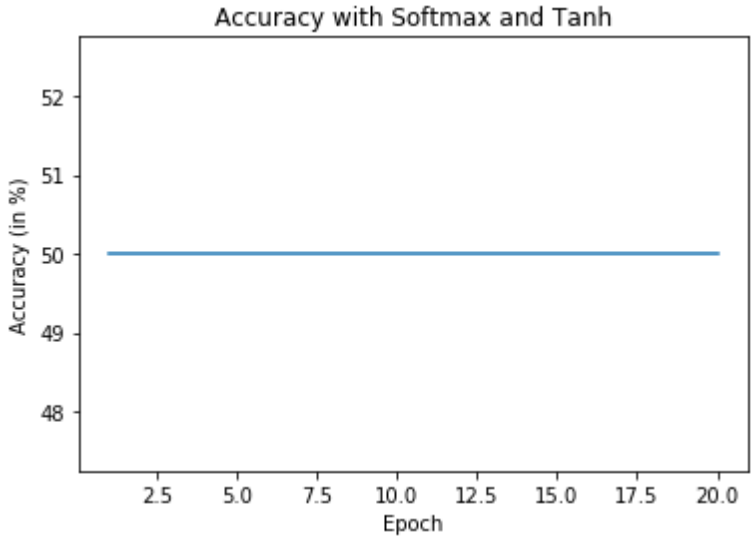
There are four characterization exactness got (as appeared in table 6.1) from above characterized CNN structures, and this correctness's are graphically speaking to in beneath diagrams (a, b, c, d), where each CNN structure appears with ages and correctness.

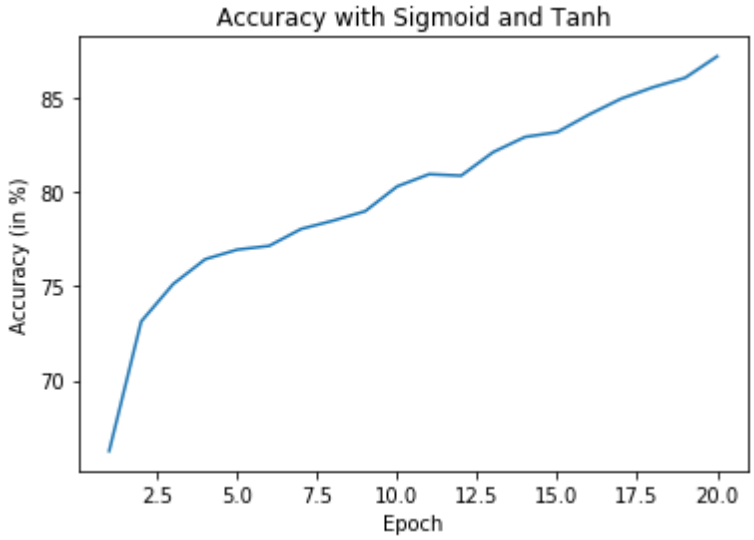
Table 6.1 Obtained accuracies with different combinations of activation function and classifier

Number of convolutional layers	Activation Function	Classifier	Classification Accuracy
2	Relu	Softmax	50%
2	Relu	Sigmoid	90.54%
2	Tanh	Softmax	50%
2	Tanh	Sigmoid	87.22%




Accuracy with Sigmoid and Relu
Accuracy (in %)
90
85
80
75
70
65
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch


Accuracy with Softmax and Tanh
Accuracy (in %)
52
51
50
49
48
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch


Accuracy with Sigmoid and Tanh
Accuracy (in %)
85
80
75
70
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch

We compare accuracies of the graph a, b, c and d and we find out that CNN with a combination of Relu activation function and Sigmoid classifier (graph b) gives better accuracy 90.54%, which is far better than accuracies of graph a and graph c (50%), and slightly better than graph d (87.22%).

We think about exactness's of diagram a, b, c, and d and we discover that CNN with a blend of Relu actuation work and Sigmoid classifier (chart b) gives better precision 90.54%, which is obviously better than correctness's of diagram an and chart c (half), and somewhat better than chart d (87.22%).

VII. CONCLUSION

Deep learning is a learning strategy for information examination and expectations; these days, it additionally turns out to be well known for picture arrangement issues. In this paper, a profound learning convolutional neural system dependent on Keras and TensorFlow is sent utilizing python for paired picture grouping. In this investigation, we look at four changed structures of CNN on the CPU framework, with various blends of classifiers and actuation capacities. With tests, we acquired outcomes for every mix and saw that for double picture arrangement, the Relu initiation work and Sigmoid classifier blend gives better order precision (90.54%) than some other mix of actuation capacity and classifier. Along these lines, we presume that on the CPU framework, the Relu enactment work and Sigmoid classifier give better order exactness for parallel picture characterization.