Artificial Neural Network – Homework 2
Analyzing the Impact of Network Depth and Width on Training Models with Artificial Neural Networks

Ammar Abdurrauf – 170421930

Introduction

In this study, various network models are designed to observe the behavior and success of each artificial neural network (ANN) model when trained using three different datasets provided by TensorFlow, namely MNIST, Fashion MNIST, and CIFAR-10 datasets. The model architectures are varied by changing the depth and width of the neural network structure. The experiment is also examined by varying the optimizer algorithm and learning rate. The results of model training is analyzed, compared, and discussed at the end of the study.

Methods and Dataset

Dataset

Three image datasets used in this study are MNIST, Fashion MNIST, and CIFAR10 datasets provided by Tensorflow. Several samples from the dataset is shown in Figure 1-3.
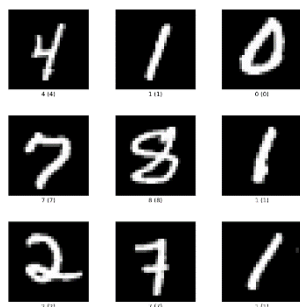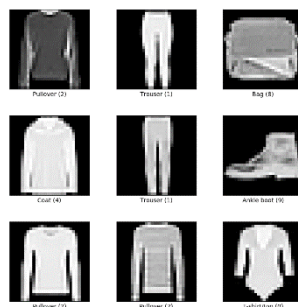

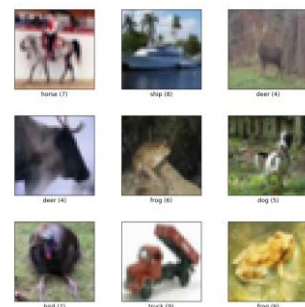Figure 1. MNIST dataset

Figure 2. Fashion MNIST dataset

Figure 3. CIFAR10 dataset

Neural Network Structures

This study examines three different structures of the model based on the depth and width of the network. The first structure of the ANN model is built using ten hidden layers with each layer having 32 nodes (neurons) and is considered a deep ANN structure. The second structure, considered a wide ANN structure, has five hidden layers with 64 nodes in each layer. The last structure is designed to be both deep and wide, featuring ten hidden layers with 64 nodes in each layer. The similarity of these models lies on the number of input layer which is the size of the input image in dataset, the Flatten layer, the Dropout layer, which drops some nodes in the network to prevent overfitting –the parameters for the Dropout layer is set to 0.2 which means 20% of the weights will be set to 0, and output layer with ten classes. Moreover, the Relu activation function is used in each hidden layer. The neural network structures proposed in this study is depicted in Figure 4-5.
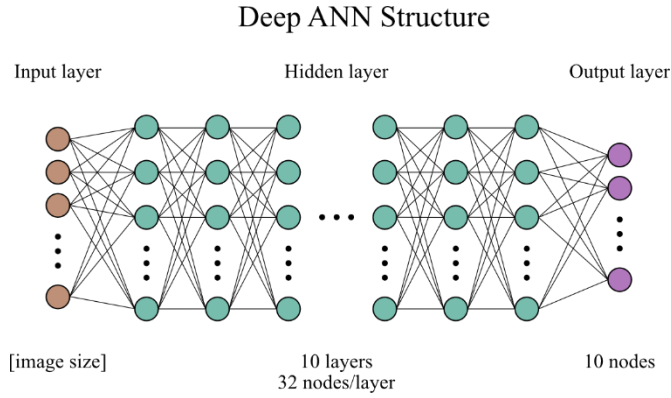
Deep ANN Structure

Figure 4. Deep ANN Structure
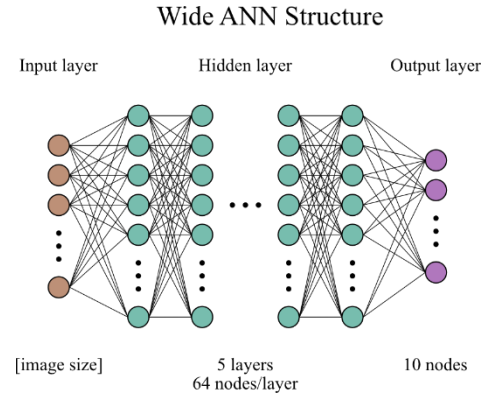
Wide ANN Structure

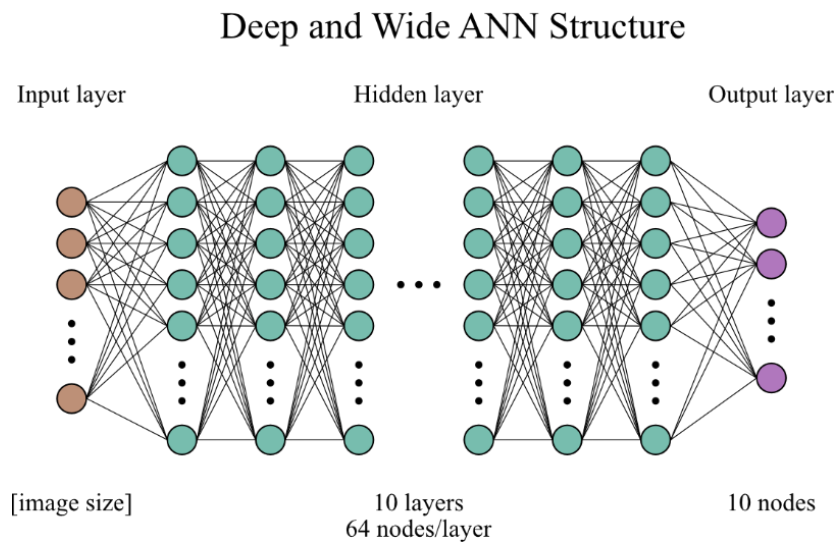Figure 5. Wide ANN Structure

Deep and Wide ANN Structure

Figure 6. Deep and Wide Structure

Two different optimizer available in Keras library, namely Adam and RMSprop, are used for training the neural network models. For each optimizer algorithm, two different learning rates, namely 0.001 and 0.0001, are employed.

Method

The K-Fold Cross Validation method is used in this study to examine the models with various structures specified earlier. The number of K is set to 5, that means five groups is split from the dataset and each group is used for validation on each training. Each ANN structure is trained using aforementioned optimizer algorithms and learning rates resulting in 36 training combinations. The number of epoch for training the models is set to three for the MNIST and Fashion MNIST datasets and two for the CIFAR10 dataset, while the batch size is fixed at 128 for each dataset. The training was conducted on two platforms, namely Google Colab and Spyder IDE to speed up the training process.
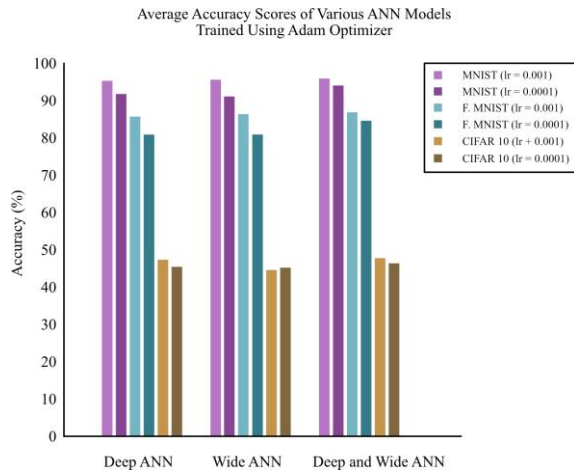
Figure 7. Average accuracy of various ANN models trained using Adam optimizer
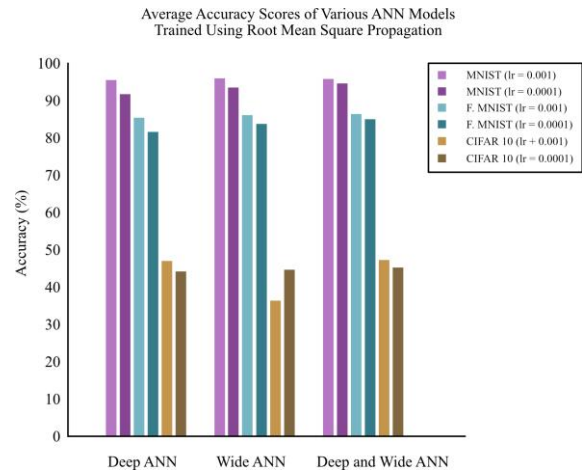

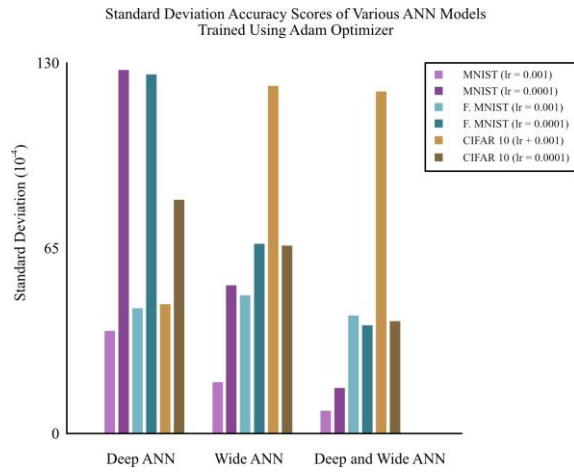Figure 8. Average accuracy of various ANN models trained using RMSprop optimizer

Figure 9. Standard Deviation Accuracy of various ANN models trained using Adam optimizer
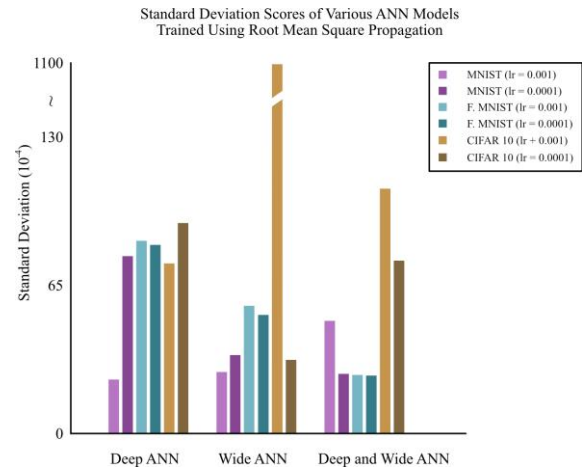
Figure 10. Standard Deviation Accuracy of various ANN models trained using RMSprop optimizer

Results and Discussion

Training Results

The result accuracy for each cross validation is saved in an Excel file and the standard deviation and average of every cross validation on a single training phase is also calculated. The results for each ANN model training along with its accuracy is fully shown in Table 1. The average accuracy and standard deviation of various ANN models is elucidated in a bar graph format as shown in Figure 7-10.

Figure 7 and 8 show the average accuracy of the models with various neural network structure. In general, the graphs indicate that every neural network structure trained with the datasets using both Adam and RMSprop optimizer resulted in similar average accuracy among each other. However, Figure 9 and 10 show that the standard deviation of each accuracy value varies between 0.00078 and 0.10812. A low standard deviation in accuracy suggests minimal variability during the cross-validation process, indicating consistent performance. On the other hand, a high standard deviation in accuracy implies a more fluctuating or unstable training process.

The models trained on MNIST and Fashion MNIST datasets did not show significant differences in accuracy. For the MNIST dataset, the accuracy is around 95% with a learning rate (lr) of 10^{-3} and between 91-94% with lr = 10^{-4}. Meanwhile, the Fashion MNIST dataset exhibited approximately 86% accuracy at lr = 10^{-3} and between 80-84% at lr = 10^{-4}. The results from training the models using CIFAR-10 dataset show a poor performance for every neural network structure. The deep and wide model, trained with the Adam optimizer and a learning rate of 0.001, achieved its average accuracy of 47.76%. When using the same model and parameters with the Fashion MNIST dataset, the model performed remarkably well, reaching an accuracy of 86.86%. The top three performing models on the MNIST dataset were the wide model with RMSprop at 0.001, the deep and wide model with Adam at 0.001, and the deep and wide model with RMSprop at 0.001, boasting accuracies of 95.96%, 95.92%, and 95.83%, respectively.

A deep neural network model trained using the three datasets with Adam optimizer algorithm at learning rate 0.0001 resulted in a high standard deviation. On the other hand, a high standard deviation was also observed after training was conducted using the CIFAR-10 dataset with wide ANN and deep & wide ANN models, utilizing both optimizers at a learning rate of 0.001. The highest standard deviation is reached by the model with wide neural network structure trained using CIFAR-10 dataset utilizing the RMSprop optimizer and learning rate of 0.001 with the standard deviation of 0.1080. In contrast, the MNIST dataset appeared to be more stable during model training, showing lower standard deviation values.

Discussion

The models proposed in this study did not show a significant difference in the accuracy values resulted from the K-Cross Validation method used in the model training. However it provides some insights through the standard deviation values from each model. The graph in Figure 9 and 10 show that the deep and wide neural network model tend to be more stable compared to other models. The accuracy values resulting from the model training using deep and wide hidden layers are also higher than several models. In conclusion, it can be inferred that a model with deep and wide neural network structure tends to perform better. Nevertheless, for simpler tasks such as classifying the MNIST dataset, both wide and deep network architectures are also effective.

Project Link

https://github.com/aabdurrauf/Artificial-Neural-Network-Marmara-University-Course/tree/main/HW2%20-%20Deep%20vs%20Wide%20ANN

DERİN (Green Table)

DERİN	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9562	0,896	0,9469	0,9258	0,9538	0,9164	0,9541	0,9244	0,9546	0,9259	0,95312	0,9177	0,00359819	0,01275069
F.MNIST	0,8513	0,8143	0,8561	0,8073	0,86	0,7891	0,8627	0,8236	0,8555	0,8099	0,85712	0,80884	0,00438657	0,01265733
CIFAR-10	0,4694	0,4466	0,4682	0,44,95	0,4737	0,4511	0,4784	0,4536	0,4772	0,4658	0,47338	0,454275	0,00454335	0,00821112

GENİŞ (Green Table)

GENİŞ	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9572	0,9164	0,9566	0,9067	0,9541	0,9167	0,9577	0,906	0,9538	0,909	0,95588	0,91096	0,00180748	0,00522331
F.MNIST	0,8692	0,8057	0,8648	0,815	0,8634	0,8173	0,8651	0,8038	0,8559	0,8031	0,86368	0,80898	0,00485664	0,00666386
CIFAR-10	0,4585	0,4407	0,4317	0,4542	0,4467	0,4521	0,4356	0,4572	0,4573	0,4559	0,44596	0,45202	0,01221753	0,00661037

DERİN VE GENİŞ (Green Table)

DERİN VE GENİŞ	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9586	0,9428	0,9585	0,9404	0,9603	0,9393	0,9597	0,94,36	0,9588	0,9396	0,95918	0,940525	0,00078549	0,00158614
F.MNIST	0,864	0,8496	0,8645	0,8451	0,8699	0,8443	0,8725	0,8408	0,8723	0,8498	0,86864	0,84592	0,00413981	0,00381143
CIFAR-10	0,4718	0,4571	0,4597	0,4638	0,4896	0,4656	0,4815	0,4666	0,4854	0,4662	0,4776	0,46386	0,01198019	0,00392785

Prop DERİN (Blue Table)

DERİN	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9527	0,9171	0,9568	0,9083	0,9568	0,9286	0,9558	0,912	0,9518	0,9196	0,95478	0,91712	0,00236685	0,00777734
F.MNIST	0,8501	0,816	0,8564	0,819	0,8418	0,8278	0,8632	0,8049	0,8596	0,8143	0,85422	0,8164	0,00844938	0,00826952
CIFAR-10	0,4722	0,4458	0,475	0,4441	0,4637	0,4442	0,4794	0,426	0,4619	0,4498	0,47044	0,44198	0,00745875	0,0092267

Prop GENİŞ (Blue Table)

GENİŞ	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9608	0,934	0,9586	0,9411	0,9608	0,9347	0,9623	0,9333	0,9554	0,9325	0,95958	0,93512	0,00268365	0,00344122
F.MNIST	0,8522	0,8366	0,8608	0,8408	0,867	0,8413	0,8641	0,8402	0,8631	0,8289	0,86144	0,83756	0,00562432	0,00518102
CIFAR-10	0,2117	0,4435	0,4059	0,4437	0,4369	0,4512	0,4002	0,4482	0,2259	0,4466	0,33612	0,44664	0,10812133	0,00323002

Prop DERİN VE GENİŞ (Blue Table)

DERİN VE GENİŞ	CV-1 0.001	CV-1 0.0001	CV-2 0.001	CV-2 0.0001	CV-3 0.001	CV-3 0.0001	CV-4 0.001	CV-4 0.0001	CV-5 0.001	CV-5 0.0001	ORT 0.001	ORT 0.0001	STD 0.001	STD 0.0001
MNIST	0,9583	0,9446	0,9601	0,944	0,9633	0,9503	0,9598	0,9442	0,9501	0,9454	0,95832	0,9457	0,00494287	0,00262679
F.MNIST	0,8609	0,8526	0,8673	0,8507	0,8621	0,8506	0,8653	0,8491	0,8649	0,8458	0,8641	0,84976	0,00257682	0,00253831
CIFAR-10	0,4829	0,4542	0,4797	0,452	0,4777	0,4566	0,4574	0,4595	0,4657	0,4398	0,47268	0,45242	0,01073695	0,0075863

Table 1.K-Cross Validation results from every model training