FASHION - MNIST CLOTHING CLASSIFICATION

Under Supervision of :
Dr. Ankit Rajpal

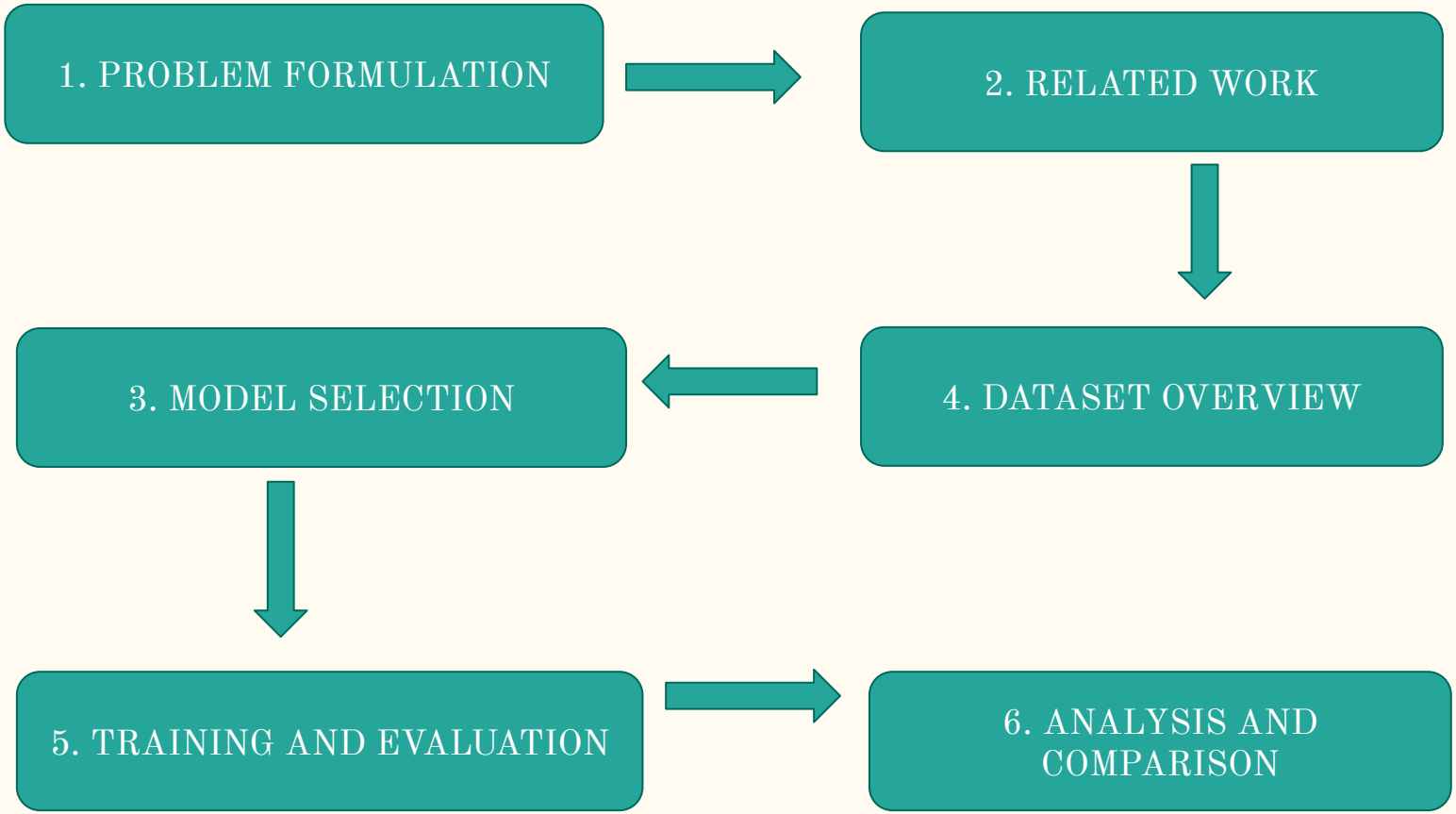
—

By:-
Aditya Sethi (07), Paras Kumar(37), Ronit Kumar(48)

PROJECT OVERVIEW

- METHODOLOGY
- PROBLEM STATEMENT
- RELATED WORK
- DATASET OVERVIEW
- MODEL SELECTION
- MODEL EVALUATION
- RESULT
- CONCLUSION
- REFERENCES

METHODOLOGY

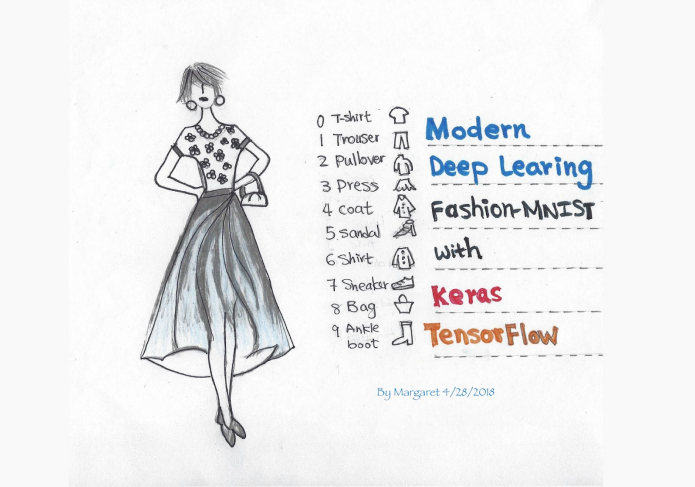


PROBLEM STATEMENT

In the fast-paced fashion industry, efficient and accurate image classification is crucial for various applications such as trend analysis, product recommendation, and inventory management.

Therefore, there is need to develop a robust and scalable image classification model tailored specifically for the fashion industry. This model should effectively classify fashion images into relevant categories such as clothing types, patterns, colors, and styles, while also accommodating variations in lighting, poses, and backgrounds.

Additionally, it should be capable of exhibiting high accuracy, scalability, and computational efficiency to meet the demands of real-world fashion applications.



REF : https://blog.tensorflow.org/2018/04/fashion-mnist-with-tfkeras.html

RELATED WORK

PAPER 1 - Valderi Reis Quietinho Leithardt, Anita Maria Rocha Fernandes, Sérgio Correia, Rodrigo Lyra, "Classifying Garments from Fashion-MNIST Dataset Through CNNs", in Advances in Science Technology and Engineering Systems Journal, February 2021. DOI: 10.25046/aj0601109. [2]

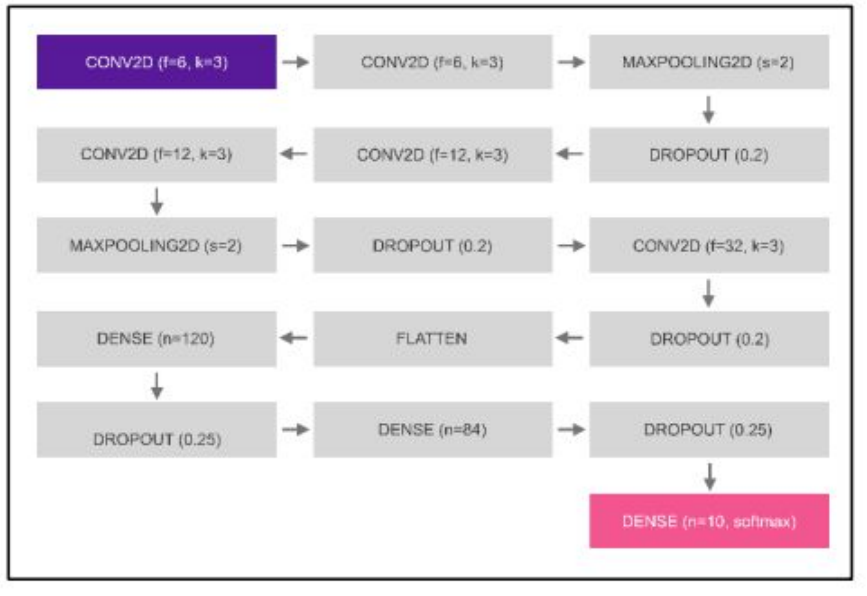
PROPOSED MODELS :

1. CNN dropout-1 and CNN dropout-3 : Both models use two consecutive blocks containing: a convolution, a max pooling, and finally a drop out. These blocks are connected to two more fully connected layers, who are connected to an output layer of ten neurons, each one representing a category.

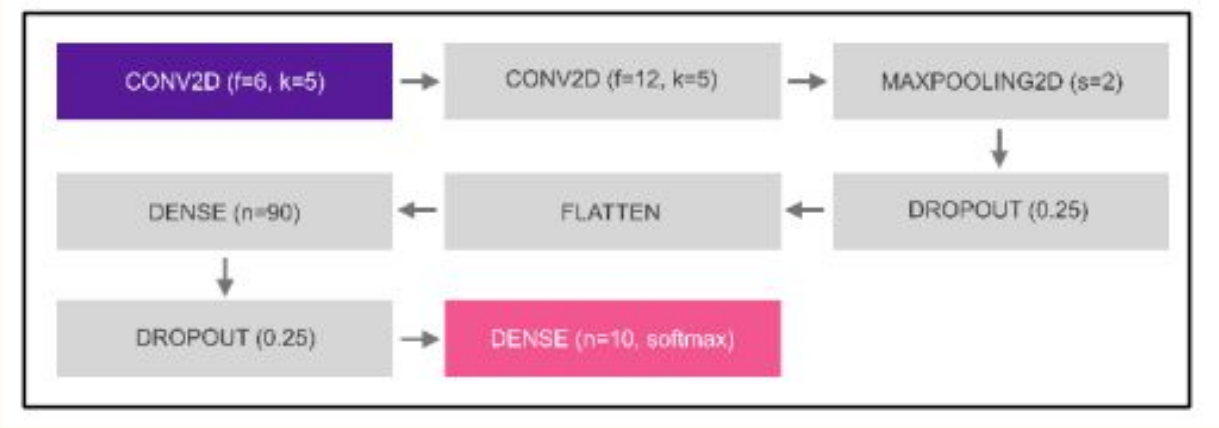
The only difference between these two models is that cnn-dropout-3 has considerably lower dropout values

2. CNN dropout -2 :

This proposed model is very similar to the cnn-dropout-1 model. However, it has two layers of convolutions before each max pooling.



3. CNN SIMPLE : Cnn-simple is a model with less layers. It has only two convolutions, followed by a fully connected layer, in addition to the respective dropout and max pooling like other models.



RESULTS

On the training dataset, the most accurate model was CNN Simple, with 98.95% of accuracy. Other models were also acceptable, based on the results obtained: 98.06% (CNN-dropout-3), 97.51% (CNN-dropout-2), and even the worst model (CNN-dropout-1), got an accuracy of 96.46%

MODEL	ACCURACY
CNN SIMPLE	98.95
CNN- DROPOUT-3	98.06
CNN - DROPOUT-2	97.51
CNN - DROPOUT-1	96.46

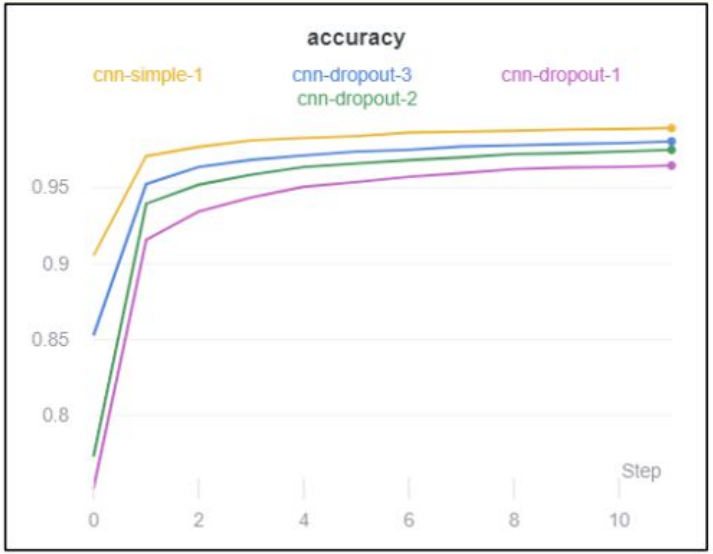


Figure 7: Accuracy

Further they compared the result of their models with the traditional non-convolutive machine learning algorithms

Model	Accuracy
cnn-dropout-3	99.10%
cnn-dropout-2	99.08%
cnn-simple-1	99.05%
cnn-dropout-1	98.69%
support vector classification	89.70%
gradient boosting	88.00%
random forest	87.30%
multilayer perceptron	87.00%
k neighbors	85.40%
logistic regression	84.20%
linear support vector classification	83.60%
stochastic gradient descent	81.90%
decision tree	79.80%
Perceptron	78.20%
passive aggressive classifier	77.60%
extra tree	77.50%
gaussian naive bayes	51.10%

Results shows that even their worst results (cnn-dropout-1) got better result than previously proposed algorithms

PAPER 2 - M. Kayed, A. Anter and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," in 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 2020, pp. 238-243, DOI: 10.1109/ITCE48509.2020.9047776.

PROPOSED MODEL :

In this paper, the author uses the last LeNet-5 for Fashion-MNIST image classification. It was used because it is simple and gives high-performance results in several domains. It is built on local receptive fields, shared weights and special subsampling.

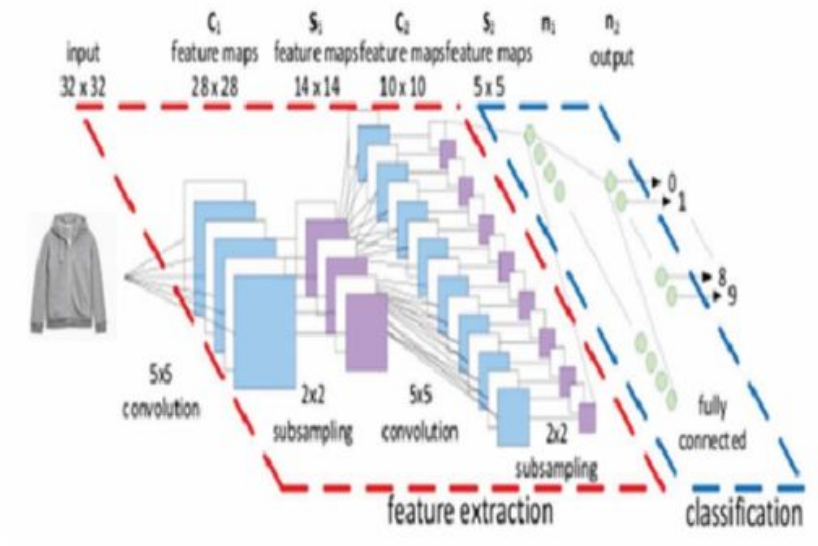


Figure 4. The architecture of our proposed LeNet-5 model.

RESULTS

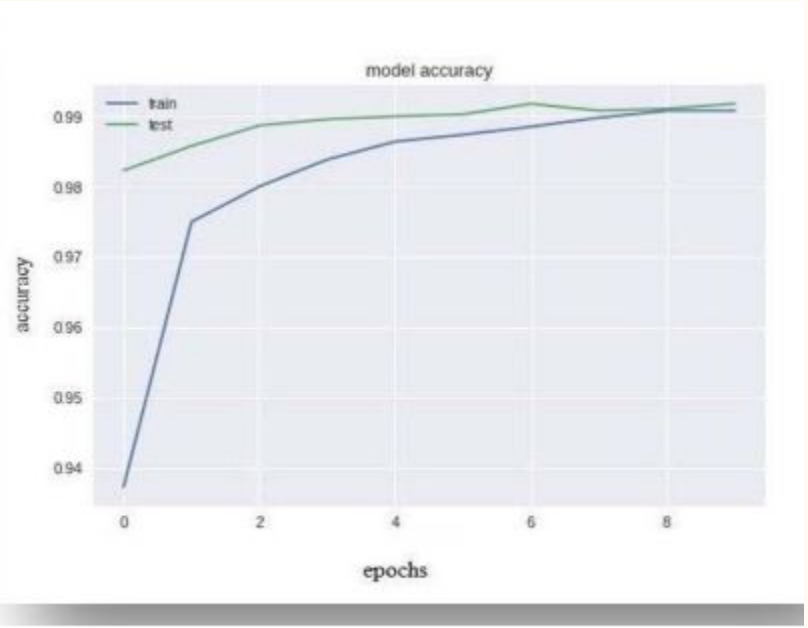


Figure 5. The model accuracy with different number of epochs.

A comparison between LeNet-5 and the other classification models on Fashion MNIST dataset.

Model(Method)	Test Accuracy
Three-layer Neural Network	87.23%
CNN using SVM activation function	90.72%
CNN using Softmax activation function	91.86%
KNeighborsClassifier	85.40%
Decision Tree Classifier	79.80%
RandomForestClassifier	87.30%
CNN-LeNet-5	98.80%

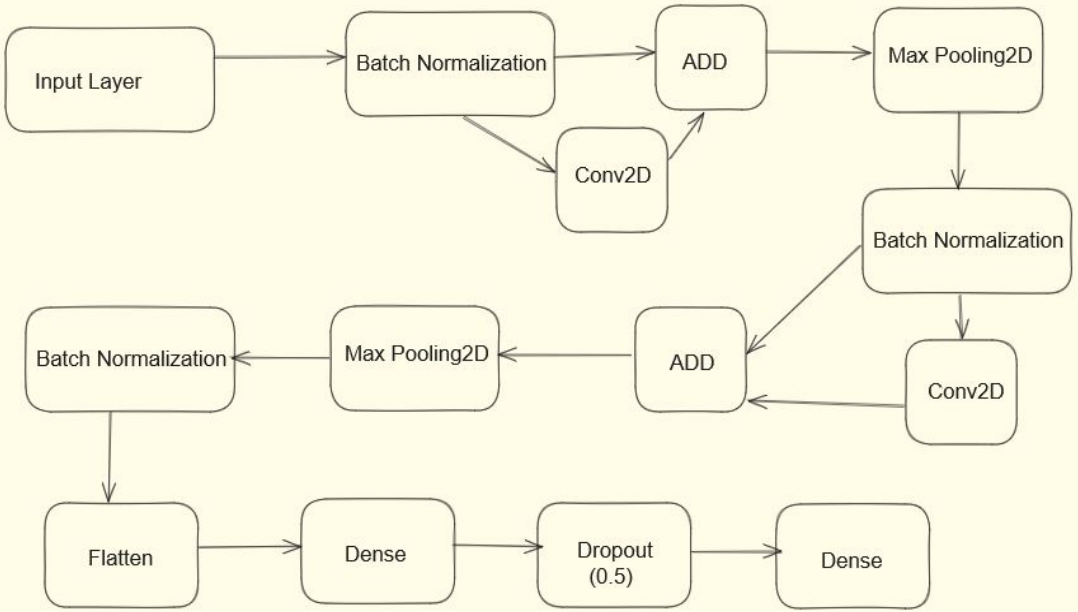
PAPER 3 - S. Bhatnagar, D. Ghosal and M. H. Kolekar, "Classification of fashion article images using convolutional neural networks," 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 2017, pp. 1-6, doi: 10.1109/ICIIP.2017.8313740.

PROPOSED MODEL :

A) 2 Convolutional and max pooling layers one after the other. Each convolutional layer has 32 filter of size 3x3, and max pooling were performed on every 2x2 pixels.

B) Batch Normalization was done before every convolutional layer to improve the training speed of the model. With Batch Normalization we were able to achieve the same loss function value after 10 epochs which was achieved after 40 epochs without Batch Normalization.

C) Residual Skip : we add the previous input and current value of convoluted output to get the final output.



RESULTS

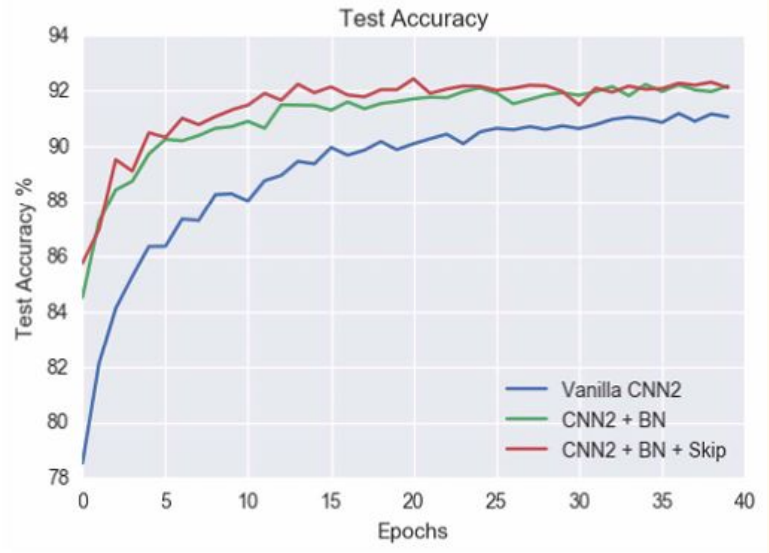


Fig. 3: Test Accuracy vs Epoch

Comparison between SVC with rbf kernel and the other classification models on Fashion MNIST dataset.

Model(Method)	**Test Accuracy**
SVC C = 10; kernel : rbf	89.70%
CNN2	91.17%
CNN2 + BatchNorm	92.22%
CNN2 + BatchNorm + Skip	92.54%

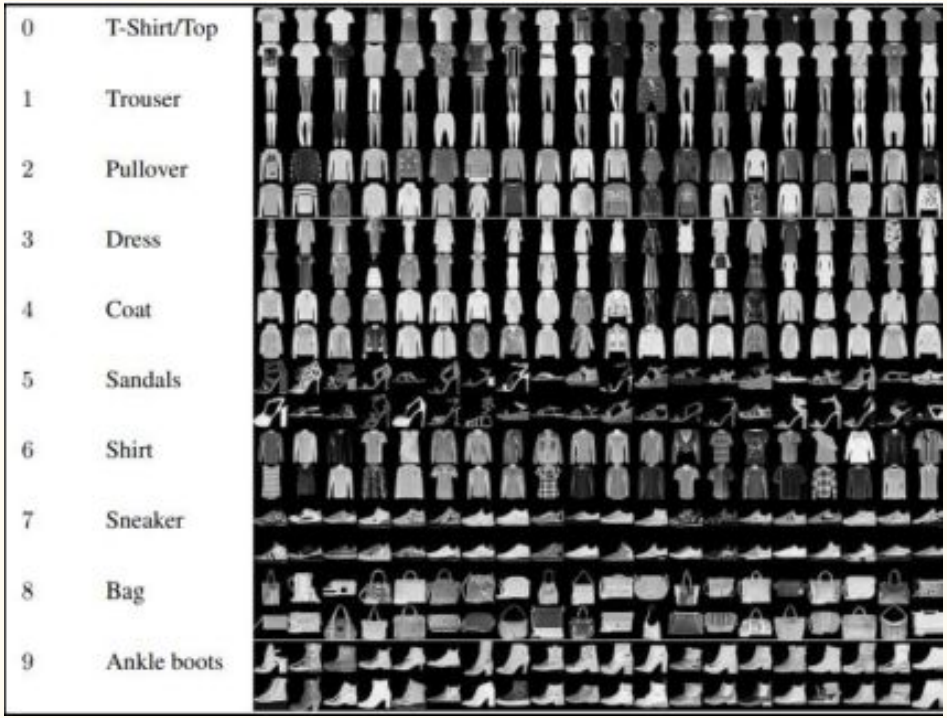
CONCLUSION : We were able to achieve an accuracy of 92.54% by using a two layer CNN along with Batch Normalization and Skip Connections. We can clearly see how Batch Normalization and Skip Connections help improve the overall accuracy and significantly reduce the training time.

DATASET OVERVIEW

DATA SET USED :

Fashion-MNIST is a direct drop-in alternative to the original MNIST dataset, for benchmarking machine learning algorithms . MNIST is a collection of handwritten digits, and contains 70000 greyscale 28x28 images, associated with 10 labels, where 60000 are part of the training set and 10000 of the testing. Fashion-MNIST has the exact same structure, but images are fashion products, not digits.

A sample of this set is presented as follows : -



Ref : https://www.kaggle.com/datasets/zalando-research/fashionmnist

Train Image shape :

```
[ ]  train_images.shape

     (60000, 28, 28, 1)
```

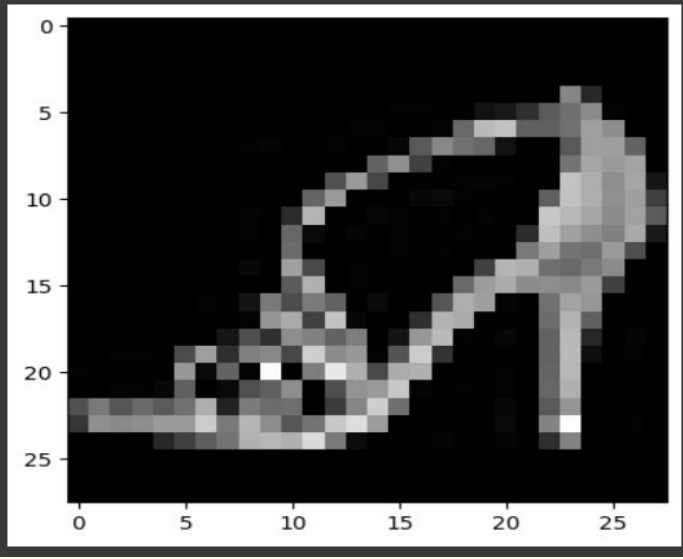
Test image Shape :

```
[ ]  test_images.shape

     (10000, 28, 28, 1)
```

Displaying an image from Dataset :

```
# Display an image from the dataset
plt.imshow(train_images[8000], cmap='gray')
plt.show()
```



Normalizing data : Normalizing the dataset to bring all the input features to a similar scale :

```
[ ]  # Normalize pixel values to be between 0 and 1
     train_images, test_images = train_images / 255.0, test_images / 255.0
```

MODEL SELECTION

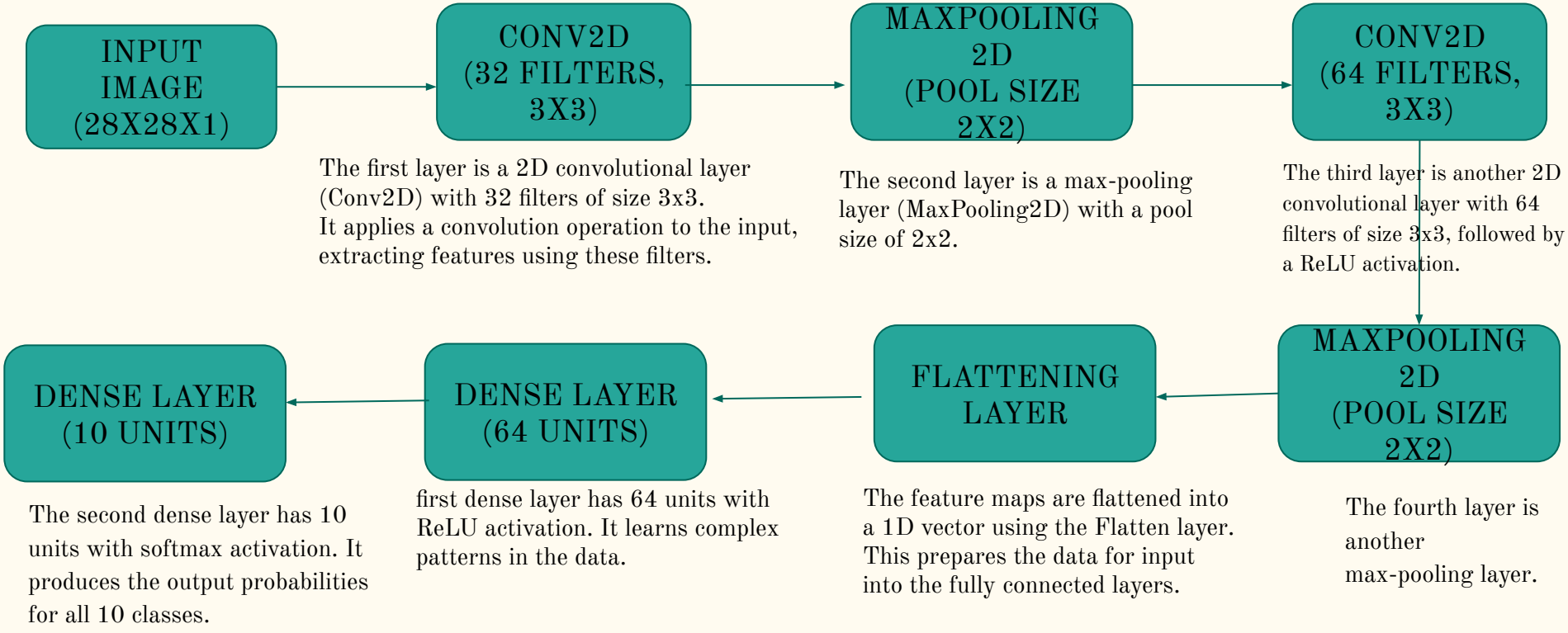
To obtain results for classification, 3 different CNN models were analyzed in python with Tensorflow and Keras. Collab Notebook was used for training using GPU-4.

Proposed Models are :

1. CNN - SIMPLE
2. CNN WITH DROPOUT
3. LENET 5 MODEL

CNN- SIMPLE

Cnn-simple is a model with less layers. It has only two convolutions, followed by a fully connected layer.



CNN WITH DROPOUT

```
# Train CNN with dropout (dropout rate 0.2)
cnn_dropout_02 = create_simple_cnn()
cnn_dropout_02.add(Dropout(0.2))
cnn_dropout_02_history = train_model(cnn_dropout_02, x_train, y_train, x_test, y_test)
cnn_dropout_02_accuracy = cnn_dropout_02.evaluate(x_test, y_test, verbose=0)[1]
print("CNN with Dropout (0.2) Accuracy:", cnn_dropout_02_accuracy)
plot_accuracy(cnn_dropout_02_history, "CNN with Dropout (0.2)")
```

- Dropout works by probabilistically “dropping out,” inputs to a layer, which may be input variables in the data sample or activations from a previous layer.
- Dropout 0.2 suggests to keep 80% and set 20% of the inputs to 0.
- We tested Dropout of 0.2 and 0.1.

LENET - 5 MODEL

In LeNet - 5, there are 7 layers in total.

- 2 Convolutional layers with filter size of 5x5 and stride of 1 also with ReLU activation
- 2 Average pooling layers with filter size of 2x2
- 2 fully connected layers with ReLU activation and
- 1 output layer with softmax activation.

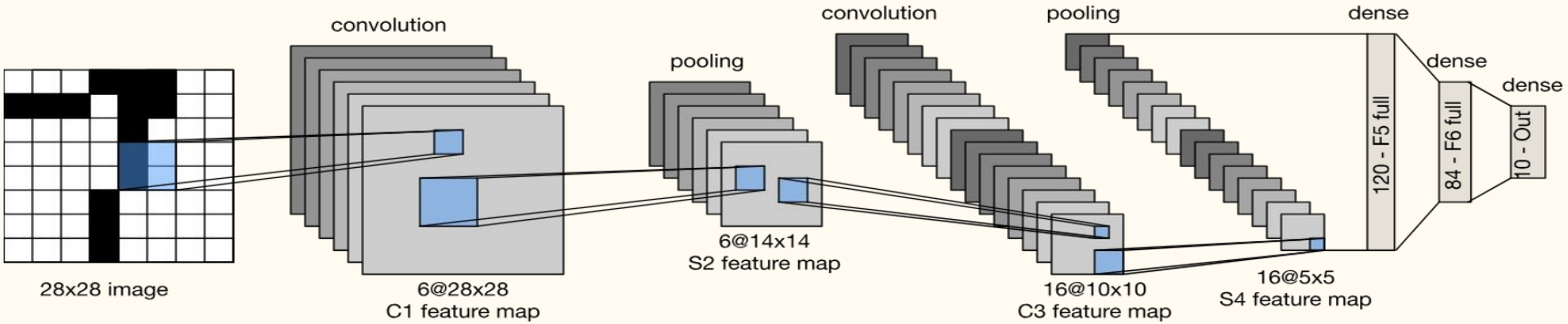
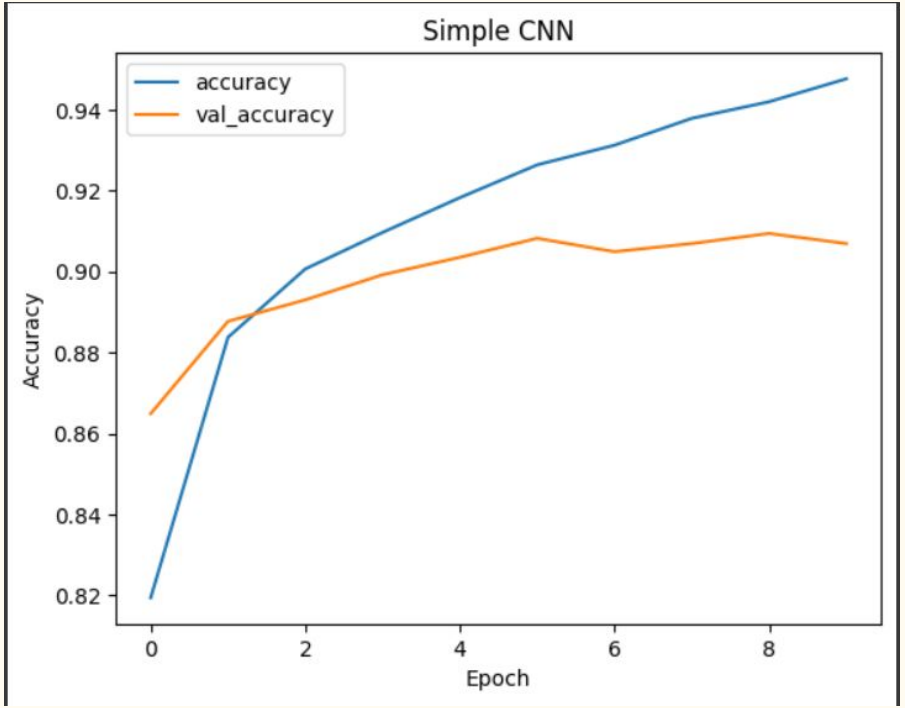


Image Ref : https://d2l.ai/chapter_convolutional-neural-networks/lenet.html

MODEL EVALUATIONS

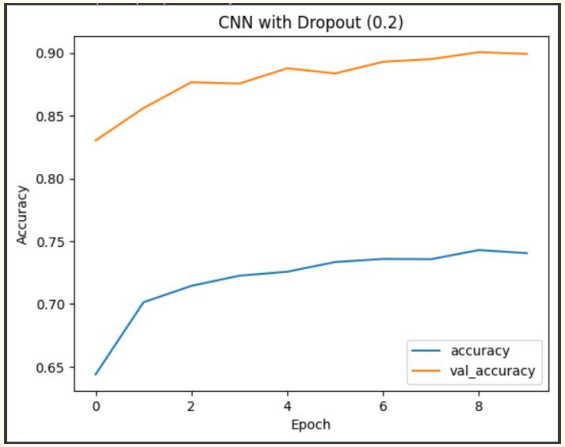
RESULT WITH **SIMPLE CNN**

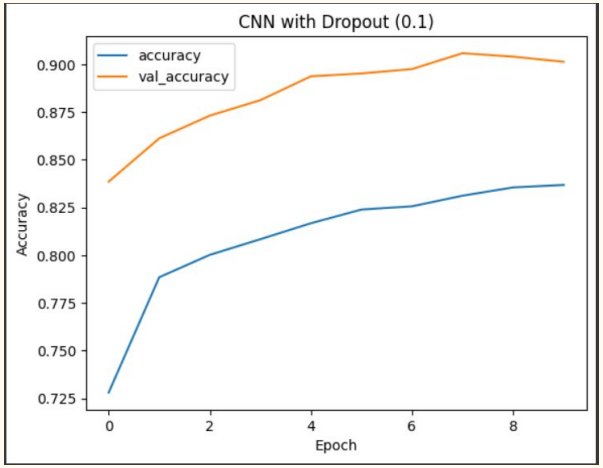


Accuracy	90.68
Epochs	10
Batch Size	32

MODEL EVALUATIONS

RESULT WITH **CNN DROPOUT**

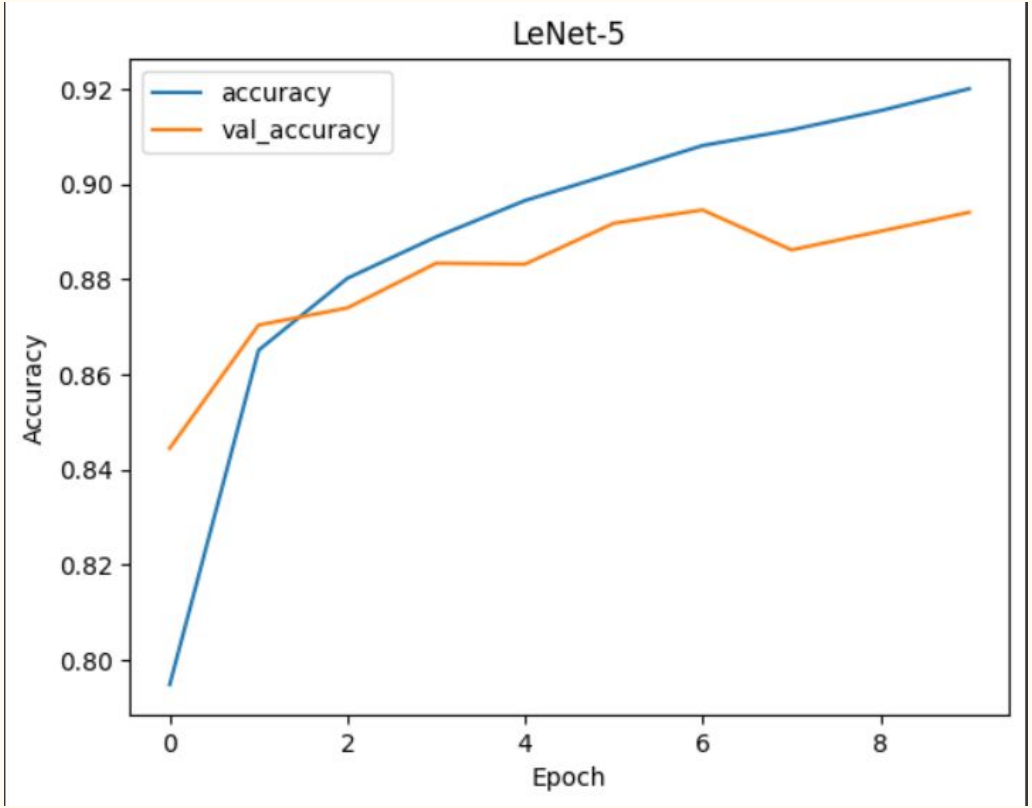




Dropout	0.2	0.1
Accuracy	89.9	90
Epochs	10	10
Batch Size	32	32

MODEL EVALUATIONS

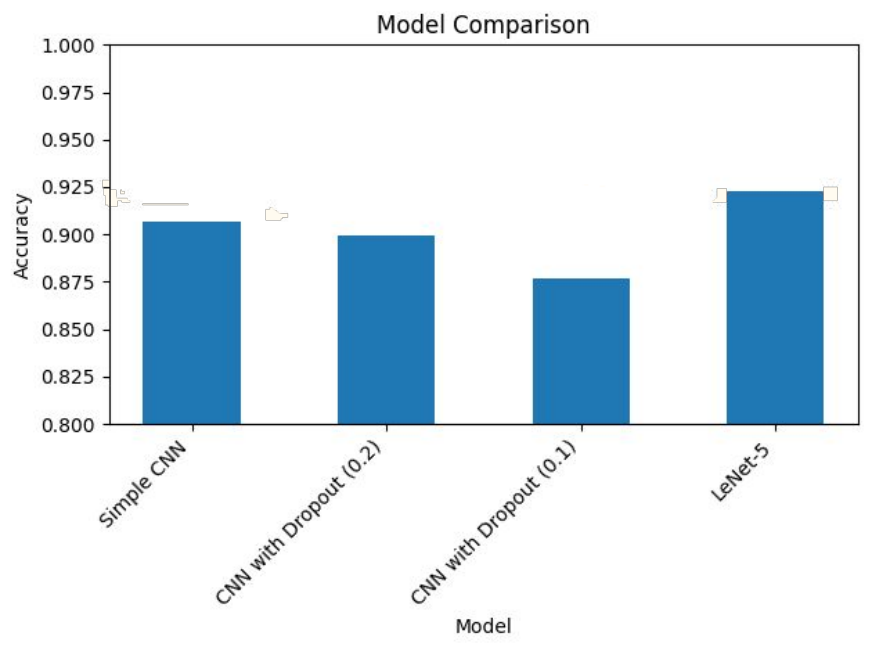
RESULT WITH CNN **LeNet-5**



Accuracy	91.9
Epochs	10
Batch Size	32

FINAL RESULT

COMPARISON BETWEEN ALL 3 MODELS



MODELS	ACCURACY	EPOCHS	BATCH SIZE
SIMPLE CNN	90.68	10	32
CNN WITH DROPOUT 0.1	90	10	32
CNN WITH DROPOUT 0.2	89.9	10	32
LE-NET 5	91.9	10	32

CONCLUSION AND FUTURE SCOPE

With the growth in deep learning methodologies, image recognition using CNN is excessively applied in fashion domains such as clothes classification, clothes retrieval and automatic clothes labeling.

In this project, we applied 3 models named Simple CNN, CNN with Dropout(0.1 & 0.2) and LeNet-5 on the Fashion MNIST dataset. LeNet-5 gave the highest performance (an accuracy over 91.1% was obtained) as compared to other models.

We plan to apply LeNet-5 and other CNN models on a dataset of real clothes images collected by ourself for the evaluation purposes.

REFERENCES

1. M. Kayed, A. Anter and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," in 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 2020, pp. 238-243, DOI: 10.1109/ITCE48509.2020.9047776.
2. Valderi Reis Quietinho Leithardt, Anita Maria Rocha Fernandes, Sérgio Correia, Rodrigo Lyra, “Classifying Garments from Fashion-MNIST Dataset Through CNNs”, in Advances in Science Technology and Engineering Systems Journal, February 2021. DOI: 10.25046/aj0601109.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
4. K V, Greeshma & Sreekumar, K.. (2019). Fashion-MNIST classification based on HOG feature descriptor using SVM. International Journal of Innovative Technology and Exploring Engineering. 8. 960-962.
5. https://blog.tensorflow.org/2018/04/fashion-mnist-with-tfkeras.html
6. https://pyimagesearch.com/2016/08/01/lenet-convolutional-neural-network-in-python/

THANK YOU

ADITYA SETHI
PARAS KUMAR
RONIT KUMAR
MSC COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF DELHI