Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

Convolutional Neural Networks for Fashion Classification

Muskan Singh*[1], Simone Shree Pathak[2], Anuska Basnet[3], Saishab Bhattarai[4]

[1,] Department of Computer Science, Deerwalk Sifal School, Kathmandu, Nepal
muskan.singh@sifal.deerwalk.edu.np

[2] Department of Computer Science, Deerwalk Sifal School, Kathmandu, Nepal
simone.Pathak@sifal.deerwalk.edu.np
[3] Department of Computer Science, Deerwalk Sifal School, Kathmandu, Nepal
anuska.basnet@sifal.deerwalk.edu.np

[4] Department of Computational Mathematics, Kathmandu University, Dhulikhel, Nepal
saishab.bhattarai.75@gmail.com

Abstract:

Neural networks, inspired by the biological structure of the brain, have become a cornerstone in various machine learning tasks, including but not limited to those within the fashion domain. Convolutional Neural Networks (CNNs) are specialized neural networks designed for processing structured grid data - particularly images. Using these neural networks, The Virtual Wardrobe Fashion Advisor was created. Additionally programming languages like Python for algorithm implementation and Django,Html, Javascript and Cascading Style Sheets for web based application development were also used. The model was approached in the traditional CNN method to detect the type of clothes users prefer. Model detects the clothes with the help of selective gender and seasons; a setting in accordance to the user preference.

Keywords: Convolutional Neural Networks, Fashion Advisor, Machine Learning, Data Collection.

1. Introduction

Over the past few years, technology has dramatically altered its relationship with fashion. Today, fashion designers are trying out new technologies that can create a difference in the fashion industry. Along with this tech which is a platform, machine learning, a part of artificial intelligence (AI), makes headlines as one of the effective ways of lifting the fashion field with all processes involved. Machine learning is the ability of computers to learn and get better with experience without being coded explicitly. Over time, it has experienced massive changes, leading to the growth of today's technology landscape and illustrating how robotics has impacted industries of such importance like healthcare or finance.

AI has its beginning in the fifties. But in the final stage of the 20th Century practically concentrated efforts were led to building plausible machine learning models. Arthur Samuel [1], the pioneer of this field, defined machine learning as the “study of computer systems that can learn without being explicitly programmed.”

The machine learning development process has had its own set of milestones. Among them are the emergence of neural networks, the appearance of the statistical learning theory and the advancement of deep learning algorithms. Networks, along with mimicking the structure of the brains, and frequent applications within machine-learning tasks have been known in the fashion domain as well. In 1943, Warren McCulloch and Walter Pitts [3] were the first to succeed in doing the study of the neural networks, as they created a mathematical model that described the behavior of artificial neurons. It became the leading factor to the development of artificial neural networks for the next few years. Artificial Neural Networks (ANNs) represent computational models that consist of the interconnected nodes, or neurons, organized as layers. This results in having a single neuron that works as an input from one initial layer and that in turn serves as a result to the next layer. The idea of artificial neural networks (ANN) has led to the creation of many new architectural and training algorithmic revisions.

Within the field of neural networks, three prominent architectures stand out; the artificial neural networks (ANN), the convolutional neural networks (CNN), and the recurrent neural networks (RNN). Each architecture is represented by a group with its own attributes that underline the best method to particular challenges of data processing and is a part of a special-set.

Artificial Neural Networks (ANN) are the traditional neural networks or the perceptron as they were already called contain interconnected layers of neurons, so that every neuron in the one layer is connected to every neuron in the subsequent layer. [4] *Convolutional Neural Networks (CNN)* are a subclass of the neural networks which are made for the processing of structured grid data, mostly images. They use convolutional layers to model hierarchical features from input images, leading them to be able to rapidly depict the spatial links (LeCun and Juvan, 1998). [5] *Recurrent Neural Networks (RNN)* have been designed for the sequential data processing tasks where the order of inputs is crucial. The same as feedforward networks, RNNs have a memory that allows information to be accessible at any moment of the flow spontaneously which, in turn, gives them the ability to capture temporal dependencies in sequential data. [6]

In the context of fashion dress classification, these neural network architectures play a crucial role in leveraging the vast amounts of fashion data available to automate and enhance various processes, ranging from product categorization to trend analysis. In the generation where purchasers are faced with the assignment of navigating through an overwhelming pool of clothing objects to discover the proper outfit, such technologies can help tremendously to ease the people with a fashion advisory system.

The primary objective of this research is to streamline and enhance the fashion decision-making procedure by implementing CNN algorithm to an advanced fashion advisory system. The main context of this paper are as follows, Section II provides Literature Review of previous works, Section III Methodology describes the mathematical use of CNN model, Section IV Results gives the summary of implementation of the CNN model and its results, Section V Discussion summarizes and gives the overview on the project and Section VI Conclusion and Future Works concludes the project with the future implications.

2. Literature Review

In the world of technology and advancement, machine learning and convolutional neural networks had been developed since the 1990s by Yann LeCun[19], along the colleagues. It got popular in the 2010s and many researchers had developed projects and given their knowledge and understanding about this topic.

2.1 Hierarchical Convolutional Neural Networks

Yian Seo , Kyung-shik Shin [7] had introduced the concept of the Hierarchical Convolutional Neural Network in 2018 A.D . The concept of Hierarchical Convolutional Network was developed in the beginning of the CNN by other researchers. Although it was not utilized by the researchers who worked on the detection of the fashion clothes and apparel.

The fashion industry is a competitive business with a greater amount of profit. The CNN models had been used in this field through different organizations to decrease the workload and make more profit. Unfortunately those models were not perfect, as the result Yian Seo and Kyung-shik Shin developed the first hierarchical convoluted network. It had a high frequency with minimal error. It can work on a huge set of data at a particular time. The apparel classification had been major for this research. The most common problems during image detection are about apparel or objects. The CNN model assumes the pants and top to be the same thing and gives inaccurate results to the users. It also faces the problem of determining whether the clothes are worn by a human figure or not. So for these types of issues they have done research and gained 91.7 percent of accuracy.

The hierarchical model reads and goes through each image and then classifies it into definite categories. They all are interconnected and it gives a perfect result. Each apparel can be classified and the object can be easily detected. In this methodology the user gets better experience and the business grows rapidly. The research made a significant impact on the detection of the clothes, apparel in the fashion field with accurate data and results.

2.2 Convolutional Neural Network Clothes Classification

Due to high accuracy and cost many businesses had not applied Convolutional Neural Networks for the clothes classification. The researchers of Gdansk University of Technology [8] have addressed this problem and use CNN to solve the problem. They had managed to get the basic information and category classification of the clothing through the input served to the model. The training and the test dataset had been a small number initially but as they got accuracy they increased the dataset to a higher number. Many ways are present for object detection, among them SDD300 was chosen. After almost 200 epochs they were able to achieve 42 percent accuracy. On increasing the training accuracy their test accuracy had a great improvement. There were many CNN architectures and layers applied in order to achieve the maximum accuracy. It contributed to a better and convenient search engine for the fashion and clothes market and business.

Alexander Schindler [9] from the Austrian Institute of Technology had also researched on this topic of fashion and apparel classification using the deep learning neural networks. In the research paper the major problem had been addressed as the inaccuracy or improper results on the classification of the apparel and the clothing materials. Due to large textures, shapes and colors the machines inability to correctly identify and give the results using the traditional model of neural networks had been clearly mentioned. Thus a smaller scale dataset with best performance which is certainly increased by the higher number of dataset for better classification.

2.3 Clothing Object Detection

A real-time clothing classification system facilitates automated fashion stylists, outfit recommendation, discovering similar fashion pieces, surveillance context, automatic annotation of images with tags or descriptions, context-aided people identification, occupation recognition, and improvement in information retrieval from various areas such as social media and consumer sites[10].

Hara, K., Jagadeesh, V., & Piramuthu, R. [11] also researched clothing object detection, and discovered a new way of classifying fashion apparel using detection based spotters in the form of bounding boxes. These boxes worked more efficiently compared to other methods as it is faster and doesn't require multiple instances like semantic segmentation. In their research, they incorporated contextual information from body poses in order to improve the detection performance. By conducting multiple experiments, they drew a conclusion that the method they discovered, Pose-dependent Object Detection,is more effective than semantic categorization.

One of the earliest constructions of a framework for robust and extremely rapid visual detection was done by Pual Viola and Micheal J.Jones [12]. It classified images based on the value of simple features. This architecture has been widely used to implement sliding window detectors for faces [12 -14], pedestrians , and cars [(Cai et al., 2016)].

3. METHODOLOGY

The Virtual Wardrobe Fashion advisor had been created through the use of the programming languages like Python for algorithm implementation and Django, HTML, JavaScript and Cascading Style Sheets(CSS) for web based application development. The dataset had been created manually. The data ranged till fifteen thousand. The dataset of cloth was categorized into male, female, and into the four seasons, namely winter, summer, spring and autumn. It also has the category of formal and informal dress. Before serving the data to the CNN model there were several preprocessing steps applied which involved resizing, and scaling the pixels of the images.

3.1 DATA COLLECTION

3.1.1 Gender**:** The dataset contains about 15000 images and among them 7000 were labeled as male and 8000 were labeled as female manually, as the ground truth annotations. It was done in order to check the predictability of the model to classify between male and females.

3.1.2 Seasons**:** At the beginning a dataset contains 40 thousand images. It has the categories of the four seasons with male, female, formal, informal and occasional. The dataset contained its ground truth annotations for each of the categories. Winter season consisted of the largest portion in our dataset. The smallest dataset was of the autumn male clothes for occasional. The complete dataset was not utilized for the training of the CNN model. About 10 percent of the data was utilized and the dataset containing less than 50 images were not taken into account. The dataset which contained an efficient amount of data was selected for the further step of the training dataset. The subset data only contained about 15000 images.

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

Fig.1: Classification of the dataset categories

3.2 The Mathematics of Convolutional Neural Network

3.2.1 Matrix Formation

The data served to the CNN model is represented in the form of a matrix. Let us assume a three by three matrix of X denoted as:

$$X = \begin{bmatrix} 10 & 20 & 30 \\ 15 & 25 & 35 \\ 5 & 10 & 15 \end{bmatrix}$$

In the matrix, each value represents specific weighting parameters in the CNN model. In CNN these parameters are required for the training of the model and to specify the features. These parameters are required in order to classify the image and predict which category this image belongs to. It goes through the binary classification in order for the detection.

3.2.2 Convolutional Layer

In the convolutional layer the images are transformed into matrix formation. It has multiple layers and can give output more accurately than other neuron networks. The convolutional is also called local connectivity as CNN connects the neurons to the local input. For example if 32 x 32 neurons are required in the next layer and 5 x 5 neurons are connected in the previous layer. It can be visualized through figure 2.

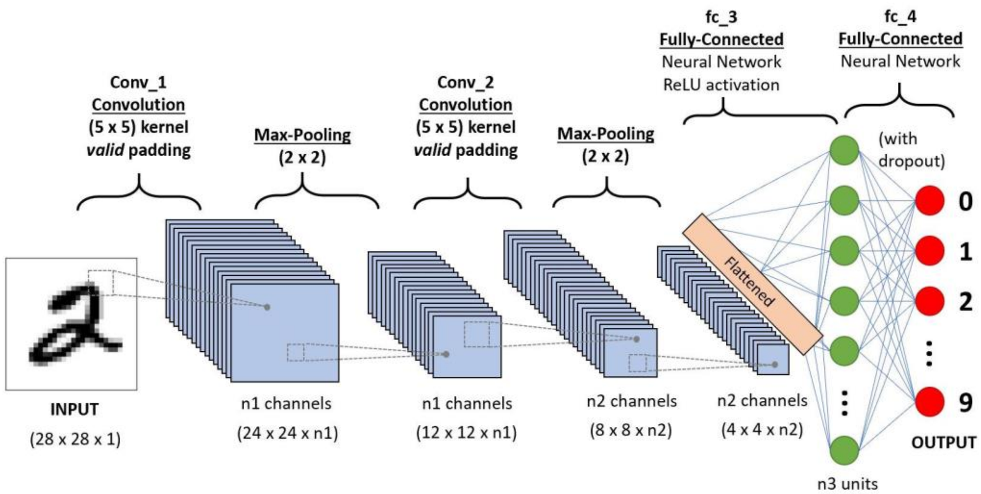

Fig.2: Fully Connected Layer of CNN [12]

Now, let’s introduce a simple 2x2 filter F for our convolution operation:

$$F = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

The convolution operation involves sliding this filter over our image matrix X and performing element-wise multiplication, producing a feature map Y :

$$Y = X * F = \begin{bmatrix} 0 & 20 \\ 5 & 10 \end{bmatrix}$$

The basic process of the training of the CNN model can be clearly visualized through the following diagram of figure 3. An image is passed through different layers which gives different output as the detection of edge, sharpened image, background, identity and blurred image. The augmented images represent the different features of the served image to the CNN model for classification.

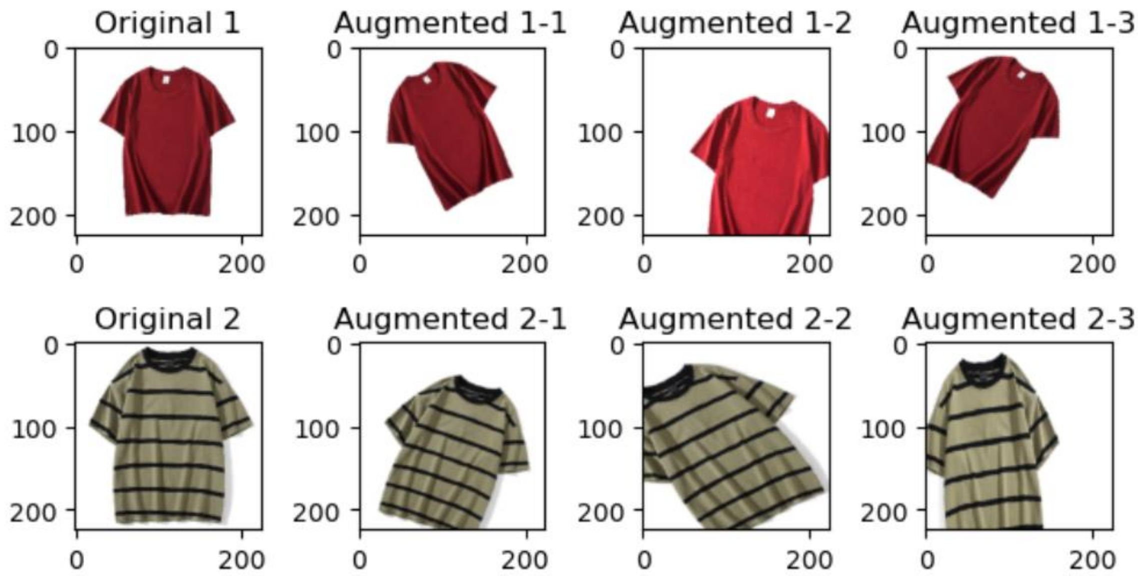

Fig.3: Effects of different Convolution

3.2.3 Non-linearity

The non -linearity indicates an optimized output through the CNN. It consists of ReLU Activation, Sigmoid function and tan hyperbolic function. The ReLU Function can be defined as : $f(x) = max(0, x)$. In other words it only passes the positive value of the matrix and if there is a negative value then it converts them to zero. The ReLU layer can be represented through the following mathematical form. If the input neuron is z, the input to the next layer is the sum of the inputs of the neuron of the previous layer. Its equational form is:

$$z = \sum_{i=1}^{n} w_i x_i + b$$

The w_i represents the weights and x_i are the inputs and b is the bias for the neuron of the inputs in this layer.

The sigmoid function similarly takes any real number as input and provides the output as 0 or 1. The hyperbolic function of tangent minimizes the output to range [-1,1]. The following graph represents the tan hyperbolic function, sigmoid and the ReLU activation of the Convolutional network in the object detection of the CNN model of fashion and clothes.

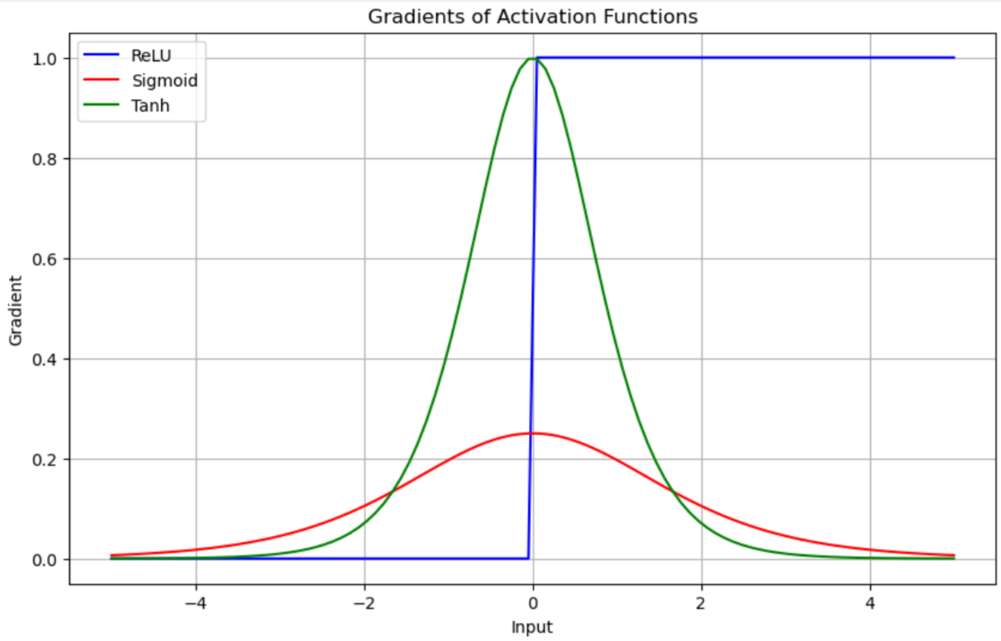

Fig.4: Uses of Non Linearity in Object Detection

3.2.4 Padding and Max-Pooling

One of the major drawbacks of the Convolution Network is that it does not include the objects at the border if we don't use zero padding. Thus we have used zero padding for the object detection of the CNN model.

This reduces the size of the image and shrinks it after it passes through each layer. Similarly the kernel touches the middle of the image more than the corners which reduces its accuracy. In order to solve this problem Padding was introduced [20]. In padding when a image of m x m matrix is utilized with f x f filter with padding p then the size of the output image is:

$$Output\ size = (m + 2p - f + 1) \times (m + 2p - f + 1)$$

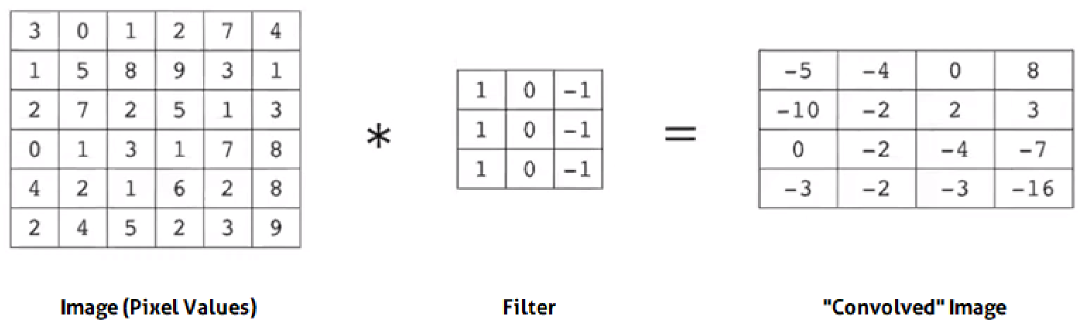

Fig.5: Convolved Image using Padding [13]

The following figure, *fig 5,* shows the stride of the filter that hovers along the image pixel. The stride simply can be defined as the filter over the input matrix during convolution. After the strides is done on the input matrix then the output image dimension can be described through the following equation:

$$Output\ size\ = (\frac{(n+2p- f+1)}{s} + 1) \times (\frac{(n+2p-f+1)}{s} + 1)$$

In the above equation: p is padding, it adds a border of p pixels around the image. Filter size is f x f , it determines the size of the kernel used for convolution. Stride s specifies the number of pixels that the filter would move after each step. Input size n x n represents the dimensions of the original image.

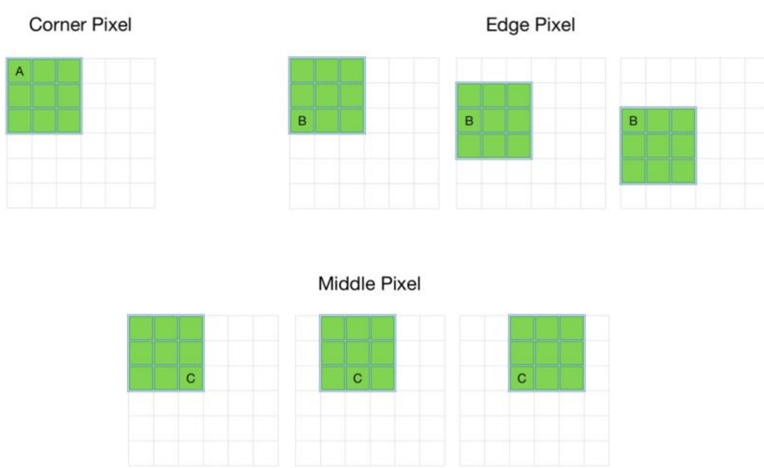

Fig.6: Padding and Sliding Filter over the image [14]

Pooling can be defined as one of the important parts of CNN as it reduces the parameters of the model and simplifies it. The maximum pooling method contributes to this work. In this process a non overlapping filter goes over the image. It only selects that matrix which has the maximum value and discards the rest. [15] It controls the overfitting of the input matrix / image. The output size of the matrix can be determined through following equation:

$$Output\ size = \frac{n - f}{s} + 1$$

Similarly Maximum Pooling had been done for example the highest number matrix had been selected after the padding when the filter layer slides over the image.

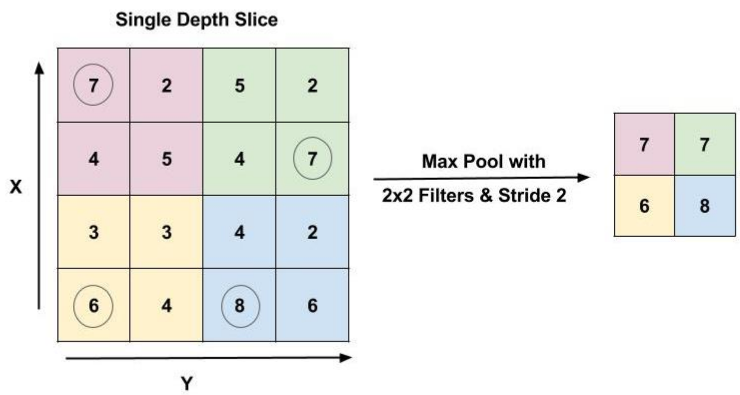

Fig.7: Max-Pooling [16]

3.2.5 Fully Activated Layer

In the fully activated and connected layer each neuron is connected to the previous layer and each neuron of the previous layer is connected to the one before them. It gives output through sending the input neuron in softmax activation. In softmax activation the exponential of the neuron is divided by the sum of all the neurons exponential. The output of this activation predicts the object of the certain category. [21]

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

The above equation represents the softmax activation of the fully connected neurons. The image of the fully connected neurons is in diagram figure 2.

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

4. RESULTS

Almost 40 thousand dataset were taken initially for training the model in order to detect the gender, season and formal, informal type through the image. After reviewing and finalizing the dataset, it was initialized to 15 thousand only. At the beginning there were 200 images in the training dataset and 100 images in the test dataset. We used binary classification and softmax activation for categorizing the data. The initial results through the dataset on prediction of the season and gender is given below:

TABLE 1: Training and Testing Accuracy of Different Attributes

INITIAL RESULTS OF PREDICTION		
Attribute	Test Accuracy	Training Accuracy
Gender	60 %	80 %
Season	55%	74.5%
Occasion	60%	79%

In table 1, the gender training had the highest accuracy at the beginning with 80 % and lowest was of the season with 55 % in the beginning and initial phase by the model. The initial data accuracy graph is plotted below(fig 7) which represents how the CNN model was recognized among these categories and what was the accuracy rate with training and test data in respect to the epochs.

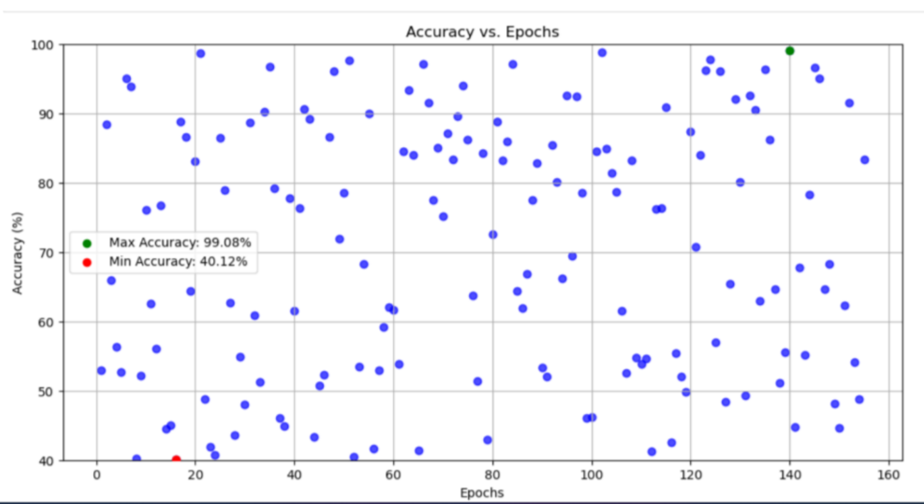

Fig.8: Accuracy vs Epochs

During the prediction one of the essential parts is how the model will be able to detect the gender and seasons. The background detection was a crucial part of the research. We have mostly preferred those images which had white background and studio lights so that it would be easier and more convenient for the model to be trained. The process of recognition of gender through the CNN model is presented through the flowchart of figure 8.

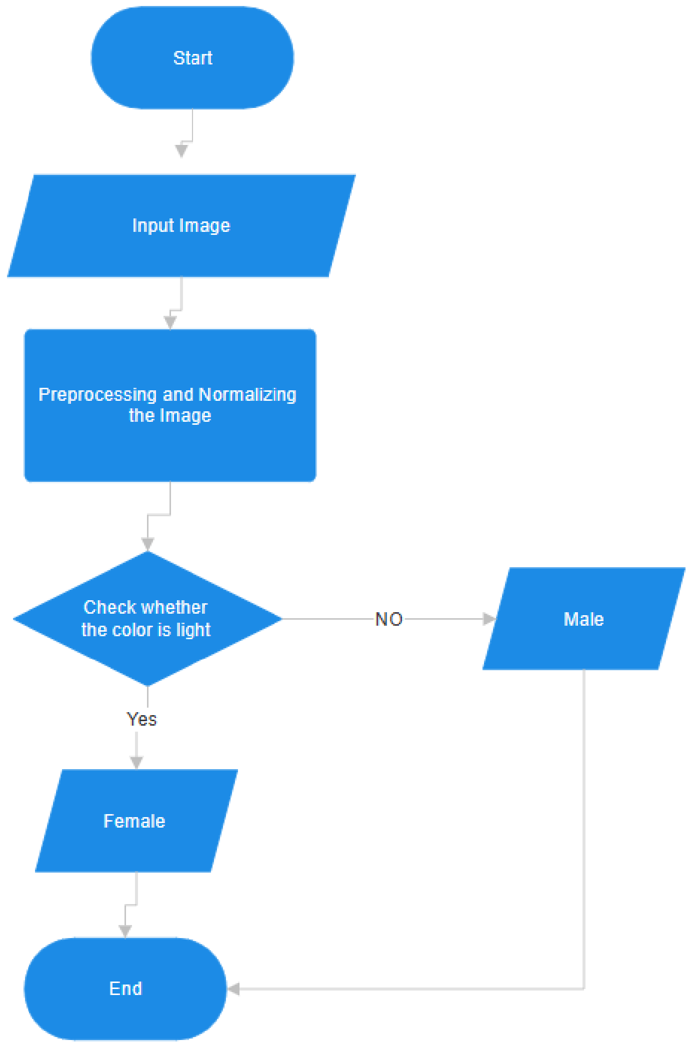

Fig.9: Flowchart of Gender Detection

In the gender detection first the input image goes through processing and normalizing the image. Then the color is checked whether it is light or not. As it's a basic model so it does not go through vivid properties for gender classification. If the color is light it considers the cloth to be of female and if the color is dark then male is considered. After the gender detection it goes through the following process shown in the flowchart (Fig 9). It depicts the season of the input image through it.

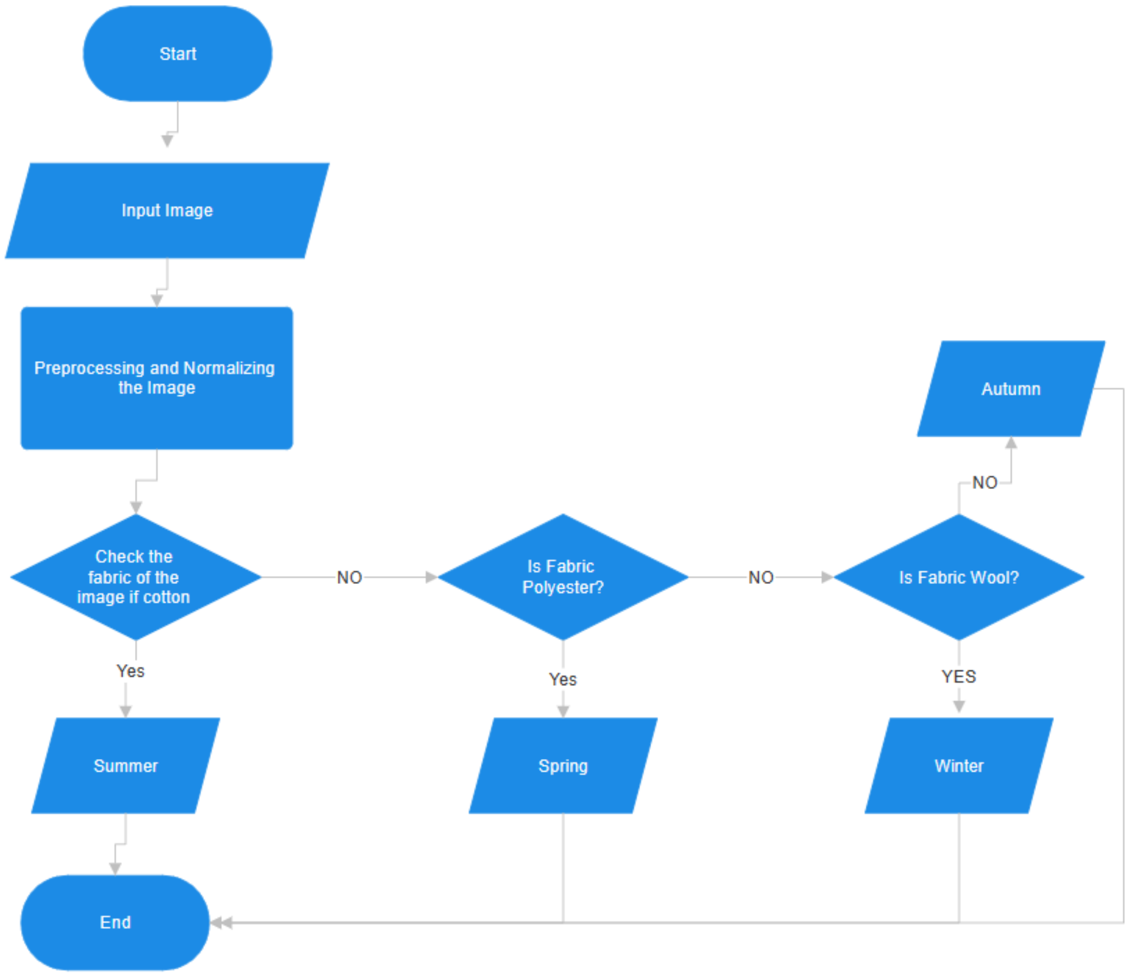

Fig.10: Flowchart of the Season Detection

After the gender detection the image is used for the season detection. In that as shown in the flowchart the conditions are applied. According to the satisfaction with the condition the results are displayed. The CNN model predicts the season for the input image. Now the occasion is left which is also vividly presented through the following flowchart(Fig: 10).

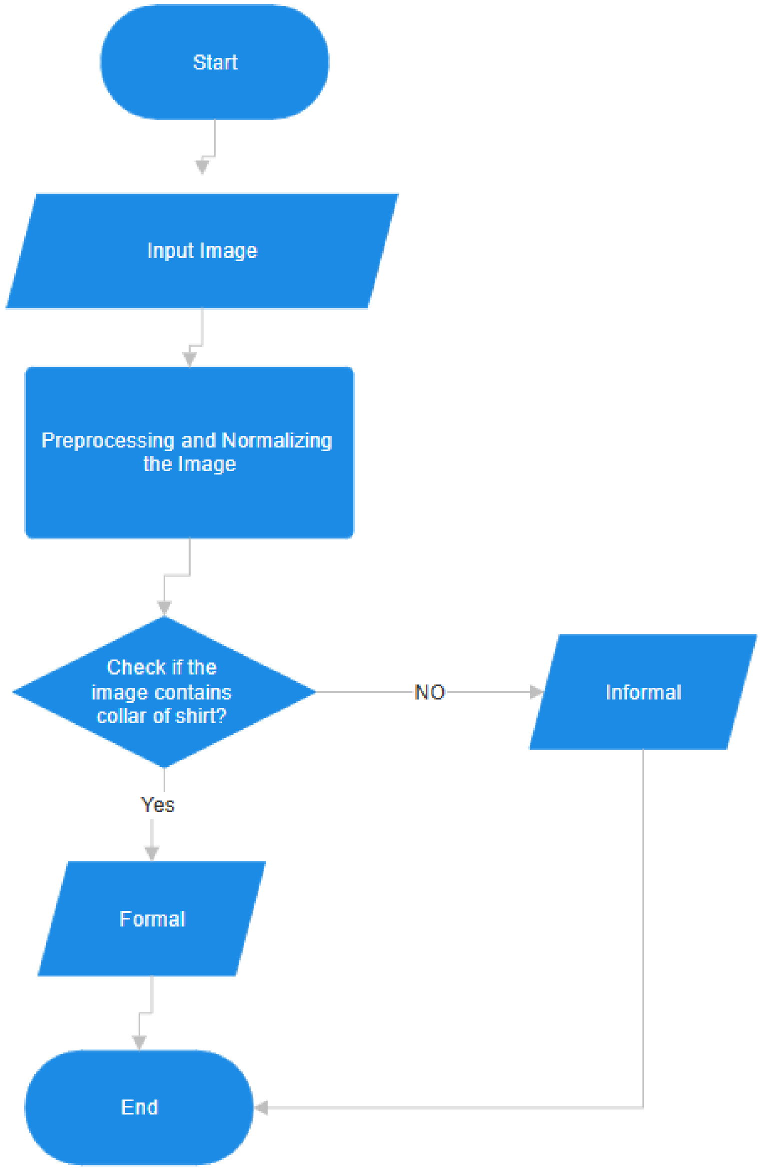

Fig.11: Flowchart of Occasion detection

When the CNN model has classified among the gender and season then it checks for occasion whether its formal or informal. It checks whether the image contains a shirt or collar and if there is a shirt or collar then it predicts the image to be formal or informal. This mechanism is also visualized using the flowchart diagram.

After the formation of the CNN model, we took in account about 200 images for the test dataset and 500 images for the training dataset. At the beginning due to imbalanced dataset, model overfitting, limited data size and data quality there were several negative impacts on the accuracy of the model. It had about 67.11% of accuracy on the dataset. It was not a practically applicable model due to loss of accuracy.

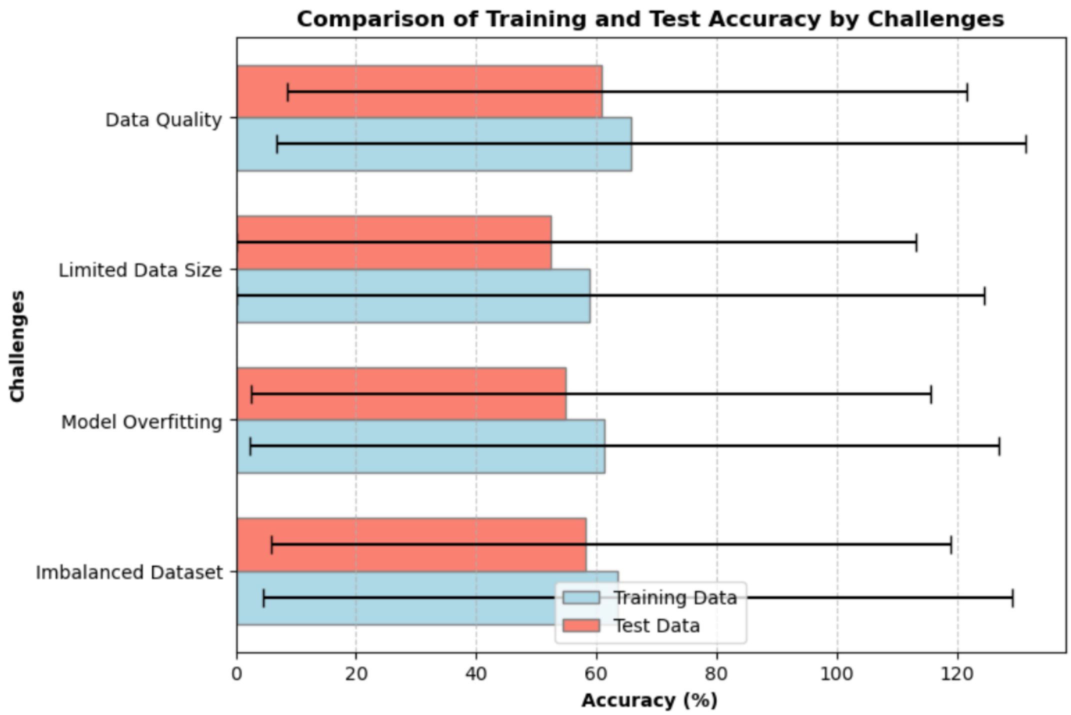

Fig.12: Challenges and Limitations Graph

4.1 Analysis of the Challenges and Limitations Graph

TABLE 2: Impact of Various Data Challenges on Model Accuracy

Challenges/Limitations	Training Accuracy (%)	Test Accuracy (%)
Imbalanced Dataset	63.50	61.25
Model Overfitting	65.75	62.50
Limited Data Size	64.00	60.75
Data Quality	62.25	59.50

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

In table 2, the limited data size had the lowest accuracy about 60.75 percent on the test data set and the data quality had the minimum accuracy in the training dataset. Similarly the maximum accuracy was received in the limited data size in the training dataset. It gave us the idea on which portions we need to improve in order to overcome the challenges and limitations to make the project more efficient. It also represented the stability of the model with the trend of accuracy when such drawbacks were not removed through the dataset. It was the initial dataset that manually contained such issues and which decreased the accuracy of the model.

As the accuracy was not as expected the model was remodified several times, for each modification the size of data was changed. We had a different accuracy rate for the prediction of the clothing images through the CNN model. The most accurate results were shown by Model C which had the highest accuracy of 89 % on the detection of the images.

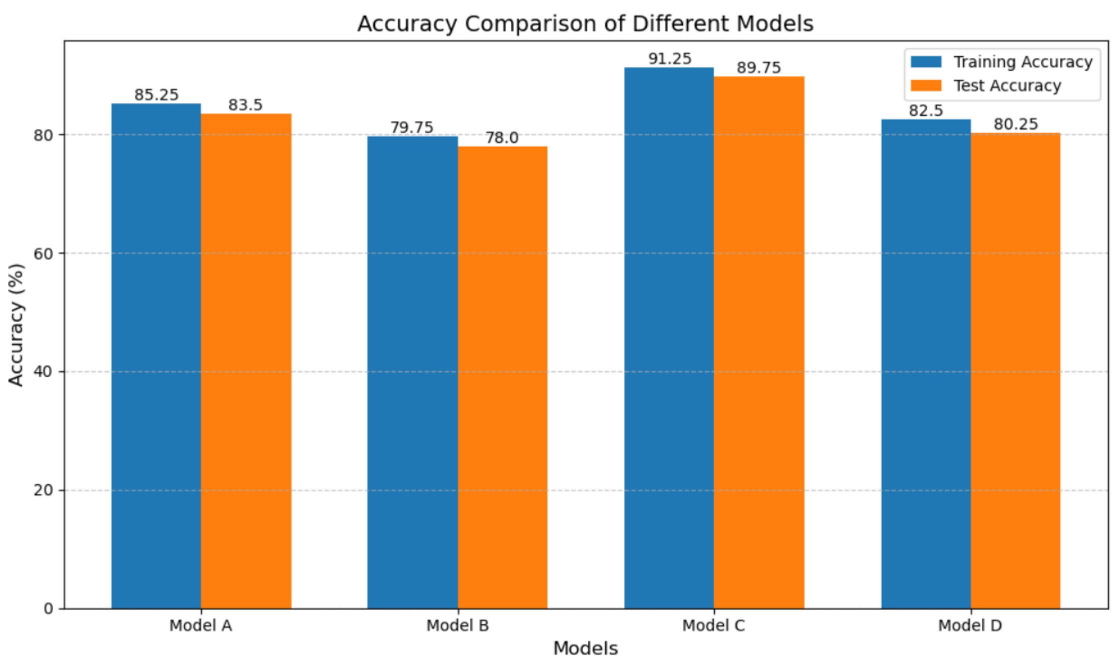

Fig.13: Comparison of the Different Models

4.2 Analysis of Comparison of Different Models

TABLE 3: Comparison of Different Models

Model	Initial Data	Data Type	Validation Data Percent	Training Data Percent	Additional Notes
A	40,000	Uncategorized from mist	83.50	77.05%	-Data Imbalance Removed -Data Overfitting Removed - Data Quality Improved
B	30,000	Uncategorized, Ground Annotated Data	77.1	92.86%	
C	15,000	Uncategorized, Ground Annotated Data	89.75	91.25%	
D	10,000	Uncategorized	80.25	93%	

In table 3, the model C had the maximum accuracy of 91.25 percent and was highly efficient. The Model B had the lowest accuracy of 79.75 percent and was not as efficient as other models. The challenges and limitations were overcome in each model and the dataset was modified. The best version of the model was C which was finalized for the final product.

Although the training data has the maximum accuracy in comparison to the test data. The loss of training data and test data was significant at the beginning. Although the prediction for the training data improved in the model improvement. The model C had almost 90% accuracy for the training data and 89% accuracy for the test data. The training and validation accuracy with the training and validation loss is presented in the following diagram.

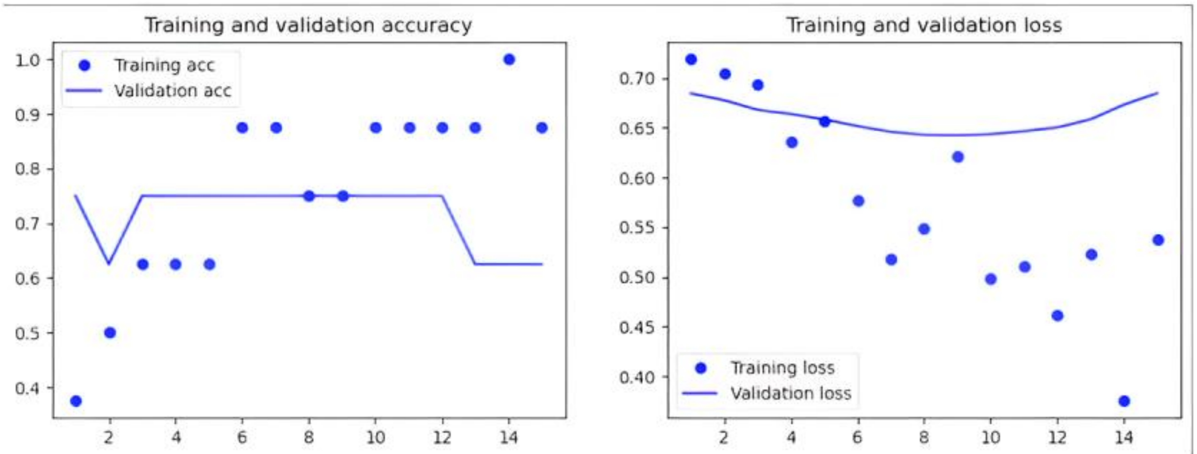

Fig.14: CNN Model Loss Metrics of Model C

4.3 Analysis of CNN Model Loss Metrics

- The above graph represents the training and validation accuracy of the final Model C which had maximum accuracy of object detection.
- From the graph of training and validation accuracy we can observe that the training accuracy is greater than the validation accuracy.
- The validation accuracy is consistent in respect to epochs but the training accuracy varies.
- The training loss is also greater than the validation loss for the initial epochs.
- The validation loss is also consistent whereas the training loss varies with respect to the epochs.

The training and validation loss of other models as Model A, B and C were also plotted in the line graph to learn the patterns and trend of prediction for those models. The table below shows their data of training and validation loss and accuracy over 20 epochs among the 155 that the model went through during the training process.

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

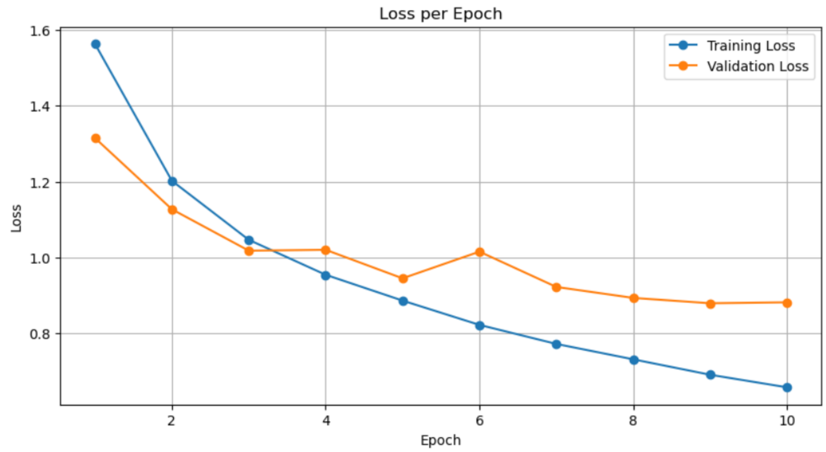

Fig.15: Loss per Epoch of Model B

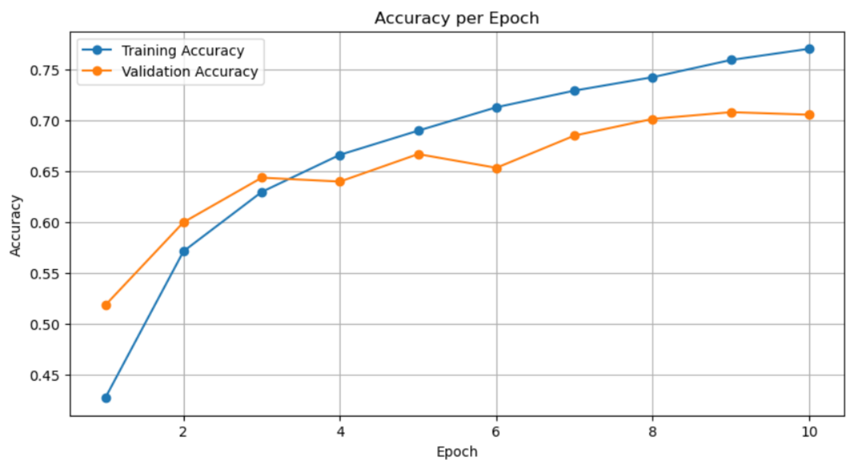

Fig.16: Accuracy per Epoch of Model B

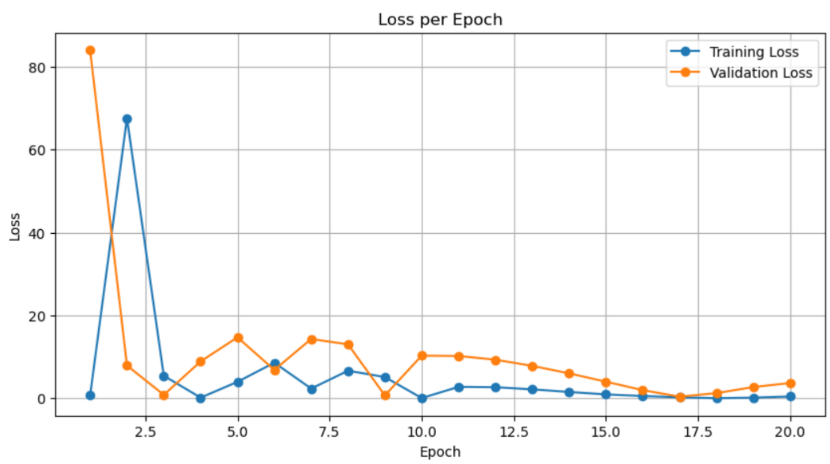

Fig.17: Loss per Epoch of Model A

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

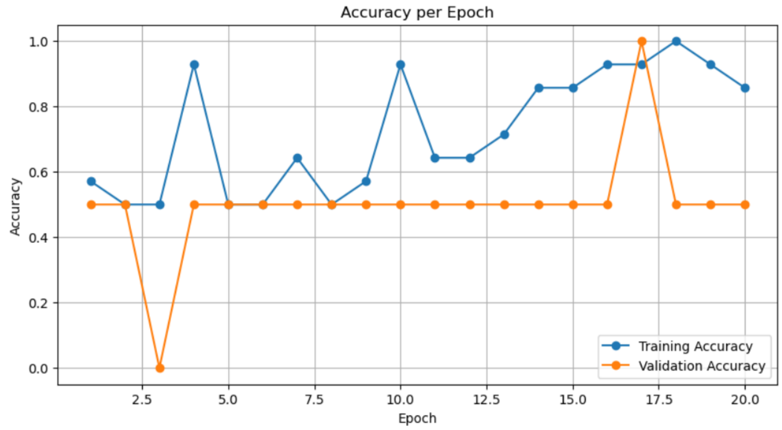

Fig.18: Accuracy per Epoch of Model A

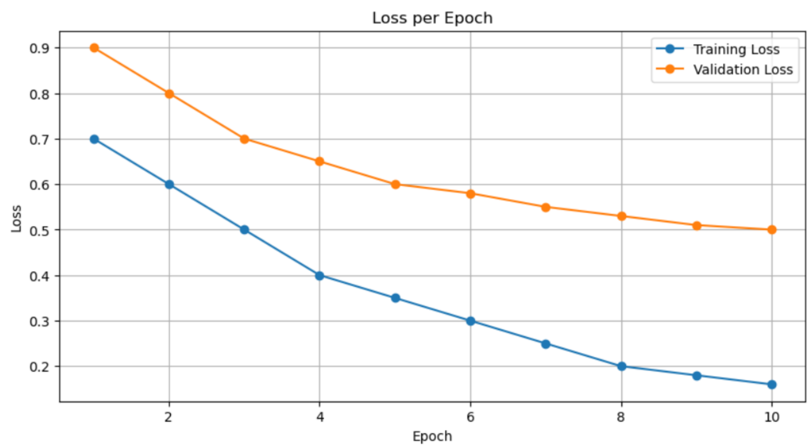

Fig.19: Loss per Epoch of Model D

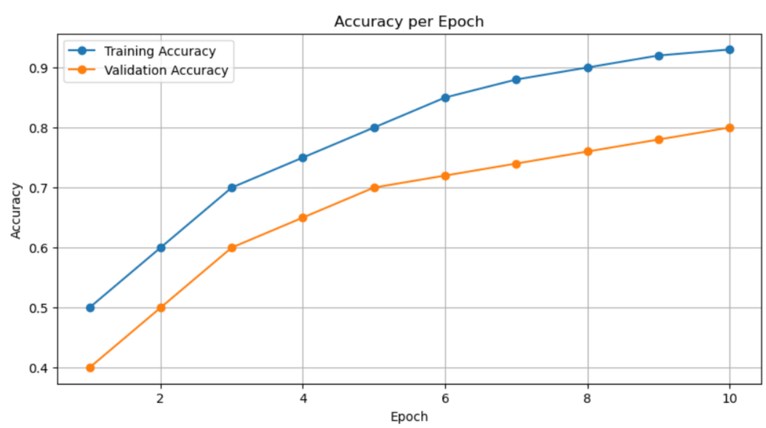

Fig.20: Accuracy of Epoch for Model D

The provided graphs visually represent the evolution of training and validation accuracy as well as training and validation loss across 155 epochs, offering a comprehensive overview of the model's

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

learning dynamics. The insights gained from these visualizations further inform potential adjustments and refinements to enhance the model's overall predictive capabilities. The following were the results shown by the model after the completion of the steps and accuracy improvement.

TABLE 4: Analysis of Model A Training and Test Data

Model A:

Epochs	Train Loss	Train Accuracy	Val Loss	Val Accuracy
1	1.5633	0.4274	1.3151	0.5185
2	1.2022	0.5715	1.1271	0.6000
3	1.0465	0.6298	1.0179	0.6438
4	0.9543	0.6663	1.0202	0.6400
5	0.8862	0.6900	0.9449	0.6670
6	0.8221	0.7130	1.0153	0.6535
7	0.7719	0.7294	0.9222	0.6852
8	0.7313	0.7426	0.8933	0.7016
9	0.6906	0.7595	0.8792	0.7082
10	0.6574	0.7705	0.8817	0.7057

TABLE 5: Analysis of Model B Training and Test Data

Model B:

Epochs	Train Loss	Train Accuracy	Val Loss	Val Accuracy
1	0.6948	0.5714	84.0172	0.5000
2	67.5981	0.5000	8.9153	0.5000

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

3	5.4514	0.5000	14.7313	0.5000
4	0.1774	0.9286	10.2294	0.5000
5	8.6057	0.5000	0.7242	0.5000
6	2.3604	0.6429	10.3088	0.5000
7	6.6716	0.5000	7.8478	0.7143
8	5.1450	0.5714	9.3287	0.7143
9	0.0741	0.9286	7.8478	0.7143
10	2.7772	0.6429	5.3287	0.7857

TABLE 6 : Analysis of Model D Training and Test Data

Model D:

Epochs	Train Loss	Train Accuracy	Val Loss	Val Accuracy
1	0.7	0.5	0.9	0.4
2	0.6	0.6	0.8	0.5
3	0.5	0.7	0.7	0.6

4	0.4	0.75	0.65	0.65
5	0.35	0.8	0.6	0.7
6	0.3	0.85	0.58	0.72
7	0.25	0.88	0.55	0.74
8	0.2	0.9	0.53	0.76
9	0.18	0.92	0.51	0.78
10	0.16	0.93	0.5	0.8

In table 4 - 7, We can observe that the model can accurately predict the male female and seasons as well as the occasion of the input image. The accessories are mostly present in the female outfit. While designing the model the accessories were also kept as the base logic for the model to identify between male and female. The vibrant colors are mostly present in the female outfits. The fabric of the clothes are checked in order to detect the season. On the basis of shirt and collar formal and informal occasion is detected. The images below are taken from the mnist dataset through kaggle. [17]

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

GENDER: MALE
SEASON: AUTUMN
OCCASION: FORMAL

GENDER: FEMALE
SEASON: SPRING
OCCASION: INFORMAL

GENDER: FEMALE
SEASON: WINTER
OCCASION: INFORMAL

GENDER: FEMALE
SEASON: SUMMER
OCCASION: INFORMAL

GENDER: MALE
SEASON: SUMMER
OCCASION: FORMAL

GENDER: MALE
SEASON: WINTER
OCCASION: INFORMAL

The following are the graphs for the original dataset training and test accuracy which represents how the original and un-modified dataset each category was depicted by the model. [18]

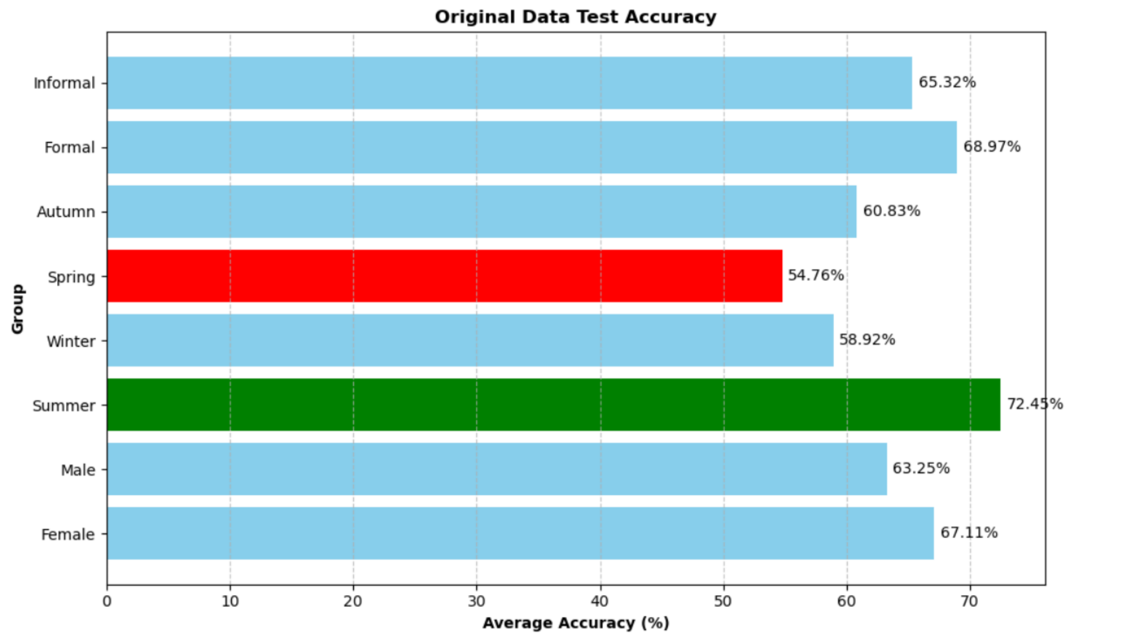

Fig.21: Original dataset Test Accuracy

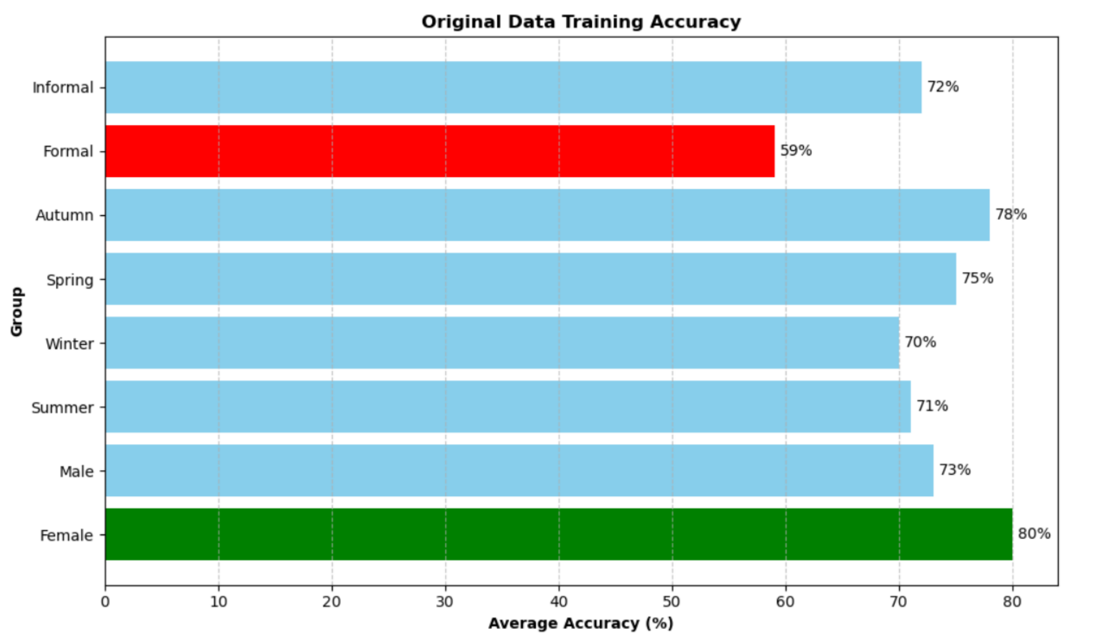

Fig.22: Original dataset Training Accuracy

4.4 Analysis of Original Dataset Accuracy

TABLE 7: Accuracy by Category of Original Dataset

Deerwalk Journal of Computer Science and Information technology, Vol. (Issue), pp. 23-32, January 2019
e-ISSN: 2018/7520

ORIGINAL DATASET		
Group	Training Dataset	Test Dataset
Female	67.11%	67.11%
Male	63.25%	63.25%
Summer	80.00%	80.00%
Winter	70.00%	70.00%
Spring	75.00%	75.00%
Autumn	78.00%	78.00%
Formal	59.00%	59.00%
Informal	72.00%	72.00%

Summary:

TABLE 8: Statistics of Original Dataset

	Training Dataset	Test Dataset
Mean	70.38%	70.38%
Minimum	59.00%	59.00%
Maximum	80.00%	80.00%

In table 8, the original dataset is not highly efficient and it needs to be improved in order to gain more efficiency. The highest amount of accuracy is not sufficient for the final product. This dataset had gone through the challenges and limitations mentioned above in the 4.2 (*Analysis of Challenges and Limitations*). It required modification and many modifications were done to get the accurate model.

The most accurate model is Model C as mentioned in 4.3 (*Analysis of The Different Models*). It had the maximum accuracy. The modified dataset with the accuracy of each category is thoroughly visualized through the bar graph. The statistics are presented in the table below:

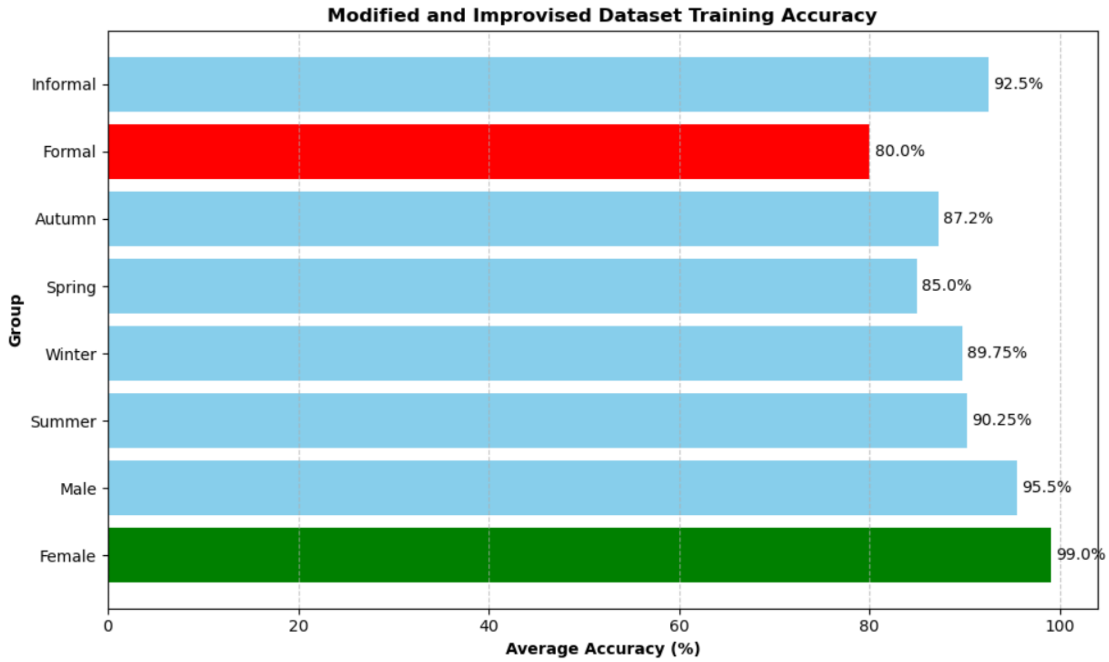

Fig.23: Modified and Improvised Dataset Training Accuracy

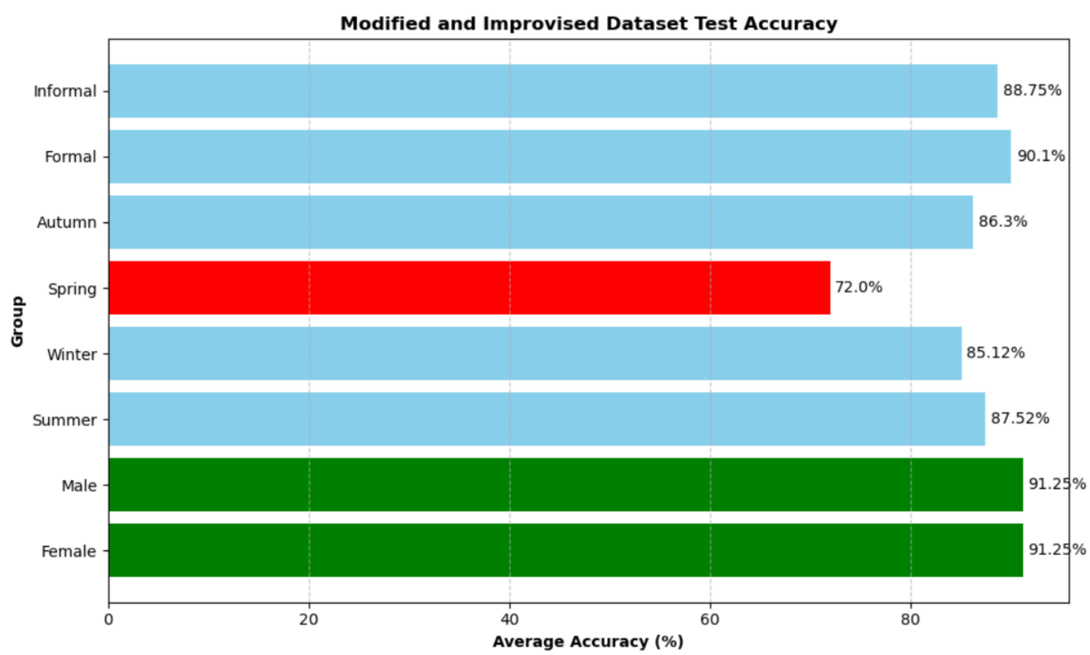

Fig.24: Modified and Improvised Dataset Test Accuracy

4.5 Analysis of Modified Dataset Accuracy

TABLE 9: Analysis of Modified Training Set

TRAINING DATASET		
Group	Modified Training Dataset	Modified Test Dataset
Female	99.00%	91.25%
Male	95.50%	91.25%
Summer	90.25%	87.52%
Winter	89.75%	85.12%
Spring	85.00%	72.00%
Autumn	87.20%	86.30%
Formal	80.00%	90.10%
Informal	92.50%	88.75%

Summary:

TABLE 10: Modified Training Dataset Statistics

	Modified Training Dataset	Modified Test Dataset
Mean	90.63%	87.70%
Minimum	80.00%	72.00%
Maximum	99.00%	91.25%

In table 10, the modified dataset is highly efficient and it is improved in order to gain more efficiency. The highest amount of accuracy is sufficient for the final product. This dataset has been mentioned in the above analysis of Model C which has the highest accuracy. It required no modification and required one was done to get this accurate model.

5 DISCUSSION

Thus we get the Model C with maximum accuracy in the training dataset to be 91.25 % which was finalized as the main product for our system. It could accurately detect the seasons, gender and occasion of the input image. The model followed the traditional method of CNN in order to detect the clothes. It had ground truth annotations corresponding to the dataset. It was trained with the dataset of about 15 thousand images with 155 epochs. It was used in the product of Virtual Wardrobe where users can get fashion advice in respect to their preferences. They could choose the gender, seasons and occasions and get advice on fashion.CNN model uses the users preferences as output and detects the image and gives the output to the users.

6 CONCLUSION AND FUTURE WORK

Deep learning technologies like CNN, widely used for image recognition can be integrated in various fields; one of them being apparel detection which is useful in advancing consumer-focused virtual wardrobe solutions. This research showed the efficiency and accuracy of Convolution Neural Networks in detecting and classifying fashion apparel. The model successfully handled large datasets and correctly classified images for categories such as gender, seasons, and occasions. This allows the users to solve the dilemmas they might incur fashion-wise and get recommendations accordingly based on the selections they make.

For the future phase of the project, we envision increasing the dataset and defining more parameters with various styles of clothes, employing advanced Machine Learning algorithms like the Generative Adversarial Networks (GANs), creating more accurate recommendations and integrating user input for personalization and interactivity. Moreover, the universe of real-time image processing for mobile devices will be discovered, and that will ensure greater accessibility and practicality of the technology for outdoor use. These advancements will boost the range and functionality of automated fashion advice systems in the realm of the digital fashion space.

7 REFERENCES

[1] Wiederhold, Gio & McCarthy, John. (1992). Arthur Samuel: Pioneer in Machine Learning. IBM Journal of Research and Development. 36. 329 - 331. 10.1147/rd.363.0329.

[2] Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, *28*, 525-552.

[3] Mcculloch, Warren S. & Pitts, Walter (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Journal of Symbolic Logic* 9 (2):49-50.

[4] Mehrotra, K., Mohan, C. K., & Ranka, S. (1997). *Elements of artificial neural networks*. MIT press.

[5] LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010, May). Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems* (pp. 253-256). IEEE.

[6] Ma, Y., Ding, Y., Yang, X., Liao, L., Wong, W. K., & Chua, T. S. (2020, June). Knowledge enhanced neural fashion trend forecasting. In *Proceedings of the 2020 international conference on multimedia retrieval* (pp. 82-90).

[7] Seo, Y., & Shin, K. (2018). Hierarchical Convolutional Neural Networks for Fashion Image Classification. Expert Systems with Applications. doi:10.1016/j.eswa.2018.09.022

[8] Cychnerski, J., Brzeski, A., Boguszewski, A., Marmolowski, M., & Trojanowicz, M. (2017). Clothes detection and classification using convolutional neural networks. 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). doi:10.1109/etfa.2017.8247638

[9] Schindler, A., Lidy, T., Karner, S., & Hecker, M. (2018). Fashion and apparel classification using convolutional neural networks. *arXiv preprint arXiv:1811.04374*.

[10] Shajini, M., & Ramanan, A. (2021). An improved landmark-driven and spatial–channel attentive convolutional neural network for fashion clothes classification. The Visual Computer, 37(6), 1517-1526.

[11] Hara, Kota, et al. “Fashion Apparel Detection: The Role of Deep Convolutional Neural Network and Pose-Dependent Priors.” 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1–9. arXiv.org, https://doi.org/10.1109/WACV.2016.7477611.

[12] Viola, P., Jones, M.J. Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154 (2004). https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[13] Dwarampudi, Mahidhar, and N. V. Reddy. "Effects of padding on LSTMs and CNNs." *arXiv preprint arXiv:1903.07288* (2019).

[14] Cai, Z., Fan, Q., Feris, R., & Vasconcelos, N. (2016, January 1). *A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection*. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-46493-0_22

[15] Janjua, Juhi, and Archana Patankar. "Framework of CNN Architecture for Fashion Image Classification." *International Conference on Computing in Engineering & Technology*. Singapore: Springer Nature Singapore, 2022.

[16] He, M., Li, B., & Chen, H. (2017, September). Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3904-3908). IEEE.

[17] Meshkini, Khatereh, Jan Platos, and Hassan Ghassemain. "An analysis of convolutional neural networks for fashion images classification (fashion-mnist)." *Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry”(IITI’19) 4*. Springer International Publishing, 2020.

[18] He, M., Li, B., & Chen, H. (2017, September). Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3904-3908). IEEE.

[19] LeCun, Yann, Koray Kavukcuoglu, and Clément Farabet. "Convolutional networks and applications in vision." *Proceedings of 2010 IEEE international symposium on circuits and systems*. IEEE, 2010.

[20] Hashemi, Mahdi. "Enlarging smaller images before inputting into a convolutional neural network: zero-padding vs. interpolation." *Journal of Big Data* 6.1 (2019): 1-13.

[21] Basha, SH Shabbeer, et al. "Impact of fully connected layers on performance of convolutional neural networks for image classification." *Neurocomputing* 378 (2020): 112-119.