# An Image-based Transfer Learning Framework for Classification of E - Commerce Products

MSc Research Project

Data Analytics

Vrushali Surve

Student ID: x19212712

School of Computing

National College of Ireland

Supervisor: Paul Stynes, Pramod Pathak and Rejwanul Haque

**National College of Ireland**
**MSc Project Submission Sheet**
**School of Computing**

| | | | |
|---|---|---|---|
| **Student Name:** | Vrushali Atul Surve | | |
| **Student ID:** | X19212712 | | |
| **Programme:** | Data Analytics | **Year:** | 2021-22 |
| **Module:** | MSc Research Project | | |
| **Supervisor:** | Paul Stynes, Pramod Pathak, Rejwanul Haque | | |
| **Submission Due Date:** | 16/12/2021 | | |
| **Project Title:** | An Image-based Transfer Learning Framework for Classification of E-Commerce Products | | |
| **Word Count:** | 5710 | **Page Count:** | 20 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** Vrushali Atul Surve

**Date:** 16/12/2021

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# INDEX

# An Image-based Transfer Learning Framework for Classification of E-Commerce Products

MSc in Data Analytics, MSCDADJAN21A_I

Student Name: Vrushali Atul Surve
Student ID: x19212712

Lecturer Name: Paul Stynes

**Abstract.** Classification of ecommerce products involves identifying the products and placing those products into the correct category, for example, men's Nike Air Max will be in the men's category shoes on an e-Commerce platform. Identifying the correct classifications of products is time-consuming for businesses. **This research proposes an** Image-based Transfer learning framework to classify the images into the correct category in the shortest time. **The framework combines** Image-based algorithms with Transfer Learning. This research compares the time to predict the category and accuracy of traditional CNN and transfer learning models such as VGG19, InceptionV3, ResNet50 and MobileNet. A visual classifier is trained on CNN and transfer learning models such as VGG19, InceptionV3, ResNet50 and MobileNet. Theis models are trained on an e-commerce product dataset that combines the ImageNet dataset with pre-trained weights. The dataset consists of 15000 images scraped from the web. **Results** of these five models shown in this paper are based on accuracy, loss, and each model takes time to identify the correct category. CNN was found to be 51% accurate, whereas vgg19 and ResNet50 were found to be 55% and 76 percent accurate, respectively. With 85 percent accuracy and timing 0.1 seconds, Inception V3 and MobileNet outperform other models. This research helps the e-commerce websites to classify their product data.

**Keywords**
- Deep Learning, CNN, VGG19, InceptionV3, ResNet50, MobileNet, ImageNet, Image classification, Transfer Learning.

## 1. Introduction

Classification of products is a time-consuming procedure for both service giver and clients, and E-commerce has expanded fast in recent years, but it is necessary to categorize products. Images provide vital information, as they lead people in proper way. On the other hand, misleading visuals can frustrate users. Analyzing the correctness of an image requires an explanation of it in the same way that a user does. Significant time is being lost in sorting and labeling the items. Deep learning methods(Jha et al., 2021) like CNN can be used for image classification as it gives better accuracy and is the best model to deal with images.

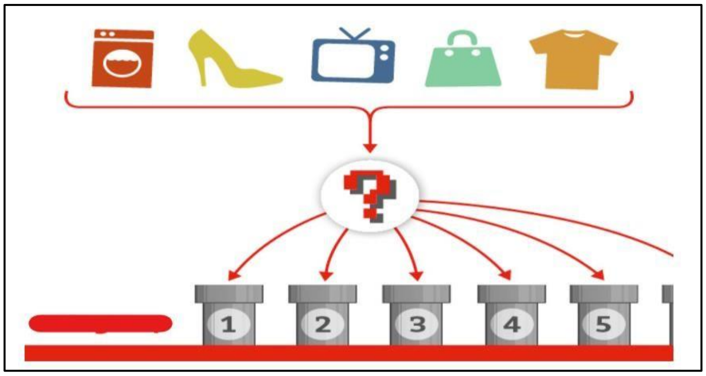



Figure 1 Ecommerce Product Classification

**This research aims** to determine how well an Image-based Transfer Learning Framework can categorize E-commerce products and post the images to an e-commerce portal in a timely manner. **The major contribution of this research** is a deep learning framework that combines and categorizes all product classes well with transfer learning models like CNN, VGG19, InceptionV3 and ResNet50, MobileNet within the shortest time and with accurate accuracy. The minor contribution is creating the e-commerce product by using the data web scraping method and then applying pre-trained weights to e-commerce product datasets. A model for the deep learning framework was chosen based on accuracy, loss, latency, and size.

**This paper covered** Various deep learning methods for image product classification, are shown in Section 2. The strategies adopted will be described in Section 3. Sections 4 and 5 deal with implementation and evaluation, respectively. Finally, Sections 6 and 7 of the study bring the research effort to a close with a conclusion and recommendations for future work

# 2. Literature survey

This section describes numerous research in image classification undertaken in the past. There are three subsections based on reviews. The first section focuses on product categorization, the second on convolutional neural network models, and the third on machine and deep learning models.

## 2.1 Review of Product Categorization

Product categorization refers to the arrangement and organization of products in their correct classifications. The power of an image to narrate concepts has captivated humanity, and with half of its population mind committed explicitly or implicitly to visuals, images have become our primary sense for understanding the world around us. Jain, (2019) demonstrates how visuals are used in item classification. Customers may now purchase any goods with the press of a button thanks to the fast growth of e-Commerce.

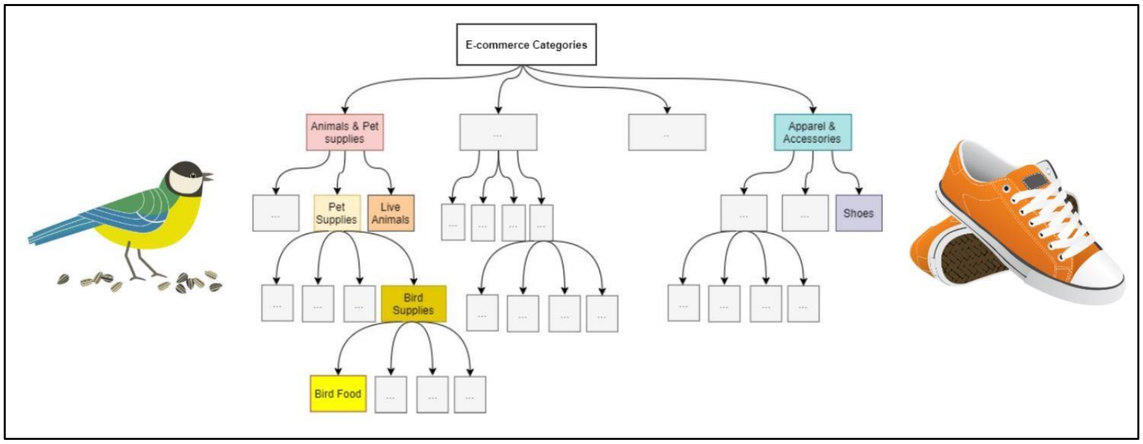



Figure 2 Product Classification

Figure 2 shows, this image represents product categorization into many categories. This article (Gossec, 2013) uses TensorFlow on GPU to demonstrate a deep learning strategy with extensive labelled data of several millions of images. The research paper (Wagh and Mahajan, 2019) examines the e-commerce product image classification approaches used in machine learning to categorize e-commerce product pictures. The major advanced classification methodologies are summarized, and the techniques used to improve classification accuracy. Different discriminating feature extraction approaches will be investigated in the future.

## 2.2 Review on Convolutional Neural Network

CNN is an abbreviation for convolutional neural network, which is a deep learning method. CNN is commonly regarded as the best image classification models (Russakovsky et al., 2015). However, as excellent as machine learning algorithms are, their classification accuracy is typically restricted in many object types of issues. This article (Krizhevsky et al., 2017) examines the e-commerce product image classification approaches used in machine learning to categorize e-commerce product pictures. The primary advanced classification methodologies and techniques used to improve classification accuracy are summarised. In 2012, a CNN was employed for the first time to reach a top 5 test error rate of 15.4 percent, while the best research paper the following year had a rate of 26.2 percent. The paper's technique was innovative, and it introduced several deep learning principles into the limelight. (Koirala et al., 2019). Transfer learning approaches are the most effective way to overcome CNN's faults. Transfer learning on a network pre-trained on a large dataset is conventional technique for more efficient and robust deep learning model training, as explained in this report. Transfer learning allows original parameters from a model learned on big datasets to be reused to evaluate new models using smaller training pictures. Available datasets, like ImageNet, give free annotated picture data of standard items, which may be used to train and optimize object identification algorithms. The research (Jha et al., 2021) demonstrates that image search is proper when there are little details about the goods. This paper suggested VGG19 and Inception V3 of transfer learning approached are good to solve challenges which are based on image classification/search/identification/suggestion. It is also beneficial to enhance the seller and consumer experience in product categorization and search.

## 2.3 Review on Machine and Deep Learning methods

Many machine learning categorization models were used in this study (He et al., 2020). There are seven models in total. SVM has the low precision of 65 per cent, whereas Ensemble' Neural Network' and 'Gradient Boosting Machine' have the most excellent precision value 90 per cent.

However, this is a time-consuming operation with hardware constraints. For picture categorization and prediction (Gujjar et al., 2021), pre-trained models such as MobileNet, MobileNetV2, VGG16, VGG19, and ResNet50 were utilized. Google Colab notebook was used for picture categorization and prediction. Compared to other pre-trained models, the result reveals that MobileNetV2 performs substantially better and occupies lower disk space. Transfer learning is one of the most effective DL approaches. Traditional machine learning algorithms attempt to learn each task from the beginning when a target task provides fewer high-quality training data; however, transfer learning transfers knowledge from a primary task to the target task. This article (Zia et al., 2018) is divided into six sections, each of which contains a high-level introduction of transfer learning when to use transfer learning, a few real-world applications, a case study comparing transfer learning to deep formal learning, and some domains where transfer learning may be employed.

# 3. Research Methodology

**The research methodology consists of** five **steps** methodology represented in the figure 3 below.

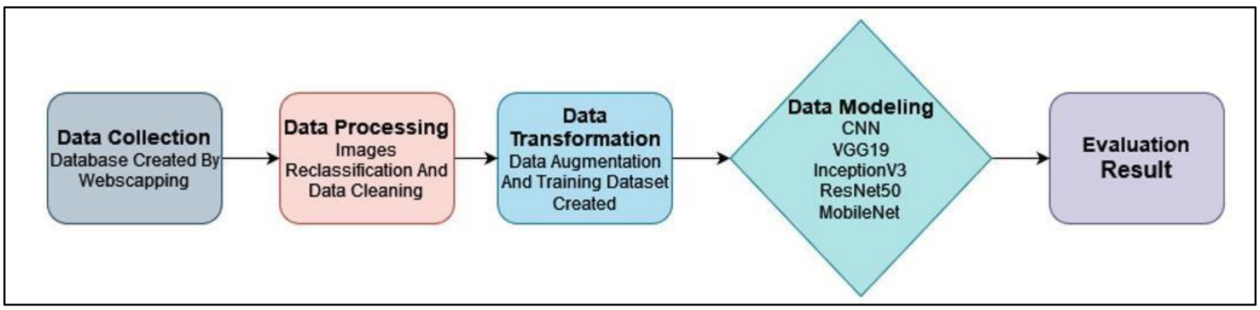



Figure 3 Research Methodology

**The first step,** *Data collection,*. Where data is gathered to create a dataset for study. This research was based on a big dataset of Flipkart E-commerce goods (PromptCloud, 2017), and it started by getting the data source from Kaggle, which contained the raw data file. This CSV dataset includes around 20000 rows, 15 columns, and no duplicate records. Image is scrapped using the web scraping method.

**The second step**, *Data processing*, involves image reclassification and cleaning. Images are reclassified as there are multiple subcategories. For example, there are around 6-7 subcategories in the category path for 'Rute Casual Sleeveless Embellished Women's Top.' Categories were manually split for better comprehension because there were numerous levels of categorisation. It denotes that the 'Rute Casual Sleeveless Embellished Women's Top' is classified under Women's Clothing. The categories with fewer data are removed like Role Play Toys, Video Players & Accessories, and misclassified images are classified in proper categories. After cleaning, there are 15 categories with 7755 images after the cleaning procedure.

**The third step**, *Data Transformation*, consists of the image transformation process, which uses the Image augmentation method. Image augmentation helps to reshape and reconstruct data images. There were 7755 images with 15 classes. After this process, there were 16679 images in the collection. These augmented datasets are used to create training datasets.

**The fourth step**, *Data Building*, is the creation of models based on the training dataset. The suggested method focuses on how image data may be used to categorize objects using a transfer learning strategy, making classification simpler and more convenient. Five models are performed.

The Convolutional Neural Networks (CNN) model was trained using a split 70:30 for training and testing. Other deep learning models are trained to utilize the transfer learning approach. Pre-trained models' weights were utilized in the Visual Geometry Group 19 (VGG19), Inception V3, ResNet50, and MobileNet models. This section explains the approaches and models used to categorize the images. The data modelling stage begins after partitioning data into training and testing sets and developing various models based on the data used. The ImageNet dataset, a massive database containing millions of images of separate classes, was used to train this model. VGG16 generalizes effectively and can produce cutting-edge findings. ResNet, most popular deep learning models, having also won the ImageNet challenge. The Inception V3 includes 48 layers and is trained on a dataset containing over a million pictures. MobileNet is a lightweight network that requires less maintenance and works admirably at high speeds.

**The fifth step**, *Evaluation and Results*, involves assessing built models' performance using criteria like execution time. Accuracy, precision, recall, and the F1-score are some of the evaluation variables utilized to measure the models' effectiveness in this research project. For the outcomes stage, which focused on analyzing the obtained information and findings, plots of Epoch vs Accuracy and Epoch versus Loss are shown.

# 4. Design Specification

This section describes the architecture. In this part, the model design specifications will be explained using deep learning framework **architecture** for Image classifications. This architecture represents the processes and tools utilized to construct this system. The developed architecture adheres to the below design architecture, which includes a business logic layer and a data persistent layer. The data persistence layer specifies the source of the data, and the business logic layer describes data preparation, training, and evaluation of classification models.

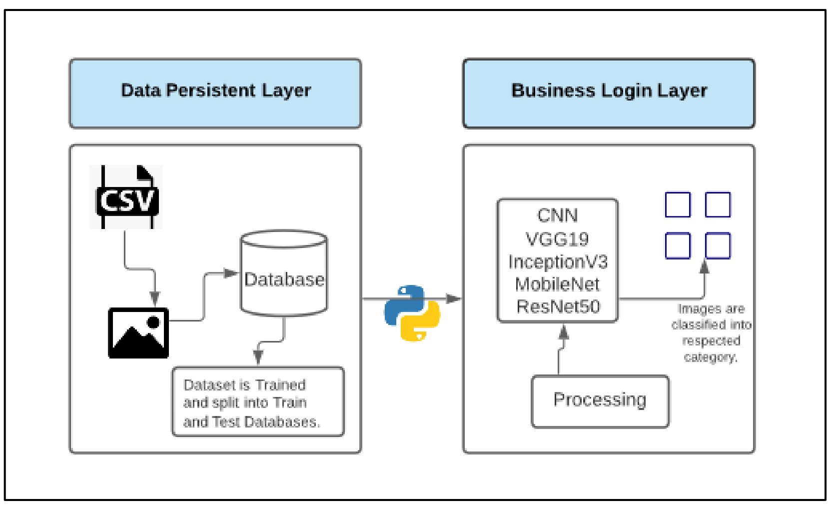



Figure 4 Framework Design

# 5. Implementation

The deep learning framework **implemented** image classification algorithms.

## 5.1 Environment Setup

Figure 5 shows various technologies used in the process from start to end. On a Google collaborative machine with a 78.19 GB hard drive, 12.69 GB RAM, and a 48.97 GB run-time GPU, the experiment is carried out. Because these models have additional layers, they take extra time to execute.

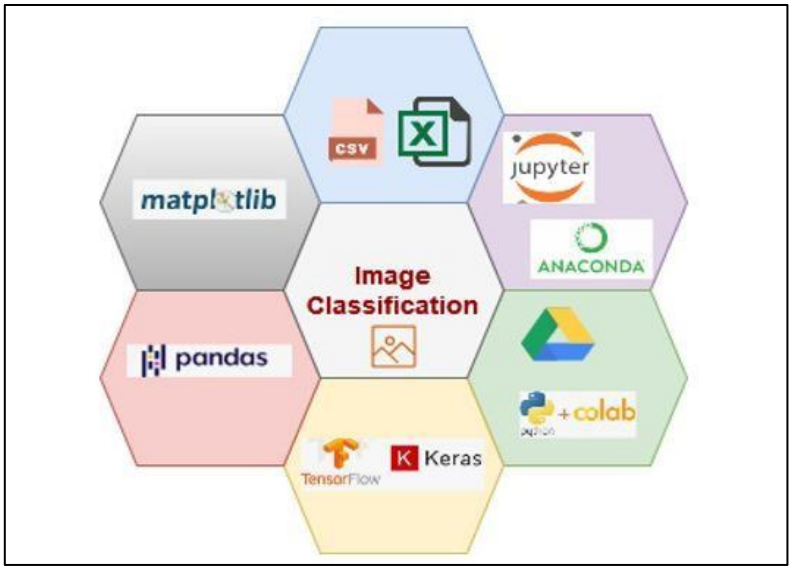



Figure 5 Environment

Data is posted to Google Drive and accessible using the Python API by mounting Google Drive. BeautifulSoup is a Python package that aids in the extraction of data from CSV files. BeautifulSoup extracted picture URLs from the CSV derived from Kaggle in this research. Excel is used to check the counts of data. The Python programming language was used to implement the project, and the code was run using the Jupyter Notebook which is based on Anaconda Platform. It is a Python-based platform for large-scale data processing, predictive analysis, data mining, and other Python-based operations that is free and open-source. Jupyter notebook is a open-source tool good for graphics, data processing, and other capabilities. Keras is a user-centric rather than a computer-centric API. Keras encourages the application of best practices to reduce computational load. It provides trustworthy and quick APIs, decreasing the number of user engagements required in typical cases. TensorFlow is a free artificial intelligence framework that uses data flow graphs to aid numerical computations. Pandas' library gives a fast and efficient data frame with customized indexing, and the evaluation matrix was produced using matplotlib, a Python visualization tool.

## 5.2 Data Transformation

The dataset was taken from Kaggle (PromptCloud, 2017). It has contained raw data csv file. Products images were crawled from this file. Web scraping tools like BeautifulSoup were used. The crawlers collected information from this CSV's column image URL content and imported the image path. That image was stored in the appropriate category folder. It was using logger function a log folder with log file created which containing failure warnings while downloading images. A total of 15015 photos were saved in the 27 categories. This stage entails working on the database before moving on to the implementation phase. It is critical to pre-process data before modelling in any research project. Each product has its category after importing images. In CSV, for example, the category route for' Rute Casual Sleeveless Embellished Women's Top' has around 6-7 subcategory. Because there were multiple levels of categorization, categories were manually separated for good understanding. It represents that the 'Rute Casual Sleeveless Embellished Women's Top' falls under the Women's Clothing category. There are 15 categories with 7755 records after the cleaning procedure. The image augmentation procedure entails altering the images in the training data. Horizontal flip, vertical flip, rotate, and blur is some of the transformation operations used in this procedure. Image augmentation is unlike data pre-processing, which entails rescaling photos, requiring specific image alterations. It may also be used for modelling. Renumbered images were saved in JPG format and bundled into an a.zip file to upload on google drive. There are 16679 photos in this collection, divided into 15 classes. This study applies project normalization to images in the correct format, allowing for faster modelling. The critical benefit of picture resizing is that it allows the model to train more quickly since it can learn quicker on smaller images. The photos are scaled to 224*224 pixels, and the BGR image is converted to RGB. Then training dataset is created. The dataset was divided into a 70:30 split for training and testing.

## 5.3 Convolutional Neural Network (CNN)

CNN is used for analysing images because of its great accuracy ("Basic CNN Architecture," 2020). This method uses a hierarchical model that builds a funnel-shaped network, then creates a fully connected layer in which all neurons are linked, and the output is processed. The CNN architecture comprises layers such as the dense layer, convolutional layer, max-pooling layer, attending layer, and dropout layer. The convolution layers algorithm uses the 'relu' activation and dropout parameters. To improve the model's execution performance, the Rectified Linear Unit (ReLU) is employed as an activation function in the layers of this model. A thick layer is required to link all the layers. The max-pooling layer provides the feature map from the photos. The data will be flattened into a one-dimensional array via the Flatten layer. The Softmax function of the output layer is used to classify images.

## 5.4 Transfer Learning methods based on VGG19, InceptionV3, MobileNet and ResNet50

### 5.4.1 VGG19

CNN model training from scratch has benefits and loss. (Jaworek-Korjakowska et al., 2019) Many grid parameters must be run to identify the ideal setting in the neural network. The most effective technique to address this problem is to remake an already optimized and learned model. VGG-19 (Visual Geometry Group) was trained on a 1000-class ImageNet and has been shown to operate well with smaller datasets. It has 19 layers, which has max pooling, dropout, convolutional, and so on. VGG-19 is mainly for training layer classification and includes both a dropout and a dense layer.

### 5.4.2 InceptionV3

This is Google's Inception CNN version 3 and initially using at the ImageNet Recognition Challenge (Szegedy et al., 2016). This model suggests a step forward from prior versions, which mainly focused on reducing processing power by factorizing convolutions. Special care must ensure that the computational advantages are not lost if an Inception Network is altered. The architecture consists of 9 inception modules to create 22 layers, global average pooling.

### 5.4.3 ResNet50

In the Net 50 architecture, there are three kinds of layers ("Understanding ResNet50 architecture," 2020) 1 max pool, 48 convolutional, and one average pool. This process can be used to computer vision and non-computer vision classification problems to improve accuracy. ResNet employs the concept of a deep residual network to lower the training error rate. This 50layer deep neural network can work over billion images with more than 1000 categories datasets, on this ResNet network loaded.

### 5.4.4 MobileNet

MobileNet is a low-maintenance network that provides outstanding performance at high speeds (Kaiser et al., 2017). Using depthwise convolution, a particular filter is employed to each input channel, and then the outputs are combined using 1 x 1 convolution using pointwise convolution. The inputs are filtered and combined into a new batch of outcomes in a one step. After that, a conventional convolution splits the outputs into two layers. Separate layers for filtering and merging are developed. Astride, two convolutions, a depthwise, a pointwise layer that double the channels, and so on comprise the MobileNet's 30 layers. The Mobilenet's final layer is removed and replaced by two Denser layers. It also has two SoftMax-activated neurons connecting to the last layer.

## 6. Results and Discussion

**This experiment aims to** how effectively CNN, VGG19, InceptionV3, and ResNet50, MobileNet models perform image classification in terms of accuracy and average time is taken for classifying sample data set images. Using the transfer learning method and the ImageNet pre-trained weights, the models were trained to detect and categorize photos into 15 different eCommerce product categories. For the evaluation phase as per epoch, the model's loss and accuracy of training and validation data calculated with its graphs. Evaluation measures for performance and confusion matrix shown in Figure 6 (Rothe et al., 2016) are calculated to get a genuinely positive and negative true value. A confusion table is a matrix that demonstrates how well a classification model works on a test data set with known actual values. Although it is easy to understand but described confusingly.

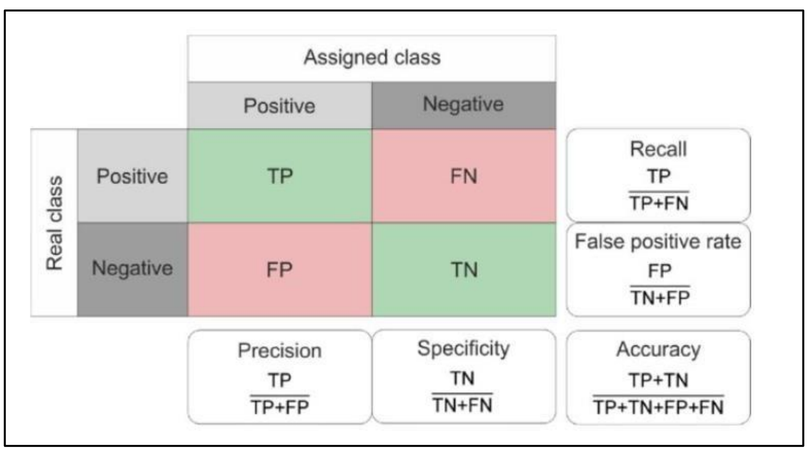



Figure 6 Confusion matrix & performance Measures

The studies detailed here improve the accuracy and value of true negatives, which are essential in this research. The number of epochs and other image enhancement procedures can be experimented with to improve the model's performance. There were three experiments taken from this research.

## 6.1 Experiment 1: Effectivity of CNN model in classification E-commerce products

It takes over 3 hours to obtain 51% accuracy while training the model on a subset of data consisting of 16679 photos. There is one major issue in this. The time required for even a small quantity of data is considerable, and if the size of the data grows, this time will only increase. Accuracy performance depend on epoch number. The model reached 38% accuracy in the 10th epoch, then dropped significantly to 30% at the 12th epoch before remaining stable with slight increases and decreases. As shown in figure 7, The CNN model considerably improves training accuracies for the accuracy vs. epoch graph. Whereas validation accuracy was also greatly improved, there were minor ups and downs after the tenth epoch. On the loss vs. epoch graph, a learning curve illustrates this model for training loss, which looks to be a reasonably good match, and a learning curve for validation loss indicates noise movement all-around training loss. The validation loss remains constant at 2.5.

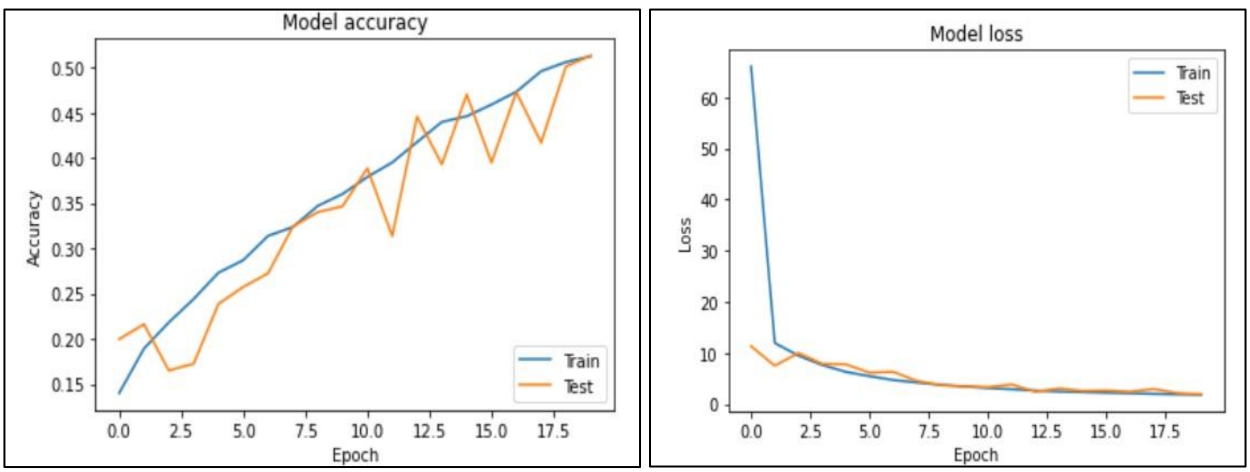
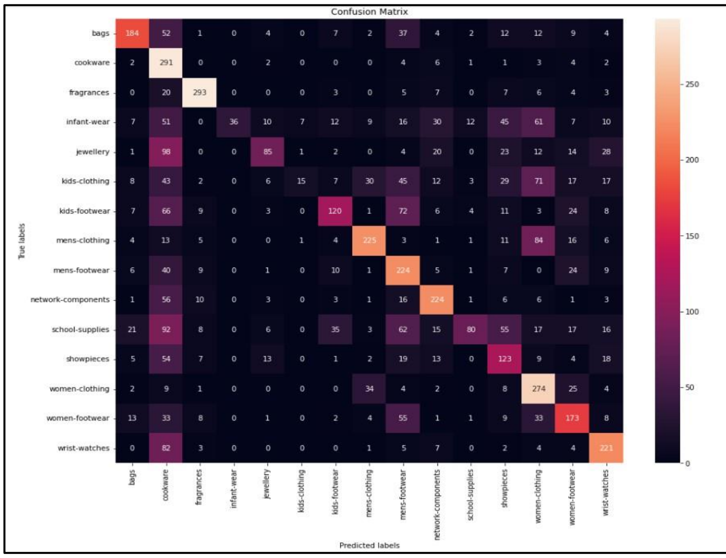



Figure 7 CNN Accuracy/Loss versus Epoch

The confusion matrix in Figure 8 is used to show the classification model, which summarizes the prediction outcomes. Log these predicted classes vertically and the valid classes horizontally. When a model predicts the correct class, it will be logged in the diagonal. It despite a higher number prediction for "fragrance". Furthermore, accuracy is calculated on the f1-score value in the categorization report. This accuracy is determined by how the model is trained for that particular class. It is also determined by the size of the photographs in each class. Precision is calculated as the ratio of genuine positive to total expected positive values. The percentage range is between 0 (Low) and 1 (High). The recall is calculated by dividing the total number of true positives by the total number of actual positives. The percentage range is between 0 (Low) and 1 (High). Precision is related to False Positives, whereas Recall is related to False Negatives. The F1 score is a weighted harmonic mean of accuracy and recalls, with 1.0 being the highest and 0.0 being the poorest. Precision values are appropriate for category '3,' i.e., one, and recall is good for category one at 0.92. The f1-score for category two is respectable, at 0.83. The CNN model has an overall accuracy of 51%. shown in Figure 8.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.70 | 0.56 | 0.62 | 330 |
| 1 | 0.29 | 0.92 | 0.44 | 316 |
| 2 | 0.82 | 0.84 | 0.83 | 348 |
| 3 | 1.00 | 0.12 | 0.21 | 313 |
| 4 | 0.63 | 0.30 | 0.40 | 288 |
| 5 | 0.62 | 0.05 | 0.09 | 305 |
| 6 | 0.58 | 0.36 | 0.44 | 334 |
| 7 | 0.72 | 0.60 | 0.66 | 374 |
| 8 | 0.39 | 0.66 | 0.49 | 337 |
| 9 | 0.63 | 0.68 | 0.65 | 331 |
| 10 | 0.75 | 0.19 | 0.30 | 427 |
| 11 | 0.35 | 0.46 | 0.40 | 268 |
| 12 | 0.46 | 0.75 | 0.57 | 363 |
| 13 | 0.50 | 0.51 | 0.51 | 341 |
| 14 | 0.62 | 0.67 | 0.64 | 329 |
| accuracy | | | 0.51 | 5004 |
| macro avg | 0.61 | 0.51 | 0.48 | 5004 |
| weighted avg | 0.61 | 0.51 | 0.49 | 5004 |

Figure 8 CNN Confusion Matrix & Classification Report

## 6.2 Experiment 2: Effectiveness of transfer learning models like VGG19 and InceptionV3 for classification of E-commerce products

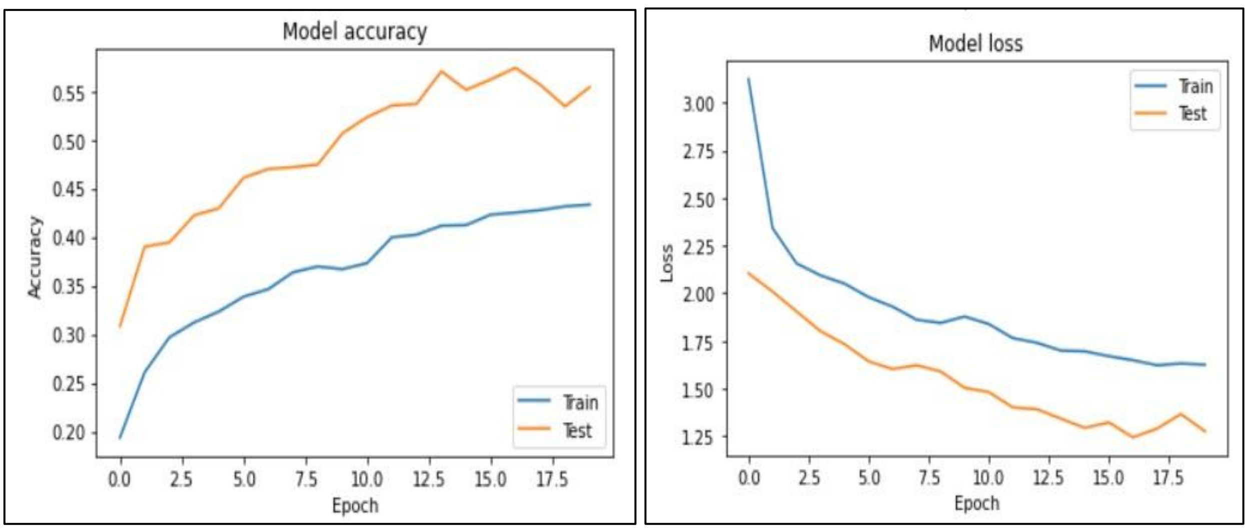
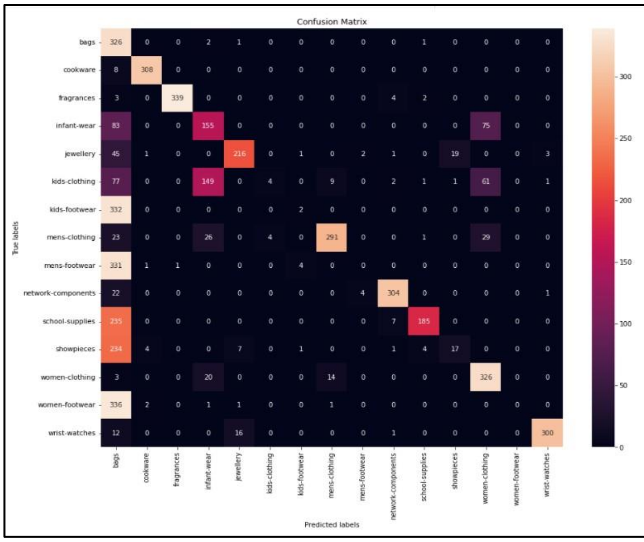



Figure 9 VGG19 Accuracy/Loss versus Epoch

In the VGG19 model, a model is trained on 20 epochs. In Figure 9, the loss value is decreased from 2.1 to 1.25, and in Figure 9, the Model accuracy 0.55 and it gradually increased till the end of epoch 20. The confusion matrix in Figure 10 is shows the performance of the classification report and prediction outcomes. The VGG accuracy model is 0.55, shown in Figure 10.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.16 | 0.99 | 0.27 | 330 |
| 1 | 0.97 | 0.97 | 0.97 | 316 |
| 2 | 1.00 | 0.97 | 0.99 | 348 |
| 3 | 0.44 | 0.50 | 0.47 | 313 |
| 4 | 0.90 | 0.75 | 0.82 | 288 |
| 5 | 0.50 | 0.01 | 0.03 | 305 |
| 6 | 0.25 | 0.01 | 0.01 | 334 |
| 7 | 0.92 | 0.78 | 0.84 | 374 |
| 8 | 0.00 | 0.00 | 0.00 | 337 |
| 9 | 0.95 | 0.92 | 0.93 | 331 |
| 10 | 0.95 | 0.43 | 0.60 | 427 |
| 11 | 0.46 | 0.06 | 0.11 | 268 |
| 12 | 0.66 | 0.90 | 0.76 | 363 |
| 13 | 0.00 | 0.00 | 0.00 | 341 |
| 14 | 0.98 | 0.91 | 0.95 | 329 |
| accuracy | | | 0.55 | 5004 |
| macro avg | 0.61 | 0.55 | 0.52 | 5004 |
| weighted avg | 0.62 | 0.55 | 0.53 | 5004 |

Figure 10 VGG19 Confusion Matrix & Classification Report

In the InceptionV3 model, the model is trained on ten epochs. This model has given the best accuracy among all models. The accuracy and Confusion matrix of CNN model are 0.85, shown in Figure 12, and the loss and accuracy model is shown in Figure 11.

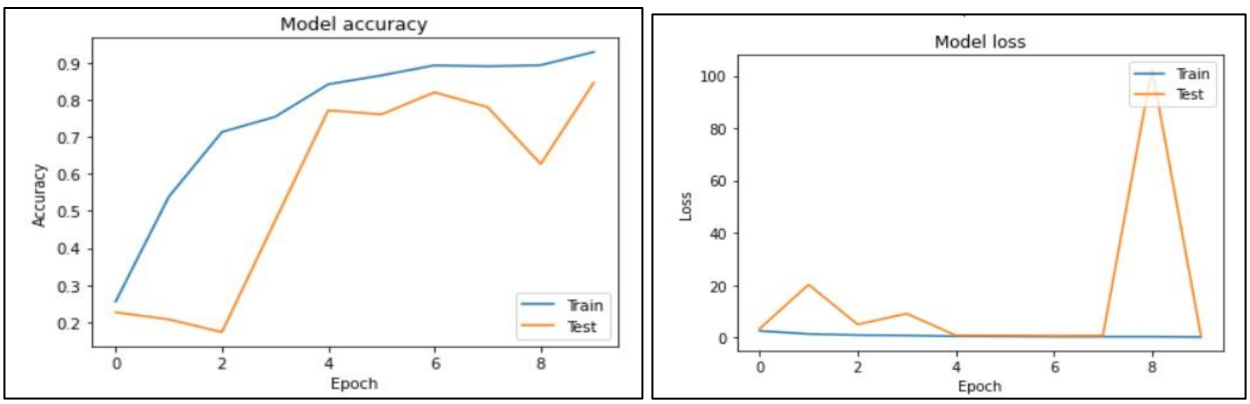
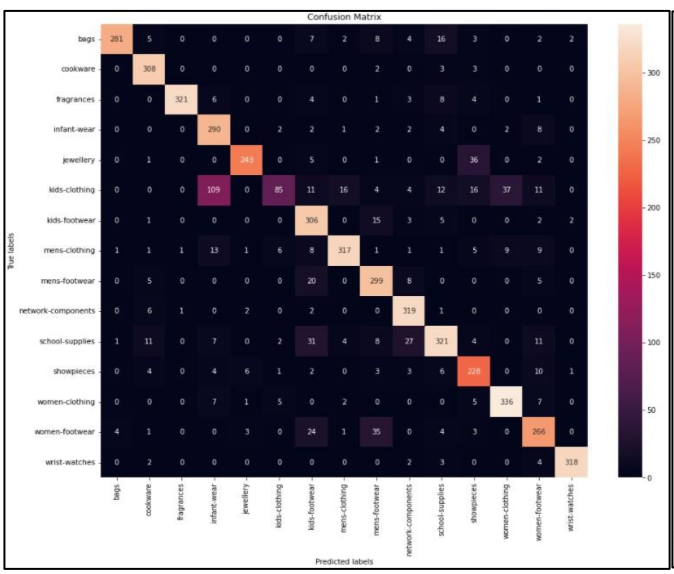



Figure 11 InceptionV3 Accuracy/Loss versus Epoch




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 0.85 | 0.91 | 330 |
| 1 | 0.89 | 0.97 | 0.93 | 316 |
| 2 | 0.99 | 0.92 | 0.96 | 348 |
| 3 | 0.67 | 0.93 | 0.77 | 313 |
| 4 | 0.95 | 0.84 | 0.89 | 288 |
| 5 | 0.84 | 0.28 | 0.42 | 305 |
| 6 | 0.73 | 0.92 | 0.81 | 334 |
| 7 | 0.92 | 0.85 | 0.88 | 374 |
| 8 | 0.79 | 0.89 | 0.84 | 337 |
| 9 | 0.85 | 0.96 | 0.90 | 331 |
| 10 | 0.84 | 0.75 | 0.79 | 427 |
| 11 | 0.74 | 0.85 | 0.79 | 268 |
| 12 | 0.88 | 0.93 | 0.90 | 363 |
| 13 | 0.79 | 0.78 | 0.78 | 341 |
| 14 | 0.98 | 0.97 | 0.98 | 329 |
| accuracy | | | 0.85 | 5004 |
| macro avg | 0.86 | 0.85 | 0.84 | 5004 |
| weighted avg | 0.86 | 0.85 | 0.84 | 5004 |

Figure 12 InceptionV3 Confusion Matrix & Classification Report

### 6.3 Experiment 3: Effectivity of transfer learning models like ResNet50 and MobileNet model in classification E-commerce products

In the ResNet50 and MobileNet models, both the models are trained on ten epochs. In ResNet50 Model, as figure 12 shown the training accuracy is higher than validation accuracy, and it goes from 0.45 to 0.91. Validation accuracy goes down after two epochs, then at six epochs goes high, then eight epoch it goes down, then suddenly goes high. For loss vs. epoch graph, Training loss starts with 2.5 then goes down till 0.2, whereas Validation loss value also starts with 2.5 but has a lot of ups and downs and goes down till 0.9. the accuracy for this model is 0.76 as shown in figure 13.

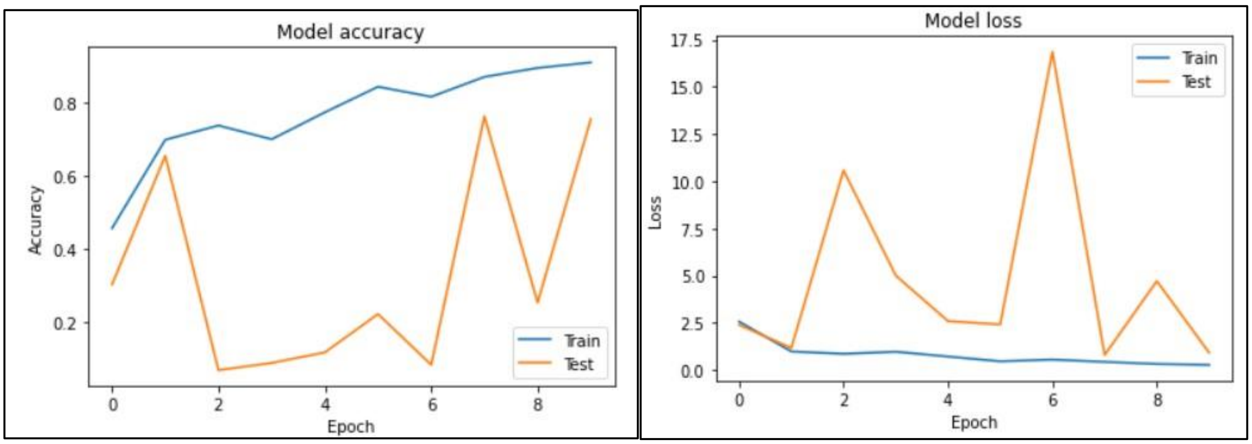
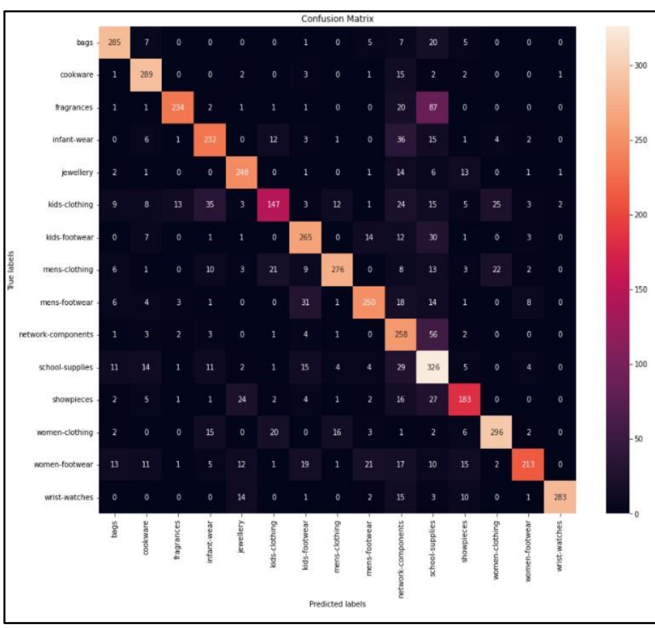



Figure 12 ResNet50 Accuracy/Loss versus Epoch




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.84 | 0.86 | 0.85 | 330 |
| 1 | 0.81 | 0.91 | 0.86 | 316 |
| 2 | 0.91 | 0.67 | 0.77 | 348 |
| 3 | 0.73 | 0.74 | 0.74 | 313 |
| 4 | 0.80 | 0.86 | 0.83 | 288 |
| 5 | 0.71 | 0.48 | 0.58 | 305 |
| 6 | 0.74 | 0.79 | 0.76 | 334 |
| 7 | 0.88 | 0.74 | 0.80 | 374 |
| 8 | 0.82 | 0.74 | 0.78 | 337 |
| 9 | 0.53 | 0.78 | 0.63 | 331 |
| 10 | 0.52 | 0.76 | 0.62 | 427 |
| 11 | 0.73 | 0.68 | 0.70 | 268 |
| 12 | 0.85 | 0.82 | 0.83 | 363 |
| 13 | 0.89 | 0.62 | 0.73 | 341 |
| 14 | 0.99 | 0.86 | 0.92 | 329 |
| accuracy | | | 0.76 | 5004 |
| macro avg | 0.78 | 0.76 | 0.76 | 5004 |
| weighted avg | 0.78 | 0.76 | 0.76 | 5004 |

Figure 13 ResNet50 Confusion Matrix & Classification Report

In MobileNet Model, The Graph in figure 14 "Accuracy Vs. Epoch" of the MobileNet Classification Model illustrates a stable and smooth increase in the Accuracy of the Trained

Dataset starting just above 0.5 level of accuracy to reaching the perfect one as the number of iterations are increased. The Maximum change is observed while training the dataset is from an epoch between 1 - 2, and accuracy is constantly rising till epoch six, after which the accuracy is maintained to almost one from epoch seven and further on. As for the Testing Dataset observed a quite irregular pattern of rising to a sure accuracy and then some steep downfall towards the final epochs and again going upwards. The maximum accuracy observed during the testing period is 0.88 during the seventh Iteration. The "Loss vs. Epoch" graph of the MobileNet Model sees a higher rate of loss at the start of the epoch passes, which gradually decreases till the sixth epoch, after which the loss rate was utmost 0.30 or less for the Training dataset. Furthermore, for the testing Phase of the Dataset, the loss rates increase from 1.75 to almost reaching two from the 0th Iteration to the 1st Iteration, respectively. Also, it observed a steep downfall after the first pass from almost 2 to 0.75 from 1st to 3rd Iteration, moving forward to the fourth epoch till the last one it is observing some bearish and some bullish behaviour in the graph as it is not moving forward in a linear way which was in the case of the training dataset. The accuracy in figure 15 shown for MobileNet model is 0.85.

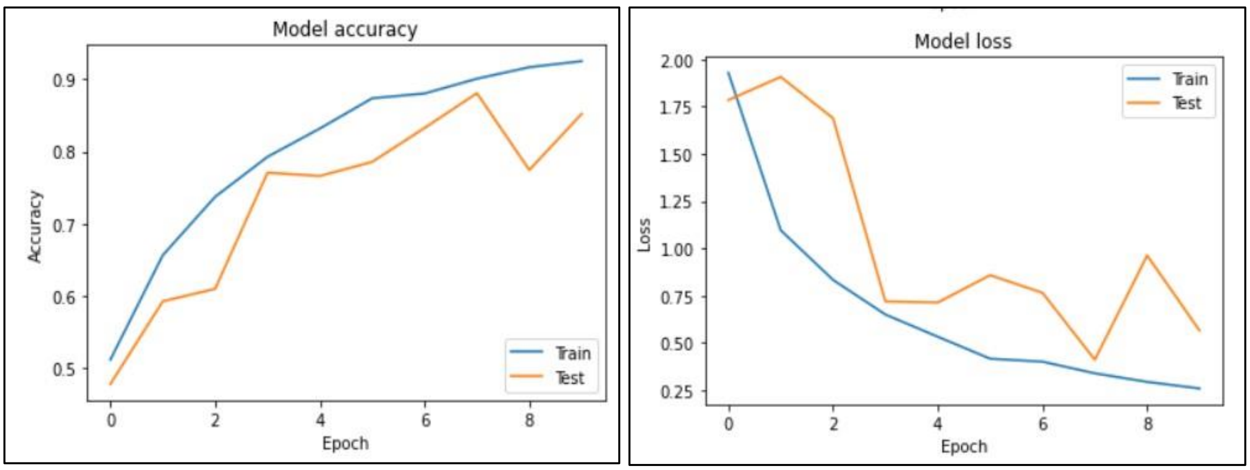
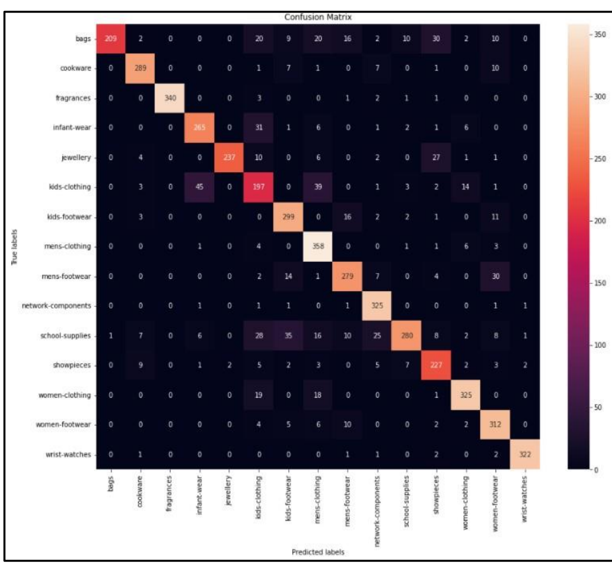



Figure 14 MobileNet Accuracy/Loss versus Epoch




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.63 | 0.77 | 330 |
| 1 | 0.91 | 0.91 | 0.91 | 316 |
| 2 | 1.00 | 0.98 | 0.99 | 348 |
| 3 | 0.83 | 0.85 | 0.84 | 313 |
| 4 | 0.99 | 0.82 | 0.90 | 288 |
| 5 | 0.61 | 0.65 | 0.63 | 305 |
| 6 | 0.80 | 0.90 | 0.85 | 334 |
| 7 | 0.76 | 0.96 | 0.84 | 374 |
| 8 | 0.84 | 0.83 | 0.83 | 337 |
| 9 | 0.86 | 0.98 | 0.91 | 331 |
| 10 | 0.92 | 0.66 | 0.76 | 427 |
| 11 | 0.74 | 0.85 | 0.79 | 268 |
| 12 | 0.90 | 0.90 | 0.90 | 363 |
| 13 | 0.80 | 0.91 | 0.85 | 341 |
| 14 | 0.99 | 0.98 | 0.98 | 329 |
| accuracy | | | 0.85 | 5004 |
| macro avg | 0.86 | 0.85 | 0.85 | 5004 |
| weighted avg | 0.86 | 0.85 | 0.85 | 5004 |

Figure 15 MobileNet Cconfusion Matrix & Classification Report

## 6.4 Result

**This result indicates all the good values from the graphs and** the accuracy summary for all the models implemented in this research work.

| Model Name | Good Predicted Category | Good Precision Value | Good Recall Value | Good F1-Score Value |
|---|---|---|---|---|
| CNN | Fragrance | 100 | 92 | 83 |
| VGG19 | Fragrance | 100 | 99 | 99 |
| InceptionV3 | Women-clothing | 99 | 97 | 98 |
| ResNet50 | School-supplies | 99 | 91 | 92 |
| MobileNet | Men's-clothing | 100 | 98 | 99 |

As seen in the above table, there are good, predicted values from the confusion matrix and calculation report. The fragrance is the most indicated category among all the categories. In contrast, all the models have the highest precision, which is near to 1 VGG19 has a good recall value, and VGG19 and MobileNet has a good f1-score among all the model categories.

| Model Name | Accuracy | Average time |
|---|---|---|
| CNN | 51 | 0.426345857 |
| VGG19 | 55 | 0.146205902 |
| InceptionV3 | 85 | 0.137190358 |
| ResNet50 | 76 | 0.126409419 |
| MobileNet | 85 | 0.101487923 |

The first model, CNN, performed with an accuracy of 51%, and VGG19 had 55% accuracy. Except for these two models, the other three models' accuracy is quite well. ResNet50 has 76, whereas MobileNet and InceptionV3 have 85% accuracy. Inferencing did for each model on the testing dataset to check the time taken to predict the category to a given image from the point of accuracy it can say as well as average time taken to predict categories as per above table it can say that InceptionV3 and MobileNet model is best in accuracy as well in time.

## 6.5 Discussion

In this study, the thesis work begins with data collection. The dataset utilized in this study was obtained by web scraping. Then, after cleaning, the data was less so with augmentation data. The CNN model is then built from scratch, while the VGG19, InceptionV3, ResNet50, and MobileNet models are made using a transfer learning technique with weights from pre-trained ImageNet datasets. The total number of photos is 16679, and the batch size and epoch for the CNN model are 100 and 20, respectively. It takes 16679/100 iterations to go through all 16679 photos, one epoch. This procedure will continue for the following 20 epochs [time]. The epoch number determines the accuracy/loss performance. The batch size for CNN, VGG19, and inceptionV3 are 100, while ResNet50 and MobileNet are 50, epochs are 20 for CNN and VGG19, and 10 for the others. MobileNet outperforms all other models regarding the average time taken by all photos in testing data. According to the table above, two models, Inception V3 and MobileNet outperform the other models with an accuracy of 85%. However, the average time taken by InceptionV3 to identify the photographs is 3 seconds greater than that of MobileNet.

# 7. Conclusion and Future Work

**The aim of this study is** with transfer learning to categorize e-commerce items into different classes. Ecommerce and transfer learning are two industries that are rapidly expanding. **This research proposes** to develop the best model for calculating less time with more accuracy. Creating a consistent category list with so many things coming from different places is challenging. Only the CNN model requires more than 2.5 hours to train. Other models can be trained in less than an hour. The accuracy of all models is then compared against the time is taken. InceptionV3 and MobileNet accuracy are more when compared with another trained model. CNN has been observed to have 51 percent accuracy, whereas vgg19 and ResNet50 have 55 and 76 percent accuracy, respectively. **Results demonstrate that** the InceptionV3 and Mobilenet are the best models for categorizing image-based e-commerce items.

**This research can potentially enhance** e-commerce websites should organize their products into appropriate categories so that products may be easily accessed via category pages. **This work can be improved by** optimising the model and changing epochs as per the model. All product-related visual material was used in this study, whereas text content was excluded. As a result, future studies will include researching fusion with multiple images and fusion with text modality. This Investigation of multimodal fusion may be archived using models like InceptionV3-based for pictures, BERT LSTM for texts, Late Fusion, and Early Fusion.

# 8. Acknowledgement

I would want to express my sincere gratitude to my supervisor Paul Stynes, Pramod Pathak and Rejwanul Haque for mentoring and supporting me on this journey and my sincerest gratitude to all the NCI faculty members for imparting knowledge guiding me during the Data Analytics course. They guided me through every step of this challenging thing, making it feasible and straightforward. I am grateful to my family for support and encouragement in helping me achieve my goals.

# References

Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network [WWW Document], 2020. UpGrad Blog. URL https://www.upgrad.com/blog/basic-cnn-architecture/ (accessed 12.15.21).

Gossec, C., 2013. Image-Based ML Techniques To Classify Billions Of E-Commerce Products Into Thousands Of Categories.

Gujjar, P., Kumar, P., Chiplunkar, N., 2021. Image classification and prediction using transfer learning in colab notebook | Elsevier Enhanced Reader [WWW Document]. https://doi.org/10.1016/j.gltp.2021.08.068

Jain, P., 2019. Image Classification for E-commerce [Part I] [WWW Document]. Medium. URL https://towardsdatascience.com/product-image-classification-with-deep-learning-part-i5bc4e8dccf41 (accessed 12.8.21).

Jaworek-Korjakowska, J., Kleczek, P., Gorgon, M., 2019. Melanoma Thickness Prediction Based on Convolutional Neural Network With VGG-19 Model Transfer Learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Long Beach, CA, USA, pp. 2748–2756. https://doi.org/10.1109/CVPRW.2019.00333

Jha, B.K., G, S.G., R, V.K., 2021. E-Commerce Product Image Classification using Transfer Learning, in: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). Presented at the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 904–912. https://doi.org/10.1109/ICCMC51019.2021.9418371

Kaiser, L., Gomez, A.N., Chollet, F., 2017. Depthwise Separable Convolutions for Neural Machine Translation. ArXiv170603059 Cs.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning – Method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric. 162, 219–234. https://doi.org/10.1016/j.compag.2019.04.017

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386 PromptCloud, 2017. Flipkart Products [WWW Document]. URL https://kaggle.com/PromptCloudHQ/flipkart-products (accessed 12.9.21).

Rothe, S., Wirtz, S., Söffker, D., 2016. About the reliability of diagnostic statements: fundamentals about detection rates, false alarms, and technical requirements.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-0150816-y

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Understanding ResNet50 architecture [WWW Document], 2020. . OpenGenus IQ Comput. Expert. Leg. URL https://iq.opengenus.org/resnet50-architecture/ (accessed 12.15.21).

Wagh, D., Mahajan, S., 2019. Product Image Classification Techniques.

Zia, B., Illikkal, R., Rogers, B., 2018. Use Transfer Learning for Efficient Deep Learning Training on Intel® Xeon® Processors 11.