Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776887

FTXNET: FASHION PRODUCT IMAGE CLASSIFICATION USING FINE-TUNED PRETRAINED XCEPTION NET ARCHITECTURE

Subhash Chandra [1], Priyanka [2]
[1]Associate Professor,Department of Computer Science & Engineering, RCEW, Jaipur
[2]Department of Computer Science & Engineering, RCEW, Jaipur
subhashccjat@yahoo.com

Abstract
Deep learning (DL) can be used to improve performance in a wide range of commercial fields, according to researchers. Fashion-related enterprises, in particular, have begun to incorporate DL methods into their e-commerce operations, such as clothing identification, a clothing search, and retrieval engine, & automated product suggestion systems. The problem of image classification serves as the most significant structural component of these applications. Fashion product classification, on the other hand, might be challenging owing to the huge no. of product categories and the absence of tagged image data for each category. Additionally, the degree of classification is complicated. To put it another way, multi-class fashion product categorization might be difficult and unclear to distinguish across different classes. When it comes to visual classification, there are many different methodologies available, ranging from traditional image processing to DL approaches. For this reason, we suggest that the XceptionNet architecture be pre-trained on a small dataset of Fashion product images before being fine-tuned on our fine-grained fashion dataset depending upon design features. This helps to compensate for the small size of the dataset & shortens training time. Final accuracy findings for Gender + masterCategory are 91.72 percent on average, 98.04 percent on average for subCategory, and 91.93 percent on average for articleType once the studies are completed. It has been possible to create a unique image recognition and feature extraction system that has high classification accuracy while still being stable and durable enough to be employed in a commercial environment. As well as presenting the key challenges and possible solutions encountered during the creation of this system, this article also includes findings from experiments performed on the Fashion Product Images (Small) dataset.
Index Terms— Image classification, Fashion products, Deep learning, Transfer learning, XceptionNet architecture.

1. INTRODUCTION

VISUALclassification of commercial items is a subset of broader areas of object recognition & feature extraction in computer vision, also it is particularly significant in the creative process of the fashion industry. Automatically classifying clothing aspects helps both designers and data professionals be aware of their total production is essential to coordinate marketing campaigns, eliminate duplication, categories clothing goods for e-commerce reasons, and so on.

Finding a visual analysis of the total production in the fashion industry is critical for generating marketing strategies & assisting fashion designers in the development of novel goods

throughout the creative process. The collection and categorization of the many outputs of the designers' labor is a crucial first step to continue with visual analysis. This is particularly true if there are a significant number of distinct designer teams that are engaged via outsourcing contracts and are dispersed across the globe, as is often the case with major corporations. Furthermore, the designers' output consists of a steady number of images that depict an extremely broad range of things, ranging from apparel to footwear. Furthermore, these works are often distinct from one another. Then, while articles are of various forms & originate from a variety of sources, their attributes should be assessed & classified as a whole by data professionals & analysts. As a result, one of the most important steps in visual analysis is the recognition, classification, and extraction of characteristics from final images or 3D renderings of items that have been acquired before the actual manufacture of the apparel. In addition to e-commerce applications, classifying clothes goods is beneficial for a variety of other reasons, such as preventing duplication, rating product kinds, and doing statistical analysis[1]. To date, the classification of garment goods has been done manually since it takes both subject expertise and a thorough understanding of the whole spectrum of products. To classify images in this way by hand involves a high risk of inaccuracy, which might lead to misclassifications during the visual analysis. In reality, the number of articles produced has risen dramatically over the years.

Each of these categories accepts data from a separate domain and spans a diverse range of potential classes and subclasses. As was predicted, executing the classification of attributes for every article by hand has become an impractical undertaking. In part because of the enormous quantity of data and the variety of picture sources, the process of categorizing and detecting elements of clothing pictures is not only difficult to accomplish manually but is also complex to automate effectively. It was decided to focus on a subset of seven critical traits, which we shall refer to like features in the next section: In addition to

I. the logo's kind, and size,
II. the three-stripes presence & colors,
III. three major color palette,
IV. prints & patterns
V. the neck shape,
VI. sleeve shape,
VII. garment material.

For both business reasons (for example, detecting logos on clothing is critical information for a brand) and their significance in identifying a specific garment, these features are the most important, as they can prevent or lessen the number of duplicates or similar products on the market. The characteristics of each feature, as well as the reasons for their relevance, are examined in more depth later in the paper. The underlying concept is to automate the classification process of these 7 variables via the use of machine learning algorithms and computer vision, however, there are other facets to such a task that pose specific research questions and challenges. In reality, every clothing feature may ultimately need a distinct classification approach, with methods ranging from segmentation [2][3], image retrieval [4], and ML (Machine Learning) approaches like DL [5][6] to be used in conjunction. Moreover, these algos should be applied to the under-researched field of fashion, also they must be fine-tuned for the particular domain. Furthermore, for automatization to be beneficial for

commercial objectives, it should have accuracy similar to that of manual approaches. So, the subject of automated classification of garment traits turned out to be a difficult application of computer vision & ML methods, requiring much research. The objective of this article is to explain work that was built by authors in order to provide efficient solution to aforementioned issue via the use of computer vision software system, as well as to examine the obstacles that arose during the creation of that system, feature by feature.

The following is the outline for the paper. Firstly, Section II presents an assessment of the current state-of-the-art for each of the approaches utilized, followed by a discussion of related works. Finally, Section III provides an overview of the developed system, with a particular focus on fashion product identification, which is then explained in depth. Section IV presents experimental data, and Section V finishes the study with a review of the paper's primary accomplishments and directions for future research.

2. LITERATURE REVIEW

Clothing fashion is a reflection of an aesthetic trend in wearing that lasts for a length of time. Recognizing the fashion period of a piece of clothing is significant for both individuals & industry. The fashion-time recognition issue is translated into the clothes-fashion classification problem, which is based on the notion that clothing fashion varies year by year. The detection and classification of human clothes is a significant difficulty in the field of computer vision. The accurate & speedy recognition of clothing categories in videos or images aids in the development of intelligent social interaction systems that are by human thought processes. It may also be used to automatically categorize large amounts of data in a network and to enhance the trend analysis of fashion trends as well as other trends. AI target identification & detection technology has benefited enormously from the development of DCNNs (deep convolution neural networks), which significantly enhances their accuracy.

Bhatnagar et al. [7] published a study in 2017 in which they provided an advanced model for the categorization of fashion article images, which is still in use today. They developed DL architectures based on convolutional neural networks to categorize images from the Fashion-MNIST dataset using the training data. The authors have presented three distinct CNN topologies, and they have made use of batch normalization & residual skip connections to make the learning process easier and faster. Using the Fashion-MNIST benchmark dataset as a testbed, their model produces outstanding results. According to the results of the comparisons, their suggested model achieves an improvement in accuracy of about 2 percent over existing advanced systems in the literature.

Seo & Shin [8], The GoogLeNet architecture was pre-trained on the ImageNet dataset, and the final tuning was done on their fine-grained fashion dataset depending upon design characteristics, as described in 2018. This will help to compensate for the limited size of the dataset & to minimize the training time required for the dataset. The average final test accuracy is 62 percent after tenfold experiments are conducted.

Liu et al. [9], It was suggested in 2019 that a new Hierarchical Classification Model (HCNN) depends upon neural networks be used to classify data. The coarse category-fine category classification layer is constructed up from the lower layer of the network to the higher layer of the network, starting at the lower layer. To classify apparel photos, the experiment demonstrates the requirement of a hierarchical classification framework. For the first time,

hierarchical CNN is used in this work to try to categorize clothing datasets. The proposed model is a classifier with knowledge-embedded features that transmits some hierarchical information. They used the Fashion-MNIST dataset to create HCNN, with VGGNet serving as the underlying foundation. When compared to the base model without hierarchy, the findings reveal that the loss has been minimized and the accuracy has been improved.

According to this study, Di [10], published in 2020, employed 4 neural network models to the categorization of apparel images in the Fashion-MNIST dataset: CNN, Fully Connected Neural Network (FCNN), MobileNetV2, & MobileNetV1, and the results were compared. MobileNetV2, the deep learning network that outperforms the others in the area of picture recognition & classification, is most effective. Compared to CNN & FCNN, the testing findings demonstrate that MobileNet is a more efficient method of apparel picture classification, with an accuracy rate of 91 percent.

A DL-based network was proposed by Zhang et al. [11], published in 2020, that achieves precise human body segmentation by fusing multi-scale convolutional features in a fully convolutional network, & afterward feature learning as well as fashion classification are performed on segmented parts without taking into account the impact of the image background. To test the validity of the proposed model, 9,339 fashion pictures from eight consecutive years are used in the tests. The results suggest that both body segmentation and fashion classification approaches are successful in identifying trends in fashion.

In this study, Kayed et al. [12], in 2020, introduced CNN-based LeNet-5 architecture to train parameters of CNN on the Fashion MNIST dataset. According to experimental data, the LeNet-5 model attained an accuracy of more than 98 percent.

The goal of Dhariwal et al. [13], who published their findings in 2020, was to investigate and assess the feasibility and usefulness of leveraging huge amounts of unlabeled data to train DL models. They compare the performance of these models to the performance of models that were generated using labeled data in the previous section. To be more specific, they compare fully supervised DL with 2 DL approaches that need no prior supervised training. Their pre-trainings are focused on a feature clustering technique called DeepCluster, as well as rotation as a task for self-supervision. The DeepFashion dataset is used for this comparison. The results of their experiments have demonstrated that when compared to fully supervised models, unsupervised pre-training may achieve similar classification accuracy (a difference of 1-4 percent) on image classification.

A broad number of fields have benefited from the widespread use of DL in recent years. When it comes to computer vision tasks like image classification, object detection, & segmentation, DL-based algorithms have shown exceptional results in recent years. Training deep neural networks on labeled data has been the most successful method of achieving much of this accomplishment. In general, the greater the amount of labeled data that is fed into a DL model, the more accurate the model is expected to become. But labeling takes a long time and is often hard to do properly. It is possible to find a vast quantity of unlabeled data in the fashion & e-commerce industries. These data without labels are very valuable, and there is an urgent need to capitalize on them. When it comes to tackling real-world challenges, CNNs are the kind of DNN (Deep Neural Networks) that provide the most rigorous results. On their e-commerce sites, fashion companies have employed CNN to tackle a variety of challenges, including clothing identification, clothing search, and clothing recommendations, among others. Image

classification is a critical step in all of these methods. To be sure, classifying garments are difficult to work since clothes have a wide range of characteristics and the depth of classification is quite complex. As a result of this convoluted depth, distinct classes have very similar features, making the classification issue very difficult to resolve.

2.1 RESEARCH METHODOLOGY

This section is discussing the proposed methods to solve the followed identified problem definition for fashion image classification and recommendation.

2.1.1 Problem Definition

Image classification is considered to be the most fundamental issue in computer vision, and it has a wide range of practical applications, including video and photo indexing [14]. Although the challenge of recognizing a visual item from a picture is a pretty simple one for a person to solve, it is very difficult for a computer program to execute the same task with the same degree of accuracy as a human [15]. To correctly recognize and categorize the images, the method must be invariant to a large number of alterations & iterations. For example, various lighting conditions, varied size and perspective changes, deformations, and occlusions may all cause the algorithm to anticipate the erroneous picture class, causing the image to be classified incorrectly. The use of deep neural networks has increased in current years, and they have been applied to a wide range of issues, with excellent results. For example, in image classification [16], image segmentation [17], computer vision challenges [18], as well as natural language processing difficulties [19], CNNs have shown excellent results. Bayesian Belief Networks [20]& Hidden Markov Models [21] have also been used to classify images using attributes depending upon color, grey level, depth, texture, and motion [22].

Incorrect labeling may result in unfavorable search results, which may harm the consumer experience. To avoid such instances, businesses must ensure that their products are accurately classified across categories like gender, product category, product kind, and so on. An effective and reliable picture classification model may also assist e-commerce in automating online product management, lowering operating costs, and eliminating anomalies in the system.

In this work, we explore the idea of classifying Fashion product images (small) with fine-tuned deep transfer learning-based pretrained neural networks.

2.1.2 Proposed Methodology

In this research paper, a novel fine-tuned transfer learning model is proposed for classifying fashion products images and recommending a fashion product. This is done by following the data gathering of fashion product images, data preprocessing in that data filtration is done, then feature extraction and classification is done using the XceptionNet transfer learning model. Finally, fashion products are recommended based on these classification results of images.

The detailed process of this proposed methodology is described below:

1. Data Pre-processing: Data filtration

Even though the Fashion Product dataset is cleanly labeled, it is very uneven when it comes to the application of ML techniques to picture classification (Fig. 2). To solve the data imbalance,

we use a data filtering procedure for data pre-processing.

Our classification objective is achieved by the use of three picture attributes, which are concatenated Gender & Master Category, Article Type, & Sub-Category, chosen from among the several image attributes available. There are a large number of classes that include just a few images. It is vital to balance the datasets to prevent such misclassifications from occurring. To do this, we eliminate the classes that contain less than 500 pictures. Table 1 depicts a snapshot of the filtered dataset as it now exists. Apart from this, incorrectly labeled images are discarded from the entire dataset.

Product category data	Before Filtering		After Filtering	
	No. of Classes	Total images	No. of Classes	Total images
Gender+masterCategory	45	44419	12	44018
subCategory	45	44419	15	40847
articleType	142	44419	23	34719

TABLE 1 :BALANCING OF DATA BY FILTERING CLASSES THAT HAVE LESS THAN 500 IMAGES

2. Feature Extraction

It is a process of extracting relevant features from fresh data by using the representations learned by a prior network. We simply layer a new classifier on top of the pretrained model, which will be trained from scratch, to repurpose the feature maps that were previously learned for the dataset. The concept of transfer learning is derived from a fascinating phenomenon in which several deep neural networks trained on real pictures acquire features that are similar to one other over time. In the first few levels, there is texture, corners, edges, & color blobs to be found. In contrast to being unique to a particular data set or task, such initial-layer features seem to be generic in the sense that they apply to a wide range of data sets and activities. There seems to be no difference between the precise cost function and the fashion product picture data-set that produces the typical characteristics seen on the first layers. These initial-layer characteristics are referred to as general features, and they may be transferred to a fashion product picture data-set for learning. The XceptionNet model does the feature extraction in this case by using some different convolution layers. The Xception architecture consistes of 36 convolutional layers , which serve as the network's feature extraction foundation.

3. Data Splitting

One of the most straightforward and likely most prevalent strategies for splitting a large dataset is to randomly pick a portion of the dataset. For example, if the size of our dataset is between 100 and 100,000 records, we divide it into three parts in the ratio of 60:20:20. That example, 60 percent of the data will be used for the Training Set, 20 percent will be used for the Validation Set, and the remaining will be used for the Test Set. Training datasets are very necessary to prevent incorrect predictions. Validation datasets should be utilized to check for

overfitting, and if overfitting is detected, dropout should be used to lessen the overfitting.

Product category	Training (60%)	Validation (20%)	Testing (20%)	No. of Classes	Total images
Gender+masterCategory	28171	8804	7043	12	44018
subCategory	26141	8170	6536	15	40847
articleType	22220	6944	5555	23	34719

TABLE 2: DATA SPLIT INFORMATION

4. Image Classification

Image classification is the process of determining what an image depicts and how it was created. An image classification model is developed to recognize different types of pictures. For example, you could want to train a model to identify photos of three distinct sorts of fashion product categories so that she can distinguish between them.Well, Transfer learning is effective for image classification tasks because Neural Networks learn in a more complicated fashion as they gain experience. Therefore, the more you go through the network, the more image-specific features are learned. It is necessary to understand the underlying architecture of the pre-trained model to properly appreciate the model construction process. The XceptionNet transfer learning model is used to classify the images in this example.It is an extension of the Inception architecture in which the fundamental Inception modules are replaced with depth-wise separable convolutions instead of the basic Inception modules. XCeption is a high-performance architecture that is founded on two essential principles:

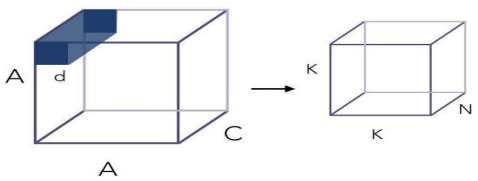
- Depthwise Separable Convolution
- Shortcuts between Convolution blocks as in ResNet

4.1 **Depthwise Separable Convolution**

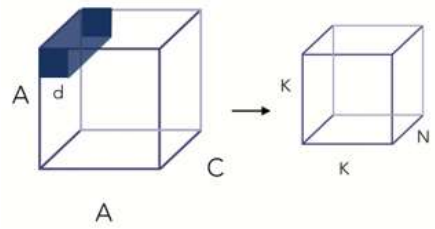
A new kind of convolution called Depthwise Separable Convolutions is being promoted as a time-saving alternative to traditional convolutions.

4.2 Limits of convolutions:

Let's start with a look at convolutions and how they work. Convolution is a time-consuming and costly procedure. Let us use the following example:



The input picture has a certain no. of channels C, for example, 3 for a color image. It also has a certain size A, for example, 100 * 100 pixels. We then apply a convolution filter of size d*d, for example, 3*3, to it. The following diagram illustrates the convolution process:



Then, how many operations are involved?

That is, given 1 Kernel, the following is true:

$$K^2 \times d^2 \times C$$

After convolution, K denotes the resultant dimension, which is dependent on the padding used (for example, padding "same" will result in A = K).

As a result, for N Kernels (depth of convolution):

$$K^2 \times d^2 \times C \times N$$

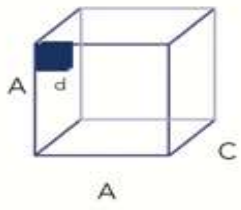
To reduce the cost of such processes, depthwise separable convolutions have been developed. Those phases are further subdivided into two major categories:

- Depthwise Convolution
- Pointwise Convolution

4.3 The Depthwise Convolution

As a 1st stage, we apply convolution of size d×d×1 to the data, rather than the usual convolution of size d×d×C, as in the previous step. In additional words, we don't do convolution calculation over all of the channels, nonetheless rather one channel at a time.

As an example of the depthwise convolution process, consider the following:

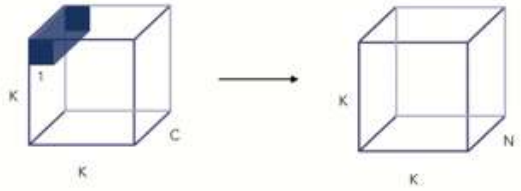


This results in 1st volume of the size K×K×C, rather than K×K×N, as was the case before. Indeed, thus far, we have only performed convolution operation on a single kernel /filter of the convolution, rather than on a set of N kernels /filters. This brings us to the second phase in our process.

4.4 Pointwise Convolution

It performs classical convolution of size 1×1×N across the K×K×C volume, as perceived in the following diagram. As previously stated, this enables the creation of volume of form K×K×N.

To demonstrate the Pointwise Convolution, consider the following example:



Implementation of the XCeption- It goes through entering flow first, then middle flow, which is frequent 8 times, & lastly exit flow before being sent to the ultimate destination. It should be noted that batch normalization is applied after each Convolution & Separable Convolution layer. XCeption provides an architecture that is composed of Depthwise Separable Convolution blocks with Maxpooling, all of which are connected by shortcuts in a similar way that ResNet implementations are linked together. XCeption differs from other convolutions in that Depthwise Convolution is not followed by Pointwise Convolution, nonetheless rather order is reversed. When compared to a depth of similar depth in traditional convolutions, this design produces a smaller no. of trainable parameters. The model summary for the proposed Fine-tuned XceptionNet is defined in Table III, which can be found below.

Model: "model"

Layer (type)	Output Shape	Param #
image_input_lr (InputLayer)	[(None, 96, 96, 3)]	0
xception (Functional)	(None, None, None, 2048)	20861480
flatten (Flatten)	(None, 18432)	0
dense_layer_1 (Dense)	(None, 1024)	18875392
dropout (Dropout)	(None, 1024)	0
dense_layer_2 (Dense)	(None, 1024)	1049600
dropout_1 (Dropout)	(None, 1024)	0
output_layer (Dense)	(None, 23)	23575

Total params: 40,810,047
Trainable params: 40,755,519
Non-trainable params: 54,528

TABLE 3:MODEL SUMMARY OF XCEPTIONNET

Proposed Algorithm

1. Import Library
2. Reading Data: Fashion product image small data,Sample Images, Create labels
3. Data preprocessing: Data filtration, Discard incorrectly labeled images
4. Filter classes that have less than 500 images
5. Features Extraction

6. Train, Validation, and Test Distribution
7. Transfer Learning CNN: XceptionNet Features
8. Train model with training & validation dataset
9. Then test the trained model if it is not forecasting well means try to change the training dataset also retrain it again.
10. Predict the original dataset using the trained model and store the predicted value in a Numpy file after loading the trained model & original dataset
11. Now remove the image using the Numpy file.
12. Test Evaluation
13. Classification of Fashion product categories
14. Performance evaluate
15. Recommend the fashion product

3. Result Analysis and Discussion

An enormous dataset, generated by the expanding e-commerce business, is ready to be scraped and investigated. Additionally, we have a variety of label attributes describing a product that was manually input throughout the cataloging process in addition to professionally captured high definition product images. In addition, we have descriptive text that provides information about the product's attributes. It is preferable to use a smaller picture dataset that comprises a similar set of images nonetheless at a lower resolution to save computing resources & run-time. Some hyperparameters are set to this fine-tuned model with dense and dropout layer in that Batch size is 32, Epochs are 10, with categorical cross-entropy Loss function, adam optimizer & Learning rate is 0.0001. Additionally, a pre-trained model is readily available using the Keras Python library, which is written in Python.

A. Dataset

There are 44000 goods with category labels & images in the Fashion Product Images (Small) dataset. There is a total of 10 columns, which include the following: id, gender, subcategory, masterCategory, baseColour, articleType, year, season, use, as well as productDisplayName. Each product is assigned a unique ID, such as 42431. There is a map that shows all of the goods organized by style. csv. The picture for this product may be obtained from images/42431.jpg, which is located here. It is possible to download photos in jpg format, with each image having a size of 80 by 60 pixels and three color channels. In styles, each product picture may be linked to its associated information based on its numeric id. csv. We discover 44,419 instances where picture IDs are matched with the IDs in the metadata file after merging the two sets of data. We've also revealed some of the most important product categories, as well as their display names in several styles, to make things even easier to get started with. csv. Figure 1 shows an example of a picture for articleType.

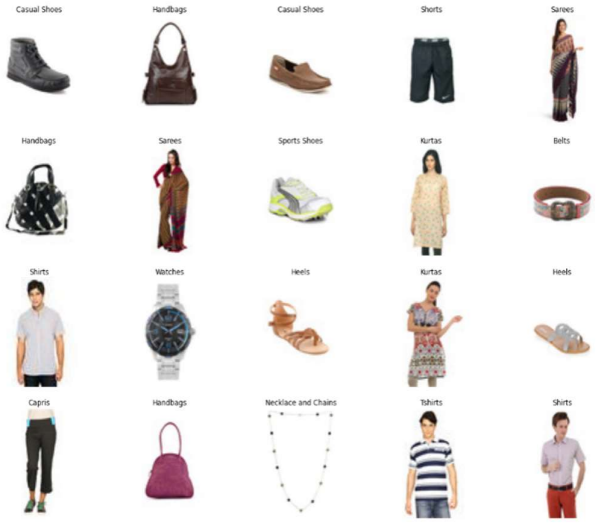


Figure. 1: Sample images of articleType

B. Performance Evaluation Metrics

The classification report presented in Table IV makes this error pattern much more obvious. Precision, Recall, & F1-Score are all measures that are often employed in classification tasks. These are the ones that are described as:

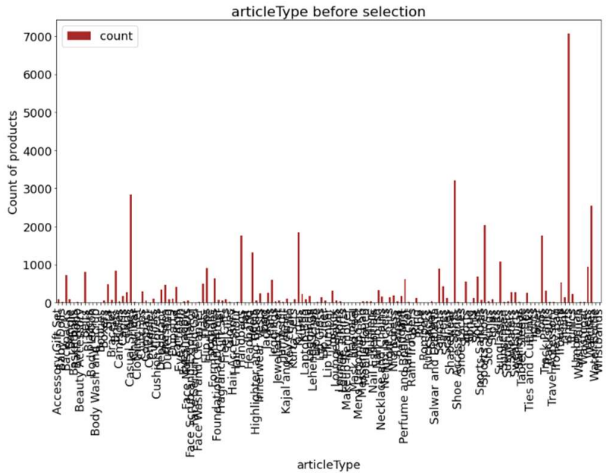
$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

$$Recall = \frac{T\,ruePositives}{TruePositives + FalseNegatives}$$

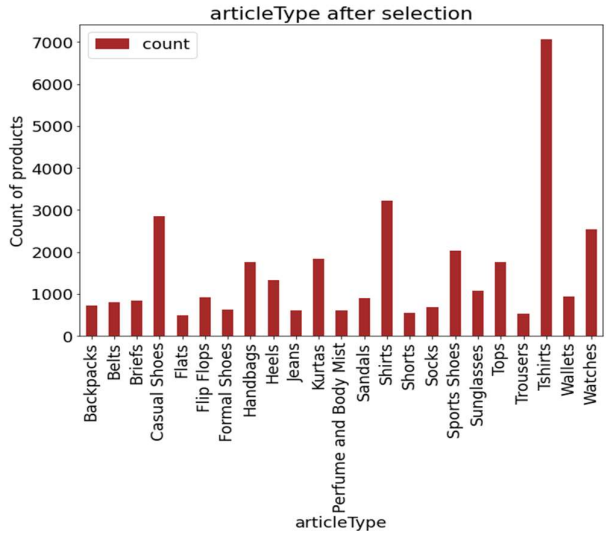
$$F1Score = 2 * \frac{P\,recision * Recall}{Precision + Recall}$$

C. Results Evaluation

We classify products using three schemes: concatenated masterCategory & Gender, articleType, and Sub Category; they are neatly labeled and have no missing variables. This section discusses the results of these three categories of data but the representation is done only for articleType category.



(a) before data filtration



(b) After data filtration

Figure. 2: articleType categories result for data filtration

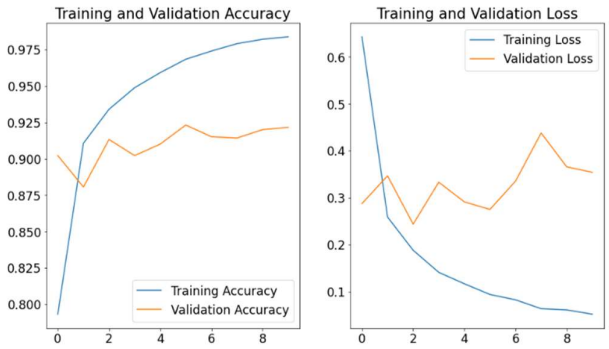


Figure. 3: articleType accuracy graph

The accuracy & loss for both training & validation is shown in Fig. 3. An interpretable accuracy

metric is used to evaluate the algorithm's overall performance measurably. The loss value of a model reflects how poorly or well it performs after each iteration of optimization is carried out, and is expressed as a percentage. During training on the training dataset, the Training Accuracy parameter represents how well the model performed. Valid Accuracy describes how well the model performs when compared to the validation dataset. We train the model with training data and then assess its performance on both training & validation data sets (evaluation metric is accuracy). Training accuracy is around 98 percent, but the validation accuracy is approximately 91 percent. Similarly, the training loss is less than 0.1, however, the validation loss is around 0.35.

	precision	recall	f1-score	support
Backpacks	0.97	0.94	0.96	121
Belts	1.00	0.98	0.99	124
Briefs	0.99	0.98	0.99	128
Casual Shoes	0.81	0.92	0.87	452
Flats	0.59	0.24	0.35	90
Flip Flops	0.85	0.90	0.87	159
Formal Shoes	0.88	0.84	0.86	101
Handbags	0.96	0.98	0.97	279
Heels	0.75	0.92	0.82	220
Jeans	0.92	0.91	0.91	87
Kurtas	0.95	0.96	0.96	282
Perfume and Body Mist	0.98	0.98	0.98	109
Sandals	0.89	0.72	0.79	156
Shirts	0.96	0.97	0.97	515
Shorts	0.96	0.99	0.98	82
Socks	1.00	0.99	1.00	107
Sports Shoes	0.94	0.84	0.89	334
Sunglasses	1.00	1.00	1.00	143
Tops	0.84	0.69	0.76	294
Trousers	0.89	0.92	0.91	90
Tshirts	0.93	0.97	0.95	1113
Wallets	0.98	0.98	0.98	143
Watches	1.00	1.00	1.00	426
accuracy			0.92	5555
macro avg	0.92	0.90	0.90	5555
weighted avg	0.92	0.92	0.92	5555

TABLE 4: ARTICLETYPE CLASSIFICATION REPORT

Table IV displayed the classification report for all classes of the articleType category. There are a total of 23 classes for this category that represent their evaluation results for each class including precision, recall, f1-score. The classification accuracy for this articleType category is 0.92 which is equal to the weighted average accuracy whereas the macro average is 0.90.

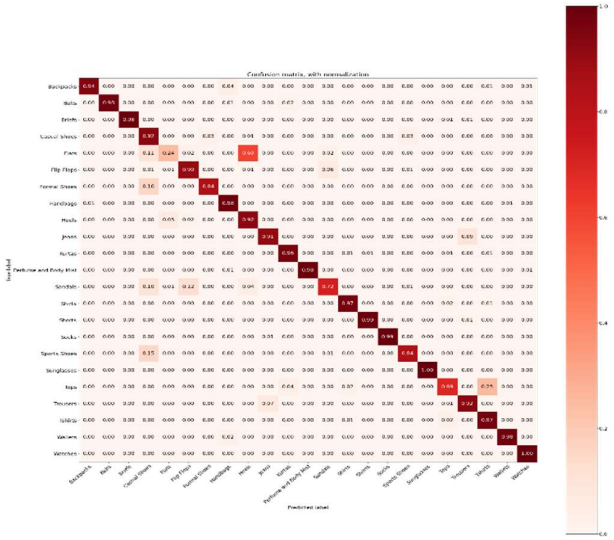


Figure. 4: articleType confusion matrix from classification

We utilize the confusion matrix as well as the precision-recall curve to assess the classification predictions to get a more in-depth understanding of the model's performance. Figure 4 depicts simply the results of the articleType classification model, which is meant to serve as an example. Although the general accuracy of product classification has been reported to be high, the confusion matrix reveals that there have been some predicted misclassifications as well. 60 percent of flats are misclassified as heels, 15 percent of sports shoes are misclassified as casual shoes, and 25% of tops are misclassified as T-shirts, to name a few examples. As a result of the similarities between these classes, these misclassifications fall within a tolerable tolerance range.

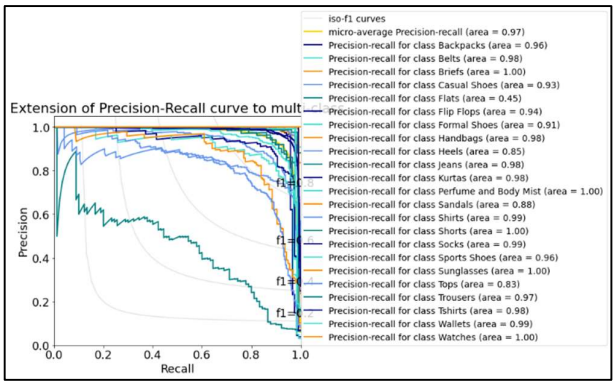


Figure. 5: Precision-Recall Curve from classification by articleType

Figure 5 depicts precision-recall-curve for articleType under consideration. It provides a clearer demonstration of the predictions provided by the model for every class. This is supported by the confusion matrix plot, which shows that precision-recall for class 4 has an area of 0.45 and belongs to the Flats class.

Product Category	Training accuracy	Validation accuracy	Testing accuracy
Gender + masterCategory	98.47	92.07	91.72
subCategory	99.75	98.35	98.04
articleType	99.27	92.16	91.93

TABLE 5: XCEPTIONNET MODEL PERFORMANCE

The findings of the XceptionNet-based product classification model have been reported in Table V. Throughout product hierarchy, the model was evaluated at every granularity level. When a model is trained to forecast sub-categories of products, we notice high overall performance, with the model accuracy being the best, at 98.04 percent. Even though many of the articles are fairly similar to one another, the model's accuracy in predicting the articleType is 91.93 percent, which is worth mentioning.

4. CONCLUSION

As a new approach that has attracted the attention of researchers in artificial intelligence, deep learning has developed, and it has the potential to be utilized in a variety of business domains to improve performance. Deep learning, namely CNN, is being utilized in the fashion industry to classify garment images. for classification. In this study, we offer an advanced model for the categorization of fashion article pictures that are based on ML techniques. CNN-based deep learning architectures were trained on the Fashion-product-images-small dataset to identify pictures for classification. The authors have presented a deep XceptionNet transfer learning architecture that makes use of batch normalization & residual skip connections to make the learning process easier and more efficient. We get outstanding results with our model when applied to the benchmark dataset Fashion-product-images-small. Comparing our proposed model to the existing state-of-the-art systems in literature, we find that it has an enhanced testing accuracy of around 3.75 percent for Gender + masterCategory, 1.49 percent for subcategory, and 5.3 percent for articleType. In most traditional classification tests, the Xception architecture outperformed the VGG-16, ResNet, and Inception V3 architectures.

5. FUTURE RESEARCH

With the fast developments in processing power & ML methods, we may even be able to use a Generative Adversarial Network (GAN) to come up with novel ideas for fashion accessories, therefore reducing our reliance on human creativity and increasing our independence from it. Even though training the GAN model is very difficult, GAN models in the fashion industry have the potential to provide considerable economic value in near future.

6. REFERENCES

[1] Y. Jing *et al.*, “Visual search at pinterest,” *Proc. ACM SIGKDD Int. Conf. Knowl.*

Discov. Data Min., vol. 2015-August, pp. 1889–1898, 2015, doi: 10.1145/2783258.2788621.

[2] S. Abdulateef and M. Salman, "A Comprehensive Review of Image Segmentation Techniques," *Iraqi J. Electr. Electron. Eng.*, vol. 17, pp. 166–175, 2021, doi: 10.37917/ijeee.17.2.18.

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000, doi: 10.1109/34.868688.

[4] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000, doi: 10.1109/34.895972.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[6] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," 2016.

[7] S. Bhatnagar, D. Ghosal, and M. H. Kolekar, "Classification of fashion article images using convolutional neural networks," in *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 2017, pp. 1–6, doi: 10.1109/ICIIP.2017.8313740.

[8] Y. Seo and K. Shin, "Image classification of fine-grained fashion image based on style using pre-trained convolutional neural network," in *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018, pp. 387–390, doi: 10.1109/ICBDA.2018.8367713.

[9] Y. Liu, G. Luo, and F. Dong, "Convolutional Network Model using Hierarchical Prediction and its Application in Clothing Image Classification," in *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*, 2019, pp. 157–160, doi: 10.1109/ICDSBA48748.2019.00041.

[10] W. Di, "A comparative research on clothing images classification based on neural network models," 2020, doi: 10.1109/ICCASIT50869.2020.9368530.

[11] X. Zhang *et al.*, "Deep learning based human body segmentation for clothing fashion classification," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 7544–7549, doi: 10.1109/CAC51589.2020.9327016.

[12] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," 2020, doi: 10.1109/ITCE48509.2020.9047776.

[13] S. Dhariwal, Y. Liu, A. Karali, and V. Vlassov, "Clothing Classification using Unsupervised Pre-Training," 2020, doi: 10.1109/MCNA50957.2020.9264287.

[14] M. H. Kolekar, "Bayesian Belief Network Based Broadcast Sports Video Indexing," *Multimed. Tools Appl.*, vol. 54, no. 1, pp. 27–54, Aug. 2011, doi: 10.1007/s11042-010-0544-9.

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, 2017, doi: 10.1145/3065386.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014, doi: 10.1109/CVPR.2014.81.

[18] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet

and the Impact of Residual Connections on Learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.

[19] D. Ghosal, S. Bhatnagar, M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "IITP at SemEval-2017 Task 5: An Ensemble of Deep Learning and Feature Based Models for Financial Sentiment Analysis," in *11th International Workshop on Semantic Evaluation (SemEval2017)*, 2018, pp. 899–903, doi: 10.18653/v1/s17-2154.

[20] M. H. Kolekar and S. Sengupta, "Bayesian Network-Based Customized Highlight Generation for Broadcast Soccer Videos," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 195–209, 2015, doi: 10.1109/TBC.2015.2424011.

[21] M. H. Kolekar and S. Sengupta, "Hidden Markov model based video indexing with discrete cosine transform as a likelihood function," in *Proceedings of the IEEE INDICON 2004. First India Annual Conference, 2004.*, 2004, pp. 157–159, doi: 10.1109/INDICO.2004.1497728.

[22] M. Kolekar, S. Talbar, and T. Sontakke, "Texture Segmentation using Fractal Signature," *IETE J. Res.*, vol. 46, pp. 319–323, 2015, doi: 10.1080/03772063.2000.11416172.

First A. Author (Fellow, IEEE) and all authors may include biographies if the publication allows. Biographies are often not included in conference-related papers. Please check.

Second B. Author,photograph and biography not available at the time of publication.