Neural Network Wardrobe Consultant based on ResNet50

ANMOL SINGH[1], MD. SABA HASMI[2], SANDEEP KUMAR[3]
[1, 2, 3] *School of Engineering and Technology, Sharda University, Greater Noida, India*

Abstract— People began to pay more attention to fashion, which is considered a popular form of aesthetic expression, as their level of living increased. Anything with greater visual appeal will always compel others to gravitate towards it. The evolution of the fashion business may be attributed to human propensity. Nonetheless, the abundance of clothing selections within an online stores has created additional difficulties for clients in selecting the appropriate ensemble. Accordingly, we presented a customised Fashion Based on user input, the recommender system in this study provides recommendations to the user. Instead than relying on the user's prior purchases and history like traditional systems do, this project aims to produce suggestions based on an image of a product submitted by the user. This is because people frequently find products that pique their interest when they see something they like. To analyse the photos from the DeepFashion dataset and provide the final suggestions, we employ CNN powered by a nearest neighbour recommender. Using the newest styles for both clothing and accessoriesand accessories have now evolved into an essential part of daily living. It boosts self-esteem and aids in someone's appealing. People are becoming more conscious of their looks, which is causing them to follow new trends and increase demand for trendy items. With the growing demand for stylish products, a large number of people are entering the fashion and textile sectors. However, there will also be fashion modifications on this trip. There are several types of fashion lifecycles, depending on how long a trend lasts. The typical lifespan of a certain fashion defines its "normal" lifetime. As a result, fads and fashions gain popularity quickly but do not stay forever. These cycles of fashion are called "fads." Fashion that is considered "classic" is designed to last a longer period of time. If a businessperson possesses analytical expertise, it becomes simpler for them to forecast future changes in fashion. Customers will benefit from it in this way as well. The correct fashion guidance will be given to the customer to ensure that their money is invested appropriately, and Appropriate business choices will be made to avoid suffering undue losses.. Fashion photo analysis is useful for fashion suggestions and retrieval, among other fashion-related applications. Numerous websites, like Heuritech, Wgsn, and Trendzoom, among others, forecast next fashion trends. While the study's goal remains the same, we will be taking it a step farther in this area. Because of the additional research prompted by this development in fashion forecasting, the efficacy of the website for the fashion forecasting analyzer should increase in the upcoming years.

Index Terms- Application of CNN Algorithm, ResNet50, Application of Image Processing, Content-Based Filtering for Recommendation System, CNN Algorithm-Based Feature Extraction, and Outfit Recommendation System, Ecommerce, and Applications of Machine Learning.

I. INTRODUCTION

These days, a user's decision while purchasing fashion goods online is mostly influenced by the way the things seem [1]. Nevertheless, the user's choices are also influenced by intangible elements that they are unable to see in the picture, including the clothing's substance and quality. As a result, the main goals of our work are to introduce feature extraction, alongside the image, the user's apparent and unseen preferences.

Numerous attempts have been made to utilise product images for style guidance in the previous few years. Most existing methods use pre-trained convolution models to incorporate the entire fashion picture into a fixed-length global image. [2]–[5]. However, most algorithms do not produce plausible visual explanations and disregard the user's preferred visual style.

In order to find fine-grained visual preferences and weakly oversee the learning of more comprehensive user preferences, Chen et al. [6] developed an attention model across several pre-segmented picture areas. In addition, Hou et al. [7] took into account the pre-segmented picture areas matching to the characteristics were generated from the readily understood semantic space derived from the semantic attributes of the product. could be extracted. Using the attention network, they were able to illustrate user preferences for various semantic attributes. These techniques give visual explanations and record the user's visible choices. However, because to the limits of image data, they are unable to provide invisible explanations and fail to capture hidden preferences.

In Top-N interpretable suggestions, user reviews—which are rich in auxiliary information—have drawn a lot of attention [8], [9]. Seo et al. specifically created a convolutional neural network with both local and global attention. [10] to extract more intricate information of persons and items from reviews and to explain with attention weight. Zhang and colleagues utilised sentiment analysis at the phrase level. [11] to explicitly represent user thoughts and product attributes, maintaining a high prediction accuracy while still producing suggestions that make sense. Chen et al. [12] suggested a hierarchical co-attention selector and an encoder-selector-decoder architecture to properly use the links between the explanation task and the recommendation task.

Anything that is more aesthetically pleasing to the eye will always catch the attention of humans. Over time, the fashion business has developed as a result of human propensity. As recommender systems proliferate across several industries, Modern technology is being purchased by retail companies to improve their operations. The fashion industry has been around for millennia and is only going to continue growing. Women interact with a larger variety of items and are more likely to be connected with style and fashion, which makes decision-making challenging. For modern families, it has become a crucial part of life because people are frequently assessed by the clothes they wear. Furthermore, clothing retailers want their customers to browse their whole selection before cutting it down to a few favourites, which is not possible to do by simply going into a store. The fast advancement of Internet technology has led to a surge in demand in the fashion retail industry for intelligent clothes. The earliest application of 3D technology was to create and imitate virtual storefronts. An interactive virtual fitting system using 3D technology was suggested, and it directed people in finding appropriate clothing based on their choices for accessories. Customers are frustrated by their lengthy searches when an online business lacks recommendation tools. recognised clothes based on the outfit photos seen in frontal perspective. The garments have been recommended to the buyers according to their previous purchasing habits. Consequently, recommendation systems in online stores assist consumers in locating appropriate and pertinent clothing as well as the fashion retail sector in making money from sales. In this work, we provide a unique method for personalised fashion recommender based on user preferences. More specifically, we focus on fashion components and develop a framework that accepts a single input image and outputs a list of the top-5 recommended pieces of apparel.

The remaining portions of the paper are arranged as follows. The Fashion Recommender system's associated work is covered Section 2 covers the suggested structure; Section 3 covers the proposed work's summary and conclusion; and Section 4 presents the work's conclusion.

II. RELATED WORK

In the mid-1990s, The online internet age saw the initial proposals for recommendation technologies. [5]. [6] Developed CRESA that synthesised textual characteristics, visual elements, and user paying attention to develop a profile for clothes and offer recommendations. [7] Made recommendations using images from fashion publications. To generate suggestions, Numerous components from the pictures were recovered, including the material, collar, sleeves, and so on. [8]. In order to meet the various needs of various users, an intelligent clothing recommender system based on fashion and aesthetic principles is studied in [9]. In [10], apparel and customer reviews were used to provide outfit recommendations. In order to provide recommendations, [11] took into account the local environment and the history of apparel and accessories.

Two fundamental methods of product recommendation may be distinguished between Collaborative and content-based filtering are two types of recommender systems. System for recommending [12][13].

The previous method is dependent on past user-item interactions, meaning that the user's previous the former makes recommendations based on item descriptions and user profiles, while the latter on item rating history. Lately, Deep Neural Collaborative Filtering Framework based on learning[14], which extends matrix factorization technique is often applied in systems that employ collaborative filtering and recommendering. Current Systems that provide recommendations take into account factors including past purchases, user reviews, and product characteristics, chronological data, etc., but one important aspect that the current system overlooks the visual aspect of the products is taken into consideration in ranking and recommendation methods. [15] incorporated users' views with visual signals and presented a scalable factorization approach using large, real-world datasets.

[16][17] offers an extensive analysis of deep learning-based recommendation systems. made product suggestions utilising convolutional neural networks for computer vision tasks including object segmentation, object classification, and object identification. Knowledge databases and keyword mapping are the main sources of recommendations for most e-commerce companies. But given that product descriptions differ from seller to seller, this turned out to be ineffective[18]. The significant The level of subjectivity in fashion articles led to subpar performance for generic recommender systems. [28]. Our approach uses picture data of the item and shows that it is reasonable to rely on the visual elements to deliver very appealing and closely matched item suggestions with the user's interests and preferences. Additionally, the suggested method lessens the chilly start issue with traditional collaborative filtering-based recommender systems [18][19].

We discuss two potential problems facing the e-commerce sector in the future. One has to do with the challenges merchants have when attempting to submit photos of their goods to an online marketplace and the consequent manual tagging that follows. It is no longer visible in search results due to the categorization issues [9]. For products to show up in search results, e-commerce platforms require merchants to provide photos of their goods and tag them with relevant labels. Due to the participation of humans, this process is error-prone. A misclassified product could not show up in search results, which could result in fewer or no sales at all. Machine learning-based classification methods are capable of correctly identifying photos and informing vendors on the proper tagging [10][8]. Another problem is that if clients know exactly what they want but lack the right keywords, they might run into a backlog while making their orders. His lack of knowledge about the appropriate keywords makes this task less appealing to customers. On an e-commerce website, the keywords for the product that a buyer wants to purchase are usually input. The user's keywords are compared to product labels in the database by the search algorithm, which then gives relevant and helpful results [11]. The consumer then submits her order after seeing the search results for the item she wishes to buy. When using a text-based search, it is assumed that the user is knowledgeable about the product and knows which terms to type into the search toolbar. Naturally, this isn't always the case. We encounter many things in our everyday lives that we are totally ignorant about. We are unable to do an online product search as a consequence. A visual search can be used to overcome this issue [12]. Customers won't need to enter anything to locate comparable products by clicking on an image of a product thanks to the use of an image-based search algorithm. It makes advantage of low-level picture representation in some photo classification [13]. Generally speaking, these methods view a picture as a collection of minute features such as size, shape, colour, texture, and so on. There are several methods for obtaining image numeric representation. A widely used technique is known as "bag-of-words," wherein every vector component of an image representation is associated with a visual word [14]. A numeric vector representation of an image is produced by counting the frequency of visual words inside the image. A predictive model may be made using the numerical representation of such a picture.

A representation of mid-level characteristics is used in other categorization methods [15]. An picture is seen by the aforementioned approaches as a pre-trained database of generic objects. A lot of works portray pictures and videos at the mid-level using characteristics. The authors suggested a method for creating mid-level visual cues for picture categorization. The results of several binary classifiers combine to create a vector that depicts a picture. [13][16]. We employed an object class hierarchy to train these binary classifiers. In a similar vein, the Classmen descriptor leverages the output of many badly trained object classifiers to describe images. These days, it's getting more and more common to describe images using neural networks, particularly deep neural networks. These techniques make it possible to extract certain neural networks feature layers, which subsequently serve as representations of numerical images. Neural network-based image processing has been the focus of several investigations [8–13]. Convolutional neural networks (CNNs) are used in deep learning approaches that have helped CBIR make major strides in image identification recently. Regarding [survey paper 8], an extensive analysis of deep learning and Convolutiional Network is provided]. Despite not being new technology, CNNs and neural networks have only just demonstrated competitive performance for problems like the ILSVRC2012 image classification challenge, thanks to early triumphs like LeNet. The notable decrease in previously stagnating mistake rates has led to a spike in interest in CNNs. A number of novel designs and methodologies, including Inception Architecture, Deep Residual Networks (ResNets), and GoogleNet, have been presented. Additionally, neural networks have been used in metrics learning for visual search and picture similarity estimates. Two new datasets have been released as part of recent articles. We were able to extract apparel from 161,638 distinct photographs by utilising the MVC Dataset and the DeepFashion Dataset, which contains 8,00,000 annotated real-life images for claasification.

//The use of machhine learning algos to ResNet systems is still relatively new, yet it is a quickly developing topic. Various machine learning techniques have been investigated in relation to ResNet security:

• Identifying anomalies in ResNet transactions: Machine learning methods, like as neural networks, have been used to find out-of-the-ordinary patterns or behaviours. These irregularities could be an early warning indicator of security risks or data breaches.

• Predictive Analytics: Predictive models, frequently based on past ResNet data, are able to foresee potential security issues, enabling preventative measures to be performed.

Gaps and Improvement Needs Identification:

Several holes and opportunities for development exist despite the advancements made in machine learning and ResNet security for e-commerce:

• Scalability: The problem of scaling ResNet systems, particularly in the e-commerce industry where the volume of data and transactions can be significant, has not yet been adequately addressed in many studies that have already been conducted.

• Privacy Enhancements: While ResNet offers a safe and unchangeable ledger, protecting patient privacy, especially in shared e-commerce contexts, is still difficult.

III. METHODOLOGY

This research presents a model that leverages Transfer learning from ResNet module, which is trained on imageNet dataset, and Convolutional Neural Network (CNN). As seen in figure 1, when the model has been trained, an inventory is chosen for the purpose of producing suggestions, and a database is constructed for each item in the inventory. The closest neighbour algorithm finds the most relevant goods based on the given image, and recommendations are generated.

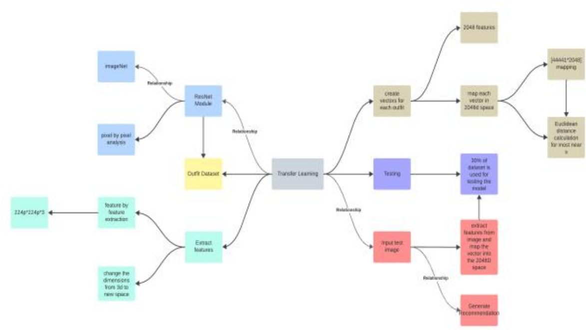

Fig. 1 Block schematic of the suggested system.

3.1 Training the Model

The model is trained using fastai and transfer learning using the ResNet50 module, which is a pre-trained model on the imageNet dataset, after the data has been pre-processed. A deep learning model called ResNet50 architecture is employed in computer vision applications. It is a design for a Convolutional Neural Network (CNN) that can accommodate thousands or even hundreds of convolutional layers. It may be located in the Python keras library.

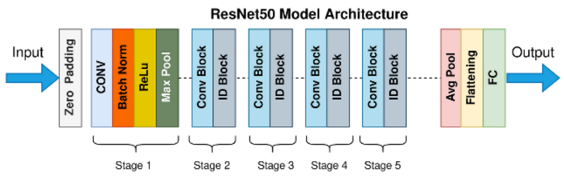

Fig. 2 ResNet50 Architecture[23]

3.2 Extracting the features

Our dataset consists of 44000 outfit images that are of size 224px224p. Since we are using transfer learning on ResNet module, we will not analyze the images pixel by pixel but by feature by feature.

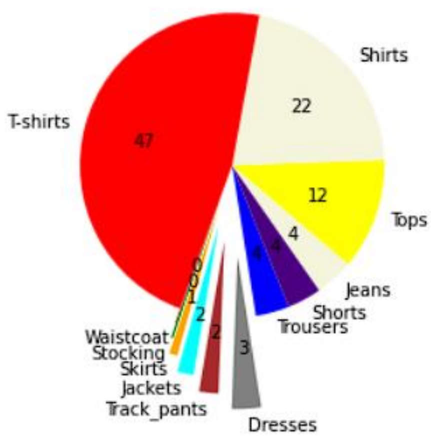

Fig. 3 Different outfits in our dataset.

Our dataset consists of various variety of outfits that are collected from different e-commerce websites and magazines.

A Residual Neural Network (ResNet) with many layers is applied to each image, extracting 2048 features in total. be it color, design, position of patterns, etc. After feature extraction a vector for that image is generated that is mapped in a 2048 dimensional space. In this way for 44000 images that many vectors are generated and mapped into the space.

Training of the model was done using 80:20 ratio of the dataset. The residual neural networks are then used to categorise and create embeddings from the inventory, and the resulting output is then utilised to produce recommendations [22][23]. In Figure 4, an example set of inventory data is shown.

Fig. 4 Sample Inventory

3.3 Recommendations Generation

The Nearest Neighbour technique is used by our suggested model to provide the suggestion. Each test input of user is passed through the layers of ResNet module and a vector for that image is also generated and that vector is mapped into the same space where training image's vectors were mapped. The database's top 5 recommendations are selected, and their images are shown. This allows us to identify the closest neighbours for the input image that was supplied. The Euclidean Distance measure is the similarity metric used in this paper. [25].

Algorithm 1: Extrraaction of images
Input: Self made Data set from wesites
Output: Extrraaction of the required data

Procedure:
n<-no_of_samples_of_a_given_data set
df<-be_a_data_frame_with_n_no_of_samples.
Df1<-
a_data_frame_that_contains_samples_belonging_to
12 classes
X=[] is an empty array to store features
Y=[] is empty to store labels of the image
for i in range(0,n):
    if i ranging in df1.index:
  image,label=extract_image and class from inventory
  pre-process image
      append_features to X
      append_label to Y
    end_if
end_for

We utilised the Euclidean distance function or cosine similarity to determine how similar the photos were to one another.

$$\cos\theta = \frac{\sum_{i=1}^{n} a_i * b_i}{\sqrt{\sum_i^n a_i^2} * \sqrt{\sum_{i=1}^{n} b_i^2}} \quad \text{.....(1)} \quad [50]$$

where b represents the characteristics of the extracted pictures of the categorised product and a is the vector of the test image features.

$$Euclidean\ distance(u,p) = \sum_i (u_i - p_i)^2 \quad (2)[52]$$

where p is the location of the displayed vector from the training data picture and u is the vector's position.

4. Experiment and Results:

The limitations of getting the accuracy of pre-trained ResNet model and complexity of analysing the images pixel by pixel can be addressed by the application of the transfer learning approach.

Thus, we pre-train the categorization models using the 44,441 garment image-rich DeepFashion dataset. The networks are trained and assessed using the gathered dataset. The training results show that the model has outstanding accuracy with minimal error, loss, and a decent f-score, as shown in Table 1.

	precision	recall	f1-score	support
0.0	0.75	0.13	0.22	94
1.0	0.53	0.20	0.29	50
2.0	0.61	0.76	0.67	132
3.0	0.77	0.56	0.64	640
4.0	0.78	0.48	0.59	94
5.0	0.00	0.00	0.00	29
6.0	0.00	0.00	0.00	2
7.0	0.50	0.52	0.51	347
8.0	0.67	0.36	0.47	50
9.0	0.68	0.43	0.53	106
10.0	0.70	0.88	0.78	1441
11.0	0.00	0.00	0.00	2
accuracy			0.68	2987
macro avg	0.50	0.36	0.39	2987
weighted avg	0.68	0.68	0.66	2987

Fig. 5 The Suggested Approach's Performance for Various Epocs.

We prepared the model for 100 epocs and in the final epochs our model reached the precision of 92%.

	precision	recall	f1-score	support
0.0	0.57	0.66	0.61	146
1.0	0.56	0.21	0.30	87
2.0	0.87	0.97	0.92	1952
3.0	0.00	0.00	0.00	19
4.0	0.92	0.73	0.82	783
accuracy			0.86	2987
macro avg	0.59	0.51	0.53	2987
weighted avg	0.86	0.86	0.85	2987

Fig. 6 Performance in the Final Epochs.

Confusion Matrix in the final epochs I given in figure 7.

$$\begin{bmatrix} 96 & 5 & 40 & 0 & 5 \\ 32 & 18 & 34 & 0 & 3 \\ 21 & 3 & 1890 & 0 & 38 \\ 11 & 0 & 6 & 0 & 2 \\ 8 & 6 & 197 & 0 & 572 \end{bmatrix}$$

Fig. 7 Confusion Matrix in the Final Epochs

The test set for evaluation consists of personally shot images in the real world and random clothes photos from the internet [26] [27]. These diverse images are utilised to assess the proposed system.

We can draw the conclusion that, although relying solely on visual inputs, this strategy is resilient and successful based on the simulation findings of the experiment. Figures 4 and 5 depict the ensembles that our method produced for the supplied input picture.

Fig. 8 Sample User Input Image

Fig. 9 Outputs Predicted By the Model

As we can see our model proposed the related outfits which are very close to the input clothes. Also for other user inputs the model recommended outfits that are very close to the input. In the above example the western outfit input showed similar western clothes.

IV. COMPARISONS

For the ResNet model's testing and training, we employed a custom pattern dataset made up of 44,000 photos. All the pictures were downsized to 256 by 256. Using an Adam, the optimizer We trained the network over a 100 period batch size.

It had a 0.5 dropout rate. The dataset was split up into five classes based on the patterns that consumers and designers preferred. Fig. 14 displays the accuracy chart for the CNN model.

While representing the categorization outcome, the accuracy graphic also successfully conveys the preferences of users and designers.

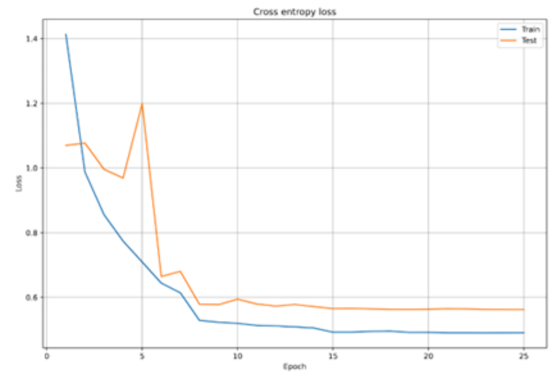

Fig. 10 Cross Entropy Loss Graphic of the CNN Approach

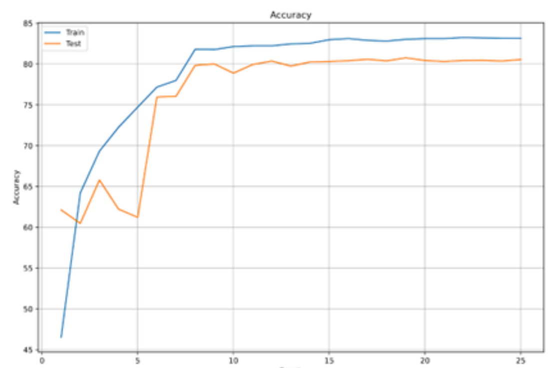

Fig. 11 Accuracy Graphic of the CNN Approach

Final train loss: 0.49078181050938263
Final test loss: 0.5623244762858923
Final train accuracy: 83.15740966796875
Final test accuracy: 80.55555725097656

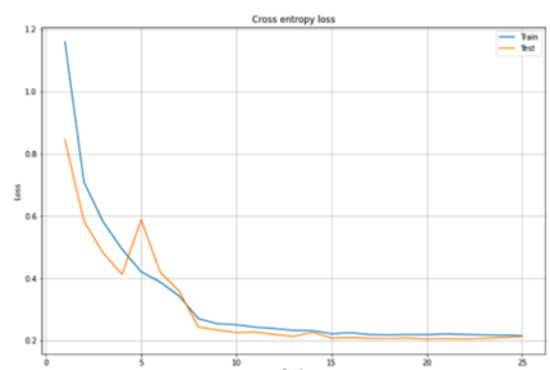

Fig. 12 An Illustration of the Suggested Model Approach's Cross-Entropy Loss

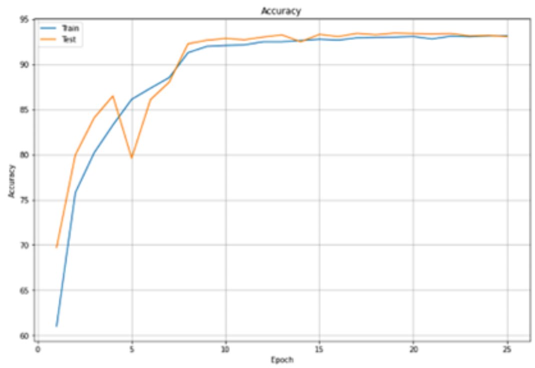

Fig. 13 Accuracy Graphic of the Proposed Model Approach

Final train loss: 0.21561731184287183
Final test loss: 0.21340482519829976
Final train accuracy: 93.14814758300781
Final test accuracy: 93.05555725097656

Results of performance measures are shown for all classes in Table II. 93.055% is the total accuracy attained. Using the following formulas, metrics for accuracy, precision, recall, and f1-score are calculated:
Accuracy : (TP+TN) / (TP+TN+FP+FN)……….(1)
Precision : TP / (TP+FP)………………………..(2)
Recall : TP / (TP+FN)…………………………..(3)
F1-score : 2×Precision×Recall / (Precision+Recall) …………………………...(4) [12]
Where TP is true positive, TN is true negative, FP is false positive, FP is false positive, and FN is false negative.

Table II Results of performance metrics

Classes	Precision	Recall	F1-score
1	0.963	0.895	0.927
2	0.763	0.708	0.734
3	0.788	0.817	0.802
4	0.765	0.962	0.852
5	0.817	0.793	0.804

5.1. The suggested model is contrasted with the current model:

The current CNN model and our suggested approach are contrasted in Table II. In this case, the suggested single CNN-Model is also doing a good job of predicting patterns and designs that are comparable. With the suggested model, accuracy is improved by 12.50%.

Table II Model comparison with other available models

Model	Accuracy
Existing System [51]	80.55%
Proposed Model	93.055%

Table III. Analyzing Loss and Model Accuracy for Outfit Detection

Model	Loss	Accuracy
Existing CNN model	0.5623	80.55%
ResNet	0.2134	93.055%

CONCLUSION

We presented a ResNet-based recommendation system in this study that uses an image stream. from the user and generates recommendation based on pattern matching using ResNet module. Our own pattern dataset, which included 44,000 photos, was utilised to train and evaluate the ResNet model. We assessed the suggested model's performance using metrics for f1 score, recall, precision, and total accuracy. 93.05% is the overall accuracy, 82.00% is the precision, 83.50% is the recall, and 82.30% is the f1-score. Recommendations seem to give suitable patterns even if the photos are categorised based on the preferences of users and designers rather than the images' inherent similarity.

Given that they express the preferences of the users, customer ratings and purchases may likewise be seen as a source for the recommendation engine. Once more, the recommendation system is trained with the feedback that is received at specific intervals. The recommendation system that receives input from the feedback will function better.

Our study has repeatedly focused on the crucial problem of e-commerce data security, highlighting the significance of safe patient-physician communication and the protection of billing and medical information.

FUTURE SCOPE

The future is bright for deep learning-based fashion recommendation systems that use transfer learning on the ResNet50 module. The following are some important domains in which they may be further enhanced and implemented:

1. More individualised suggestions: The majority of user choices used by current fashion recommendation systems are style, size, and colour. But more data and sophisticated machine learning techniques will make it feasible to create systems that can recognise and accommodate a greater variety of user preferences, including budget, occasion, and personal taste.

2. Multimodal recommendations: Fashion recommendation systems can make use of text descriptions and user evaluations in addition to visual characteristics. This can enhance the precision and applicability of suggestions.

3. Real-time suggestions: Most current fashion recommendation systems are batch-based, which means that they provide recommendations based on an inventory and user preference snapshot. But with edge computing and 5G coming along, it will be feasible to create systems that can provide suggestions in real time depending on the user's current environment, which includes their location, the time of day, and the weather.

The following are some particular uses for transfer learning on the ResNet50 module in deep learning-based fashion recommendation systems:

1. Personalised online and in-store shopping experiences: Fashion companies may utilise these technologies to provide their clients personalised online and in-store buying experiences. For instance, once a buyer tries on an article of apparel, the system can suggest matching accessories or shoes.
2. Fashion forecasting: By using these techniques, fashion firms may predict trends and create new items that are likely to be well-liked by their target market.
3. Fashion styling: These platforms allow fashion bloggers and stylists to design customised looks for their clientele.
4. Social media fashion: By using these algorithms, social media companies may provide fashion-related material to its users according to their preferences.

REFERENCES

[1] F. Cordier, H. Seo, and N. Magnenat-Thalmann Made-to-measure technologies for an online clothing store 2003 IEEE Computer Graphics and Applications23(1), pp. 38–48

[2] R. Li, K. Zou, X. Xu, Y. Li, and Z. Li Research of interactive 3D virtual fitting room on web environment 2011 in Proceedings of the Research of Interactive 3D Virtual Fitting Room on Web Environment. Symposium on Computational Intelligence and Design (Iscid'11),Hangzhou, China, 2011.1 pp. 32–35

[3] W. Zhang, B. Bo, M. Chu, J. Liu, and N. Yee, “Real-time clothes comparison based on multi-view vision,” in Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras, pp. 1–10, Stanford, Calif, USA, September 2008

[4] C. Limaksornkul, D. N. Nakorn, O. Rakmanee, and W. Viriyasitavat, “Smart closet: Statistical based apparel recommendation system,” in Proceedings of the 3rd ICT International Senior Project Conference (ICT-ISPC '14), pp. 155–158, March 2014.

[5] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, “Recommending and evaluating choices in a virtual community of use,” in Proceedings of the Conference on Human Factors in Computing Systems, pp. 194–201, Denver, Colo, USA, May 1995.

[6] V. D. M. Ernani, E. A. Nogueira, and D. Guliato, Content-Based Filtering Enhanced by Human Visual Attention Applied to Clothing Recommendation 2015 in Proceedings of the International Conference on TOOLS with Artificial Intelligence, pp. 644–651

[7] T. Iwata, S. Watanabe, and H. Sawada, Fashion coordinates recommender system using photographs from fashion magazines in Proceedings of the22nd International Joint Conference on Artificial Intelligence (IJCAI '11), pp. 2262–2267, Catalonia, Spain, July 2011.

[8] D. Sha, D. Wang, X. Zhou, S. Feng, Y. Zhang, and G. Yu, An approach for clothing recommendation based on multiple image attributes 2016 Web-Age Information Management

[9] S. Liu, T. V. Nguyen, J. Feng, M. Wang, and S. Yan, "Hi, magic closet, tell me what to wear! 2012," ACM Multimedia, 43, pp. 1333-1334

[10] X. Hu, W. Zhu, and Q. Li, "HCRS 2014 A hybrid clothes recommender system based on user," in Proceedings of the Ratings And Product Features, pp. 270–274

[11] C. Limaksornkul, D. N. Nakorn, O. Rakmanee, and W. Viriyasitavat, "Smart closet: Statistical-based apparel recommendation system," in Proceedings of the 2014 3rd ICT International Senior Project Conference (ICT-ISPC '14), pp. 155–158, March 2014.

[12] M. Sridevi and Dr. R.Rajeswara Rao A Comprehensive Survey of Personalized Recommender Systems International Journal of Computer Science and Technology, Vol. 5, Issue 3, July -Sept 2014, ISSN : 0976-8491.

[13] M.Sridevi, Dr. R.Rajeswara Rao and Dr.M.Varaprasad Rao, A Survey on Recommender System 2016 International Journal of Computer Science and Information Security,14(5),

[14] X. He, L. Liao, H. Zhang, et al. Neural collaborative filtering. In WWW2017, pages 173–182, 2017.

[15] R. He and J. McAuley. Vbpr: Visual bayesian personalized ranking from implicit feedback. In AAAI, pages 144–150, 2016.

[16] S. Zhang, L. Yao, and A. Sun. Deep learning based recommender system: A survey and new perspectives. arXiv:1707.07435, 2017.

[17] I. Goodfellow, Y. Bengio, and A. Courville. Deep learning. MIT, 2016.

[18] L. Chen, F. Yang, and H. Yang. Image-based product recommendation system with convolutional neural networks, 2017.

[19] Praveen P., Rama B(2020). "An Optimized Clustering Method To Create Clusters Efficiently" 2020 Journal Of Mechanics Of Continua And Mathematical Sciences , ISSN (Online) : 2454-7190 15(1), pp 339-348

[20] T. Zhang, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, D. Wang, G. Liu, and X. Zhou, ''Feature-level deeper self-attention network for sequential recommendation,'' in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 4320–4326.

[21] R. He and J. McAuley, ''VBPR: Visual Bayesian personalized ranking from implicit feedback,'' in Proc. AAAI, Feb. 2016, pp. 144–150.

[22] Y. Hu, X. Yi, and L. S. Davis, ''Collaborative fashion recommendation: A functional tensor factorization approach,'' in Proc. 23rd ACM Int. Conf. Multimedia, 2015, pp. 129–138.

[23] X. Song, F. Feng, J. Liu, Z. Li, L. Nie, and J. Ma, ''NeuroStylist: Neural compatibility modeling for clothing matching,'' in Proc. ACM Multimedia Conf., 2017, pp. 753–761.

[24] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou, ''Graph contextualized self-attention network for session-based recommendation,'' in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 3940–3946.

[25] X. Chen, H. Chen, H. Xu, Y. Zhang, Y. Cao, Z. Qin, and H. Zha, ''Personalized fashion recommendation with visual explanations based on multimodal attention network,'' in Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2019, pp. 765–774.

[26] M. Hou, L. Wu, E. Chen, Z. Li, V. W. Zheng, and Q. Liu, ''Explainable fashion recommendation: A semantic attribute region guided approach,'' 2019, arXiv:1905.12862. [Online]. Available: http://arxiv.org/abs/1905.12862

[27] C. Xu, P. Zhao, Y. Liu, J. Xu, V. S. S. S. Sheng, Z. Cui, X. Zhou, and H. Xiong, ''Recurrent convolutional neural network for sequential recommendation,'' in Proc. World Wide Web Conf. (WWW), 2019, pp. 3398–3404.

[28] P. Zhao, H. Zhu, Y. Liu, J. Xu, Z. Li, F. Zhuang, V. S. Sheng, and X. Zhou, ''Where to go next: A spatio-temporal gated network for next POI recommendation,'' in Proc. AAAI Conf. Artif. Intell., vol. 33, Jul. 2019, pp. 5877–5884.

[29] S. Seo, J. Huang, H. Yang, and Y. Liu, ''Interpretable convolutional neural networks with dual local and global attention for review rating prediction,'' in Proc. 11th ACM Conf. Recommender Syst. (RecSys), 2017, pp. 297–305.

[30] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, ''Explicit factor models for explainable recommendation based on phrase-level sentiment analysis,'' in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.(SIGIR), 2014, pp. 83–92.

[31] Z. Chen, X. Wang, X. Xie, T. Wu, G. Bu, Y. Wang, and E. Chen, ''Co-attentive multi-task learning for explainable recommendation,'' in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 2137–2143.

[32] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, ''Image-based recommendations on styles and substitutes,'' in Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2015, pp. 43–52.

[33] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, ''Visually-aware fashion recommendation and design with generative image models,'' in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2017, pp. 207–216.

[34] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis, ''Learning fashion compatibility with bidirectional LSTMs,'' in Proc. ACM Multimedia Conf., 2017, pp. 1078–1086.

[35] X. Chen, H. Xu, Y. Zhang, J. Tang, Y. Cao, Z. Qin, and H. Zha, ''Sequential recommendation with and data mining,'' in Proc. 18th ACM Int. Conf. Web Search Data Mining, 2018, pp. 108–116.

[36] Y. Zhang, ''Explainable recommendation: Theory and applications,'' 2017, arXiv:1708.06409. [Online]. Available: http://arxiv.org/abs/1708.06409

[37] Y. Zhang and X. Chen, ''Explainable recommendation: A survey and new perspectives,'' 2018, arXiv:1804.11192. [Online]. Available: http://arxiv.org/abs/1804.11192

[38] Y. Zhang, Y. Zhang, and M. Zhang, ''SIGIR 2018 workshop on Explainable recommendation and search (EARS 2018),'' in Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2018, pp. 1411–1413.

[39] J. McAuley and J. Leskovec, ''Hidden factors and hidden topics: Understanding rating dimensions with review text,'' in Proc. 7th ACM Conf. Recommender Syst., 2013, pp. 165–172.

[40] Y. Tan, M. Zhang, Y. Liu, and S. Ma, ''Rating-boosted latent topics: Understanding users and items with ratings and reviews,'' in Proc. IJCAI, vol. 16, 2016, pp. 2640–2646.

[41] L. Zheng, V. Noroozi, and P. S. Yu, ''Joint deep modeling of users and items using reviews for recommendation,'' in Proc. 10th ACM Int. Conf. Web Search Data Mining (WSDM), 2017, pp. 425–434.

[42] C. Chen, M. Zhang, Y. Liu, and S. Ma, ''Neural attentional rating regression with review-level explanations,'' in Proc. WWW, 2018, pp. 1583–1592.

[43] Y. Tay, A. T. Luu, and S. C. Hui, ''Multi-pointer co-attention networks for recommendation,'' in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp. 2309–2318.

[44] L. Wu, C. Quan, C. Li, Q. Wang, B. Zheng, and X. Luo, ''A context-aware user-item representation learning for item recommendation,'' ACM Trans. Inf. Syst., vol. 37, no. 2, pp. 1–29, Mar. 2019.

[45] P. Li, Z. Wang, Z. Ren, L. Bing, and W. Lam, ''Neural rating regression with abstractive tips generation for recommendation,'' in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2017, pp. 345–354.

[46] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. De Rijke, ''Explainable outfit recommendation with joint outfit matching and comment generation,'' IEEE Trans. Knowl. Data Eng., early access, Mar. 19, 2019, doi: 10.1109/TKDE.2019.2906190.

[47] J. Ni and J. McAuley, ''Personalized review generation by expanding phrases and attending on aspect-aware representations,'' in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, 2018, pp. 706–711.

[48] J. Tang, Y. Yang, S. Carton, M. Zhang, and Q. Mei, ‘‘Context-aware natural language generation with recurrent neural networks,’’ 2016, arXiv:1611.09900. [Online]. Available: http://arxiv.org/abs/1611.09900 VOLUME 8, 2020 68745 Q. Wu et al.: VTJEI Fashion Recommendation.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017.

[50] A. L. Maas, A. Y. Hannun, and A. Y. Ng, ‘‘Rectifier nonlinearities improve neural network acoustic models,’’ in Proc. ICML, vol. 30, no. 1, 2013, p. 3.

[51] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, ‘‘Attentive collaborative filtering: Multimedia recommendation with Item- and component-level attention,’’ in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2017, pp. 335–344.

[52] Q. Liu, S. Wu, and L. Wang, ‘‘DeepStyle: Learning user preferences for visual recommendation,’’ in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2017, pp. 841–844.

[53] X. He and T.-S. Chua, ‘‘Neural factorization machines for sparse predictive analytics,’’ in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2017, pp. 355–361.

[54] D. P. Kingma and J. Ba, ‘‘Adam: A method for stochastic optimization,’’ 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[55] G. Karypis, ‘‘Evaluation of item-based top-n recommendation algorithms,’’ in Proc. 10th Int. Conf. Inf. Knowl. Manage., 2001, pp. 247–254.

[56] K. Järvelin and J. Kekäläinen, ‘‘IR evaluation methods for retrieving highly relevant documents,’’ ACM SIGIR Forum, vol. 51, no. 2, pp. 243–250, Aug. 2017.

[57] Y. Zhang, Q. Ai, X. Chen, and W. B. Croft, ‘‘Joint representation learning for Top-N recommendation with heterogeneous information sources,’’ in Proc. ACM Conf. Inf. Knowl. Manage. (CIKM), 2017, pp. 1449–1458.

[58] Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan, “Deep domain adaptation for describing people based on fine-grained clothing attributes,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5315-5324.

[59] N. Yarahmadi Gharaei, C. Dadkhah and L.Daryoush," Content-based Clothing Recommender System using Deep Neural Network," 2021 26th International Computer Conference, Computer Society of Iran (CSICC), 2021,pp.1-6,doi:10.1109/CSICC52343.2021.9420544.