

Computer vision-based multi-class classification of garments using a three-level hierarchical approach

Asta Kvedaraite¹, Neda Buineviciute¹ Agne Paulauskaite-Taraseviciene¹

¹ Kaunas University of Technology, Studentu g. 50, Kaunas, LT-51368, Lithuania

Abstract

Manually collecting and measuring garment data can be a complex and time-consuming process, including garment classification, which can be a difficult task even for humans. Computer vision algorithms can be trained to classify clothes by analysing large amounts of data and identifying patterns and features specific to each class. A 3-level hierarchical garment classification model has been proposed in the paper, which classifies garments into 3, 8 and 21 classes. The model has been tested with three deep learning architectures LeNet5, AlexNet and sequential CNN model. The results obtained show that the hierarchical approach has the greatest advantage when classifying garments into three and eight classes, and allows an improvement of up to 28%.

Keywords

hierarchical model, multi-class problem, garment classification, deep learning.

1. Introduction

The garment industry is a significant sector of the retail market, encompassing both new and second-hand clothing sales. According to industry estimates, the global apparel market was valued at over \$1.53 trillion in 2022 and is expected to continue to grow in the coming years. In addition to new clothing sales, the second-hand clothing market has been rapidly expanding in recent years, driven by consumer demand for sustainable and affordable fashion option. There has been a significant growth in online platforms that sell second-hand clothes, such as Poshmark, ThredUP, Depop, and Vestiaire Collective, and others.

Manual collection and measurement of garment data can be a time-consuming and complicated process, particularly for large volumes of garments. Measuring each garment individually for size, color, material, and other attributes can be a tedious and error-prone task that requires significant time and resources. The rise of online platforms for second-hand clothes has also created new opportunities for using machine learning and artificial intelligence technologies in the garment industry. For example, computer vision algorithms can be used to automatically classify and tag second-hand garments based on their style, brand, and other attributes. This can help to improve the accuracy and efficiency of the online marketplace and provide a better experience for buyers and sellers alike.

However, classifying garments can be more challenging than identifying simple attributes like color or size, as there are many factors to consider such as style, fabric type, and etc. Additionally, the number of relevant classes for garment classification can vary depending on the context and purpose of the classification [21], [24], [25].

While classifying clothes can be a difficult task, even for humans, machine learning algorithms can be trained to identify patterns and features that may be overlooked or difficult for a human to distinguish. For example, a machine learning algorithm can analyze thousands of images of clothes and learn to recognize common patterns and features that are unique to each clothing category [1], [2]. However, it is important to note that the results of machine learning algorithms are highly data-

28th Conference on Information Society and University Studies (IVUS'2023), May 12, 2023, Kaunas, Lithuania

EMAIL: asta.kvedaraite@ktu.edu (A. Kvedaraite); neda.buineviciute@ktu.edu (N. Buineviciute); agne.paulauskaite-taraseviciene@ktu.lt (A. Paulauskaite-Taraseviciene)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

dependent. It is therefore essential to have a reliable training dataset and to continuously test and improve the algorithm to ensure its accuracy and efficiency. It has been observed that the classification of garments into 10 or more categories can indeed be complex and there is certainly room for improvement. Therefore, a hierarchical classification approach based on deep learning, which decomposes the classification process into several depth classification steps, may be beneficial for better classification accuracy.

Hierarchical classification is a method of organizing and classifying objects or data into a hierarchy, based on their relationships and similarities. Hierarchical machine-learning based approach that involves decomposing the classification process into multiple stages [5],[16]. This can be particularly useful for handling complex and variable data, such as images of garments with a wide range of features and variations. By breaking the classification process down into multiple stages, it may be possible to achieve higher levels of accuracy and efficiency [6], [17].

There are several different hierarchical approaches and techniques that can be used for image classification, including traditional machine learning methods and convolutional neural networks (CNNs). In hierarchical multi-label classification, each level of the hierarchy is represented by a local neural network, which is trained to classify the data into a specific set of labels [7]. The local neural networks at each level of the hierarchy are connected, and the output of one level is the input of the next. This allows the classification process to be decomposed into several steps, allowing the model to handle complex relationships between labels and increase accuracy [8].

One advantage of hierarchical classification is that it allows for a more intuitive and natural way of organizing and understanding data. By grouping similar objects or concepts together and nesting them within broader categories, it can be easier to understand the relationships and connections between different pieces of information. Another advantage of hierarchical classification is that it allows for a more flexible and dynamic approach to categorization. By organizing data into a hierarchy, it can be easier to locate and access specific pieces of information, as the search can be narrowed down to increasingly specific levels of the hierarchy [9],[10].

2. Related works

Garment images can have complex textures, patterns, and colors, which can make it difficult for humans to determine their class or category. However, artificial intelligence (AI) can be trained to classify garment images accurately and efficiently. The most commonly used dataset for garment classification is Fashion MNIST [4], and many experiments have been carried out to find AI-based models with high accuracy [3]. Moreover, the use of deep learning models, has revolutionized the field of garment classification and has opened up new possibilities for automation in the fashion industry [14],[15]. One of these models is VGG19 is a deep neural network that has 19 layers, including 16 convolutional layers and 3 fully connected layers. It is a powerful model that has achieved high accuracy on various computer vision tasks, including image classification. [11]. Several experiments have been performed using the VGG19 model on the Fashion-MNIST dataset for garment classification. The model has shown promising results, with high accuracy in identifying different classes of clothing items, including classification tasks based on garment type [12] or pattern [13].

More simple convolutional neural network (CNN) architectures like AlexNet and LeNet can also achieve high accuracy results for garment classification [21]. The AlexNet model [18] trained on the Fashion-MNIST dataset with 9 garment classes achieved an accuracy of 92%, but the accuracy could vary depending on the specific implementation and training process [19] therefore can vary from 90 to 93%. LeNet-5 has been widely used as a benchmark model in the field of computer vision and can achieve an F1-score accuracy of 98% when classifying garments into 10 categories [2]. Other approaches, such as using a CNN with SVM or SVM+HOG or shallow convolutional neural networks, also showed good performance with accuracies ranging from 86.53% - 94.04% [17], [18]. The improved HSR-FCN can be used for garment classification tasks, achieving high accuracy results in a shorter training time by learning from deformed garment images, and the average accuracy of the original network model R-FCN increases by about 3% to 96.69% [22]. In this paper, the authors, inspired by mask R-CNN (for segmentation) and YoloV2 for faster object detection, proposed models for detecting the location of an object with the probability of a class, and deforming the contour of the

initial boundary marker according to the shape of an object [23]. The experimental results of 11-class classification task show that such model performs better on the Deepfashion2 dataset (mAP 86.86%) compared to other recent deep learning models.

3. Methodology

Two datasets were used in this study: (1) a set of 890 manually labeled photos of clothes, all of which are on a hanger or mannequin; (2) Fashion MNIST - a popular dataset used for training and testing machine learning models in the field of computer vision. It consists of a collection of 70,000 grayscale images of size 28x28 pixels, which are divided into 60,000 training images and 10,000 testing images [4].

Our hierarchical classification methodology involves three levels of classification process, starting with the three most distinguishable groups at the highest hierarchy level (first), then dividing them into 8 smaller categories, and then further dividing the categories at second hierarchy level into 21 categories. At the first level, there are three main classes: “Top”, “Bottom” and “Full wear”. Each class is subdivided into more specific classes, i.e. “Top” is subdivided into “Shirts”, “Blouses” and “Sweaters”. The categories of second level are divided into very specific subsets of garments, which are likely to be the most mixable because the garments are very similar (e.g. class “Shirts” is divided into “Shirts-U-Neck”, “Shirts-V-Neck” and „High neck“) (see **Figure 1**).

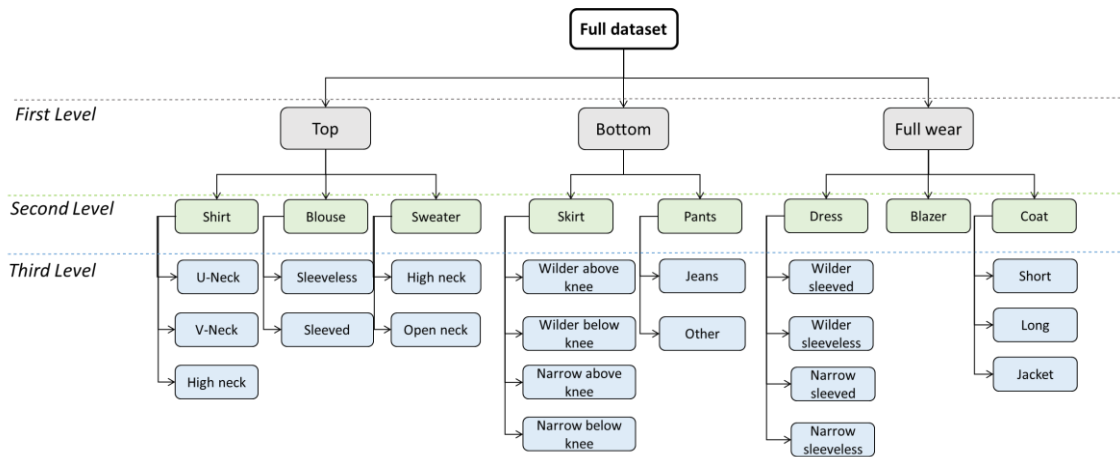


Figure 1: 3 –level hierarchical garment classification approach

Each category was manually selected, therefore the size of each category varies. In total there are 21 categories and 2012 images of garment in general that are used to train models which predict categories **Figure 2**. This hierarchical classification system allows us to accurately and efficiently classify a wide variety of items or data into groups based on their characteristics, with increasing levels of detail and specificity as we move down the hierarchy.

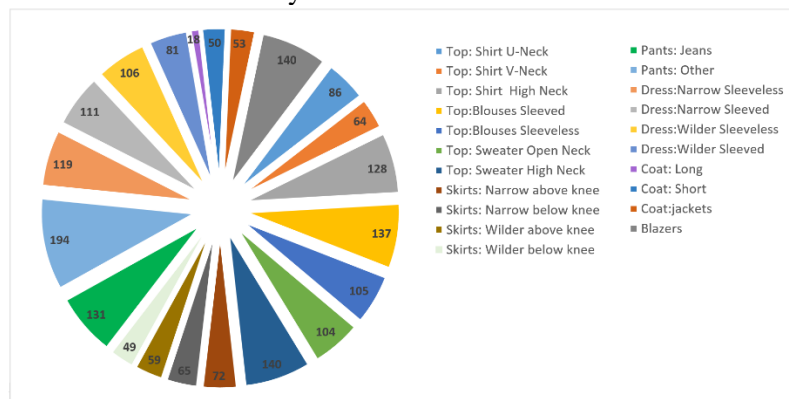


Figure 2: Distribution of testing dataset among 21 categories

Different deep learning architectures have been used for experimentation: LeNet-5 [20], AlexNet [18] and simple sequential CNN model. LeNet-5 has Conv2D layer, which applies a 3x3 filter to the input image and applies a ReLU activation function to the output. AlexNet consists of eight layers, including five convolutional layers, two fully connected layers, and one softmax output layer. Both models are trained using a sparse categorical cross entropy loss function and the Adam optimization algorithm, and the accuracy metric is used to evaluate the model's performance. CNN is created using the Sequential model type from the Keras library in Python, which allows us to add layers to the model in a linear stack. The model starts with a Conv2D layer, which applies a set of filters to the input image and applies a ReLU activation function to the output. The output is then passed through a MaxPool2D layer, which reduces the size of the feature map by taking the maximum value of a group of adjacent pixels. The output of the pooling layer is then flattened and passed through a fully connected layer, which consists of several units or neurons that are connected to all the input units and can perform classification or regression tasks. The output of the fully connected layer is passed through a final layer with a softmax activation function, which outputs a probability distribution over the possible classes. The model is then compiled using a sparse categorical cross entropy loss function, the Adam optimization algorithm, and the accuracy metric. The model can then be used for image classification tasks by passing in an input image and using the model's predict method to obtain the class probabilities.

4. Experimental results

In the initial experiments, all three models were trained with grayscale and RGB images, but it was observed that the models trained with the grayscale clothing images classify significantly worse and achieve a printability of 85% at the first level of the hierarchy, which is about 14% worse than with the RGB images. When split into 8 classes, the accuracy drops to 59-68%, while when split into 21 classes, the accuracy barely reaches 25-33%. Therefore, further experiments and results are presented for all models trained with RGB clothing images. For comparative analysis of the results, both hierarchical (namely, LeNet5_H, CNN_H, AlexNet_H) and non-hierarchical models (namely, LeNet5_NH, CNN_NH and AlexNet_NH). The table below (see Table 1) shows the results of the classification into the three clothing classes – “Top”, “Bottom” and “Full wear”. As we can see, the results are very similar and the advantage of the hierarchical model is most pronounced only in the case of the AlexNet model, where we can see that the average accuracy of the AlexNet_NH model is 58.77%, while the average accuracy of the hierarchical model - AlexNet_H, is 99.09%

Table 1. 3-class classification accuracy results of different models

	1*Top	2*Bottom	3*Full wear	Average
LeNet5_H	0.991	0.997	0.990	0.9926
LenNet5_NH	0.989	1.000	0.987	0.9920
CNN_H	0.976	0.988	0.976	0.9801
CNN_NH	0.968	0.984	0.980	0.9772
AlexNet_H	0.986	0.997	0.990	0.9909
AlexNet_NH	0.443	0.698	0.622	0.5877

Classifying into 8 classes, we see that the hierarchical model achieves better accuracy than the simple model, with 7.5% higher accuracy for LeNet5, 10.1% for CNN and 28.7% for AlexNet models (See Table 2). Sweaters were the worst classified with 59.8% accuracy, followed by Shirts (64.1%) and Blouses (73.7%). All models had the best classification of the pants, resulting in an accuracy of 9.16%.

Table 2. 8-class classification accuracy results of different models

	1*Shirts	1*Blouses	1*Sweater	2*Skirts	2*Pants	3*Blazers	3*Dress	3*Coats	Average
LeNet5_H	0.643	0.216	0.980	0.592	0.975	0.929	0.988	1.000	0.790
LenNet5_NH	0.500	0.820	0.380	0.900	1.000	0.710	0.820	0.590	0.715

CNN_H	0.768	0.796	0.551	0.714	0.631	0.917	0.840	1.000	0.777
CNN_NH	0.460	0.740	0.370	0.866	0.910	0.630	0.840	0.590	0.675
AlexNet_H	0.985	0.998	0.997	1.000	1.000	0.989	0.973	0.997	0.992
AlexNet_NH	0.490	0.850	0.310	0.980	0.980	0.740	0.792	0.500	0.705

The results of the classification of garments into 21 classes are shown below, including examples in the confusion matrices (Figure 3 - Figure 5), in order to analyze which garments are difficult to distinguish. The average accuracy for all classes are provided in the Table 3. When classifying into 21 classes, the hierarchical model classifies worse in the case of LeNet5 and AlexNet, and the superiority of the hierarchical model is only visible in the case of CNN. The results obtained for LeNet5_H show that the accuracy is 16.24% lower than LeNet5_NH, while the accuracy of AlexNet_H is 14.8% lower than AlexNet_NH. The CNN_H model is 7.57% more accurate than CNN_NH. It can be noted that the classification accuracy is relatively low and the best value of 47% is achieved with the LeNet5_NH model

Table 3. 21-class classification accuracy results of different models

Model	LeNet5_H	LenNet5_NH	CNN_H	CNN_NH	AlexNet_H	AlexNet_NH
Accuracy	0.307	0.470	0.3414	0.265	0.297	0.445

From the confusion matrices, we see that the subclasses "Shirts" and "Pants" are the most confused, and due to the small amount of data, some subclasses do not contain any data at all (e.g. "Coat:Long", see Figure 4, Figure 5). Dresses are also often classified as coats, and it is not uncommon to observe that no long-sleeved dress has been classified correctly (zero value in the confusion matrix).

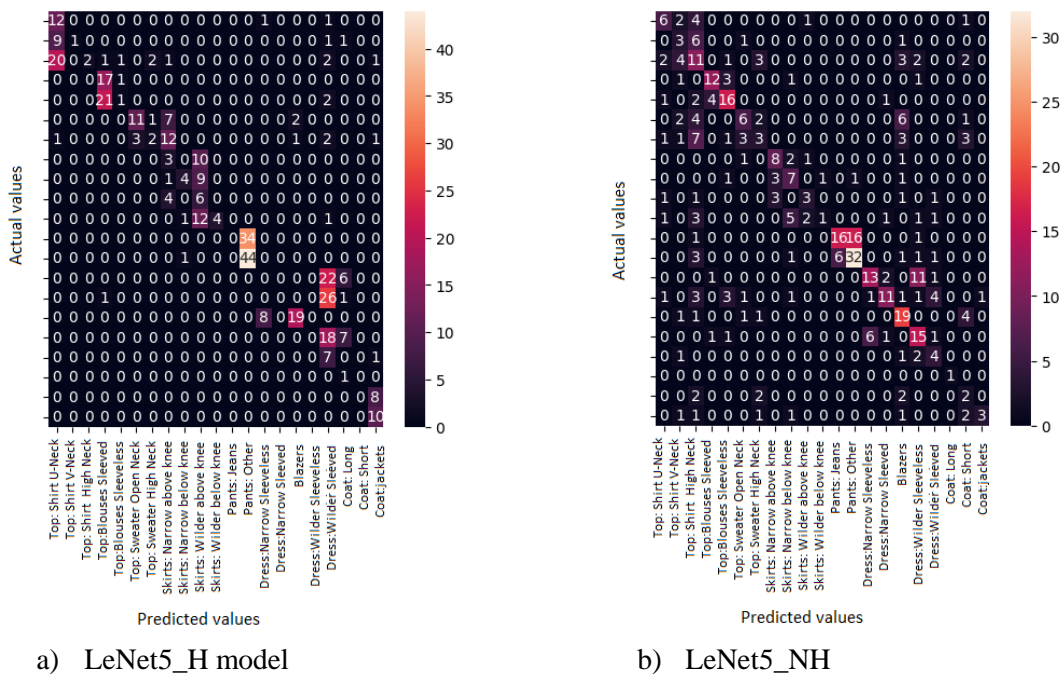


Figure 3: Heatmap of confusion matrix for LeNet-5 models

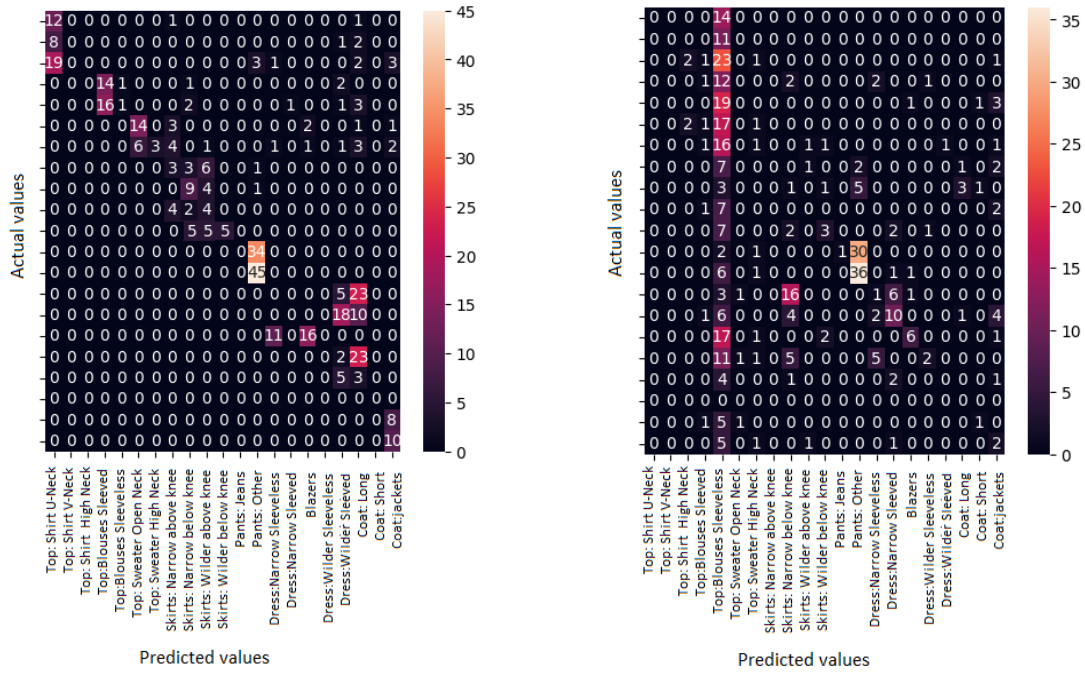


Figure 4: Heatmap of confusion matrix for CNN

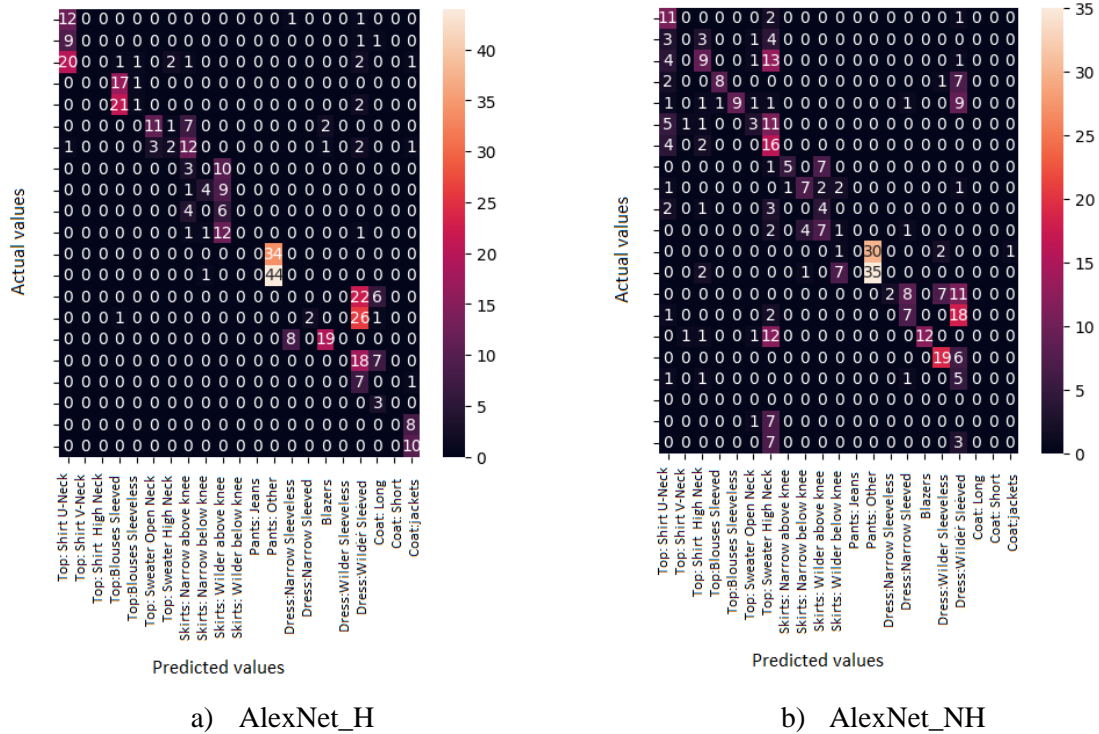


Figure 5: Heatmap of confusion matrix for AlexNet

5. Conclusion

A 3-level hierarchical garment classification model has been proposed in the paper, which classifies garments into 3, 8 and 21 classes. The model has been tested with three deep learning architectures LeNet5, AlexNet and sequential CNN model. From the results obtained, it is observed that the advantage of the hierarchical model is highest when classifying garments into eight categories and allows to

increase the average accuracy up to 28% in the case of the AlexNet model. When classifying into the three main classes - Top, Bottom and Full wear - the hierarchical model is only marginally more accurate for LeNet5 and CNN, with accuracies above 99% for all models. For the AlexNet model, the hierarchical model is significantly more accurate due to the low accuracy of the AlexNet_NH model, which is only 58.77%. The hierarchical model was found to be model-dependent in the classification of the 21 classes and 2 out of the 3 models were found to be less accurate and hence hierarchical subdividing is not appropriate for the LeNet5 and CNN architectures used in the research. From the confusion matrices, we can see that this low accuracy is due to several reasons: 1) the small sample size, which is very unbalanced; 2) the relatively high degree of intermixing between the subclasses of shirts, long and short coats as well as pants.

6. Discussion

The study has provided many insights and ideas for further work to improve the hierarchical job classification model. In particular, different garment classification methodologies can be tested, involving different numbers of hierarchy levels. Another important aspect is that we can use different deep learning architectures at different hierarchical levels to select the most accurate. This approach could combine the strengths of each model to create a more robust and accurate overall system. Therefore, further research could be done by relabeling the dataset used in the study. This could involve using a different classification system or adding more detailed labels to the existing data. This would provide a more fine-grained understanding of the data and enable the use of more specialized models. This could lead to improved performance and a greater understanding of the underlying patterns in the data. Additionally, relabeling the dataset could enable the use of more advanced techniques such as transfer learning and fine-tuning of pre-trained models, which could further improve the accuracy of the garment classification system.

It would also be appropriate to include other more sophisticated architectures (such as Yolo) in the study, but architectures such as ResNet50 and VGG-19, which were included in the first tests, did not work well. In particular, the accuracy was lower than LeNet5 and the training took significantly longer. The VGG-19 model took more than 6 hours to train and was 88% accurate in classifying garments into three classes

7. References

- [1] Sha, D., Wang, D., Zhou, X., Feng, S., Zhang, Y., & Yu, G. (2016, June). An approach for clothing recommendation based on multiple image attributes. In *International conference on web-age information management* (pp. 272-285). Springer, Cham.
- [2] Kayed, Mohammed & Anter, Ahmed & Mohamed, Hadeer. (2020). Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture. 238-243. 10.1109/ITCE48509.2020.9047776.
- [3] Nocentini, O.; Kim, J.; Bashir, M.Z.; Cavallo, F. Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset. *Sensors* 2022, 22, 9544. <https://doi.org/10.3390/s22239544>.
- [4] ZALANDO RESEARCH (2018). Fashion MNIST. <https://www.kaggle.com/datasets/zalando-research/fashionmnist>
- [5] Seo, Y., & Shin, K. S. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert systems with applications*, 116, 328-339
- [6] Papadopoulos, S. I., Koutlis, C., Sudheer, M., Pugliese, M., Rabiller, D., Papadopoulos, S., & Kompatsiaris, I. (2022, March). Attentive hierarchical label sharing for enhanced garment and attribute classification of fashion imagery. In *Recommender Systems in Fashion and Retail: Proceedings of the Third Workshop at the Recommender Systems Conference (2021)* (pp. 95-115). Cham: Springer International Publishing.
- [7] Levatić, J., Kocev, D., & Džeroski, S. (2015). The importance of the label hierarchy in hierarchical multi-label classification. *Journal of Intelligent Information Systems*, 45(2), 247-271.

- [8] Cerri, R., Barros, R. C., & de Carvalho, A. C. (2011, November). Hierarchical multi-label classification for protein function prediction: A local approach based on neural networks. In 2011 11th International Conference on Intelligent Systems Design and Applications (pp. 337-343). IEEE.
- [9] Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86-97.
- [10] Seo, Y., & Shin, K. S. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert systems with applications*, 116, 328-339.
- [11] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [12] Li, F., Kant, S., Araki, S., Bangera, S., & Shukla, S. S. (2020). Neural networks for fashion image classification and visual search. *arXiv preprint arXiv:2005.08170*.
- [13] Sreemathy, R., Turuk, M. P., & Khurana, S. (2022). Cloth Pattern Recognition Using Machine Learning and Neural Network. *Malaysian Journal of Science and Advanced Technology*, 1-8.
- [14] Vijayaraj, A. & Pt, Vasanth & Rethnaraj, Jebakumar & Senthilvel, P. & Kumar, N. & Kumar, R. & Dhanagopal, R.. (2022). Deep Learning Image Classification for Fashion Design. *Wireless Communications and Mobile Computing*. 2022. 1-13. 10.1155/2022/7549397.
- [15] Steffens, Alisson & Maria, Anita & Fernandes, Anita & Lyra, Rodrigo & Reis, Valderi & Leithardt, Valderi & Correia, Sérgio & Crocker, Paul & Luis, Rudimar & Dazzi, Scaranto. (2021). Classifying Garments from Fashion-MNIST Dataset Through CNNs. *Advances in Science Technology and Engineering Systems Journal*. 6. 989-994. 10.25046/aj0601109.
- [16] Guo, Y., Liu, Y., Bakker, E. M., Guo, Y., & Lew, M. S. (2018). CNN-RNN: a large-scale hierarchical image classification framework. *Multimedia tools and applications*, 77(8), 10251-10271.
- [17] Kolisnik, B., Hogan, I., & Zulkernine, F. (2021). Condition-CNN: A hierarchical multi-label fashion image classification model. *Expert Systems with Applications*, 182, 115195.
- [18] Krizhevsky, A., Sutskever, I., & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks.
- [19] TIIKTAK (2020). Fashion MNIST with AlexNet in Pytorch. <https://www.kaggle.com/code/tiiktak/fashion-mnist-with-alexnet-in-pytorch-92-accuracy>
- [20] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, *Proceedings of the IEEE*, 86(11):2278-2324, (1998)
- [21] Sun-Kuk Noh, Recycled Clothing Classification System Using Intelligent IoT and Deep Learning with AlexNet, *Hindawi, Computational Intelligence and Neuroscience Vol. 2021, ID 5544784*, <https://doi.org/10.1155/2021/5544784>
- [22] Wang, Jing. Classification and Identification of Garment Images Based on Deep Learning, *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 3, pp. 4223-4232, 2023.
- [23] M.Marryam, S.Muhammad, M.Yasmin, K. Seifedine. (2022). A novel approach of boundary preservative apparel detection and classification of fashion images using deep learning, *Mathematical Methods in the Applied Sciences*. <https://doi.org/10.1002/mma.8197>
- [24] Donati L, Iotti E, Mordonini G, Prati A. Fashion Product Classification through Deep Learning and Computer Vision. *Applied Sciences*. 2019; 9(7):1385. <https://doi.org/10.3390/app9071385>
- [25] Medina, Adán, Juana Isabel Méndez, Pedro Ponce, Therese Peffer, Alan Meier, and Arturo Molina. 2022. "Using Deep Learning in Real-Time for Clothing Classification with Connected Thermostats" *Energies* 15, no. 5: 1811. <https://doi.org/10.3390/en15051811>