WOMAN HIJAB DETECTION USING TRANSFER LEARNING

Hamdi R. Alabsi[1], Abdelkareem M. Alashqar[1], Ashraf Y. Maghari[1*]

[1]*Faculty of Information Technology, Islamic University of Gaza, PO Box 108, Gaza, Palestine*

**Corresponding author: amaghari@iugaza.edu.ps*

ABSTRACT: Person clothing is considered one of many important issues related to Islamic rulings (Sharia). One of these issues is the woman dress where an Islamic woman is required to keep consistent wearing of Hijab (veils) when she is outside. Based on Sharia and most of Islamic scholars, Hijab must cover woman's hair, ears, and neck along with the top of chest. Classifying woman images into one who wearing Hijab or not can be significantly facilitated using the current approaches of deep learning. In this paper we proposed a structured method consisting of multiple steps for accurately classifying women's images into Hijab and Non-Hijab by utilizing transfer learning approach of deep neural networks. We initially created a balanced and labeled dataset that includes 12,000 images from multiple sources, one half of the dataset for women wearing Hijab and the other for women not wearing Hijab. The dataset is then preprocessed for normalization techniques. After that we used three well-known pretrained models of transfer learning which are VGG16, Xception and MobileNetV2 to conduct our experiments. The feature extraction and fine-tuning strategies were used for examining the models. The selected models gave better performances when applying fine-tuning strategy where the accuracies values of 96.85%, 97.6% and 96.3% were achieved for VGG16, Xception and MobileNetV2 respectively. Our results proved the capability of transfer learning in detecting Hijab in women's images in order to help individuals, institutions and others who are interested in Islamic dress.

KEY WORDS: *Hijab, Deep Neural Networks, VGG16, Xception, MobileNetV2*

1. INTRODUCTION

People across societies have always cared about clothing to enhance their look and appearance. Further, clothing can be semantically analyzed to forecast human's dress style, characteristics and profession (Malisiewicz et al., 2011). However, the shape, style and what parts of the human body the clothes cover are affected by the commitment of Islamic rulings (Sharia). Based on Sharia, and most of Islamic scholars, a woman must not show her body parts in public (outdoor) except her face and palms. Therefore, she is required to wear Hijab (veil) that covers her hair, ears, neck and the top part of her chest.

The classification of women's images into Hijab and Non-Hijab classes has significant needs for individuals, institutions and others in Islamic societies. This classification process can be significantly simplified by computer vision approach which is an important field of deep learning. Additionally, applying convolutions through deep learning layers give better input transformations in producing considerable spatial features and knowledge for images.

Further, the availability of deep learning models that pretrained previously on very large dataset (e.g. ImageNet) gives the researchers and practitioners the capability for transferring features learned by these models to other novel computer vision problems especially when using smaller datasets (Chollet & Chollet, 2021). There are various deep learning models available in literature for transfer learning that mostly applied the concepts of convolutional neural networks (ConvNets).

Although there are significant studies found in the literature for the use of transfer learning for computer vision problems in the domain of human apparel, fewer have addressed veiling practices. To the best of our knowledge, no prior study has applied transfer learning specifically to the detection of hijab by classifying images of women into Hijab and Non-Hijab classes.

In this paper we used three well-known models which are VGG16, Xception and MobileNetV2 for classifying women's images into Hijab and Non-Hijab classes. The results showed that the selected models provide significant performances in the classification problem of this study. Additionally, the results showed the capability of transfer learning in providing good solutions for handling this novel computer vison problem. The results also help interested institutions and individuals in determining if a woman wears a hijab or not.

The rest of the paper is organized as follows: the related work is provided in Section 2. Section 3 introduces the designed method regarding the transfer learning approach and the strategies used. The experimental results and discussion are drawn in Section 4. Threats to validity are provided in Section 5. Section 6 concludes the paper and recommends for future work directions.

2. Related Work

Clothing and attire in images of people have attracted the interest of many researchers. For instance, Zhao (2001) built a vison system that is able to detect people in different shapes, sizes and clothing, whilst Liang et al. (2016) developed a clothing co-parsing system (CCP) for parsing clothing images into semantic configurations. Their system performs two main steps. In the first step which is called "image cosegmentation" it applies an exemplar-SVM approach to segment the images into refined regions, and in the second step which is called "co-labeling" the system applies a Graph Cuts algorithm to combine clothing configuration into the segmented regions. Additionally, Usmani et al. (2022) developed a framework for clothing segmentation by utilizing the feature and fusion extraction modules. They used Mask Region Convolutional Neural Network (RCNN) for extracting low-level features and used InceptionV3 for extracting high-level features. Both levels are fused by fusion module to improve clothing segmentation performance.

With regard to veil-related research such as studies on face masks, Almghraby & Elnady (2021) proposed a method for detecting face mask in real-time using MobileNetV2, while Rokhana et al. (2021) applied MobileNetV2 for classifying images to detect if the face mask is properly used or not.

With regard to the effect of veils in image processing, Sikandar et al. (2017) used RGB and YCbCr color spaces to analyze face skin with Hijab and Niqab (a veil that mostly covers the whole face except the two eyes). The authors found that YCbCr is more suitable in detecting the skin and non-skin areas when the fabric color is different from the skin tone.

In the context of research related to Hijab, Nugraha & Nasrudin (2015) utilized the approach of augmented reality for displaying hijab that virtually best fit to Islamic woman face. They applied both face and mask detection techniques. On the other hand, Oktavianti et al. (2016) built a prototype for creating a hijab model that conforms to a person face type and obeys to Sharia as well. They applied a classification process based on canny operator and matching correlation template for defining the type of face. The satisfaction of the prototype results is evaluated by analyzing users' responses which are collected via a questionnaire.

Khaliluzzaman et al. (2017) applied parallel processing approaches in the images of women to detect hijab and define if it is *Shar'i Hijab* or not where Shar'i is a known style of clothing worn by Indonesian Muslim women. To reduce the time of image processing, the authors utilized the CUDA platform developed by Nvidia that lets professional programmers to easily apply the concepts of parallel computing in Graphical Processing Unit (GPU) processor. They also applied the artificial neural network approach using Yolo feature to quickly and accurately detect objects in images.

Madkour et al. (2019) developed a model that applied fully convolutional network to segment a woman image into three semantic regions which are skin, hijab (veil) and background.

Cholissodin et al. (2020) proposed a framework including multiple steps for detecting hijab in images. They initially utilized the Viola-Jones cascade classifier to detect the regions of face and eyes of the person. Then they used the YCbCr color model to extract the surrounding face region. After that, the bounding box of the eyes region is used to detect the hair and neck. Finally, the hijab is detected based on the cover of hair and neck.

Based on the previously stated literature review, there is no such study found in the literature that applies the transfer learning for classifying women's images into Hijab and Non-Hijab classes. Therefore, this study is performed to demonstrate the capability of three different models in detecting hijab in women's images.

3. Designed Method

To achieve the objective of this paper, a structured methodology that adopts the approach of transfer learning is proposed. As shown in Fig. 1, our proposed methodology includes four main steps, which are data collection and preparation, data preprocessing, examining pretrained models with two strategies, and models evaluation and results analysis.

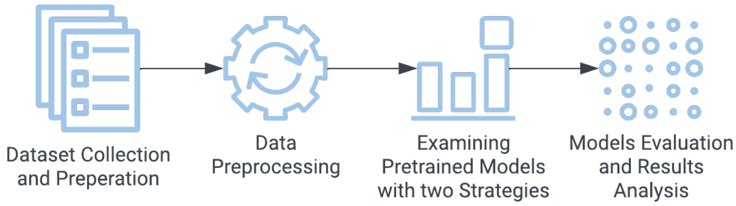


Fig. 1. Main steps of the adopted methodology

3.1. Dataset Collection and Preparation

We collected the data using two main sources. We used the "Hijab dataset" published by Najamudin Ridha (2020) as the main source that includes 10,000 images, one half of them labeled as Hijab images and the second half labeled as Non-Hijab images. An additional 2,000 images were added to the dataset, which were sourced from the Internet. These images have different characteristics than those found in the main source. Typically, 1,000 of them are Hijab images where 800 images have different backgrounds and the remaining 200 images are of women wearing niqab (where most of the head is covered except the two eyes). The other 1,000 images are Non-Hijab images with different backgrounds, other clothing situations like a woman wearing a hat or scarf. Additionally, some of the 2,000 images includes other situation such as a woman rises her hand close to head, face or neck regions. Eventually, our constructed dataset is balanced and includes 12,000 images; 6,000 for Hijab class and 6,000 for Non-Hijab class. It is noteworthy that the regions of interest in the images for the potential experiments are the head, face, neck and top of chest.

3.2. Data Preprocessing

Some preprocessing on the dataset images is performed. Each image in the dataset is resized to 224 X 224 scale so the images can be easily fed to the selected pretrained models. Additionally, each image is rotated by 90 degrees and its brightness is adapted while keeping the original colors of each image. A sample images from the dataset after preprocessing is provided in Fig. 2.

Fig. 2. Sample images from dataset after preprocessing step

3.3. Examining Pretrained Models with Two Strategies

In this study the concept of transfer learning is adopted for detecting hijab in women's images. Next, we provide the concepts and background of convolutional neural networks, transfer learning, the selected pretrained models and the strategies used.

3.3.1. Convolutional Neural Network

Convolutional neural network is a type of deep learning models that used commonly in computer vision field. It is also known as ConvNet and CNN. ConvNet is preferred over dense (fully connected) network for classifying images because it has the ability of preserving the spatial information and relationships of an image during data transformation of the learning process.

The ConvNet consists mainly of three types of layers which are convolutional, pooling and fully connected layers. It starts with a convolutional layer, followed by alternating convolutional and pooling layers, and ends up with one or more of fully connected layers. The core processing of a ConvNet is done by the convolutional layer. This layer takes an input, such as an image, and uses a filter (kernel) to detect features from a specific part (location) of this input to produce a vector of feature map. This process is called convolution, and it is repeated for the other input parts by moving the filter with a distance called a stride to finally produce a stack of feature maps. The convolution is often characterized by two main parameters; the size of the filter that is called a window, for example 3 X 3 or 5 X 5, and the depth of the feature map which is represented by the number of filters, for example 32 or 64. The pooling layer accesses the entire input and always applies one of the two types of aggregation, which are max pooling or average pooling, in order to produce an output vector. The fully connected layer is used to produce the predictions of the model based on the preceding transformations of the network. The earlier convolutional layers always use filters to learn easier and general features, while the later layers learn more sophisticated and more specific features (Chollet & Chollet, 2021; Kapoor et al., 2022).

3.3.2. Transfer Learning Using Pretrained Models

A pretrained model is a model that was previously trained on some specific deep learning task (e.g. classification task) using a very large dataset such as ImageNet which consists of 1.4 million of images labeled within 1,000 different classes. The hierarchy of features learned works as a generic model that can be effectively used in a novel computer vison task, although this novel task has different number and type of classes other than those learned originally.

We adopted two strategies (Chollet & Chollet, 2021) for applying the pretrained images, which are extracting features from pretrained model, and fine-tuning of pretrained model, which are explained in the following subsections.

3.3.3. Extracting Features from Pretrained Model

The pretrained model is mainly used for extracting the features learned from previously training task of classification in order to be reused for a new problem. Practically, it consists of two main parts, the first part is called the base convolution that consists of mostly multiple alternating convolutional and pooling layers. The second part that followed the first one in processing, consists of a fully connected classifier. In this strategy the base convolution part is executed over the dataset at hand in order to extract features, knowledge and data representations that previously learned from previous classification problem, then applying a complete new training with a suitable added fully connected classifier to produce the classification predictions.

3.3.4. Fine-tuning of Pretrained Model

In this strategy, some of the deeper (top) layers of the pretrained model especially the convolutional layers are unfrozen in order to be retrained on the dataset at hand. Deeper layers always are unfrozen because these layers hold more specific features while the earlier layers hold more general ones. In other words, while the network gets deeper and deeper, it moves from holding more general features to more specific features. So, it is not convenient to unfreeze and retrain the earlier layers because this will not produce considerable transformations from smaller dataset with comparison to the previously obtained from huge dataset. In this strategy, a newly customized fully connected layer is added and trained with the previously (already) trained base convolutional layer of the model without freezing the layers, then unfreeze few layers from the base convolutional alongside the customized added one in order to retrain the whole model on the new dataset. It is necessary to train the model with the customized layer before implementing the unfreezing process to avoid destroying the pretrained transformations of the model. It is noteworthy that this strategy implicitly includes the concepts of feature extraction strategy in addition to the capability of added customized layers.

3.3.5. The Selected Pretrained Models

A three well-known pretrained models were selected to achieve the objective of this study, which are VGG16, Xception and MobileNetV2. It is worthy to note that all of these models were trained on the well-known huge dataset, ImageNet. In the following subsections we provide an explanation for each of the selected models.

VGG16

The architecture of VGG16 was created by Simonyan & Zisserman (2014) and consisted of 16 layers that are distributed into 5 blocks. Each block includes two adjacent 2D convolutional layers followed by one pooling layer, except the fifth block that includes three 2D convolutional layers instead of two. It is noteworthy that the layers explained here are the parameterized layers holding weight and known as activation layers. Nonetheless there are other types of layers such as pooling layers.

Xception

Xception is an “Extreme” form of the Inception model that applies the technique of depthwise separable convolution. Using this technique, the convolution layer achieves spatial convolution for every channel (e.g. 3x3) on its input separately then concatenating them by pointwise convolution layer to produce 1x1 convolution. Additionally, Xception applies the idea of residual networks (ResNet) by which the output of earlier layers can be forwarded and fed fast to the deeper layers. This idea is also called fast-forward/skip connections technique and it is used to reduce gradient explosion and dying of the deeper networks.

MobileNetV2

MobileNetV2 is a lightweight deep learning network built mostly for mobile devices because of their limitation in memory size (capacity) and processing efficiency. It also applies the separable depthwise convolution technique explained previously. To reduce the heavyweight of deep learning structure, MoblieNetV2 applies the inverted residual technique. By applying this technique, the skip-connected channels become smaller and as a result a reduction is performed in layers parameters.

3.4. Models Evaluation and Results Analysis

We evaluated the performances of the selected models int terms of precision, recall, F1-score and accuracy. Precision (P) is calculated using the ratio of: $P = (\frac{TP}{TP+FP})$, where TP is the number of input instances (images) that correctly predicted to the class whereas FP is the number of input instances that predicted incorrectly and actually belongs to the other class. Precision is important for reducing the FPs.

Recall(R) is calculated using the ratio of: $R = (\frac{TP}{TP+FN})$, where FN are the number of input instances that incorrectly classified to the other class but actually belongs to the current class of input images. Recall is always used to reduce the FNs.

F1-score (F1) is calculated using the formula of: $F1 = (\frac{2\times TP}{8\times TP+FP+FN})$, which is the harmonic mean of precision and recall.

Accuracy (A) is calculated using the formula: $A = (\frac{TP+TN}{TP+TN+FP+FN})$, which represents the ratio of the total number of input instances of images that predicted correctly to the total number of instances that predicted correctly and incorrectly.

4. Experimental Results

We experimented three pretrained models on a balanced and labeled dataset consisting of 12,000 images, 6,000 images for women wearing hijab and the other 6,000 images for women not wearing hijab. The dataset was split into approximately 67%, 16.5% and 16.5% for training, validation and testing as shown in Table 1. Typically, we experimented the VGG16, Xception and MobileNetV2 models by applying both feature extraction and fine-tuning strategies.

Table 1: The distribution of the experimented dataset

Total	Training	Validation	Testing
12,000	8,044	1,978	1,978

4.1. Conducting experiments and analyzing results

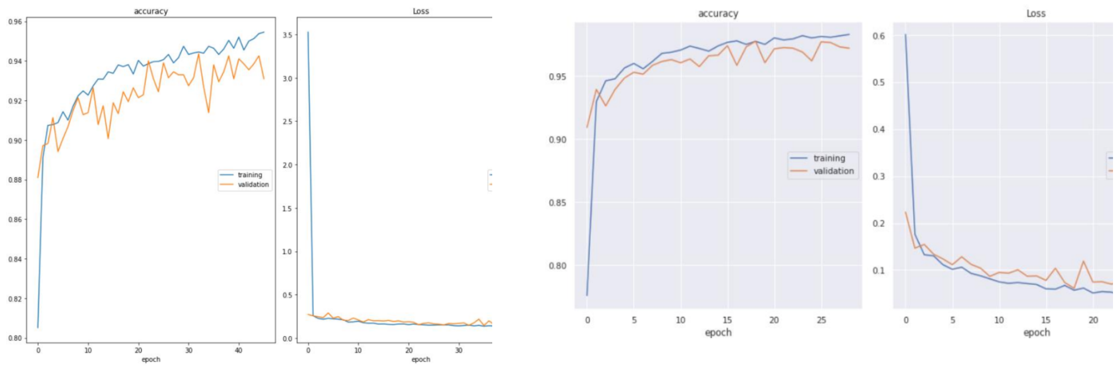
For applying both strategies in each model, we added a fully connected block on the top (end) of the model that consists of two dense layers with 4,096 and 1,024 nodes respectively, followed by a dropout layer that set to 0.25, and finally an output layer that consists of only one node that set to sigmoid activation function to appropriately classify the input to one of two classes.

Regarding the fine-tuning strategy, unfreezing blocks is applied as follows: block 5 (layers 15 to 18) in VGG16 model, blocks 11 to 16 (layers 98 to153) in

MobileNetV2 model, and blocks 13 and 14 (layers 116 to 131) in Xception model. It is noteworthy that the number of layers here are all the types of layers including activation (parameterized) and non-activation layers as built by Keras library.

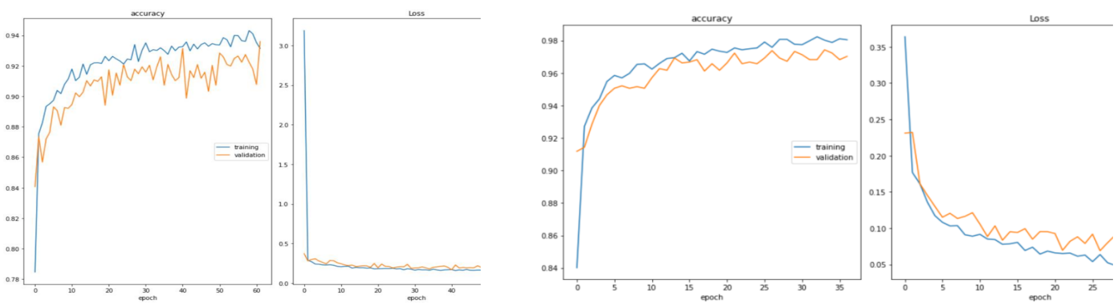
Each model is compiled using Adam optimizer and binary entropy loss function and fitted using 64 epochs. Practically, we used the approach of early stopping to improve model generalization. Early stopping is mostly used for reaching the best fit boundary between underfitting and overfitting. This situation obtained when reaching several epochs (iterations) by which there is no further increase in accuracy nor decrease in loss. Fig. 3, Fig. 4, and Fig. 5 depict the curves of accuracy and loss reached for each model when applying both training strategies. Table 2 summarizes the final number of epochs reached when the best maximum (Max) accuracy and minimum (Min) loss are obtained on the validation portion of the dataset, and the time taken by the model to reach these values. The starting (Min) values of accuracies and starting (Max) values of losses are also shown in the table.

It is noteworthy that the pretrained models were downloaded from Keras library and the code was run on Google Collab using Python 3.6 version utilizing the GPU processor.



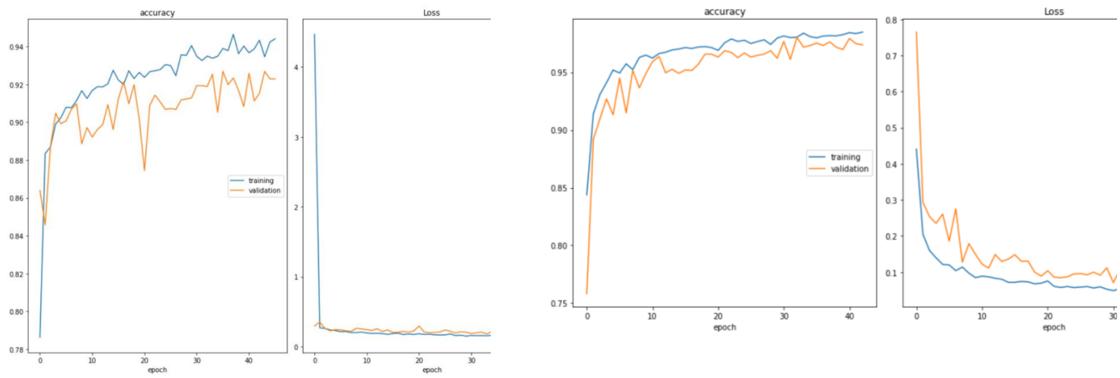
(a) **(b)**

Fig. 3. Accuracy and loss per epoch for VGG16 model with a) Feature extraction, b) Finetuning



(a) **(b)**

Fig. 4. Accuracy and loss per epoch for Xception model with a) Feature extraction, b) Finetuning



(a) **(b)**

Fig. 5. Accuracy and loss per epoch for MobileNetV2 model with a) Feature extraction, b) Finetuning

Table 2: Results summary of applying early stopping technique for experimenting pretrained models

	Feature Extraction				**Finetuning**			
	Time Taken (m)	**Number of Epochs**	**Min-Max Accuracy**	**Max-Min Loss**	**Time Taken (m)**	**Number of Epochs**	**Min-Max Accuracy**	**Max-Min Loss**
VGG16	180	45	0.88-0.95	3.53-0.12	90	30	0.91-0.97	0.60-0.05
Xception	180	60	0.84-0.94	3.19-0.14	145	40	0.91-0.97	0.36-0.05
MobileNetV2	210	45	0.85-0.93	4.47-0.14	180	42	0.76-0.97	0.44-0.04

The detailed classification results as achieved by the examined models are shown in

Table 3 and Table 4 when applying the feature extraction and finetuning strategies respectively.

Table 3: Performance of three pretrained models using feature extraction strategy

	VGG16		**Xception**		**MobileNetV2**	
	Hijab	Non-Hijab	Hijab	Non-Hijab	Hijab	Non-Hijab
Precision	0.9344	0.9680	0.9250	0.9861	0.9296	0.9626
Recall	0.9668	0.9320	0.9870	0.9200	0.9640	0.9270
F1-Score	0.9509	0.9491	0.9550	0.9519	0.9465	0.9444
Accuracy	0.9500		0.9535		0.9455	

Table 4: Performance of three pretrained models using fine-tuning strategy

	VGG16		Xception		MobileNetV2	
	Hijab	Non-Hijab	Hijab	Non-Hijab	Hijab	Non-Hijab
Precision	0.9689	0.9671	0.9675	0.9847	0.9705	0.9557
Recall	0.9670	0.9690	0.9850	0.9670	0.9550	0.9710
F1-Score	0.9680	0.9680	0.9762	0.9758	0.9627	0.9641
Accuracy	0.9685		0.9760		0.9630	

Although remarkable performances results were achieved, it is clear that the fine-tuning strategy gave slightly better performances than the feature extraction strategy. The best accuracy results obtained are 97.6%, 96.85 and 96.3 for Xception, VGG16 and MobileNetv2 models respectively when adopting fine-tuning strategy as shown in Table 4. Additionally, it is obvious that all models produced better results for all metrics when implementing the fine-tuning strategy except for the VGG16 precision value when classifying Non-Hijab class (shown in Italic format in Table 4).

Moreover, significant improvements were performed in the recall measurements for the Non-Hijab class. This ensures that adopting the fine-tuning strategy minimizes the incorrectly predictions of Non-Hijab images to other class (Hijab).

By comparing the performances of the three examined pretrained models, the Xception model performs slightly better than both VGG16 and MobileNetV2 model. The VGG16 also slightly outperforms the MobileNetV2 model.

4.2. Testing the models with real life images

To prove the effectiveness of the examined pretrained models, we tested them with new samples that were not seen in the original experimented dataset. A collection of new 60 real-life images, 30 for Hijab class and 30 Non-Hijab class were used in the testing process. Table 5 summarizes the results.

Table 5: Results of testing models using two strategies with new real-life samples

	Feature Extraction		Finetuning	
	Hijab	Non-Hijab	Hijab	Non-Hijab
VGG16	26	27	27	30
Xception	28	29	29	29
MobileNetV2	28	30	28	29

The comparison of our results to the other results found in the literature is not provided due to the differences on the datasets used and, to the best of our knowledge, no such classification problem has been addressed for detecting the Hijab in the women's images through transfer learning with the three well-known models examined in this paper.

5. Threats to Validity

This section provides the threats of validity of this research concentrating on the most important limitation.

1. This study used transfer learning approach for classifying women's images into Hijab and Non-Hijab classes. To lessen bias results due to using only

one model, we used three various models including the older one named VGG16, and the newer two models named Xception and MobileNetV2 where the later was built dedicatedly for mobile devices.

2. Although the transfer learning models provide a good classification technique for novel problems with small-sized datasets, a considerable large dataset including 12,000 images was used in order to not negatively affect the results obtained from applying the strategies of transfer learning especially fine-tuning.
3. To ensure that the classification results are not prone to overfitting, we used part of the dataset including new images (that were not used in model training) for testing the selected models.
4. Although the study showed results with very good performances in classifying women images into Hijab and Non-Hijab classes, it does not provide an approach for detecting whether the Hijab obeyed Sharia, as defined by Islamic rulings and scholars.

6. Conclusion and Future Work

In this paper, we used the approach of transfer learning provided by pretrained models for classifying women's images into Hijab and Non-Hijab classes. We created a dataset of 12,000 women's images to evaluate three well-known pretrained models which are VGG16, Xception and MobileNetV2. Both feature extraction and fine-tuning strategies were applied in performing the experiments on the created dataset where outperforming results were obtained. The examined pretrained models were evaluated in terms of precision, recall, F1-score and accuracy metrics where outperforming results were obtained when applying both strategies. However, the results showed that the fine-tuning strategy performs slightly better than feature extraction in the classification problem. When comparing the results of the three selected pretrained model, it is found that the Xception model outperforms the other two models.

For future work, a larger dataset can be used in addition to experimenting more pretrained models. The issue of detecting the *Shar'i Hijab* can be also examined.

ACKNOWLEDGEMENT

This paper is partially supported by the dean of higher studies and scientific research at Islamic University of Gaza.

REFERENCES

Almghraby, M., & Elnady, A. O. (2021). Face mask detection in real-time using MobileNetv2. *International Journal of Engineering and Advanced Technology*, *10*(6), 104–108.

Cholissodin, I., Palupi, D. E., Putra, M. Y. Y., & Aprilisia, S. (2020). Detection of hijab syar'i as smart clothes system for moslem people using high performance of parallel computing. *IOP Conference Series: Earth and Environmental Science*, *456*(1), 12074.

Chollet, F., & Chollet, F. (2021). *Deep learning with Python*. simon and schuster.

Kapoor, A., Gulli, A., Pal, S., & Chollet, F. (2022). *Deep Learning with TensorFlow and*

Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd.

Khaliluzzaman, M., Begum, A., Fatama, S. K., & Islam, M. M. (2017). Detection and analysis of hijab based on visual feature of neck and hair. *2017 20th International Conference of Computer and Information Technology (ICCIT)*, 1–6.

Liang, X., Lin, L., Yang, W., Luo, P., Huang, J., & Yan, S. (2016). Clothes co-parsing via joint image segmentation and labeling with application to clothing retrieval. *IEEE Transactions on Multimedia*, *18*(6), 1175–1186.

Madkour, D. M., Ahmed, M., & Mohamed, W. F. (2019). Automatic face and hijab segmentation using convolutional network. *International Journal of Integrated Engineering*, *11*(7), 61–66.

Malisiewicz, T., Gupta, A., & Efros, A. A. (2011). Ensemble of exemplar-svms for object detection and beyond. *2011 International Conference on Computer Vision*, 89–96.

Najamudin Ridha. (2020). *naja-dataset*. https://github.com/mnajamudinridha/naja-dataset

Nugraha, A., & Nasrudin, M. F. (2015). Augmented reality system for virtual hijab fitting. *Advances in Visual Informatics: 4th International Visual Informatics Conference, IVIC 2015, Bangi, Malaysia, November 17-19, 2015, Proceedings 4*, 454–463.

Oktavianti, A., Sugeng, W., & Agusta, A. (2016). Implementasi Aplikasi Hijab Berbasis Android dengan Metode Canny Operator dan Template Matching Correlation. *Jurnal Teknik Informatika Dan Sistem Informasi*, *2*(2).

Rokhana, R., Herulambang, W., & Indraswari, R. (2021). Multi-class image classification based on mobilenetv2 for detecting the proper use of face mask. *2021 International Electronics Symposium (IES)*, 636–641.

Sikandar, T., Ghazali, K. H., Mohd, I. I., & Rabbi, M. F. (2017). Skin color pixel classification for face detection with hijab and niqab. *Proceedings of the International Conference on Imaging, Signal Processing and Communication*, 1–4.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Usmani, U. A., Happonen, A., & Watada, J. (2022). Enhanced deep learning framework for fine-grained segmentation of fashion and apparel. *Science and Information Conference*, 29–44.

Zhao, L. (2001). *Dressed human modeling, detection, and parts localization*. Carnegie Mellon University.