Gender-Isolation Impact on The Accuracy Detecting Emotions Using CNN Models

Abstract— emotions are a vital part of human communication. Our emotions can be sent explicitly through words, but there are many ways in which we share how we feel, one of which is facial expressions. This project explores the impact of gender isolation while training various convolutional neural networks (CNN) models. This project tries to improve the initial models using several techniques, such as data augmentation, dropout layer, a combination of both, manipulation of learning rate, and using l2 weight regularization. The performance was not stable throughout. Some key findings are, first, women`s models performed slightly better on men`s images. The opposite is true. However, given that the overall accuracy of the models is not significant, no solid conclusion can be made.

Keywords—convolutional neural network (CNN), Facial emotion recognition (FER), gender-isolation

I. Introduction

1.1 Background

There are many vital parts in human communication; messages are not conveyed through words only. We tend to unconsciously send various information through non-verbal communication, which is an integral part of human interaction. We use movements and facial expressions to deliver how we feel. Observing facial expressions allows us to know how others feel. The brain analyzes the visual data to understand their mental state. Facial expressions convey emotions more accurately than language, offering a genuine glimpse into a person's true feelings during interpersonal communication.[1]

In social contexts, humans naturally express emotions, and a precise understanding of these emotions fosters mutual understanding and trust. Scholars highlight that facial expressions are the main driver of human emotion information. Research indicates that facial expression information constitutes approximately 55 percent of the transmitted data, surpassing voice (38 percent) and language (7 percent). This underscores the greater importance of facial expressions in emotional comprehension compared to language and sound. Consequently, researchers focus on studying facial expressions to gain deeper insights into human inner emotional experiences.[1]

1.2 Facial emotion recognition

Recently, there has been a growing interest in the automatic process of facial emotion recognition (FER). Therefore, it is important to define facial emotion recognition. It can be defined as a computer vision task where the input can be in static, images, or non-static form, videos, and the result is a classification of the person's emotion. Generally speaking, FER systems work to identify a person's emotions and their intensities; some of those systems classify the cause of expressions as genuine or simulated. [2]

In the most recent research change, FER models have been switched to the use of artificial neural networks rather than using traditional machine learning methods, such as like histogram of oriented gradients (HOG) or local binary pattern (LBP), coupled with data classifiers such as support vector machine (SVM), k-nearest neighbors (KNN), or random forest.[2]

1.3 Related work

The impact of sex on this task has gained some interest from computer scientists and psychologists. A study reported that women tend to be more expressive in conveying negative emotions compared to men, based on self-reporting after viewing emotion-eliciting images. However, there was no significant difference in expressing positive emotions between genders. This study solely relied on self-reporting [3]. Moreover, in another study where participants were recorded and then used machine learning to analyze these recordings. The paper did not mention the type of model used. The result suggests that "Overall, women express facial actions more frequently than men, and in particular express more actions of positive valence. However, for negative valence actions expressiveness is dependent on the discrete emotional state. Women expressed actions associated with anger less and actions associated with fear and sadness more than men."[4]

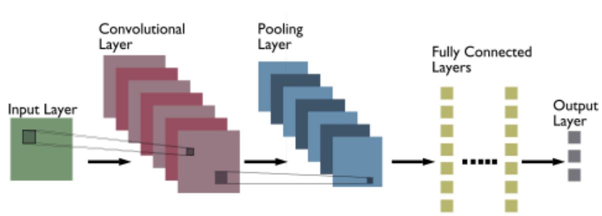
1.4 Objectives

This study aims to investigate the impact of gender-isolation during training deep learning models, Convolutional Neural Network (CNN), on the accuracy of emotion detection. Specifically, examining how CNN models trained only on one gender may exhibit biases in recognizing emotions in the opposite gender in the testing process. A range of six emotions, happiness, sadness, fear, anger, surprise, and neutral, was captured in a diverse dataset that was used in this project. Also, this research distinguishes itself by working on different methods to improve the initial CNN model. Thus, by uncovering and addressing potential biases, this work hopes to provide some insights about the impact of stereotypical thinking about gender and emotion expression.

II. METHODOLOGY

2.1 CNN General Architecture

This project uses deep learning to implement its models, more specifically CNN models. A typical CNN model usually consists of three basic components: convolutional layers, pooling layers, and fully connected layers. First, the convolutional layer, the most integral layer of CNN architecture, comprises an array of convolutional filters, also known as kernels. The input image, represented as N-dimensional matrices, undergoes convolution with these filters to produce the resulting feature map. The second part is the pooling layer with a primary role in performing subsampling on the feature maps generated by the convolutional operations. Additionally, this process reduces the size of large-scale feature maps to generate smaller ones while keeping the essential information or features at each stage of the pooling phase. Lastly, fully connected layers come after the pooling layers. The input is derived from the last pooling in the form of a vector that was generated by flattening the feature maps. The output of the fully connected layer serves as the ultimate CNN output[5]. Fig(1)



Fig(1): General CNN architecture [5]

2.2 Initial Model Desgin

My initial model uses 4 layers, each layer consists of a convolutional layer with filter size (32,64,128, 256). Each convolutional layer is followed by a batch normalization layer and a max pooling layer. Those are flattened and followed by fully connected layers. The activation function used is ReLU.

$$f(x)\text{ReLU} = \max(0, x)$$

The loss function is Softmax. This function is frequently utilized to assess the performance of CNN models and is alternatively known as the log loss function. Its output is a probability, denoted as p ∈ {0,1}.[6]

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^{N} e_k^a}$$

Two models with the same architecture are to be used for each sex, female, and male. The optimization used in both is Adaptive Moment Estimation (Adam), which can be represented by the following equation.

$$w_{ij^t} = w_{ij^{t-1}} - \frac{\eta}{\sqrt{\widehat{E[\delta^2]^t} + \epsilon}} * \widehat{E[\delta^2]^t}$$

2.3 Training and testing plan

The hypothesis is that training models exclusively on one gender's facial expressions (Model 1: Women, Model 2: Men) will lead to performance differences when these models are tested on the opposite gender's facial expressions. Any observed variation in accuracy will be analyzed to gain insights into stereotypical ideas about emotion expression between genders. This gender-isolation experiment will offer a quantitative understanding of how training data separation impacts the model's performance on the opposite gender, contributing to the broader discussion on stereotypical ideas about emotion expression with regard to gender.

2.4 Evaluation

Since the main task for this project is emotion classification, the main metric to evaluate is accuracy given by the following equation:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All Samples}}$$

III. Experiment

3.1 Dataset

The dataset for this project is a combination of two datasets taken from Kaggle. The first dataset "Rating OpenCV Emotion Images" uploaded by LUIZ BUENO [7]. It contains 32298 48X48 grayscale images for classification, divided into training and validation sets. Each consists of seven classes (emotions), angry, disgust, fear, happy, neutral, sad, surprise. The second dataset, "Facial Emotion Recognition Dataset", was uploaded by KUCEV ROMAN. This dataset is part of a commercial dataset. This project just uses the sample in Kaggle. This sample contains 20 participants each showing eight different emotions[8]. The project uses both datasets after separating them into female and male. Each sex has the six emotions that are the interest of this project. The female dataset contains 6323 images. The male dataset contains 6048 images. The validation part contains 1430, 1386 images belonging to female and male, respectively.

3.2 Tools

This project was built using Python programming language. I use TensorFlow library to build, train, and evaluate the models. It is an open-source machine learning library. It provides a framework for the development of deep neural networks. TensorFlow has several applications, including image and speech recognition, natural language processing, and more. TensorFlow provides flexibility in model building and customization. Keras is the high-level API of the TensorFlow platform which offers a user-friendly and efficient interface for addressing machine learning (ML) challenges. It provides the entire machine learning workflow, encompassing tasks ranging from data processing to hyperparameter tuning and ultimately deployment[9].

3.3 Procedure and Attempted Improvments

The dataset has already been preprocessed to some extent. ImageDataGenerator was used to load and normalize the data at first. Two identical models were created using the above-mentioned structure with 20 epochs. The models were trained and tested on the respective gender of training. Initial accuracy for female model was 90.99% and validation accuracy was 47.90%. For the male model, the training accuracy was 89.58% and validation accuracy was 46.10%. The first technique I used to try and improve the performance on my testing dataset is data augmentation, which is a technique that enlarges the original training set by creating modified copies of existing images, typically by some image transformation. For this project another ImageDataGenerator was created. The changes interdicted to the data set includes rescaling pixel values, rotating images up to 20 degrees, shifting images horizontally and vertically by up to 20% of the total width or height, applying shear transformations, zooming in and out, horizontally flipping, vertically flipping, adjusting brightness in a specified range, and using nearest-neighbor filling for any newly created pixels during augmentation. These augmentations help improve the model's generalization and robustness by exposing it to a more varied set of training examples, Fig(2). This technique resulted in training accuracy of 92.80% and validation accuracy of 49.09% for the women model and 92.68% training accuracy and 48.34% validation accuracy for the men model.

```
train_data_gen_aug= ImageDataGenerator(  rescale=1./255,
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    vertical_flip= True,
    brightness_range=[0.2,1.2],
    fill_mode='nearest'
                                        )
```

Fig(2). Data augmentation

The second attempted enhancement is trying to introduce dropout layers after each Maxpooling layer. The benefit of dropout layers is to prevent overfitting. I started with rate 0.1 for the first layer, increasing by 0.5 for the remaining layers. That is, the final layer has a dropout layer with rate 0.25. In the dense layer the dropout layer has a rate of 0.5. The training accuracy dropped in both models to 70.11% and 69.99% for women model and men model, respectively. The validation accuracies are 53.10% and 47.13% for men model and women model, respectively.

The third attempt was to combine the two techniques. That is, I trained models with the dropout layers in the augmented training datasets. The training accuracy dropped in both models to 69.75% and 69.58% for women model and men model, respectively. Validation accuracy for women is 51.61%; and validation accuracy for men model is 51.52%.

The fourth attempt is trying to manipulate the learning rate. This part is trial by error, the optimizer, as mentioned above, is ADAM. The learning rate for the women model is 1-e3. The men`s model is using weight decay by 0.01 to prevent overfitting. This model resulted in training accuracy: 70.83% with validation accuracy 53.68%. the women`s model resulted in accuracy of 83.90% and validation accuracy 50.63%.

I have made two attempts improving the women model with l2 regularization with weight decay of 0.001, and another model with decay of 0.01. These two models performed training accuracy of 72.12%, validation accuracy 50.35%, training accuracy: 56.18% and validation accuracy 55.03%, respectively; both models do not seem to be improving, so attempts stopped at that.

IV. RESULTS

In this section we will dive into the testing results for each model as well as the results of the cross-gender testing. The testing dataset contains 588 images for the female model and 820 images for the male model. The testing accuracies and losses for the women model are shown in Fig (3), Fig(4). The baseline model achieved an accuracy of 45.58%. The introduction of data augmentation resulted in a decrease of accuracy to 36.56%. Adding dropout layers improved the model's robustness, resulting in an accuracy of 41.50%. The mix of dropout and data augmentation further increased accuracy to 48.81%. The attempt of learning rate manipulation resulted in accuracy of 45.92%, which suggests a worsening in the model`s performance. Moreover, models using a convolutional neural network with dropout and regularization techniques achieved accuracies of 50.85% and demonstrated the efficacy of these architectural modifications.

The testing results for the men's models are summarized as follows. To begin with, the baseline model achieved an accuracy of 38.54%. The introduction of data augmentation significantly improved accuracy to 50.24%, highlighting the positive impact of increased diversity in the training dataset. Incorporating dropout layers yields an accuracy of 49.76%. However, experimenting with both dropout and data augmentation resulted in a slightly decreased accuracy of 45.24. Optimizing the learning rate leaded to an accuracy of 48.66%. Fig (5) and Fig (6) show the accuracies and losses of the men`s models.

The second part of this section covers the results of testing female's images on the male models and vise-versa. Testing male`s images seems to be consistent with an increase in accuracy in most models. The model with the most amount of increase in accuracy is the model using l2 regularization with an accuracy of 53.17%. That is, there has been about 3% increase in accuracy. This increase in accuracy might suggest that when models trained in women images could detect men`s emotion more accurately. However, based on how low the accuracy is, it is hard to say that the stereotypical idea is true. That is, we cannot conclude women`s strength in facial expression to be generally true. Model`s performance has to be overall better to be able to make some solid conclusions. Fig(7) and Fig(8) show the accuracies and losses of the women`s model when being tested on men`s images.

Furthermore, looking at the performance of the men`s models, we can see a trend of dropping in accuracy, generally. Some models have a more significant decrease compared to the rest. For example, the model adopting the data augmentation technique has dropped 10% in its original accuracy. It is important to note here that this model achieved the highest accuracy among the five models created for this project. However, as said in the cross testing conducted using the women`s models, it is hard to make conclusions since we do not have a good-enough models. Fig(9) and Fig(10) present the accuracies and losses for the men`s model when being tested on women`s images.

V. CONCLUSION

In conclusion, this project explores the impact of gender isolation in the training process on the efficiency of detection facial emotions. The project makes use of convolutional neural networks CNN to build and test its hypothesis. There are seven models implemented in total and trained on female`s images only whereas there are five models trained on men`s images only. The testing process consisted of two stages. The first stage is testing each model on the model of the training; and the second stage is testing on the opposite gender. between the models` implementation and testing, there have been several attempts to enhance the initial models with respect to accuracy as it was the metric chosen to compare models` performances. It is crucial to note that the dataset for this project was a part of a relatively wide and diverse one; it would make a huge difference if this project had utilized the entire dataset. One challenge is that the mentioned dataset was not fully supporting to the idea of this project, so part of the dataset was randomly selected and separated into men and women images. For future work, it might be a sound improvement if a diverse dataset is used. Given the analysis above, it is unwise to conclude the correctness of any stereotypical norm about the expressiveness of one gender over the other. It might be helpful for future work to adopt the structure of models that are proven to produce substantial accuracy. From an analytical perspective, it might be useful to consider other metrics such as precision and recall evaluating the models` performances.

REFERANCES

[1] Song, Z. (2021). Facial Expression Emotion Recognition model integrating philosophy and machine learning theory. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.759485

[2] Cîrneanu, A., Popescu, D., & Iordache, D. D. (2023). New Trends in Emotion Recognition using Image Analysis by Neural Networks, A Systematic review. *Sensors*, *23*(16), 7092. https://doi.org/10.3390/s23167092

[3] Givon, E., Berkovich, R., Oz-Cohen, E. et al. Are women truly "more emotional" than men? Sex differences in an indirect model-based measure of emotional feelings. Curr Psychol (2023). https://doi.org/10.1007/s12144-022-04227-z

[4] McDuff D, Kodra E, Kaliouby RE, LaFrance M. A large-scale analysis of sex differences in facial expressions. PLoS One. 2017 Apr 19;12(4):e0173942. doi: 10.1371/journal.pone.0173942
PMID: 28422963; PMCID: PMC5396880.

[5] Abyaneh, A. H. Y., Gharari, A. H., & Pourahmadi, V. (2018). Deep neural networks meet CSI-Based authentication. *ResearchGate*. https://www.researchgate.net/publication/329608053_Deep_Neural_Networks_Meet_CSI-Based_Authentication

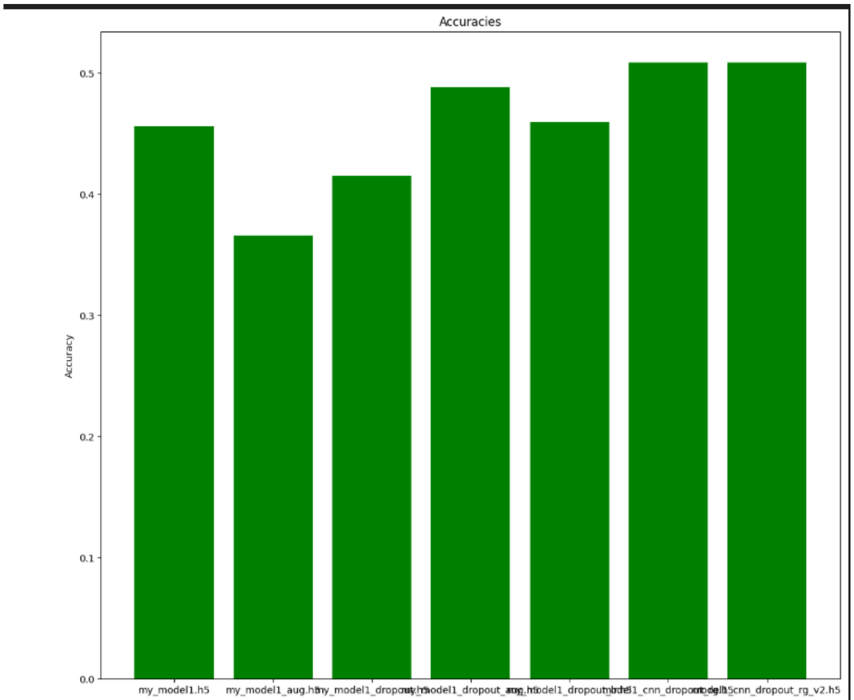
[6] Alzubaidi, L., Zhang, J., Humaidi, A.J. *et al*. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* **8**, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[7] *Rating OpenCV Emotion Images*. (2021, December 4). Kaggle. https://www.kaggle.com/datasets/juniorbueno/rating-opencv-emotion-images

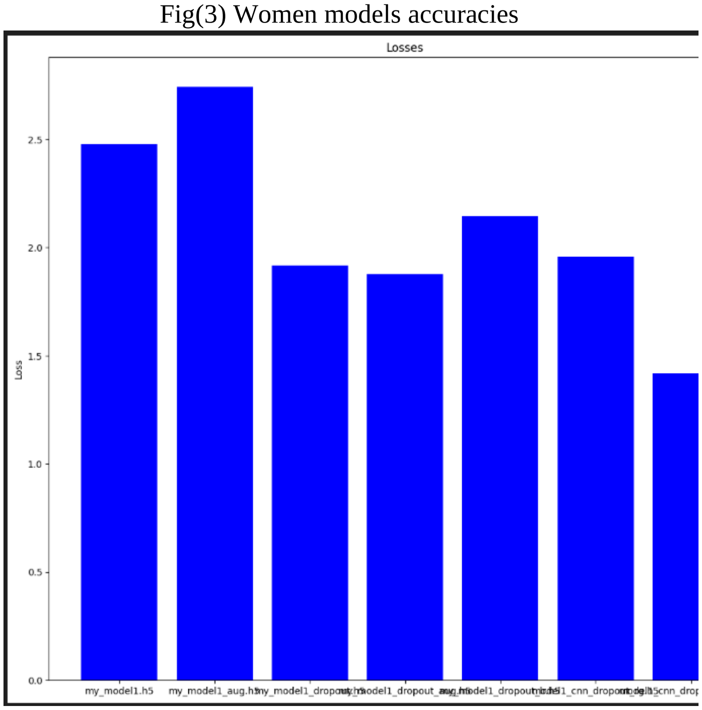
[8] *Facial Emotion Recognition Dataset*. (2023, August 1). Kaggle. https://www.kaggle.com/datasets/tapakah68/facial-emotion-recognition

[9] *Keras: The high-level API for TensorFlow*. (n.d.). TensorFlow. https://www.tensorflow.org/guide/keras

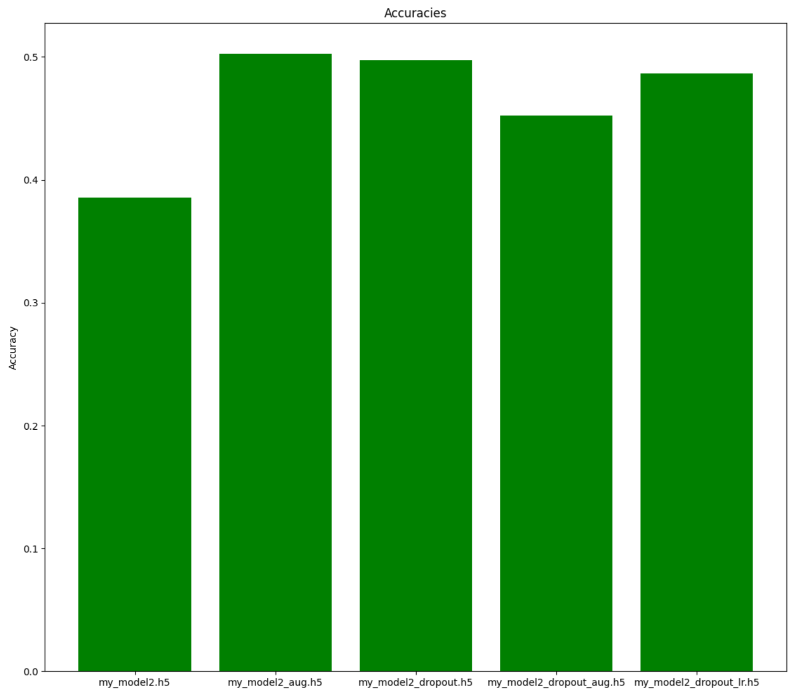
This section includes clear figures mentioned in the result section of this paper.



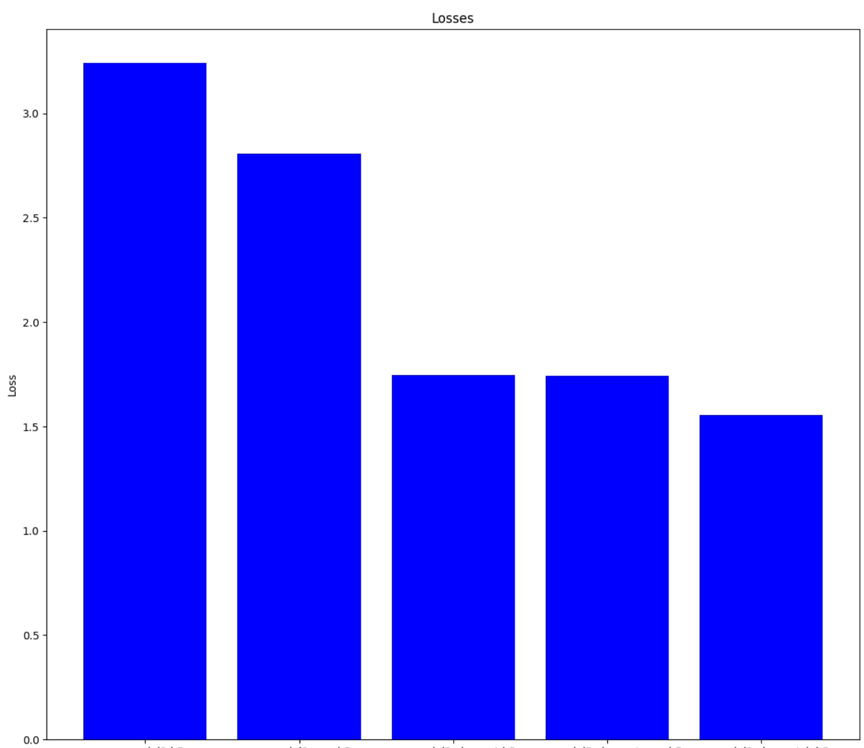
Fig(3) Women models accuracies



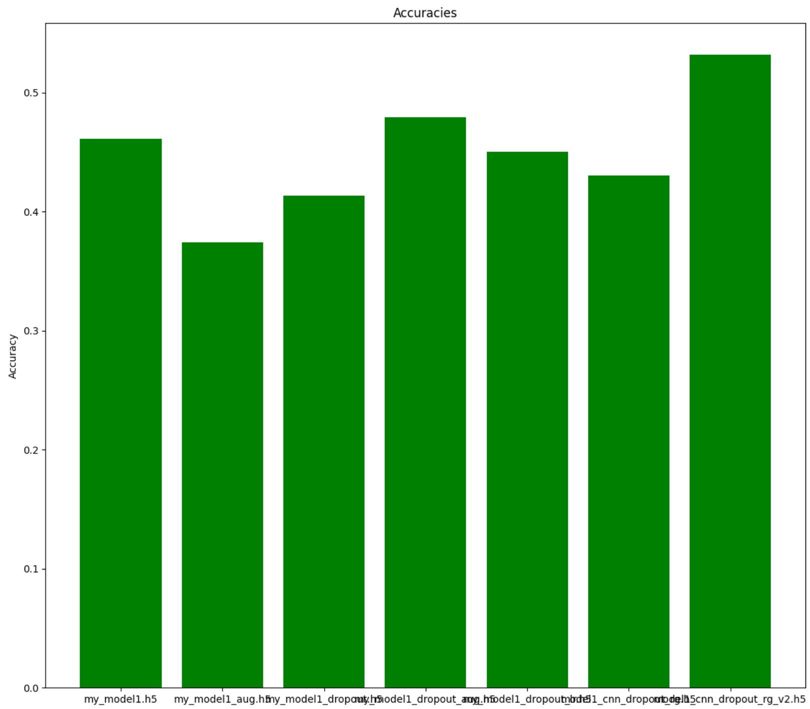
Fig(4)women`s model losses



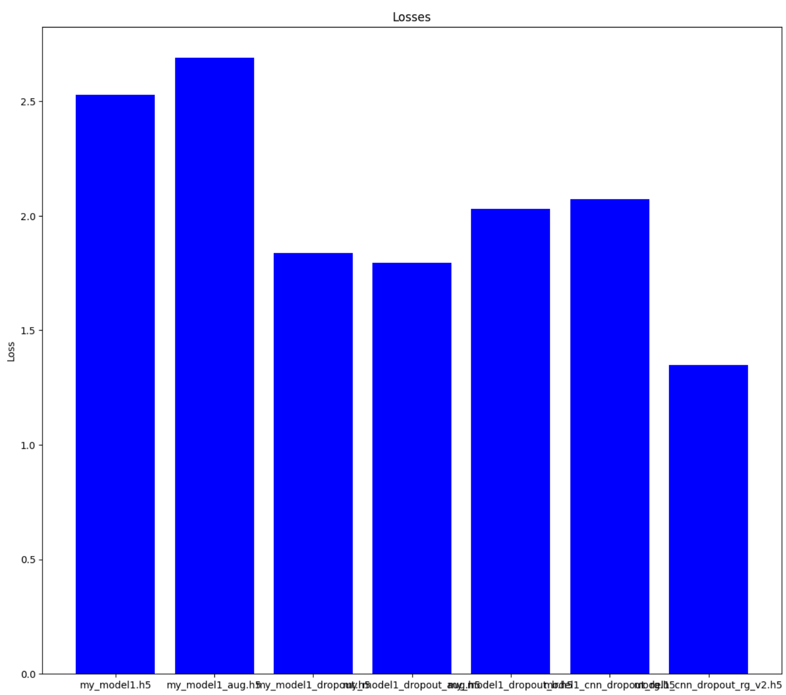
Fig(5) Men`s models accuracies



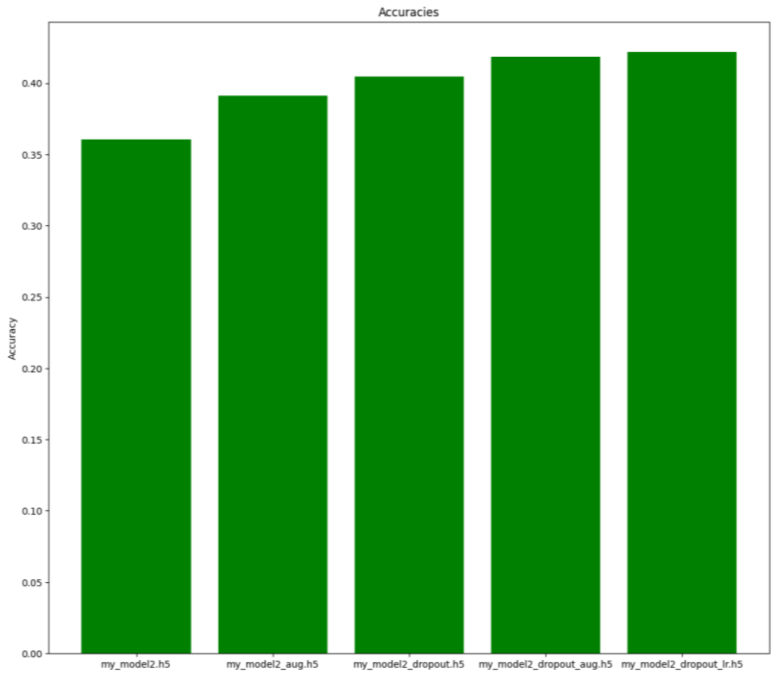
Fig(6) men`s model losses



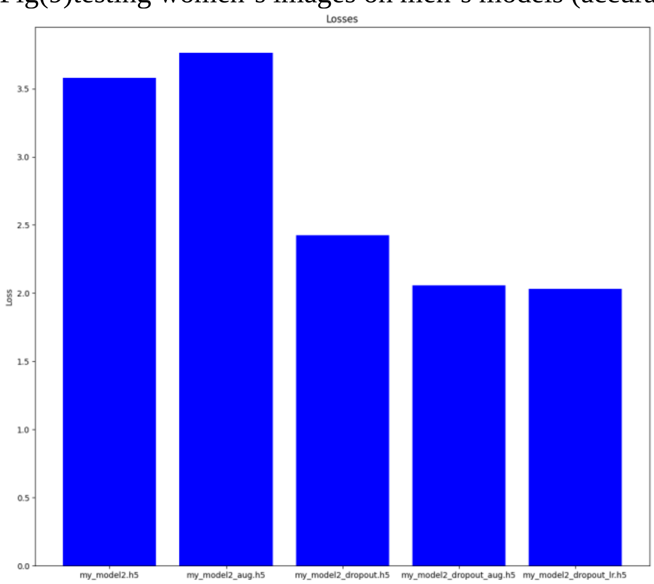
Fig(7) accuracy when testing men`s images on women`s models



Fig(8)loss when testing men`s images on women`s models



Fig(9)testing women`s images on men`s models (accuracy)



Fig(10) testing women`s images on men`s models (loss)