Implementation and comparison of gender classification models

Oscar Chen, Savith Jayasekera, Prince Okoli
Department of Electrical and Computer Engineering,
University of Calgary
Calgary, Alberta, Canada
Email: {oscar.chen1, scjayase, prince.okoli}@ucalgary.ca

Abstract—We propose the implementation and comparison of multiple machine learning models focused on predicting the gender from a data set of face images. The models will be based on five different supervised classification algorithms along with a model based on an artificial neural network (ANN). The algorithms include K-nearest neighbor (KNN), support vector machines (SVM), tree-based techniques, logistic regression, and convolutional neural network (CNN). With the trained models, the project aims to determine the best performing model for gender prediction and explore differences in model performance when predicting male and female faces.

Index Terms—Gender classification, Supervised classification, Artificial neural network

I. Introduction

With the increase in available user information through social media, automated gender prediction has become relevant to a variety of domains; these include human computer interaction, targeted advertising, and content-based searching [1]. Gender classification from facial images is therefore a thoroughly researched area in machine learning. Existing research have implemented gender classifiers based on traditional machine learning algorithms and ANNs. In this project, we will implement multiple machine learning models with the goal of exploring the differences in model performance. Specifically, the following topics will be explored:

- The accuracy of each model when trained on a data set of similar size. We aim to determine the performance and "how quickly" the model learns.
- The performance of the models when predicting male faces compared to predicting female faces. We aim to explore any possible differences between the predictions

The remainder of the proposal is structured in the following order: Section II reviews relevant studies and briefly discuss their methods and findings. Section III outlines the research questions of this study and describes the process of data collection, preparation, and analysis. Lastly, the expected results will be discussed in Section IV along with our key findings, limitations, and conclusions in Sections V and VI.

II. Related Work

Gender classification using a hybrid structure which includes convolution neural network and extreme learning machine was implemented by Duan et al. [2]. The model was trained on the MORPH-II image database and resulted in an accuracy of 88.2% which was higher than all the previous models trained on the same database.

Nazir et al. [3] proposed a method of using discrete cosine transform for feature extraction and training a KNN model with the extracted features. Using the Stanford university medical student frontal facial images database, the proposed technique was able to achieve an accuracy of 99.3% for gender classification.

Moghaddam and Yang [4] investigated the performance of a gender classifier based on SVM algorithm. It was trained on a low-resolution image set from the FERET face database. The paper concluded that the SVM model yielded 3.4% error rate and was shown to be superior to traditional pattern classifiers such as linear, quadratic and nearest-neighbor.

Basha and Jahangeer [5] explored a method for gender classification of face images using a random forest model. The model was trained on the ORL database containing 400 face images. They reported a classification accuracy of 100% outperforming other techniques such as SVM, linear discriminate analysis, KNN, and fuzzy c-means.

Gender classification has been an area of interest since the beginning of the field of artificial intelligence dating back to the 1950s; Rosenblatt, a pioneer of the formative years of neural networks, used gender classification based on images of men and women to showcase his revolutionary perceptron [6]. While the perceptron failed due to its simplicity of being one-layer neural network, the core theory of behind the device was the foundation of ANNs. With the increase in available data and computing power, a resurgence of the use of image processing techniques for gender classification has occurred since the mid-2000s. The authors of the project acknowledge that the proposed techniques have already been explored in detail; our goal is to gain an appreciation of the practical aspects of machine learning techniques by delving deeper into the implementation of different models.

III. Methodology

A. Research Questions

This project is focused on understanding and practicing the implementation of different machine learning algorithms. As such, the following Research Questions (RQs) revolving around implementation and performance of the models will be answered.

RQ1 – How accurate are the machine learning models at predicting gender?

This question aims to explore the accuracy of the models used during the project. The models will be trained and tested on the same training and testing sets from the data set. Each model will be evaluated and a direct comparison between the models will be provided.

RQ2 – How do the models perform when detecting male faces compared to female faces?

This question aims to find if the models perform better at detecting either gender. Two important metrics of binary classification will be used to answer this research question. Namely, the metrics are specificity and precision which can be calculated as shown below:

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

where:

TP	number of males correctly classified
TN	number of females correctly classified
FP	number of females incorrectly classified as male
FN	number of males incorrectly classified as female

B. Data Collection and Preparation Plan

For the project, the IMDB-WIKI face images data set will be used [7]. Within the data set, the group will focus on the images of the cropped faces. The entire data set contains 523,051 images which were gathered from IMDB and Wikipedia. The creators of the data set used a pre trained face detector to find faces within the images and create cropped images of the faces. The data set also contains the gender labels which will be utilized for our project. The group will use the cropped face images to train and test our machine learning models. The image data set will be uploaded to the University of Calgary Spark cluster where the data manipulation and the model training will be performed.

As the project is based on image processing, there is minimal data clean up required before training the models. The only step needed to clean up the data set will be to remove the data points that does not contain gender labels. For preliminary training, unique data sets will be created by randomly selecting 1%, 5%, 10%, 20% of the original data set. Since the project uses a large data set, even a 1% sample will contain a quantity large enough to train our models. The models will be tuned based on the performance results of the models trained using the smaller data sets before retraining the updated models with a larger data set. Once all the models are trained, the analysis will be performed to achieve the goals of the project.

C. Data Analysis Plan

Using the IMDB face images data set, the group aims to create machine learning models which will be able to predict the gender of the person; from the created models, our goal is to answer the above-mentioned research questions. The following supervised machine learning classification techniques along with one artificial neural network technique will be used to create the models:

- CNN
- SVM
- KNN
- Tree-based techniques
- Logistic regression
- Naïve Bayes classifier

With the smaller sample data sets that will be created as explained in the data collection and preparation section, multiple models will be trained to determine the best model for each of the above categories.

IV. EXPECTED RESULTS

RQ1 –The accuracy of each technique is expected to vary significantly. The linear regression classifier, Naïve Bayes classifier, trees, and the logistic regression models are expected to perform poorly given the nature of the features used in image processing. It is expected that the CNN and SVM models will outperform the other models given the group is able implement a robust model.

RQ2 – It is expected that the models will perform with similar accuracy for predicting either gender. An initial exploratory data analysis will provide more information about the distribution of the genders within the data set. With a skewed data set, the models may perform better for one gender compared to the other. The expectation is based upon the data set being equally distributed within the two genders; any difference in performance will be further explored should they occur.

V. DISCUSSION

A. Limitations

One limitation of the model is that it will be trained on facial images. For training, we will use the cropped facial images so that facial detector is not needed as pre-processing step of data preparation. This means that the model will need facial images with specific requirements to increase the accuracy of the prediction.

An area of interest in the field of computer vision is the use of video sequences as the data set. This requires both facial detection and an optimized techniques that can perform gender classification on streaming video. A possible technique to perform fast gender prediction is to treat the problem as a regression deep learning. Since the proposed CNN will be focused on gender prediction on facial images as a classification problem, the resulting models of the project will be unsuitable to be used with video streams.

B. Possible Extensions [Future Work]

The models created in this project will be trained on static images. A possible extension of this project will be to create a model that will work on video streams to determine gender.

Applications in surveillance and targeted advertising can utilize the models that work on video streams. Combining the gender classification with a facial detector as a pre-processing step is also another extension for this project. A facial detector can be used prior to the use of the trained model so that gender prediction can be performed on varied types of images such as full body images.

VI. CONCLUSION

Gender classification is an important area of interest in the field of image processing. The use of the gender classifier have many applications in fields. Our project is focused on creating gender classifiers based on both traditional machine learning classification and ANNs. Results of the project will be used to determine the best technique to be used for gender prediction. While the proposed techniques in the project has been researched previously, the knowledge gained from the implementation of the project can be used to extend the models to be used with video streaming to create a fast gender classifier which has the potential to have applications within multiple domains.

REFERENCES

[1] S. A. Khan, M. Ahmad, M. Nazir, and N. Riaz, "A comparative analysis of gender classification techniques," *Middle-East Journal of Scientific Research*, vol. 20, no. 1, pp. 1–13, 2014.

[2] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning cnn–elm for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, 2018.

[3] M. Nazir, M. Ishtiaq, A. Batool, M. A. Jaffar, and A. M. Mirza, "Feature selection for efficient gender classification," in *Proceedings of the 11th WSEAS international conference*, 2010, pp. 70–75.

[4] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, 2002.

[5] A. F. Basha and G. S. B. Jahangeer, "Exploring a novel method for face image gender classification using random forest and comparing with other machine learning techniques," *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 6, p. 8, 2014.

[6] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990.

[7] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 10–15.