TOPIC: COMPARATIVE ANALYSIS OF MODEL FOR MULTI-TASK LEARNING FOR AGE GROUP ESTIMATION AND GENDER RECOGNITION USING FACIAL FEATURES

MODELS: CONVOLUTIONAL NEURAL NETWORK (CNN) AND SUPPORT VECTOR MACHINE (SVM)

Akinrele Taye Oluwabusayo

Computer Science Department

Email: [tayeakinrele@gmail.com](mailto:tayeakinrele@gmail.com)

ABSTRACT

Human facial image processing has been an active and interesting research issue for years. Since human faces provide a lot of information, many topics have drawn lots of attentions and thus have been studied intensively.

# **CHAPTER ONE**

# **INTRODUCTION**

## **Background of Study**

The human face conveys much information, which people have an astonishing ability to extract, analyze and decipher (*Ali Maina, et al, 2016*). The main characteristic feature of the human being is the face (*Sayantani, 2015*), which exhibits different emotions that can be determined and easily predicted by the several facial expressions. Just by glancing at a person’s face, one can estimate or predict the age and gender of that person. Identifying human faces and modeling the distinguishing features of human faces that contribute most toward face recognition are some of the challenges faced by computer vision and psychophysics researchers (*M. R. Dileep & Ajit Danti, 2018*).

Age and gender are significant properties regarded as a crucial biological characteristics, which plays a fundamental role in human social interaction. The human face contains a wide range of information for gender perception and age estimation. (*Ke Zhang et al, 2019*). Classification of age and gender is an important visual task for human beings, since many social interactions critically depend on the correct age and gender perception.

As technologies such as visual surveillance and human computer interaction evolve, computer vision systems for age and gender classification plays an important role in our lives, it is therefore not surprising that a lot of researches has been done to investigate age and gender classification from face perception in humans, proposing various methods in order for a machine to attain human level of accuracy (*Md. Nurul & Emon Kumar, 2018*). Predictably, it is challenging for machines to identify these visual information, since discriminative feature extraction is easily affected by various factors like large variations in facial gestures, lighting, background, etc. Perceived age and gender classification is a topic with a high application potential in areas like surveillance, face recognition, video indexing and dynamic marketing surveys.

This research attempts to compare two multi-task learning models Convolutional Neural Network (CNN) and Support Vector Machine (SVM), in terms of performance while classifying humans into age groups and gender. The proposed approach treats gender prediction and age estimation as a classification problem, adopting datasets which composes of images labeled for age and gender. It uses eight classes of age group (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+) and classifies gender into one of the two sex labels, Male (M) and Female (F).

## **Research Motivation**

The primary motivation behind this research is due to the observation that the amount of data available for the study of computer vision problem, can have an enormous impact on the machine capabilities developed to solve it (*Prajakta, Dr. G. S. Sable, 2018*). In view of this, this approach makes use of different datasets to train two multi-task learning models Convolutional Neural Network (CNN) and Support Vector Machine (SVM), comparing their performances, in other to generate an accurate result in the prediction of gender and age groups from input images. Some datasets that will be used in this paper includes IMDB-Wiki dataset (523,051 face images), Adience dataset (34,795 images) and UTKFace dataset (over 20,000 images).

This research aims to develop a web-based assessment system that will solve the problem of collection of data from manual sources like mobile phones or computer.

## **Research Objectives**

The specific objectives of this research are to:

1. Compare these two models of multi-task learning Convolutional Neural Network (CNN) and Support Vector Machine (SVM).
2. Design a CNN and SVM model to jointly recognize the gender and age group of a person through an input image.
3. Compare the models using Adience image datasets.
4. Design and implement a web-based model evaluation system.

## **Research Methodology**

This research proposes an effective method for human gender and age prediction/classification from the given facial images using the Convolutional Neural Network and Support Vector Machine models. The methodology includes: data preparation, data pre-processing, feature extraction and classification/prediction.

### **Data Preparation**

This is the process of transforming raw data so that data scientists and analytics can run it through machine learning algorithms to uncover insights or make predictions (*Wiki*). In this research, datasets are needed for training and testing the models. The organization of these dataset is also very essential since each of them must follow the same pattern of age group classification group (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+) and gender classification (Male and Female).

In this research, the following datasets are used to train the CNN and SVM models:

1. **IMDB-WIKI dataset**: this is the largest publicly available dataset of face images with gender and age labels for training, it is composed of 2 datasets that were created by the same authors, so they both have similar structure in terms of available metadata. It provides pretrained models for both age and gender prediction containing 460,723 images in IMDB and about 62,328 in Wikipedia, giving a total of 523,051 facial images. (*Rasmus Rothe et al, 2015*)
2. **Adience Dataset**: is composed of pictures taken by camera from smartphones or tablets. The images of this dataset capture extreme variations, including exreme blur (low-resolutions), occlusions, out-of-plane pose variations, expressions. The entire Adience dataset contains 34,795 images of 2,284 subjects. (*Jia-Hong Lee, et al, 2018*). Each image is annotated with gender and one of 8 age groups.
3. **UTKFace dataset**: is a large-scale face dataset with long age span (ranging from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occusion, resolution (*Kaggle.com*). (*UTKFace, 2018*).

### **Data Pre-processing**

Before the face can be classified into its appropriate age group and gender, some preprocessing is needed because the classifiers are usually sensitive to imperfection of the image. Therefore, the face image is processed to obtain a transformed face image to increase the quality of the face image, retaining the important characteristics.

In this stage, brightness and contrast are normalized, the face image geometric features are improved and also the image size (number of pixels) is reduced (*Zofia Stawska, 2016*).

Pre-processing includes three steps: resizing the image, converting to gray scale and noise reduced image. The input image is resized and the color images are converted to gray scale.

### **Feature Extraction**

This is considered to be an indispensable process which is needed to be followed and implemented to accomplish the task of age and gender detection and its classification.

Extraction of features from images can be achieved by adapting the following approaches:

1. **Geometry-based Approach**: which uses geometric information such as features relative positions and sizes of the face components as a features measure. It requires finding the face characteristic points like nose, mouth, eyes, ears or hair. These points are called fiducial points.
2. **Template-based Approach:** in which previously designed standard face pattern template is used to match with the located face components. It works on the image pixels that were previously transformed on the local or global level. At the local level, image can be divided into lower windows or specific face regions such as mouth, nose or eyes. This method preserves natural geometric relationships which can be used as a naïve feature.

### **Model Implementation**

A Convolutional Neural Network and a Support Vector Machine model is implemented to perform the gender and age classification. The gender has two classes- male and female and age has eight classes- [(0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 – 32), (38 – 43), (48 – 53), (60 – 100)].

1. **Convolutional Neural Network Model:** The CNN model to be implemented contains three convolutional layers, each followed by a rectified linear operation and pooling layer. The first two layers also follow normalization using local response normalization. The first Convolutional Layer conatins 96 filters of 7 x 7 pixels, the second Convolutional Layer contains 256 filters of 5 x 5 pixels. The final Convolutional Layer contains 384 filters of 3 x 3 pixels. Finally, two fully connected layes are added, each containing 512 neurons (*Gil Levi, Tal Hassner, 2015* ) . (After the input image is fed into the network, a filter (3 x 3) is passed through it and a convolved featured is generated after which an activation function is passed, then it is pooled, flattened and lastly the accurate gender and age group is predicted.)
2. **Support Vector Machine Model:** The SVM model to be implemented for the classification of gender the binary classifier SVM and for age, multi-class SVM is used.

Input Layer

Convolution Layer

Convolution Layer

Convolution Layer

Flatten/Global Pool

Dense Layer

Dense Layer

Dense Layer

Dense Layer

Age

Gender

Figure 1.1. Model Developed for Multi-task Learning Convolutional Neural Network

Figure 1.2. Model Developed for Multi-task Learning Support Vector Machine

START

Input Jpeg Image Set of Male and Female

Convert Each Image to Grey Scale

Perform Feature Extraction to extract ‘lip’ from individual image

Shuffle the content of the feature vector

Convert the dimension of the extracted feature from 2D to 1D and form

Cross validate the data set

Associate a class label with each row vector and form feature vector

Train a part of the Input data set along

with the class label

Test data along with the trained model to classify the image set

Output the class labels of classified data set

END

## **Contributions to Knowledge**

It is expected that this work will do the following:

1. Show whether a Multi-task Convolutional Neural Network performs better and more efficiently than a Multi-task SVM classifier in recognizing the gender and estimating the age group of humans given the facial features in their pictures or not.
2. Add to the collections of unfiltered images dataset through the web-based evaluation system.

# **CHAPTER TWO**

# **LITERATURE REVIEW**

## **Background of Study**

Automatic gender and age classification from facial image has become an attractive research area in the field of machine learning, it is one of the most promising research areas since the couple of decades. Various methods have been proposed for gender and age recognition in both controlled and uncontrolled situations. In uncontrolled situations, many problems arises such as high rate of noises, lack of illumination etc. which results in distorted and blurry images (*Md. Nurul, 2018*). These methods try to mitigate these problems, however, some of this methods are computationally expensive and also complex.

This research makes use of two methods already established, Convolutional Neural Network (CNN) and Support Vector Machine (SVM), comparing the accuracy of their results while predicting age and gender from input images. Some crucial topics and keywords, associated with this research will be discussed below.

## **Gender Prediction**

Recognition and classification of gender automatically is a fundamental task in computer vision, it plays an essential role in a wide range of real-world applications such as targeted advertisement, forensic science, visual surveillance, content-based searching, human-computer interaction systems, etc. (*Khalil Khan, et al, 2019*). Gender prediction is the gateway tool for any application that takes advantage of it to improve their functionality by doing targeted interaction with the gender of its preference and it can also be useful in other applications, where eliminating false identification on the basis of mismatched gender could be of use to the motive of that application (*Aakash, et al, 2019*). However, gender classification is still an arduous task due to various changes in visual angles, face expressions, pose, background, and face image appearance (*Khalil Khan, et al, 2019*) as shown in Figure 2.1.

One common limitation of many gender classification systems is that they cannot account for individuals who do not identify as either a woman or a man, and they have no concept of gender identity as separate from physical appearance (*Stefan, Emma, 2019*). <https://www.pewresearch.org/internet/2019/09/05/the-challenges-of-using-machine-learning-to-identify-gender-in-images/>

Generally, gender recognition is treated as a binary classification problem, in which the obtained facial image is assigned to one of two classes (male and female), (*Andrey V. Savchenko, 2019*).

The output layer in the gender prediction network is a type of softmax with two nodes indicating the ‘Male’ and ‘Female’. (https://www.learnopencv.com/age-gender-classification-using-opencv-deep-learning-c-python)

## **Age Prediction**

Age estimation has many useful applications, such as age-based face classification, finding lost children, surveillance monitoring, and face recognition invariant to age progression (*Young-Sik, et al, 2014*). Age has been investigated as a soft biometric and facial attribute, and age prediction has historically been one of the most challenging problems within the field of facial analysis (*Ke Zhang, et al 2019*).

Normally, Age Predication should be approached as a Regression problem, since a real number (such as 3, 9, 20, 48, 67 etc.) is expected, but, accurately estimating age using regression is very challenging. Even human beings cannot accurately predict the age of a person just by looking at that person. However, we can easily predict whether they are in their 20s or 30s, Due to this reason, treating age estimation as a classification problem is better, since estimating the age group a person is in is easier to predict. The Adience dataset (used to train the models), has 8 classes divided into the following age group [(0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (60-100)]. Thus, the age prediction network has eight (8) nodes in the final softmax layer indicating the age ranges above. (https://www.learnopencv.com/age-gender-classification-using-opencv-deep-learning-c-python).

There are several factors that affects accuracy in age prediction, they include varying illumination conditions, pose variations, blurring, image resolution and modality, noise and so on.( <https://www.intechopen.com/>) An illustration is shown in Figure 2.2.

## **Classification Methods**

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e., learning where a training set of correctly identified observations is available. (Wikipedia) “<https://en.wikipedia.org/wiki/Statistical_classification>”, It involves computer program learning from the data input given to it and using this data to classify new observation. (Mandy, 2017) “<https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14>”.

Classification may be binary or multi-class where the former can be used to identify whether a person is male or female (two classes), and the latter is used to identify the age group of a person (more than two classes). Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, age recognition, document classification, gender prediction, etc. (Mandy, 2017).

There are different types of classification algorithms in machine learning, they include: Linear Classifiers (Logistic Regression, Naïve Bayes Classifier), Nearest Neighbor, Support Vector Machines, Decision Trees, Boosted Trees, Random Forest, Neural Networks.

In this research, two of these algorithms are used and compared, Convolutional Neural Network (CNN) and Support Vector Machine (SVM), in the classification of age and gender.

### **Convolutional Neural Network**

A neural network consists of units (neurons), arranged in layers, which converts an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally, the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer (Mandy, 2017).

Convolutional Neural Network (CNN or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery (Wikipedia). A ConvNet is a deep learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods, filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics (<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>).

They have applications in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing and financial time series. (Wikipedia). (<https://en.wikipedia.org/wiki/Convolutional_neural_network>).The role of the ConvNet is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.

CNN image classifications takes an input image, process it and classify it under certain categories. Computers sees an input image as array of pixels and it depends on the image resolution, based on the image resolution, it will see h x w x d (h = Height, w = Width, d = Dimension). E.g., an image pf 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.

Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, Fully Connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. (<https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>). (Prabhu, 2018)

1. Convolution Layer – The Kernel: is the first layer to extract features from an input image. It preserves the relationship between pixels by learning image features using squares of input data. This layer is a mathematical operation that takes two input such as image matrix and a filter or kernel. The image below illustrates what this layer does.



Figure 2. Convoluting a 5x5x1 image with a 3x3x1 kernel to get a 3x3x1 convolved feature

Some key terms in this layer are stride, padding and ReLU. Stride is the number of pixels shifts over the input matrix (Prabhu, 2018), padding is used when filter does not fit perfectly into the input image, it is of two types (same padding or valid padding) (Sumit Saha, 2018) (<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>), ReLU (Rectified Linear Unit) for a non-linear operation, it is an activation function, its purpose is to introduce non-linearity in the ConvNet, since the real world data would want the ConvNet to learn non-negative linear values.

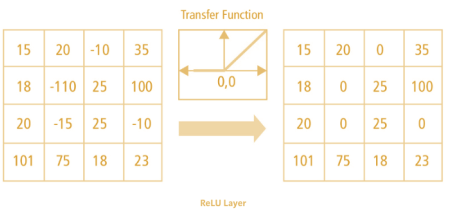


Figure 2. ReLU Operation

1. Pooling Layer: this layer is responsible for decreasing the special size of the convolved Feature in other to decrease the computational power required to process the data through dimensionality reduction. It is also useful for extracting dominant features which are rotational and positional invariant, thereby maintaining the process of effectively training of the model. They are of two types: Max Pooling and Average Pooling, the former returns the maximum value from the portion of the image covered by the kernel, while the later returns the average of all the values from the portion of the image covered by the kernel. This terms are illustrated in the figure below.



Figure 2.

### **Support Vector Machine**

Support Vector Machine is a linear model for classification and regression problems. The Figure 2. below describes how the SVM classifier works. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is straightforward: the algorithm creates a line or a hyperplane which separates the data into classes. (<https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>)

SVM at first approximation, finds a separating line (or hyperplane) between data of two classes. This algorithm takes the data as an input and outputs a line that separates those classes if possible. A typical illustration is shown in Figure 2. below.

(<https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>).

For age and gender classification, SVM takes the training datasets and finds a hyperplane between the eight classes, for age - [(0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 – 32), (38 – 43), (48 – 53), (60 – 100)] and two classes for gender- Male and Female, it tries to make a decision boundary in such a way that the separation between the classes is as wide as possible and then it predicts the accurate age and gender based on its result.

**Advantages of SVM**

1. It is effective in high dimension spaces.
2. It is still effective in cases where the number of dimensions is greater than the number of samples.
3. It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
4. It is versatile, i.e. different Kernel functions can be specified for the decision function, common kernels are provided, but it is also possible to specify custom kernels.

**Disadvantages of SVM**

1. If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
2. SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

(<https://scikit-learn.org/stable/modules/svm.html>), (<https://www.simplilearn.com/classification-machine-learning-tutorial>).

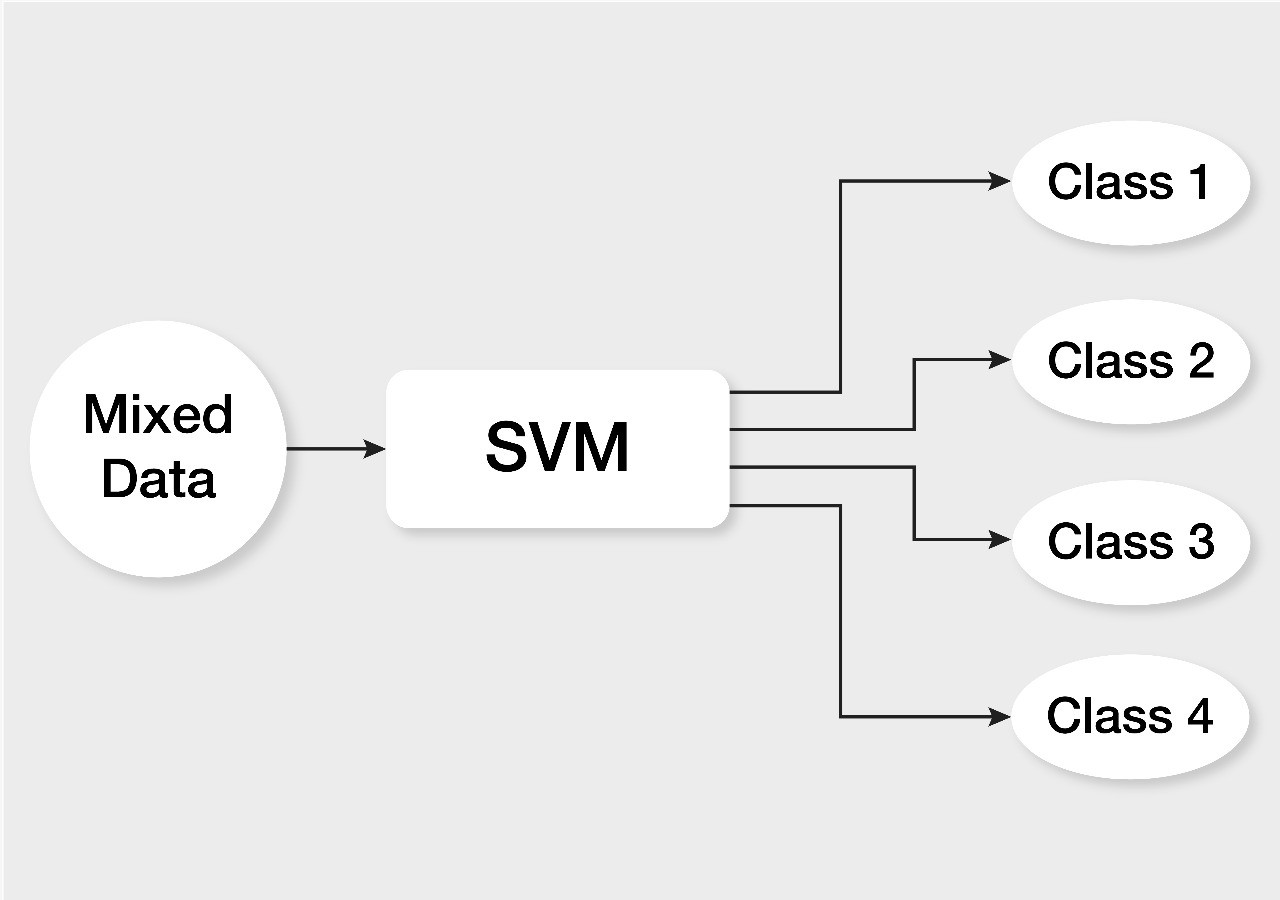


Figure 2. The Support Vector Machine Classifier

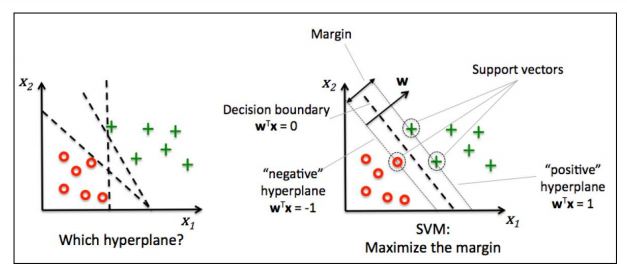


Figure 2. SVM illustrating hyperplane for data of two classes.

## **Differences between CNN and SVM**