**Chapter Two**

**Literature Review**

**2.1 Introduction**

This chapter talks about the review of related works on age and gender estimation method. Also it makes a discussion on all the reviewed literatures.

**2.2 Review of Related Works**

The application of machine learning has rapidly advanced in various fields such as speech recognition, text and image recognition and so on.

An example of image classification using machine learning is demonstrated by Pratik Devikar (2016). He used the pretrained model, Inception v3 to do transfer learning. The dataset used to retrain the model is obtained from Google images. The aim of this research was to train a model which can recognize and differentiate 11 types of dog breeds. Hence, he prepared 11 types of datasets, each datasets comprised of 25 slightly different images of particular dog breeds. To ensure the uniformity of the datasets, the images were set to a resolution of 100×100 pixels. Throughout the experiment, he implemented Python programming language and import TensorFlow library to conduct classification task. The accuracy score was generated by using SoftMax algorithm. The resulting accuracy of testing he achieved reached 96% (Devikar, 2016).

Another example that utilised machine learning in image recognition is demonstrated by Tapas (2016). The aim of this experiment is to classify plant phenotyping. He used the pretrained model, GoogleNet to do retraining. The dataset was extracted from the database of Computer Vision Problems in Plant Phenotyping (CVPPP 2014) database. The dataset comprised of 3 categories, 2 on Arabidopsis, with 161 images and 40 images respectively. Another category of the dataset was Tobacco species, which consisted of 83 images. The retraining process is conducted via TensorFlow library and python as a programming language. The output was displayed in probability using the computation of SoftMax function. The results of accuracy based on testing image reached 98%.

In addition, a similar study on the flower classification using Inception v3 worth a consideration. The study was based on Inception v3 model of TensorFlow platform. The experimental datasets were acquired from two sources, Oxford-17 database, which consist of 17 categories of flowers and Oxford-102, which consist of 102 categories of flowers. The results depicted by SoftMax function regarding the possible output with the input of testing mages are compared according to the two types of dataset. The result shows model trained under Oxford-17 dataset reach 95% of accuracy whereas Oxford- 102 dataset gives an accuracy of 94% (Xia & Nan, 2017).

According to Chin et al. (2017), a research on intelligent image recognition system for marine fouling using SoftMax transfer learning and the deep convolutional neural network was done. They implemented transfer learning by retraining Google's Inception v3 model and SoftMax as an output of prediction based on image input. The images were processed by Open Source Computer Vision Library (OpenCV) and the retraining process is done with the help of TensorFlow Library. At the beginning of the process, Raspberry Pi 3 captured image of the marine fouling. The image was then uploaded to cloud to be classified by the retrained Inception V3 model and convolutional neural network. Then, the image was processed and the percentage of the area of macro fouling organisms was determined. The percentage in the range of 25-40% was considered as heavy fouling and cleaning process must be conducted. The datasets were obtained from captured images from the web. The model was retrained to classify 10 classes of fouling species, with dataset size in the range of 82-228 images. In order to enhance the accuracy of the model, the model was trained twice. Results show the lowest improvement in percentage is 10.302% where the highest can reach 41.398% of improvement. Upon testing on the reliability of the trained model, the highest accuracy achieved among the 10 classes of fouling species was rock oysters, which can reach 99.703% correct prediction. On the other hand, finger sponge species possessed the lowest accuracy, which is 76.617%.

According to Tamkin et al (2013), they claimed that diabetic retinopathy can be detected with the application of deep Convolutional Neural. They extracted a dataset from Kaggle competition database. The database was chosen because the images are taken in various conditions, including different cameras, colours, lighting and orientation. The more variety was the images sources, the higher the robustness of the trained model. A total of 35,126 images, with the size of more than 38 gigabytes is separated in the ratio 0f 8:2, whereby 80% of the images are used as training set, 20% are used as testing set. All images were resized to 256 pixels x 256 pixels. The highest accuracy achieved at the end of the experiment is 92.59%.

According to Thukral, et al (2012), they proposed a hierarchical method to estimate human age. The datasets were obtained from FG-Net website. Upon gathering the dataset, they grouped the images into 3 major groups, in the ranges of 0-15, 15-30 and 30+years old respectively. The experiment can be divided into 3 steps, feature extraction, regression and classification. In feature extraction, facial landmarks points at corners or extremities of eyes, mouth and nose are extracted. Regression was conducted by determining the independent variable, x and the dependent variable, y. Next, they used the Relevance Vector Machine (RVM) regression model conduct machine learning according to the age groups. After that, in classification phase, they utilised 5 types of classifiers, including μ-SVC, Partial Least Squares (PLS), Fisher Linear Discriminant, Nearest Neighbor, and Naïve Bayes to classify the images into the correct age group. Results showed that if the classifiers were able to classify the images into correct age group, the age estimation task by RVM can perform more accurate, which can reach 70% of accuracy.

According to Jana, et al (2013), they claimed that facial features can be used to estimate age group. Their experiment involved 3 stages: pre-processing, feature extraction and classification. During the pre-processing phase, they prepared datasets by taking images of 50 persons by using a digital camera (Nikon Coolpix L10). The face images were cropped, and the positions of eye pair, mouth, nose and chin were detected. During feature extraction, global features such as distances between 2 eye balls, eye to nose tip, eye to chin and eye to lip were determined. 6 types of ratios are then computed by referring to the distances obtained. After that, classification is carried out by using K-means clustering algorithm. Results showed ratio obtained using pixels (F5) was most reliable, with an accuracy of 96% when the samples are separated into 2 age groups, 84% of accuracy is obtained for 3 age groups and 4 age groups had accuracy of 62%.

**2.3 Discussion**

From the above, the hence mentioned authors and their respective journals, they worked toward the objects estimation on the digital images. Some of them uses and implemented transfer learning that is pretrained models which includes Inception V3, GoogleNet and using SoftMax, many of them achieved accuracy over 90%. Some Authors proposed a related work to our study area by hierarchically method to estimate human age. They obtained the datasets in FG-Net website. They group the images in 3 major groups for training the model ranges (0-15), (15-30) and (30+). Their experiment also were performed in 3 steps, feature extraction, regression & classification.