

1. Introduction

In Korean Medicine, a human body type has been classified into four types such as Jung, Ki, Shin and Hyul according to the shape of face, which is geometrically close to circle, rectangle, triangle, and ellipse, respectively. In this paper, one of typical machine learning algorithm, Random Forest, is exploited to generate the classification model which can classify the body type based on the front face image.

2. Methods and materials

2.1. Data

In 2019, Nation Information Society Agency(NIA) photographed 1,000 Korean front facial image and opened to the public(See Figure1). Jeju Institute of Korean Medicine obtained the data and labelled each image into one of four body types. The resolution of each image is 864×576 pixels.

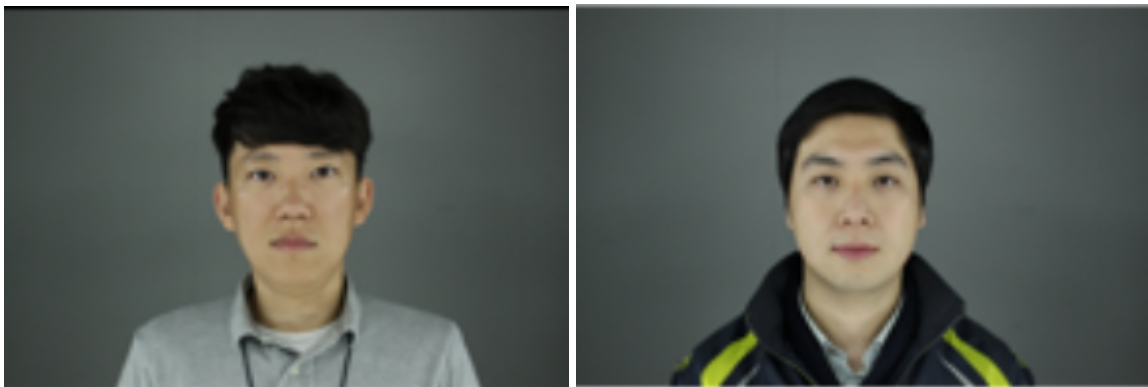


Figure 1. Sample images of 1,000 Korean front faces

2.2. Data Preparation

A machine learning algorithm requires explanatory variables, so called, features and response variable. In this research, features are extracted from distances among landmarks detected by the face detection algorithm of Shiqui Yu's on the statistical package R environment which is most widely used one due to its efficient and highly accurate precision, 99% even under the heavy occlusion and extreme illumination(See Figure 2).



Figure 2: Five landmarks(Left/right eyes, nose tip, left/right ends of lips) detection

Shiqui Yu's algorithm is known as YuFaceDetectNet based one CNN(Convolutional Neural Network) which is modified to use depth-wise and point-wise convolutions. It

extracts the X/Y coordinates of the five landmarks.

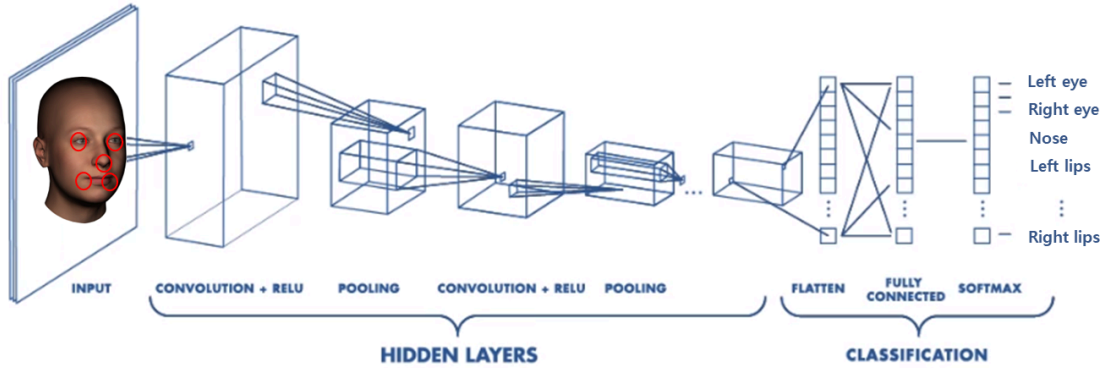


Figure 3. Shiqui Yu's algorithm

Ten features(= $_5C_2$) are derived by calculating the distances among five landmarks using Euclidean distance. Therefore, the distance between landmarks i and j is calculated by the following equation:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

2.3. Body type classification model

A Random Forest is a well-known machine learning algorithm to avoid the over-fitting with fast data training and ensembles of independent decision trees.

Decision tree method is a supervised learning method used for classification and regression. The objective is to learn straightforward decision rules derived from the data features such as explanatory variables in order to build a model that predicts the value of a target variable(See Figure 3).

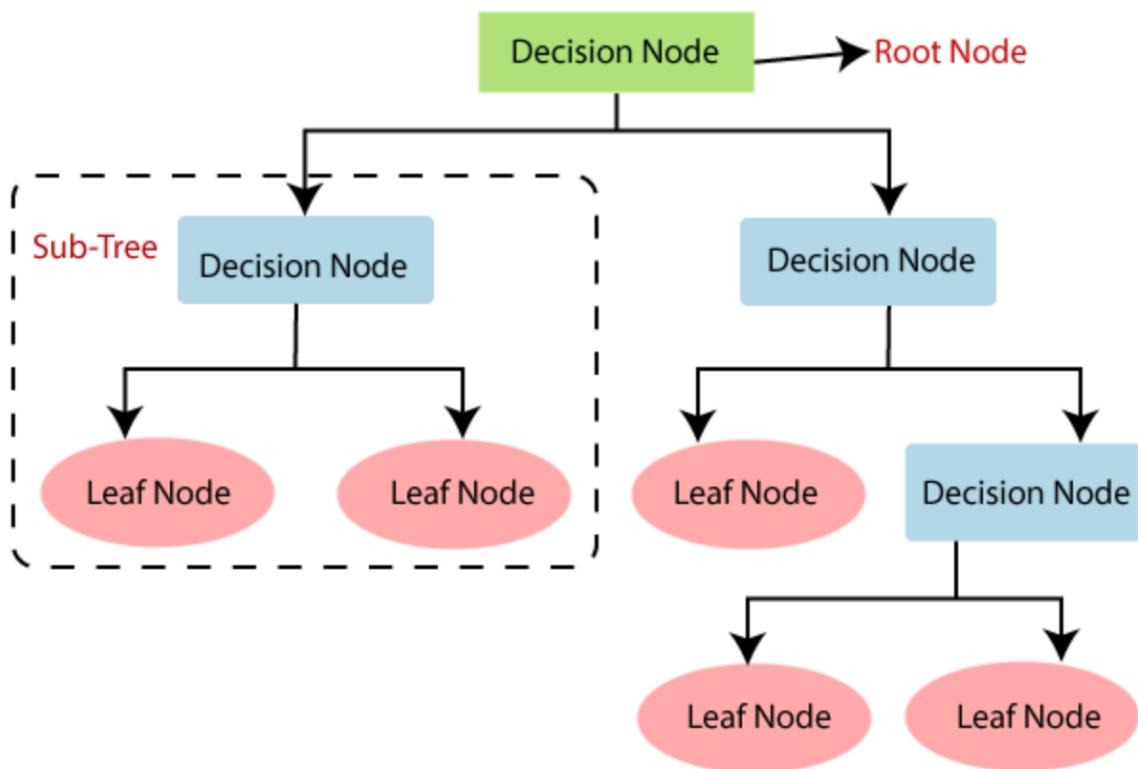


Figure 3: Decision tree finds out the rules to predict the class of target variable by the use of explanatory variables

2.3. Environment for computations

All the calculations need for the analysis is executed by the statistical package, R and mac i3 processors.