National College of Ireland

Multi-label Image Classification to Detect Air Traffic Controllers' Drowsiness Using Facial Features

MSc Research Project
Data Analytics

Nevin Saini
x18132260

School of Computing
National College of Ireland

Supervisor: Dr. Cristina Muntean

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Nevin Saini		
Student ID:	x18132260		
Programme:	Data Analytics	**Year:**	2018-19
Module:	Research Project		
Supervisor:	Dr. Cristina Muntean		
Submission Due Date:	12/12/2019		
Project Title:	Multi-label Image Classification to Detect Air Traffic Controllers' Drowsiness Using Facial Features		
Word Count:	9084	**Page Count**	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Nevin Saini

Date: 12/12/2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	x
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	x
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	x

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multi-label Image Classification to Detect Air Traffic Controllers' Drowsiness Using Facial Features

Nevin Saini
x18132260

Abstract

With the increase in globalization, there is a rise in the frequency of airplanes activities and air traffic request causing immense financial profits to the air carrier. To maintain the same progression, it is necessary to consider the safety of passengers that require highly skilled Air Traffic Controller who can clearly communicate with the pilots and sustain the safety of crew and passengers. Also, working as an ATC (Air Traffic Controller) involves extreme pressure due to the safety of passengers and over load of work which leads to drowsiness and fatigue among the staff affecting the quality of work and endangering the life of travellers. This study intends to design a drowsiness detection model to overcome the above issue by using numerous modeling techniques like Convolutional Neural Network, Support Vector Machine, K-Nearest Neighbor and some ensemble techniques like XGBoost and Random Forest which can detect and classify facial features like Closed eyes, Yawning and Open eyes of an ATC. Two different models: Baseline and Tuned/Improved model after hyperparameter tuning are designed to improve the performance of an baseline algorithm and finally, these techniques are assessed on the basis of performance measures like Accuracy, Precision, Recall and F1 Score on the dataset used in this study and, it is observed that Convolutional Neural Network has outperformed other machine learning models and obtained an accuracy of 98.6% without overfitting.

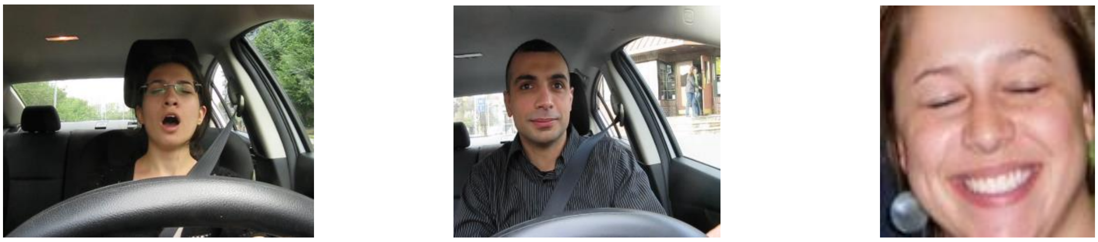
Keywords: Drowsiness Detection System, Aviation, Convolutional Neural Network, Hyperparameter Tuning

1 Introduction

The aviation industry, one of the most regulated businesses in the world, is an incessant industry which is subjected to supervisory decisions in the area of safety and security. ATC (Air Traffic Controller) has a vital role to play in averting the collisions among aircrafts, maintaining strategic air traffic flows and supplying pilots with important information for safe flight landings and take offs (Moon, Yoo and Choi, 2011). Also, it is the airplane controllers who secure several activities of airplane and they get drowsy, tired and more susceptible to stress due to the intricacy of job (Vogt et al., 2004).

Numerous sophisticated drowsiness detection systems are designed for top-notch cars to alert the sleepy driver using sensors. Also, the possibility of designing a model which can detect sleepiness and avert dangerous accidents with a decent accuracy is there by utilizing the cutting edge technologies like data mining, machine learning and deep learning. This research revolves around securing the innocent lives of aircraft crews and passengers by keeping a track of sleepiness of Air Traffic Controllers by designing a cost-effective model to detect sleepiness of Controller using machine learning algorithms. In addition, the drowsiness detection system will have significant effects on the well-being of airplane passengers and aircrew as it will dodge aviation disasters and also increases accountability and concentration of ATC towards their job.

The Air Traffic staffs continuously work late nights, on weekends and holidays without regular break from the job. As per Vogt et al. (2004), blend of extensive job hours, high amount of work and great responsibility for safety makes aviation controllers a distressing and demanding profession. As a result, ATC feel drowsy, sluggish and often loses focus on workstation screen resulting in decline in communication level causing a possibility of change of information moved to pilots while important airplane activities which prompts calamitous effects. Also, according to the study conducted by NTSC (National Transportation Safety Board), in the midnight shifts, 60% of controllers felt somnolence and lack of focus. In addition, multitude of concerns about collisions and near-misses due to drowsiness of ATC (Air Traffic Controller) has been posed in the past; including the four close misses that occurred in the United States and an instance of a fatal air crash in Kentucky due to the loss of air traffic controller's situational alertness. Chunks of these issues happened because of extensive work load and overtime of representatives initiating lack of sleep and exhaustion (Monk, 2007). Furthermore, it is evident that sleepiness and fatigue causes diminished performance and concentration at work which appears to pose a peril to the safety of the travellers and it needs to be addressed which needs a few countermeasures.



(a) Yawning (b) Open Eyes (c) Closed Eyes

Figure 1: Three different classes used in the study to detect drowsiness of an Individual

The main goal of this project is to find solution to this research question:

"How can several modeling techniques (SVM, KNN, Random Forest, XGBoost and Convolutional Neural Networks) can be used to perform multi-label image classification of some facial characteristics for detection of an Air Traffic Controls' drowsiness which can prevent catastrophic airplane accidents and near-misses?"

This inquiry of research question looks to save various lives by averting aircraft glitches and accidents due to somnolence and absence of thought of Air Traffic Controllers. For solving this problem, a well-designed model is created which is specified in section 5 that involves the use of many classification models to correctly identify sleepiness by considering facial features like closed eyes, yawning and open eyes as represented in Figure 1 with a decent accuracy.

To solve the above research question, facial attributes like eyes and mouth are considered for solving the problem. All the image datasets are divided into training and testing ratio before image pre-processing where several techniques like rescaling, rotation, brightness and canny edge detection are applied to improve the performance of the model and further avoid the over-fitting problem. Following this, several classification algorithms like Convolutional Neural Network, Support Vector Machine, K-nearest neighbor, XGBoost and Random Forest are designed on the pre-processed training images before applying the models on the validation dataset.

This study elucidates the different parts of the research project including the literature review in Section 2 followed by the steps required to execute the project in the form of Methodology described section 3. Also, a well-designed architecture is presented in section 4 which clearly describes the structural design of this research project. In addition, Implementation of the image pre-processing and application of classification techniques on the project are described in section 5. Furthermore, all the modeling techniques are evaluated in the form performance measure in section 6 followed by the critical analysis of the observations in the section 7 and 8.

2 Related Work

Several studies have been carried out to detect sleepiness of a person (especially drivers) using numerous machine/deep learning and face recognition techniques and a few of these highly effective solutions were never used in the aviation industry which is a critical area.

This subsection includes critical analysis of methods used in all the research papers studied to detect drowsiness using several classification models.

2.1 Classification Techniques to detect drowsiness

According to Alioua, Amine and Rziza (2014), 98% accuracy has been observed while detecting yawning to check the sleepiness and fatigue of a driver. They have realized the pros of SVM and thus, used it in the initial stage for the face detection followed by gradient edge detector to successfully detect mouth region and eliminate other facial characteristics. Lastly, a large amount of sequential frames were observed to identify an open mouth using CHT (Circular Hough Transform) technique. In another study, Fan, Yin and Sun (2007) have extracted 400 images from 20 videos for detection of yawning in individuals. Gravity-Center template is used for the face detection stage as it saves impressive sum of time in detection of human faces in rotated images. Grey projections is implemented for the identification of left and right corners of mouth whose change in texture is later extracted by Gabor wavelets technique. They have considered LDA (Linear Discriminant Analysis) for feature classification to accurately identify yawning, but Gabor coefficients seemed to be more robust and efficient as there is an improvement of 20% in the classification rate.

Besides detecting driver drowsiness, some researchers have also contributed to forecast sleepy state time of the individuals like Jacobé de Naurois et al. (2019) who predicted the time needed to reach a state of sleepiness and identified the drowsiness of a person. Two Artificial Network models have been used to detect somnolence every 60 seconds and forecast the necessary time required to achieve a moderate level of sleepiness in every minute. It was observed that behavioral features like eye closure and gaze are more efficient than physiological dataset in identifying drowsiness which can be detected by using ANN. Eye-blinking frequency is an important characteristic for sleep detection which is used as a primary characteristic by Picot, Charbonnier and Caplier (2010). They have done deep analysis on blinking rate by capturing 60 hours of videos from 20 drivers. Also, quantification of drowsiness is designed on a scale of 0-4 with 4 being the somnolence phase on the basis of features like eye blinking rate and percentage of eye closure. For the above-mentioned attributes, different features were set like percentage when at least 80% of eyes are closed is used along with Eye blinking rate. They observed an accuracy of 81.7% by using fuzzy fusion which is slightly better than that of individuals' features.

A few investigators like Vural et al. (2007) gave more attention to the combined results of facial attributes such as yawning and blinking rates of both eyes by using numerous machine learning models. Boosting techniques were performed initially to identify both the face and eyes in real time. Scaling is done on detected faces to 96 X 96 pixels and further transmitted the scaled faces to Gabor Filters with a normalized output and finally sent to the Support Vector Machine which outputs a facial feature on a continuous scale. Also, they have used classifiers like Adaboost and MLR (Multinomial Ridge Regression) which takes the facial attributes as inputs. It is clearly noticed that MLR performs better than Adaboost by achieving overall accuracy of 94%. Ngxande, Tapamo and Burke (2018) have added eye closure analysis to the features used by Vural et al. (2007) like frequency of blinking and yawning to identify driver's somnolence . They have trained large set of images before using classification techniques like Support Vector Machine, Convolutional Network and Hidden Markov Model to categorise each class and beep an alarm if the driver is sleeping. Finally, a Meta-analysis is performed by them on several papers to check the performance in terms of accuracy of various models and it is observed that CNN performs better than both HMM and SVM, but SVM is a model which is usually used.

Lin et al. (2013) used numerous data mining models to check the sleepiness of train drivers. Out of all the facial detection techniques available, they have performed Adaboost as it can be used in any weather settings and is highly efficient in night mode. Moreover, Haar classifier is chosen to execute AdaBoost technique and a general eye pattern is considered for identifying the eyes. Following these steps, a time difference is applied on the frames taken from video to locate the eye position. They have used PERCLOS for the eye closure rate that shows 81% accuracy on drivers wearing spectacles. Introduction of deep learning has improved the image classification and computer vision fields by giving high performing algorithms like Convolutional Neural Network which is performed on the video input by Dwivedi, Biswaranjan and Sethi (2014) for accurate classification. The system initiates the process by extracting frames from videos which is fed to Voila and Jones algorithm for detecting faces. Resizing of the faces is set to 48 X 48 before its normalization. In addition, Neural Network takes these input normalized images and outputs the extracted features via Hidden layer. A softmax layer is used at the end of the network to correctly classify the attributes and it is noticed that Neural Network performed exceptionally well on the testing images with an accuracy of 92%.

A few investigations were focused on the combination of pupil detection and other facial attributes. Azim et al. (2009) performed similar study by using features like yawning and pupil identification and performing Support Vector Machine as a classifier. Firstly, they have performed Voila and Jonas technique to accurately detect faces and seeking the bright blobs to locate eyes due to its cost-effectiveness. Also, a mouth is detected which is given to Fuzzy c-means classifier to provide the segmentation of the lips followed by its detection using eccentricity analysis which is 1 when it is actually lips. In addition, SVM model is employed which receives the width to height ratios of the images and beeps an alarm if closed eyes are detected. Results from the investigation suggest that SVM is able to correctly detect the drowsiness of drivers with a high accuracy of 92%. As a part of their future scope, the system should identify the pupils for the individuals wearing glasses and correctly detect yawning when a person puts a hand on the mouth. This gap is addressed by Assari and Rahmati (2011) who have used characteristics like eyes and mouth to check the sleepiness of drivers wearing spectacles. Face detection is performed on the input frame by accurate identification of background image which is later subtracted from the image to classify the faces. To locate eyes and eyebrows, Horizontal projection approach is performed and template matching is used to detect mouth. They have detected sleepiness by taking the vertical coordinates of the extracted characteristics which issues a warning message if there is a change in these facial attributes due to drowsiness.

A non-intrusive methodology is employed by García et al. (2012) to identify the drowsiness by using the field of computer vision. The have initiated the process by doing face detection by using Viola and Jonas to locate faces and used kalman filter to detect eyes. For each eye, ROI is calculated and some image pre-processing is done like morphologic transformation is done to completely eliminate the high intensity object. They have used PERCLOS for detecting eye closure having integral projections to calculate the center of IRIS. Results showed that this system has achieved a sensitivity of 92% and specifity close to 80%.

2.2 Machine learning models for image classification

Due to high dimensionality of data, it becomes difficult for the algorithms to deal with content based image classification. To deal with high dimensionality, Agrawal et al. (2011) have used Support Vector Machine and trained 500 images which is divided into four classes i.e. Car, Fire, Lion and Elephant that are further compared with one another on 6 histogram levels. Also, 3 different versions of SVM were considered like OAO (One Against One), OAO (One against All) and DAG (Directed Acyclic Graph) SVM for numerous color spaces by keeping fixed histogram level of 10 were considered. They have considered color image histograms due to better performance and insentivity to translations and rotations. Finally, desired results were obtained in the form of accuracy for different levels of histograms and several color spaces. Saradadevi (2008) designed a system to detect fatigue of driver using Support Vector Machine. Initially, they have converted videos into images where 10 normal frames and 10 yawning frames were taken from each video. Following this, mouth is located from the collected frames using Viola Jonas technique due to its higher performance over other models followed by training of several images consisting of yawning and normal frames on the SVM classifier which gives a decent accuracy of 81% for yawning and 86% of normal faces.

Velasco-forero, Member and Manian (2009) have used K-nearest Neighbor for accurate classification of remote sensing images due to simplicity in execution and excellent performance. They have collected large data of remote sensing image and using the concept of image segmentation and have obtained the required image objects. Also, three different versions of KNN which are original KNN, clipping KNN and improved KNN are applied on the dataset to classify the sample points which is compared with one another after the successful implementation. It is found that the improved KNN model outperformed the other two models by achieving higher accuracy

2.3 Critical Analysis & Limitations of literature review

After the extensive review of all the journals and conference papers, a few challenges were encountered while checking the sleepiness of a driver by classifying facial characteristics. Fan, Yin and Sun (2007) used a single facial attribute to detect the somnolence of the drivers which should not be considering keeping other features into consideration. Also, Nur et al.

(2017) failed to include the lighting conditions in the designed model to calculate the duration of closed eyes which is an important environment settings in ATC Tower. Moreover, ethnicities of people and use of spectacles were not taken into consideration in some of the research papers. So, it becomes extremely important to consider lightning conditions, diverse angles, several nationalities and both genders when using this concept on Air Traffic Controller which is done in the designed model.

These issues and gaps are solved using this study which can identify drowsiness of an ATC by detecting behavioral characteristics like Closed eyes, open mouth while yawning and also differentiating between open and closed eyes using multi-label image classification through highly efficient classification models. The system is trained on diverse image data set with both the genders, people wearing glasses, individuals of different ethnicities through many angles in several environment conditions like Lightning settings in ATC Tower.

3 Research Methodology

For this study, numerous methodologies are compared to be used on the image dataset to solve the business problem. Also, after doing the broad literature review, many authors have done their research utilizing different machine learning techniques and CRISP-DM (Cross-Industry Standard Process for Data Mining) is considered after doing broad investigation of those conference papers and journals.

CRISM-DM provides a organized blueprint to implement a data mining project that has six vital phase as shown in Figure 2 like gathering data; comprehending data, exploratory data analysis, designing modeling techniques and finally, its deployment. (Palacios et al., 2017)

There are six stages of CRISP-DM projects which are as follows:

3.1 Business Understanding

It is the first stage that gives a clear comprehension of the objectives to be accomplished from a business perspective. In addition, It incorporates evaluating the present situation and gathering resources needed to arrive at a business objective. As per the research question mentioned in section 1, the business aim is to plan an economical model which can identify sleepiness of an Air Traffic Controller in the aviation business using image classification models. Executing this cost-effective model will offer assistance to capture the drowsiness and inattentiveness of an Air Traffic Controller which will escape disastrous mishaps and plane crashes thereby sparing numerous innocent lives.

3.2 Data Understanding

This is the second phase of the Cross-Industry Standard Process for Data Mining (CRISP-DM) which involves collection of data and checking if it is complete and aligned with the business requirements specified in 1 by confirming the quality of data. Two datasets are used for the project:

<u>Closed eye dataset</u>: Originally extracted from CEW (Closed Eye in the Wild) Database, it contains around 1189 images of people from different ethnicities with closed eyes and

different environmental conditions like blur and lighting which are acquired from https://faithfull.me/blinks-dataset/.

Yawning video dataset: It includes 2 datasets in video format which are collected from http://www.site.uottawa.ca/~shervin/yawning/.

Both the datasets includes male and female participants from different ethnicities, with and without glasses who are talking/singing and yawning while driving the car. It contains 322 Recordings of diverse volunteers with several facial attributes like talking and yawning on the driver seat of a car. Besides that, 29 videos of both male and female yawning and talking with and without glasses with different ethnicities while driving using a different angle of the camera.

These videos are converted into frames and divided into two classes of yawning and open eyes. Almost 1222 frames of closed eyes and 1479 images of yawning faces are extracted from the videos.

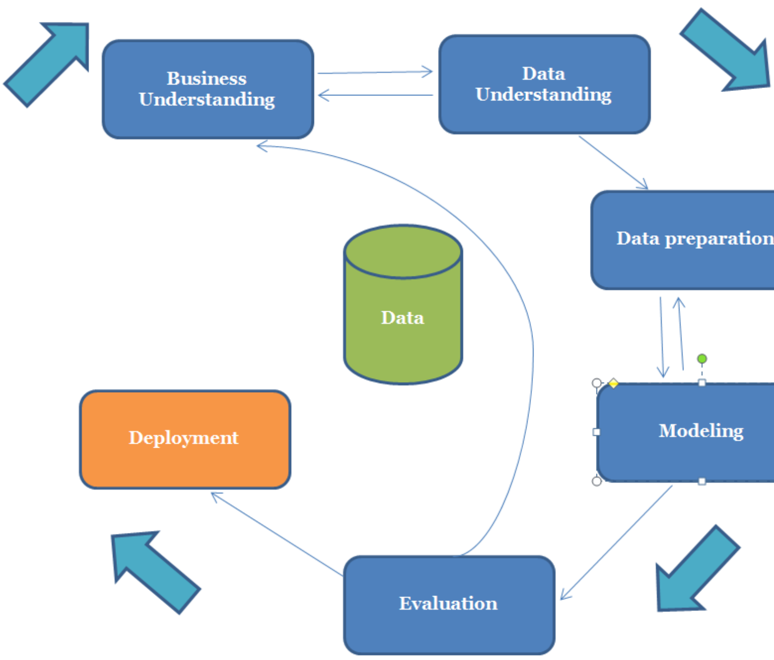


Figure 2: CRISP-DM Architecture

Closed eye dataset consisting of around 1200 facial images in non-uniform size and jpg format of many volunteers having closed eyes with many surrounding conditions like lightings and occlusion. In addition, yawning dataset comprising of 640X480 of 351 videos in avi format of 5 Gigabytes size presenting various facial expressions while talking, singing and yawning, with and without glasses, different Lighting conditions and nationalities.

3.3 Data Preparation

Data Preparation is the most vital phase of CRISP methodology which requires additional time and consideration to details within the timeline of the research project. Data selection is an integral fragment of data preparation phase which has shown the presence of many features in both yawning and closed eye datasets, but some highlights like Facial hair will not be taken into attention keeping the research question in consideration.

Image pre-processing is a crucial step which consists of preparing the images to be modeling techniques. It involves many concepts like Image Augmentation represented in Figure 3 which is done on the dataset before the implementation of a particular model. It includes rescaling to maintain uniformity where all the images/frame captured from both closed eyes and yawning video dataset are set to a base size, rotation of all the frames to a default angle of 90 degree, increased brightness to include the images having high intensity of light and edge detection to closely observe the edge of mouth and eyes.

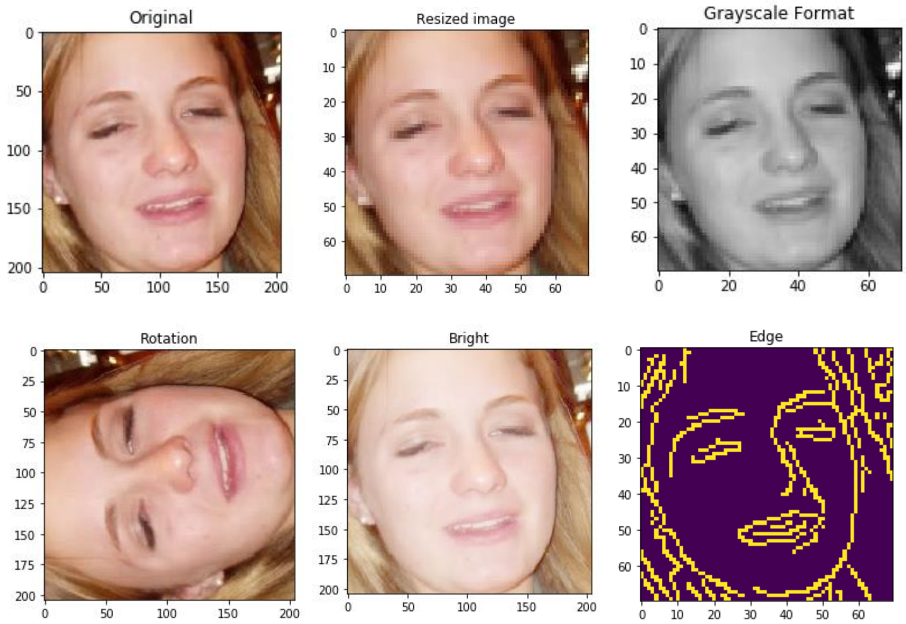


Figure 3: Pre-processed image using different transformations

3.4 Modeling

Modeling is fundamental step in CRISP methodology which includes creation of machine and deep learning models and its application on testing image data. There are 4 different classification models presented below which are applied on the testing image dataset:

3.4.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is a highly efficient technique in deep learning that accepts an input image; converts it into low dimensional feature map followed by pooling to further lowering the pixels in the image and flattening to reduce the feature map in a single vector which is the input to a Neural Network. Also, each Neuron in the hidden layer has attached weights to differentiate objects in the learning images to classify them from one other.

Moreover, these powerful networks are able to acquire important features with adequate training of the dataset unlike other basic methods which requires manual effort. CNN is an efficient technique which is capable of extracting numerous facial characteristics from an input image without performing much pre-processing with high accuracy. Being a computationally effective model, it has outperformed other highly performing algorithms like SVM classifiers and Deep Neural Networks in visual-related area. (Akar and Güngör, 2015)

3.4.2 Support Vector Machines

A Support Vector Machine (SVM) is a classification model that works on linear and non-linear data and is popularly used for image classification due to its high performance. In SVM, the plane is divided into two portions by a separating hyper plane and each category lies on the either side of the plane. The closest values to the hyper plane are known as support vectors which maximizes the distance between the support vectors and hyper plane.

It is heavily used by the researchers in their research for image classification problems as it avoids over-fitting and yields a good performance over other traditional models. Due to its high performance, it can also work well with high dimensional data (Thai, Hai and Thuy, 2012).

3.4.3 K-Nearest Neighbor

KNN is a machine learning and non- parametric technique which consists of training data i.e. it simply learns from the labelled training set and predicts the output class with the majority in the K-nearest neighbours calculated using Euclidean distance. Also, no data points are required and this makes training stage faster but somehow slows down the testing of images as it takes more time and requires more memory for capturing the training data. It is a method for classifying objects based on closest training examples in the feature space when there is no information about the distribution of data (Akar and Güngör, 2015)

Each data point is classified corresponding to the neighbouring sample and which can be forecasted by taking the classes of surrounding data points. Given an unknown sample and a training set, all the distances between the unknown sample and all the samples in the training set can be computed. The distance with the smallest value corresponds to the sample in the training set closest to the unknown sample. KNN works effectively well on multi modal classification problems as its decision is based on the neighbourhood of same objects and thus, the technique produces good accuracy if target class has many categories (Kim et al., 2012).

3.4.4 Random Forest

Decision Trees classifier is considered to be a decent technique for both regression and classification problems but it is vulnerable to many issues like lack of generalization and over-fitting.

To mitigate these problems, Random Forest is considered which is a robust ensemble technique involving randomly selection of samples for training decision trees from the training set, calculating the performance of every learning tree and merging the results from the all the decision trees into a single output category by using the voting principle. Also, It has been shown that Random Forests results in lower testing errors than the conventional decision trees (Yin et al., 2007) and shows comparable performance with high computational efficiency against SVMs in classification problems (Schroff, Criminisi and Zisserman, 2008).

It is one of the mostly used technique as it solves the over fitting problems and prevents pruning of trees as it cancels the biases by taking the average of all the predictions or majority voting. In addition, due to its capability of automatic detection of outliers, it generates results with higher performance and accuracy.

3.4.5 XGBoost

XGBoost is an ensemble technique which uses the principle of boosting and the gradient boosting (GBM) framework. It is developed to show improved performance of boosting techniques and consists of multitude of decision trees designed in a sequential order with diverse weights are given to each feature before feeding them to trees where misclassified features are given higher weights and the one corrected predicted is assigned lower weight. Wrongly forecasted weights from the first tree is fed to another in sequence and all of these classifiers merged together to give accurate observations.

XGBoost is a faster method due to the high power of parallel computing and it can use large datasets with least number of resources. Also, it handles sparse data using a tree learning technique and performance can further be improved using hyper parameter tuning.

3.5 Evaluation

Evaluation phase involves evaluating the five designed machine/deep models in the form of different performance metrics like Accuracy, Precision, Recall and F1 score. After the assessment and comparison of the machine/deep models specified in section 6, the results are verified to confirm if the business requirements are met and findings are summarised. Finally, the model that has outperformed other algorithms at a specified gets approval. Process reviewing involve list of steps to be taken to get an detailed look of all the tasks which are carried out in the implementation of the project to check for any missing part which can be considered as a part of Quality Assurance.

Overall, Four performance measures like Accuracy, Precision, Recall and F1 score are considered for the research project. These measures are calculated and compared with one another to assess the performance of classification models.

4 Design Specification

A well designed architecture is created to implement the study which can solve the research question mentioned in section 1 by carrying out tasks specified in objectives mentioned in section 1 by using economical models and algorithms.

As per Figure 4, the system will work from initial loading of dataset followed by conversion of videos (avi format) into frames in jpg format and then combining both the image datasets into a merged folder containing 3890 images of individuals with closed and opened eyes along with yawning faces. As the data is collected from multiple sources, it is messy and non-uniform and needs to be standardized before using it in learning models.

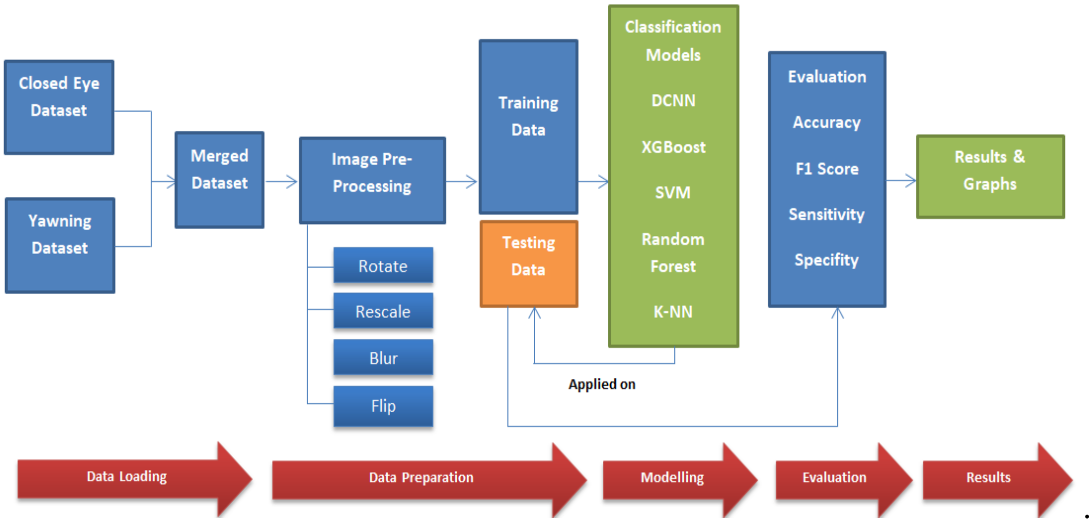


Figure 4: Designed Architecture of Drowsiness Detection system

Image Pre-processing needs to be done before dividing the dataset into training and testing set to get an improved performance of the algorithms and thus, avoiding overfitting problem. It includes rescaling the images to a non-uniform size which leads to an identical height and width. In addition, another popular image pre-processing technique, Data augmentation, is used in the dataset to augment the images to reduce the problem of over fitting by creating multiple versions of same image with some amendments like Scaling, Rotating/Flipping, Blurring and Brightness.

After the successful data preparation by using image pre-processing, the merged dataset is divided into the ratio of 75 Training and 25 Testing as it is the standard splitting proportion used in data mining projects. Many studies have been done to check for the usage of classification techniques for the dataset and after the extensive literature review, models like CNN, SVM, KNN and ensemble techniques like Random Forest and XGBoost are considered for this study. The above mentioned models are designed on the training dataset of 2917 images and finally it is tested on the 972 testing images after successful creation of the models. Following this, it is necessary to assess the performance of each algorithm by using performance measures like specified in Evaluation stage in section 3.5. Finally, the results are obtained and presented in the form of captivating visualizations to conclude the best model among all to be used on the dataset.

5 Implementation

Data captured from Yawning dataset is a video in avi format which is first converted into frames using openCV library and then merged with closed eye dataset into a single repository that is then ready for image Pre-processing. All the pre-processed images are added into the data frame to increase the dataset and improve the performance by learning images having different angles, diverse size, and blurring, flipped and high brightness. This will further avoid the overfitting problem in the image dataset.

5.1 Image Pre-Processing

The closed eye dataset contains 1189 images with varying size and color space, 1479 yawning face images and 1222 images with open eyes. These images are converted into a numpy array which is the used by the algorithms to design models. Before this, Image augmentation is done as part of pre-processing on the dataset using many steps which are as follows:

5.1.1 Rescaling

Due to the non-uniformity of images, the system was not able to perform well as it got less accurate results and hence, it requires scaling to make each frame to be uniform in terms of base size. In this study, 3890 images were utilized with distorted sizes which is scaled to 20 x 20 size using resize() function of openCV library. Out of all interpolation methods, cv2.INTER_AREA is performed for zooming as both the width and height are even numbers and used imshow() function of matplotlib library to plot the changed visualizations.

5.1.2 Translation

Translation is the technique which is used to shift the location of object in the image using geometric transformation that changes the location of an object to a new position in the resultant image. Translation is represented in terms of transformation Matrix which is taken into numpy array before passing it to warpAffine() function in OpenCV to change the position of the object.

5.1.3 Rotation

This transformation is responsible for rotating images in either clockwise or anti-clockwise direction. For rotating the images at a specified angle, a transformation matrix is used by getRotationMatrix2D() and further it is fed into warpAffine() function to get the rotated image.

5.1.4 Exposure and Blurring

All the Images are filtered to remove the unwanted noise, blurring the image and increasing the exposure and intensity. Blurring and whitening is achieved for excluding the high frequency and increase the overall brightness of an image respectively where Gaussian Filtering is used to blur the images and adjust_gamma() function is used to increase the exposure of the image.

5.1.5 Edge Detection:

Canny edge detector is multi-stage edge detection technique used for clear detection of edges. As it is vulnerable to noises, we have applied Gaussian filters are applied to remove the noises followed by finding the magnitude of intensity gradient and thresholding to differentiate the edges in the image. All the above steps are included in a single canny() function which clearly represents edges from an input image.

5.2 Classification Models

For this study, several research papers were investigated to choose the best fitted models for the dataset used in the study and numerous machine learning and ensemble techniques along with deep learning algorithm were performed on the dataset which is divided into 75%-25% training and testing ratio to solve the research objectives. Five classification techniques such as CNN, SVM, KNN, Random Forest and XGBoost are used where two models; baseline and tuned model are made on the four machine learning techniques and the later one is designed after doing hyperparameter tuning to improve the performance over the baseline model.

5.2.1 Convolutional Neural Network

Before images are fed to the Convolutional Neural Network, a separate pre-processing is done through data augmentation in which images are rotated by 90 degree, zoomed, translated and flipped. This is done to avoid over-fitting by training the increased number of images.

Convolutional layer is used to decrease the size of image by preserving the vital features of the training images which is done by performing ReLU (Rectified Linear Unit) activation function is considered for the convolutional layer having 32 feature maps and a small filter size of (3, 3) which reduces the size of image. A Convolutional layer is followed by a pooling layer of size 2x2 where max pooling is used to reduce down sampling and preserve all the relevant features in a reduce sized image without the loss of information which further prevents over-fitting. Also, it helped to correctly identify the features in manipulated images like rotation, translation and blurring. In addition, a dropout layer is used after the pooling layer with a rate of 0.3. To reduce the 3 dimensional data to 1 dimension, flattening is used which ensured the pooled features are mapped into a single vector representing all the vital features which is later fed into Artificial Neural Network as an input. Also, a drop layer is applied with 15% of rate to regularize the model followed by a softmax layer to standardize the output on a scale of 0-1. Finally, a compiler block is applied with loss function as cross entropy for the output having three classes. Finally, the above model is then trained on the training data by using an epoch of 50 and 32 images are used as batch sizes to be fed into the network. It was observed that after both forward and backward propagation of the network for 20 times, its error late has reduced and accuracy has increased.

5.2.2 SVM (Support Vector Machine)

Support Vector Machine is implemented on the training datasets after the successful image pre-processing and division of dataset into training and testing data. Initially, an SVM classifier is built using sklearn python library by setting C=1. The baseline modeling initiates by training on the pre-processed training data points which is observed to have a decent accuracy as represented in the following section 6. Also, Hyper-parameter tuning is done to design a tuned model by finding the best parameters in the algorithm using GridSearchCV function which has improved the performance of the technique by giving an optimized value of gamma and Cost. The tuned model is trained on the learning images using the extracted values of gamma and Cost which improves the performance of the model.

5.2.3 K-Nearest Neighbor

After the availability of training data points, a baseline KNN classifier is designed with number of neighbors set to 10 which is then fitted on the learning data. It is observed that all the testing images were tested one by one using k = 10 and the algorithm performed well on the dataset by achieving an excellent accuracy specified in section 6.1. Keeping hyper parameter tuning into consideration, different values of k are used to check the optimal value which optimised the model and its prediction rate by finding an ideal value of K and weights to be used.

5.2.4 Random Forest

Random Forest is considered for this research due to its robustness, reliability, improved performance and its ability to avoid over-fitting. After taking the pre-processed learning images as input, the processing of Random Forest technique by using the function RandomForestClassfier to design a base model having maximum depth of 3 and minmum samples split of 2 is available in ensemble package of sklearn library. It has shown a good performance which later on improved further after tuning the model by optimising the value of maximum depth, maximum features and minimum sample split using GridSearchCV function.

5.2.5 XGBoost

XGBoost is considered to be one of the models to be used in the project due its high computational efficiency as it uses distributed processing of the large datasets. The steps starts from creation of XGBoost model with learning rate is set to 0.05 on the training data using XGBClassfier function and its validation on the remaining testing set to calculate the performance which is later observed to have an very high accuracy and F1 score. To optimise it further using the concept of hyper parameter tuning, an optimized value of gamma and learning rate is extracted by using Grid search. Following this, its performance showed a jump in terms of accuracy, precision and F1 Score.

6 Evaluation

This study has 3980 images of people including 1189 frames with closed eyes, 1479 images doing yawning and 1222 images with the open eyes as shown in Figure 5. After successful image pre-processing followed by data splitting into 75%:25% training and testing ratio and implementation of classification algorithms, It is a necessity to evaluate the performance of each machine/deep learning model on the validation image dataset using a few measures and compare them after the successful implementation of the technique.

Precision, Recall and F1 Score is utilized to find the two best techniques amongst the KNN, SVM, XGboost and Random Forest. These two machine learning algorithms are compared in terms of training time and the best is selected for further comparison. Finally, Accuracy and training time is used for comparing CNN against the best machine learning technique. This is clearly discussed in section 7.

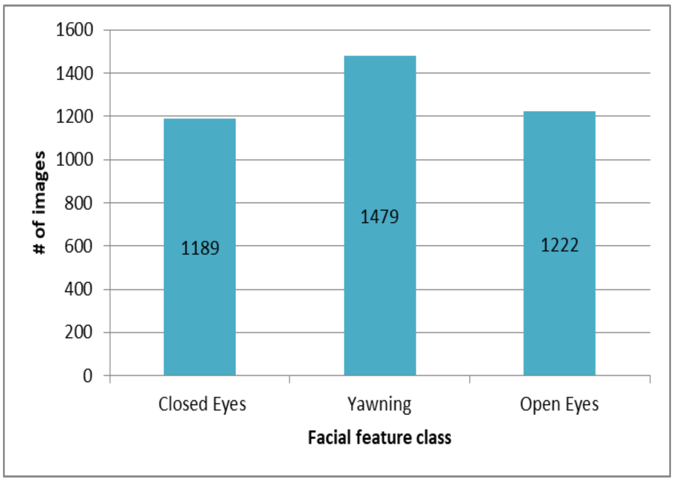


Figure 5: Data distribution in each class

6.1 Accuracy

Accuracy is the total number of correct prediction of all three classes (open eyes, closed eyes and open mouth) out of all the classification done by the drowsiness detection model. Accuracy, as per our business requirements, is mentioned below.

$$Accuracy = \frac{Correct\ Prediction\ of\ \ eyes, yawning\ and\ open\ eyes}{Total\ number\ of\ prediction\ done\ bydrowsy\ detection\ system}$$

6.1.1 Machine learning Models

Table 1: Accuracy of Baseline Model and Tuned Model

Techniques	Baseline Model	Tuned Model
SVM	71%	88.89%
Random Forest	73.14%	89.67%
XGBoost	87.10%	93.03%
KNN	95.53%	97.12%

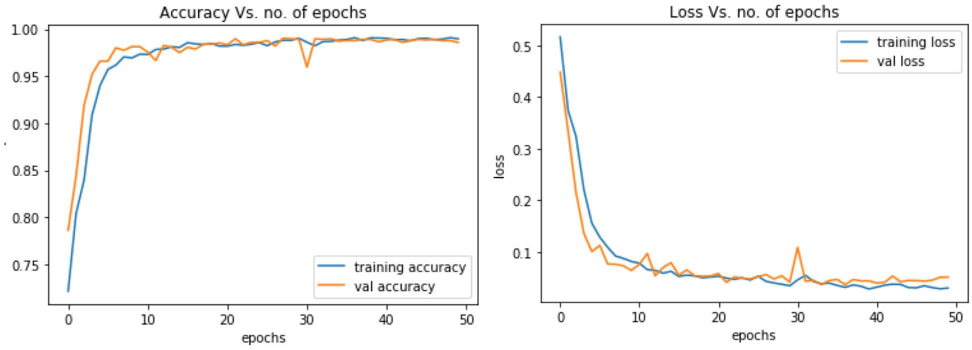
As per table 1, the performance of all the classification techniques is impressive as a baseline model but SVM has the lowest accuracy amongst all other techniques and it followed a similar trend in tuned model when it jumps to 88.89% which is also lower when compared the other models. However, K-Nearest Neighbors dominates all the algorithms as it is able to correctly identify the classes both in the baseline and tuned model and shows an overall accuracy of 97% in improved version.

6.1.2 Deep learning model (Convolutional Neural Network)

Table 2: Validation accuracy of CNN after 50 epochs

Techniques	Validation Accuracy	Testing Accuracy
CNN	98.59%	98.48%

CNN model is trained on the large training and validation data after the pre-processing step; it is clearly evident from table 2 that 98.6% validation and 98.5% of training accuracy is achieved by CNN. As, as both the accuracies are close to each other with training correctness is less than validation accuracy, the model is a good fit.



(a) (b)

Figure: 6 (a). Learning curve to compare both the training and validation accuracy against number of epochs; (b). Loss curve representing both the training and testing loss compared against the number of epochs

After closely looking at Figure 6(a) for the CNN model, It represents a good fit as both training and testing loss has reduced with increase in number of epochs with loss in training is slightly lower than testing loss, both the losses have a very small generalization gap.

6.2 Precision

Precision is parameter which describes the correct identification of a class by the model. Our goal is to minimise the number of open eyes instances when it is actually closed to decrease the risk of accidents**.**

6.2.1 Comparison amongst machine learning models

Table 3: Comparison of four machine learning techniques with two modelling levels

	Baseline Model		Tuned Model	
Technique	Precision	Recall	Precision	Recall
SVM	75%	71%	91%	89%
Random Forest	75%	73%	90%	90%
XGBoost	87%	87%	93%	93%
KNN	96%	96%	97%	97%

It is important to identify instances when the ATC is sleeping but the model is showing awake. As per table 3, it is shown to have highest precision for KNN model which is 97% and most amongst all the techniques used in the research followed by 93% obtained by XGBoost.

6.3 Recall

Recall is a performance measure which minimizes the instances of false negatives. In this study, it is important to minimize the instances of closed eyes instances when it is actually open and the number of yawning faces when the ATC is either singing or smiling.

It is evident from the table 3, K-Nearest Neighbor outperformed other algorithms by getting a recall and precision of 97% and reducing the number of false closed eye and yawning instances. Also, XGBoost has shown potential in its performance due to decent precision and recall of 93%.

6.4 F1 Score

As both Precision and Recall are identical in nature, F1 Score is a better measure to seek a balance between the two which can be calculated as a cumulative score of Precision and Recall that ranges between 0 and 1.

It is clearly evident from the Figure 7 that KNN technique has the best F1 score both in the base model and improved one among four machine learning models.

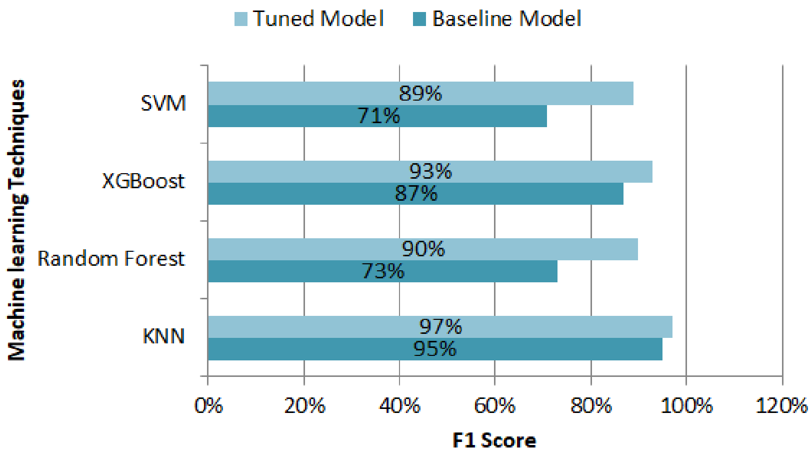


Figure 7: F1 Score of machine learning algorithms

6.5 Training Time

Training time is the time required to train a model and is a vital parameter to consider besides the above four as it explains the executional and computational speed of an algorithm. In this study, it is performed on all the five modeling techniques on both the baseline and tuned/improved model to find the algorithm which requires the least and most time for training the model.

After looking at table 4, It seems that Random Forest and KNN is outperforming other techniques in terms of training time as both the algorithms are taking significantly less time in training than other machine learning models.

Table 4: Training time of Machine learning techniques

Machine learning Techniques	Training time of Tuned Model (in sec.)
SVM	3500
Random Forest	130
XGBoost	1664
KNN	496

6.6 Results & Discussion

In this paper, a drowsiness detection model is designed to accurately identify different facial characteristics like closed eyes, yawning for sleepiness and open eyes for non-drowsy mode to detect sleepiness of an Air Traffic Controller which is going to avert many airplane accidents and near-misses. After doing literature review, many studies have been done to identify the sleepiness of driver, however this work is done on ATC sleepiness detection which is not done and the combination of the applied algorithms makes this as a novel work. In addition, Due to the variety of images, the system can work accurately on both the genders, different ethnicities, individuals wearing glasses and under different lightning conditions.

To solve the business problem, several classification techniques are applied on the dataset which provided some amazing findings. All machine learning techniques are compared with one another by using performance measures such as precision, recall and F1 score to get the two best algorithms and it was found that KNN (precision – 97% and recall – 97%) and XGBoost gave 93% of precise results and recall of 93%. are the best techniques with have shown decent improvement in terms of accuracy after the application hyper parameter tuning for searching the best parameters where KNN has shown 97.12% accurate results which is highest among all the four modelling techniques followed by 93% in XGBoost. In addition, SVM and Random Forest achieved the lowest accuracy of 89% and 90% respectively and these two will not be considered further. One model is designed in CNN which has shown an accuracy of 98.6% which is highest amongst all the five modelling techniques performed in the study. To further investigate the performance of machine learning algorithms, Precision and Recall is used as a performance measure which clearly indicates that KNN has the highest precision and recall percentage that can minimize the number of wrong classifications done by the system. Moreover, combined effect of both the above measures is shown in the form of F1 Score where SVM has achieved the least score of 89% and KNN represented the highest F1 value of 97%. Keeping the observations into consideration, it is noticed that Random Forest, SVM and XGBoost is not up to the mark in terms of their performance and hence, cannot be the best models for the study. Also, both Training time is vital component which also needs to be explored for both KNN and CNN. After calculating and deep analysis of obtained outputs, training time of Random Forest is found to be 130 seconds which is very less compared to any other machine learning technique used in the study.

After analysing the results, it is clear that CNN is the best suited technique for solving the research question as it can solve the business problem with the highest accuracy, low training time and very high level of confidence. However, in this study, only two facial features like eyes and mouth were considered and it did not take other behavioural characteristics into consideration which can further improve the performance. Also, this model is not a real time detector which is a significant point and all of these limitations can be achieved as a part of future scope discussed in the following section.

7 Conclusion and Future Work

In this study, a multi-class image classification is done to check drowsiness of an Air Traffic Controller on the basis of many facial features like open eyes, closed eyes and yawning by using several modelling techniques like CNN, SVM, KNN and ensemble techniques like XGBoost and Random Forest. The images are pre-processed after extracting the frames of videos where translation, rotation, shearing and blurring is done to increase the size of dataset and avoid the over fitting problem. These pre-processed images are fed to the image classifiers which are divided into two models; the baseline model with some parameters and the second one is an improved model using hyper parameter tuning which uses the best fitted parameters. It is observed that both the models performed well, but the tuned/improved model seemed to have higher accuracy in classifying the facial attributes of sleepiness. In the improved models like KNN showed a jump of 1.6% and reached accuracy 97% and achieved highest precision, recall and F1 score than any machine learning models. In addition, Both SVM and Random forest showed the least accuracy close to 89% and lowest F1 score of 71% and 73% respectively. Training time is also assessed for all four machine learning models. Overall, CNN outperformed all the machine learning models as it has achieved better accuracy than all the other algorithms.

As a future reference, this model can use other behavioural characteristics like head nodding and eye blinking frequency and can be integrated with an AI application showing the detection of drowsy state of an Air Traffic Controller in real time..

6 Acknowledgement

I would like to thank and express my deep gratitude to my supervisor Dr. Cristina Muntean for the continuous feedback and support through the learning phase of research project and allow me to learn different technologies used in the thesis. She spends much time in understanding the project and has given useful suggestions where she shared her views and tips regarding documentation and implementation for the research study.
Also, I would like to thanks my parents for supporting me in pursuing my dream of doing masters in Data Analytics with their interminable support throughout this wonderful journey.

References

Agrawal, S. at al. (2011) 'Content Based Color Image Classification using SVM', *Eighth International Conference on Information Technology: New Generations*, pp. 1090–1094. doi:

10.1109/ITNG.2011.202.

Akar, Ö. and Güngör, O. (2015) 'Classification of Multispectral Images Using Random Forest Algorithm', vol. 2, no. 2, pp. 105-112, doi: 10.9733/jgg.241212.1.

Alioua, N., Amine, A. and Rziza, M. (2014) 'Driver's fatigue detection based on yawning extraction', *International Journal of Vehicular Technolog*y, doi: 10.1155/2014/678786.

Assari, M. A. and Rahmati, M. (2011) 'Driver drowsiness detection using face expression recognition', *International Conference on Signal and Image Processing Applications*, Kuala Lumpur, Malaysia, 16-18 November. doi: 10.1109/ICSIPA.2011.6144162.

Azim, T. et al. (2009) 'Automatic fatigue detection of drivers through yawning analysis', *Communications in Computer and Information Science*, vol. 61, doi: 10.1007/978-3-642-10546-3_16.

Choi, M. and Kim, S. W. (2017) 'Online SVM-based personalizing method for the drowsiness detection of drivers', *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, doi: 10.1109/EMBC.2017.8037781.

Dwivedi, K., Biswaranjan, K. and Sethi, A. (2014) 'Drowsy driver detection using representation learning', *IEEE International Advance Computing Conference*, Gurgaon, India, 21-22 Feb., doi: 10.1109/IAdCC.2014.6779459.

Fan, X., Yin, B. C. and Sun, Y. F. (2007) 'Yawning detection for monitoring driver fatigue', *Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, China, 19-22 August, doi: 10.1109/ICMLC.2007.4370228.

García, I. et al. (2012) 'Vision-based drowsiness detector for real driving conditions', *IEEE Intelligent Vehicles Symposium*, Alcala de Henares, Spain, 3-7 June, doi: 10.1109/IVS.2012.6232222.

Jacobé de Naurois, C. et al. (2019) 'Detection and prediction of driver drowsiness using artificial neural network models', *Accident Analysis and Prevention*. doi: 10.1016/j.aap.2017.11.038.

Kim, J. et al. (2012) 'Comparing Image Classification Methods : K-Nearest-Neighbor and Support-Vector-Machines', *6th WSEAS international conference on Computer Engineering and Applications*, pp. 133–138.

Lin, Z. et al. (2013) 'Efficient Train Driver Drowsiness Detection on Machine Vision Algorithms', *Indonesian Journal of Electrical Engineering*, vol. 11, no. 5, pp. 2566-2573, doi: 10.11591/telkomnika.v11i5.2488.

Monk, T. H. (2007) 'Practical consequences of fatigue-related performance failures', *Sleep*, vol. 30, pp. 1402-1403, doi: 10.1093/sleep/30.11.1402.

Moon, W.-C., Yoo, K.-E. and Choi, Y.-C. (2011) ‘Air Traffic Volume and Air Traffic Control Human Errors’, *Journal of Transportation Technologies*, vol. 1, pp. 47-53, doi: 10.4236/jtts.2011.13007.

Ngxande, M., Tapamo, J. R. and Burke, M. (2018) ‘Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques’, *Pattern Recognition Association of South Africa and Robotics and Mechatronics*, Bloemfontein, South Africa, doi: 10.1109/RoboMech.2017.8261140.

Nur, F. I. Y. et al. (2017) ‘Analysis of eye closure duration based on the height of iris’, *6th IEEE International Conference on Control System*, Batu Ferringhi, Malaysia, 25-27 Nov. 2016, doi: 10.1109/ICCSCE.2016.7893610.

Palacios, H. J. G. et al. (2017) ‘A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change’, *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 598-604, doi: 10.25046/aj020376.

Picot, A., Charbonnier, S. and Caplier, A. (2010) ‘Drowsiness detection based on visual signs: Blinking analysis based on high frame rate video’, *IEEE International Instrumentation and Measurement Technology Conference*, Austin, USA, 3-6 May, doi: 10.1109/IMTC.2010.5488257.

Schroff, F., Criminisi, A. and Zisserman, A. (2008) ‘Object Class Segmentation using Random Forests’, *British Machine Vision Conference*, Leeds, doi: 10.5244/C.22.54.

Saradadevi, M. (2008) ‘Driver Fatigue Detection Using Mouth and Yawning Analysis’, *International Journal of Computer Science and Network Security*, vol.8, n0.6, pp. 183–188.

Thai, L. H., Hai, T. S. and Thuy, N. T. (2012) ‘Image Classification using Support Vector Machine and Artificial Neural Network’, *International Journal of Information Technology and Computer Science*, vol. 5, pp. 32-38, doi: 10.5815/ijitcs.2012.05.05.

Velasco-forero, S., Member, S. and Manian, V. (2009) ‘Improving Hyperspectral Image Classification Using Spatial Preprocessing’, *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 297–301. doi: 10.1109/LGRS.2009.2012443.

Vogt, J. *et al.* (2004) ‘Economic Evaluation of CISM--A Pilot Study.’, *International Journal of Emergency Mental Health*, University of Copenhagen, Denmark, vol. 6, no. 4, pp. 185–196.

Vural, E. et al. (2007) ‘Drowsy Driver Detection Through Facial Movement Analysis’, *Human–Computer Interaction*, pp. 6-18, doi: 10.1007/978-3-540-75773-3_2.

Yin, P. et al. (2007) ‘Tree-based Classifiers for Bilayer Video Segmentation’, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, doi: 10.1109/CVPR.2007.383008.

Zhang, C. et al. (2018) 'A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133-144, doi: 10.1016/j.isprsjprs.2017.07.014.