Development of a computer vision-based system for recognising fatigue of truck drivers

AGATA SOWA

SUPERVISOR:
JOB ZWIERS

CRITICAL OBSERVER:
DANIEL DAVISON

CLIENT:
TECHSPREAD

July 2022

Department of EEMCS
Faculty of Electrical Engineering,
Mathematics and Computer Science

Abstract

Fatigue is inextricable from the environment of truck drivers. Despite the advances in fatigue detection systems, fatigue continues to be one of the leading causes of traffic accidents. Introducing a computer vision-based system for fatigue detection could divert that trend. This graduation project researched various computer vision techniques and facial-related fatigue parameters to develop such a system. Through the course of this research combination of yawning, head nodding, and OpenFace 2 as software is examined in detail, and finally, a system is proposed to detect the yawning and fatigue state of the driver. The machine learning algorithm used to determine the yawning state of a face given values obtained from Facial Action Units has the accuracy of 0.961, recall of 0.899 and precision of 0.651. A low precision score is partially corrected by introducing various mitigation approaches. Subsequently, a fuzzy logic model based on yawning and head-nodding frequency is proposed as a model to determine driver's fatigue.

Acknowledgements

First and foremost, I would like to thank my supervisor, Job Zwiers, for his guidance and support during this graduation project and my critical observer Daniel Davison for his help with research documents and insights into the thesis. Subsequently, I am very grateful to Techspread and Fleur Bake for the opportunity to work on an interesting project which allowed me to develop my skillset further. Moreover, I would like to thank all the research participants for their help and time. Lastly, I would like to thank my friends and family for their invaluable support.

Table of Contents

Figures

Tables

1 Introduction

Sixty percent of truck drivers admitted to being fatigued every ninth or lesser drive. Furthermore, 30% of them stated that they had fallen asleep behind the wheel within 12 months before the study conducted by Vitols [1]. Current safety systems implemented in the trucks to recognise driver fatigue mostly take into account truck position between the lanes and steering patterns [2]. This thesis explores the possibility of detecting fatigue based on computer vision techniques applied to the facial features of drivers. Following statistics indicate that there is still space and need for introducing new technology that could further improve fatigue detection in truck drivers. Of all road accidents involving heavy goods vehicles, 90% are caused by human factors. The human factors are split into more detailed categories. One of the most prominent ones is drivers' fatigue – 20% of all accidents caused by human factors result from driver's fatigue [3].

Techspread is a data-oriented start-up that intensively collaborates with the technological sector. One of their projects is to develop in-car systems that can detect fatigue in truck drivers by the data obtained through computer vision and face-recognition technology. That approach can also enhance current fatigue detection systems, thus reducing the number of accidents. Therefore, the main research question of this paper can be formulated as follows:

- **RQ1:** Developing a camera-based system to recognise fatigue in truck drivers.

When devising such a system sub research questions must be taken into account:

- **subRQ1:** Determining which facial characteristics are related to fatigue.
- **subRQ2:** Determining which technique related to computer vision and facial recognition is the most optimal for recognising features derived from subRQ1

The description of the process of designing this system is separated into the following chapters. The second chapter is devoted to existing knowledge regarding fatigue, symptoms, and computer vision technology suitable for face detection. The third chapter describes the requirements that have to be met by these projects and lists the possible solutions. The fourth chapter, split into iterations, showcases the steps necessary to develop this system. Subsequently, evaluations of the proposed system are described in the fifth chapter, followed by a discussion and recommendations in chapter six. Finally, the thesis is concluded in the seventh chapter.

2 Background research

Background research is necessary to create a concept derived from theory. Therefore, with the support of relevant papers, this section begins with creating a definition of fatigue, followed by a listing of the possible fatigue detection methods. Moreover, a subsection describing the self-assessment technique is also included. Finally, selected computer-vision techniques are described, and state of the art is presented.

2.1 Fatigue

Publication by European Transport Workers' Federation (ETF) from 2021 [1] suggests using a definition of fatigue proposed by Phillips [4]:

"Fatigue is a suboptimal psychophysiological condition caused by exertion. The degree and dimensional character of the condition depends on the form, dynamics, and context of exertion."

Furthermore, both sources point out that fatigue can impact a person in a cognitive, physiological, and emotional way. Those impacts can affect a person's performance in either a long- or short-term scenario. Lupova [5] argues that sleepiness and drowsiness also fall under the term fatigue. Furthermore, the same study identifies feeling of weariness, lack of motivation, decrease in awareness, brain activity and heart rate, and yawning and head nodding as symptoms of fatigue. Increased reaction time, reduced vigilance, loss of situational awareness, poor decision-making and judgment are also listed as symptoms by Phillips [4]. ETF [1] states that fatigue in the driving scenario can result in drivers having no memory of the last few kilometres, poor speed and steering control and reduced attention. This results in overall reduced performance of the driver, which can lead to an accident.

According to May and Baldwin [6], fatigue can be split into active, passive and sleep-related categories. Active fatigue is task-related, it occurs when there is an increased task load or in road conditions: poor visibility or traffic. Passive fatigue is also task-related. When a particular task is repetitive, predictable, or monotone – for a driver, that could be an extended driving period, monotonous drive or high reliability on automated systems present in the vehicle. The last type of fatigue, sleep-related one, is related to circadian rhythm and sleep deprivation. Sikander and Anwar [2] indicate that sleep-related fatigue is most likely to occur during the night (20:00 – 5:00), in the morning (6:00 – 10:00) and midday (13:00 – 15:00). Moreover, Caldwell [7] points out that

"fatigue is a physiological problem that cannot be overcome by motivation, training, or willpower"

and that

"people cannot reliably self-judge their own level of fatigue-related impairment"

Those two statements show that developing a fatigue detection system is not a waste of time.

2.2 Fatigue detection methods

Shi [8] proposes to split fatigue detection methods into three categories – physiological parameters, vehicle driving parameters and facial features-based methods. A similar categorisation is also present in a review by Sikander and Anwar [2]. The difference is that Sikander and Anwar [2] also propose parameters derived from the driver's posture as a fourth category. In both papers, within the three listed categories, there are subcategories used to describe particular methods in greater detail. Those methods can be combined both within the category and outside of it. That creates multifactor methods; this research only focuses on methods in isolation without describing their combinations. Moreover, this section generally describes the vehicle and physiological parameters-based method and focuses on the facial features-based method, as the latter is crucial for the development of a facial-recognition based system.

2.2.1 Vehicle Driving Parameters

With increasing fatigue in drivers, their ability to control a vehicle and perceive the environment decreases. This reflects on the vehicle's behaviour on the road. Consequently, the first parameter that can be examined is the vehicle position in the lane. The inability of driving a car in the middle of the line and swaying from left to right can be interpreted as a sign of fatigue [8]. That parameter relates to the angle of the steering wheel. Shi [8] states that with progressing fatigue in a driver, the amount and amplitude of movement of the steering wheel increases. The grip of the steering wheel can also be used as an indicator of fatigue. Contrary to the increasing steering wheel angle amplitude, the strength of the grip will lessen with the progressing fatigue [8]. Sikander and Anwar [2] also mention the possibility of installing sensors in a driver's chair to detect posture changes. However, it is simultaneously stated that that method still needs further examination.

2.2.2 Physiological Parameters

Brain, heart, and skin-based methods are the sub methods of detecting drivers' fatigue using physiological parameters. During the brain's activity, certain brain waves appear and can be detected by electroencephalography (EEG). A study conducted by Borghini [9] showed that the presence of alpha waves in the EEG is an indicator of fatigue. The heart sub-method is related to the driver's pulse, which decreases as the fatigue increases; the same principle can also be used when analysing the breathing frequency of the driver [8]. Skin based method is connected with the use of surface electromyogram (sEMG). The signal of sEMG can be used to interpret muscle fatigue, which is related to the general fatigue of the driver [8]. Sikander and Anwar [2] argue that those methods can indicate whether there are signs of fatigue early on. However, physiological methods are hard to implement in the truck on a regular basis due to their intrusive sensors.

2.2.3 Facial Features

According to Sikander and Anwar [2], the changes appearing on a face caused by fatigue are the most apparent symptoms. The regions from which the signs of fatigue can be extracted on a face are the whole head, mouth, and eyes.

2.2.3.1 Head

Fatigue signs can be extracted from the head position. The head of a driver should be oriented frontal to the road. Looking in another direction for a prolonged period is a sign of distraction and fatigue that can lead to fatal accidents [10]. Another vital sign of fatigue derived from the head is its tilt to the front. Bergasa [10] also claims that the frequency in which the head tilts occur and their angle are other indicators of fatigue, which corresponds to the findings of Ahmed [11]. Moreover, head tilt to the front is also a signal of fatigue. This head movement occurs most often when a driver has dozed off [12]. The forward head tilting, also called nodding off, occurs when a driver is no longer conscious – in a state of microsleep, points out Lic and Summala [13]. Sigari [14] mentions that a driver having their head in a fixed position for a prolonged time can also be a sign of distraction-related fatigue. Those findings are confirmed by study conducted by He [15]. Moreover, a threshold is estimated that the head nods should occur with frequency higher than 0.05 Hz and lower than 0.2Hz in order to classify a person as not fatigued. The head is a significant body part that can depict multiple signs of fatigue by its movement.

2.2.3.2 Mouth

Observing the mouth region of a driver's face can also be used to determine their fatigue. After analysing different characteristics of fatigued faces, Sundelin [16] observed that droopy corners of a mouth could be attributed to being tired. Another sign of drivers being tired is their yawning. Yawning itself is one of the most telling signs of fatigue. Furthermore, Sigari [14] and Jin [17] agree that yawn distribution over time can determine whether the driver is fatigued and how tired they are.

Daquin [18] categorises yawn based on time duration: long expiratory phase (4 to 6s), brief acme (2 to 4s) and rapid expiration. Daquin [18] also states that the yawn can last maximally 10 s and, once started, cannot be prematurely terminated but can be modulated – changes in mouth shape, facial contraction, and thoracic participation can occur. A singular yawn consists of the inspiratory phase, acme and expiration. The inspiratory phase happens at the beginning of the yawn when the lower jaw opens widely to allow the superior airway (mouth) to be maximally open. After the tongue is displaced to the back of the mouth and vocal cords are maximally stretched, the acme phase begins. During the acme phase, the position obtained in the inspiratory phase is maintained, and mimic muscles are contracted. The expiration phase starts when the muscles involved in the two previous phases relax, and quick passive (the muscle does not have to be contracted) expiration occurs [18]. Furthermore, Daquin [18] refers to a study conducted by Schino and Aureli [19] to point out that men and women yawn equally often. Daquin [18] also states that yawning cannon be prematurely finished; once started, it cannot be stopped till it comes to the natural end.

Zilli [20] conducted a yawn-oriented study on two sets of participants. The first set consisted of people who were used to waking up early and going to bed early. The second set was the opposite; people used to waking up later and going to bed late. The observations showed that people who belong to the first category yawn less (an average of 11.08 yawns per day between 10:00 and 21:00) than people from the second category (average of 22.63 yawns per day between 10:00 and 21:00).

Unfortunately, a driver talking can often be mistaken for yawning. This is the case as yawning is recognised as having one's mouth open, which can also occur while talking. Therefore, Sigari [14] proposes using yawn detection in combination with other fatigue signs for the best results. The fatigue indicators can be derived from a driver's mouth region. However, they are not always unambiguous.

2.2.3.3 Eyes

The state of a person in terms of fatigue can be derived from eye movement and the percentage of eyelid closure over pupil in time. Schleicher [21] claims that blinking frequency and duration are essential variables when detecting one's fatigue. The same study states that while the study participants reported growing fatigue, the intervals between blinks decreased. Subsequently, blink duration increased, and the time the eye would stay closed. A lid closure covering the whole pupil and lasting more than 500 milliseconds can be considered a microsleep. A non-fatigued driver should blink 15-30 times per minute, and the blink should not take more than 300 milliseconds [22]. The faster the person blinks, the more fatigued they are [21]. Another piece of information that can be derived from the lid is the Percentage Of Eyelid Closure Over The Pupil Over Time (PERCLOS). According to Zhang [22], the more fatigued a person is, the higher the PERCLOS value. Not only the lid movement is essential when determining one's fatigue, but the eye movement itself can also be considered an indication. A short and quick movement of an eye is called a saccade. The frequency of saccades can be used to obtain information about a person's tiredness. The more tired a person is, the smaller the frequency at which the saccades occur, as stated by Catalbas [23]. Gaze direction, especially when considering a driver's alertness, is another variable worth considering. Wang [24] suggests that the percentage of time the driver spends looking at the road while driving can be used to investigate their fatigue further. Observing how the driver's eyes behave over time gives a lot of information about their alertness.

2.2.4 Self-reporting

Instead of measuring fatigue as described in the previous sections, individuals can also indicate their fatigue by the use of a scale or questionnaire. Lupova [5] lists Swedish Occupational Fatigue Index (SOFI), Karolinska Sleepiness Scale, and Stanford Sleepiness Scale as a form of self-assessment.

SOFI was created to measure work-related fatigue [25]. The original SOFI, proposed by Åhsberg [26], consisted of 95 expressions gradable on a scale from 0 to 11. Later on, it was revised to 20 expressions gradable on a scale of 0 to 6, where 0 means "not at all" and 6 "to a very high degree" [27]. The 20 expressions are divided into five factors: Lack of Energy (worn out, spent, drained, overworked), Physical Exertion (palpitations, sweaty, out of breath, breathing heavily), Physical Discomfort (tense muscles, numbness, stiff joints, aching), Lack of Motivation (lack of concern, passive, indifferent, uninterested), and Sleepiness (falling asleep, drowsy, yawning, sleepy) [27]. Åhsberg [27] determined that the Lack of Energy is the general factor. Moreover, in the same study, next to the Lack of Energy expression, bus drivers (which is the closest profession to the truck driver examined in that study) gave high score to the expression present in Sleepiness category when fatigued. In his study [21], Johansson used Fatigue Severity Scale (FSS) next to SOFI. FSS consists of 9 questions, each one of which can be answered using a scale from 1 to 7, where 1 is "Strongly disagree" and 7 is "Strongly agree" [28]. The questions are present in Table 1.

		Scores 1 = Strongly Disagree. 7 = Strongly Agree						
1	My motivation is lower when I am fatigued	1	2	3	4	5	6	7
2	Exercise brings on my fatigue	1	2	3	4	5	6	7
3	I am easily fatigued	1	2	3	4	5	6	7
4	Fatigue interferes with my physical functioning	1	2	3	4	5	6	7
5	Fatigue causes frequent problems for me	1	2	3	4	5	6	7
6	My fatigue prevents sustained physical functioning	1	2	3	4	5	6	7
7	Fatigue interferes with carrying out certain duties and responsibilities	1	2	3	4	5	6	7
8	Fatigue is among my three most disabling symptoms	1	2	3	4	5	6	7
9	Fatigue interferes with my work, family, or social life	1	2	3	4	5	6	7

Table 1 Fatigue Severity Scale with questions

Johansson [25] created three fatigue categories based on the mean FSS score: non-fatigued (FSS ≤ 4.0), borderline fatigue (4.0 < FSS < 5.0) or fatigue (FSS ≥ 5.0). However, both SOFI and FSS scales only create information about how fatigued a person is or how fatigued a person was after completing a task [27] [28].

Karolinska Sleepiness Scale determines how fatigued a person was in the last 10 minutes [29]. It consists of 9 states shown in Table 2. However, contrary to FSS and SOFI, this scale is not filled by individuals experiencing fatigue but by people observing them [30].

Extremely alert	1
Very alert	2
Alert	3
Fairly alert	4
Neither alert nor sleepy	5
Some signs of sleepiness	6
Sleepy, but no effort to keep alert	7
Sleepy, some effort to keep alert	8
Very sleepy, great effort to keep alert, fighting sleep	9

Table 2 Karolinska Sleepiness Scale

Stanford Sleepiness Scale consists of only one question that is meant to be answered by an individual that could be experiencing fatigue [31]. Because the scale consists of only one question, it can be answered quickly (1-2 minutes with reading each score description) [31]. According to Shadid [31], this scale can be used to determine how fatigue of a person fluctuates over a period of time. The scale is shown in Table 3.

Degree of Sleepiness	**Scale Rating**
Feeling active, vital, alert, or wide awake	1
Functioning at high levels, but no at peak; able to concentrate	2
Awake, but relaxed; responsive but not fully alert	3
Somewhat foggy, let down	4
Foggy; losing interest in remaining awake; slowed down	5
Sleepy, woozy, fighting sleep; prefer to lie down	6
No longer fighting sleep, sleep onset soon; having dream-like thoughts	7
Asleep	X

Table 3 Stanford Sleepiness Scale

The Borg CR10 scale developed by Borg is used to measure perceived exertion during an individual's physical activity [32]. Using this scale makes it possible to measure how hard it feels on the body to perform a specific task. Similarly to the Stanford Sleepiness Scale, participants are asked to answer the form by themselves [32]. The scale is shown in Table 4.

Borg CR10 Scale	
Score	**Level of exertion**
0	No exertion at all
0.5	Very, very slight (just noticeable)
1	Very, slight
2	Slight
3	Moderate
4	Somewhat severe
5	Severe
6	
7	Very severe
8	
9	Very, very severe (almost maximal)
10	Maximal

Table 4 Borg CR10 Scale

2.3 Facial features recognition techniques

Development of face recognition system can be divided into three phases: face detection, feature extraction, and face recognition itself. Face detection, as identified by Kortli [33], is the phase of deciding whether there is a human face in a particular picture and localising it within the frame. Once the face is detected, the feature extraction can take place. This phase is based on locating ever-present facial landmarks: eyes, nose, mouth and determining the general geometry of the face. Finally, a person can be recognized or distinguished from other people during the face recognition phase. Using the features extracted in the previous phase, a face can be compared with other faces from a given database to find similarities or recurring patterns. Subsequently, the state of the facial features, whether eyes or mouth are open or closed, can be derived by calculating the geometry of the extracted features. This section of the chapter is devoted to state the art in the fields of face detection. The aspect of facial feature detection is part of chapter 3.

As stated in the requirements section, a system should operate in real-time or close to it to detect fatigue in drivers. Kumar [34], in his review, claims that it is possible for some portion of facial-recognition algorithms to operate in real-time. This claim is also supported by the review of Kortli [33]. Nevertheless, the face and facial features recognition techniques do not always work perfectly. Kortli [33] lists the resolution of the given input, illumination, complexity of the image background, and percentage of the face that is visible as factors that can hinder the performance of the algorithms. Subsequently, Kumar [34] adds unique, nonstandard facial expressions and face orientation within the frame to the list of challenges given by Kortli [33].

2.3.1 Face presence detection

Kumar [34] proposes dividing face detection into two categories: image and feature-based approach. The same division is also present in a review by Dang and Sharma [35].

2.3.1.1 Feature-based approach

The feature-based approach can be further split into Active Shape Model (ASM), Low Level Analysis, and Feature Analysis. The ASM search an image for the occurrence of a shape that fits an earlier created model [34]. The earlier created model can be an extracted facial feature such as an eye. The ASM is again split into three categories: snakes – detects and identifies face boundaries, Point Distribution Model (PDM) – uses vector representation of shapes (such as the example of eyes), and Deformable Templates – improves performance of the snakes as they can locate facial feature boundary [34].

Skin colour is a crucial parameter for Low Level Analysis. Tracking the face can be done by processing the camera input regarding skin colour [34]. Kumar [34] also states that processing other facial features is slower than just processing the skin colour. In Low Level Analysis, first, the colour of one's face is determined. Then, a threshold is applied to compensate at least a bit for changes in lighting or movement, and then the face image is extracted from the area in which colour lies within the threshold. The Low Level Analysis can be done for RGB, HSV and YCbCr colour models. These colour models differ in the way they describe colour. For RGB, the colour is described in terms of three main colours Red, Green, and Blue [34]. The HSV model takes into account attributes of the colour – Hue, Saturation and Intensity (V) [34]. The YCbCr model uses luminance (Y) and chrominance (Cb and Cr) to describe the colour [34]. Although colour is a base of all three listed colour models there are difference between them. The RGB model can be converted into normalized vector of colour which allows for fast processing of for example skin detection. However, RGB model is also light sensitive, and cannot set a clear distinction between Chroma and intensity of a pixel. Therefore, RGB colour detection is the least favourable of the colour models. The colour values of pixels belonging to the skin have similar Cr and Cb values for different races. Therefore, it is possible to create thresholds [Cr1, Cr2] and [Cb1, Cb2] that would adequately categorize region with skin colour as skin. However when there is more than one skin region present in the picture the RGB and YCbCr model fail [34]. The HSV model also creates thresholds with [H1, S1] and [H2, S2]. The Low Level Analysis method is a fast method for detecting faces; however, it falls short under the conditions of ambient lighting on the object [34].

One of the methods of Feature Analysis in Feature Searching. The Feature Searching method can use the Viola-Jones algorithm. This algorithm had 95% accuracy when released [36]. The Viola-Jones algorithm works on a greyscale image on which, firstly, Haar-like features (Figure 1) are used to detect regions of the face [37].

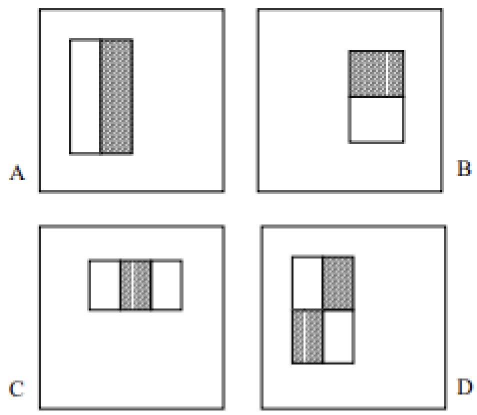

Figure 1 Examples of Haar-like features

Subsequently, an integral image is created by calculating rectangles neighbouring with the given rectangle. This is followed by training classifiers and finally combining the classifier to discard the background region quicker (so that only the face region is the region on which the calculations should be performed) [35]. The Viola-Jones is a quick and accurate face detection method that, even though developed in 2001 [37], still remains a relevant face detection method. However, as it operates on greyscale images, it does not detect black faces [34].

The classifiers for Viola-Jones algorithms are obtained by using the AdaBoost algorithm [35]. Nevertheless, AdaBoost can be used as a standalone algorithm for face detection [34]. AdaBoost was the first algorithm that was used for boosting. In machine learning, boosting refers to a technique of combining weak rules to create a strong prediction rule [34]. As described by Kumar [22], AdaBoost algorithms take a family of a simple classifiers to generate a strong classifier. The simple classifiers are combined linearly. The AdaBoost can generate a classifier that is close to the true classifier by using multiple iterations.

2.3.1.2 Image-based approach

The structure of the human brain inspires Neural Networks. They consist of artificial neurons connected to each other, forming multiple layers. For a signal to be passed through a neuron, the signal input needs to fit within a threshold defined on that neuron. The data carried by that signal can be changed by mathematical functions beforehand [35]. For detecting a face, a neural system goes through the image and examines the small windows on that image to determine whether that window has a face or not. This is done by two parallel subnetworks [34]. Afterwards, the outputs of the subnetworks are processed again to eliminate false detection from the previous stage. The face presence information is the output of this process [34].

2.3.1.3 Comparison between features and image-based approaches

The main difference between a feature and an image-based approach is the simplicity of implementation. The feature-based approaches are easier to implement than the image-based ones [34]. However, image-based approaches perform better with a cluttered background. The feature-based approaches are prone not to succeed under changing illumination and occlusion circumstances [34].

2.4 State of the Art for face recognition and face feature recognition

The studies in the field of face recognition have been conducted for over 40 years [38] yielding multitude of results and advancements. This section describes existing tools and approaches which features could be used to detect fatigue based on the symptoms described in section 2.2.3.

2.4.1 Mouth detection based on colour

Abtahi [39] proposes a solution based on Low Level Analysis to detect the mouth and its state. Firstly, the face is seen by Low Level Analysis and using bounding rules of three colour models (RBG, HSV, YCbCr) so that the detection is invariant to lighting and skin colour. This means that instead of focusing and using only one colour model information from all three of them are used to create a threshold for face detection. The mouth is detected using the same principle. After the face is already derived from the picture, it is possible to derive the mouth as the colour of the mouth differs from the colour of the skin. Lastly, a yawn can be detected again by looking at the colour difference. The inside of the open mouth has a different colour than the lips or skin. Therefore, it can be detected in the same way as the face and then the mouth. Moreover, when the detected area of the yawn falls inside the area between the lips and reaches a certain threshold size wise, it can be categorised as yawn [39].

2.4.1.1 Hand detection

During yawning mouth region is sometimes occluded by the hand of a person yawning [40]. This hinders mouth detection, as the algorithm might not recognise yawning when a person covers their mouth with a hand [40]. Monitoring the mouth state before and after the occlusion could help determine the yawning. Jie [41] proposes including other parts of the face in that evaluation. Detecting wrinkles created near the eye region while yawning was one of the proposed parts of the face. Moreover, detecting the hand itself is also possible. Viola-Jones algorithms and Haar-like features can be used again to detect hands [42]. However, if initial face detection is done by using the Viola-Jones algorithm, using the same method for detecting hands might yield poor results.

2.4.2 A Boundary-Aware Face Alignment Algorithm

There are many variations in the way in which the face can be oriented to the camera. As the face orientation changes regarding the camera, so do the position of facial landmarks; with the orientation of the head changing, the illumination of different landmarks changes as well. Moreover, some facial landmarks can be fully or partially occluded due to the head positioning, which makes it hard to detect the landmarks. The Boundary-Aware Face Alignment Algorithm first determines the head orientation by estimating facial boundary heatmaps. Afterwards, by using those boundaries, the facial landmarks are derived. With already placed boundaries that determine the shape of the face, it is easier for the facial landmarks to be derived as the area where they could fall is narrowed down [45]. Wu [45] points out that the boundary heat maps are used in multiple stages to improve the facial landmarks prediction further. The model of the face with landmarks is presented in Figure 2. The process of using the heatmaps is described in Figure 3.

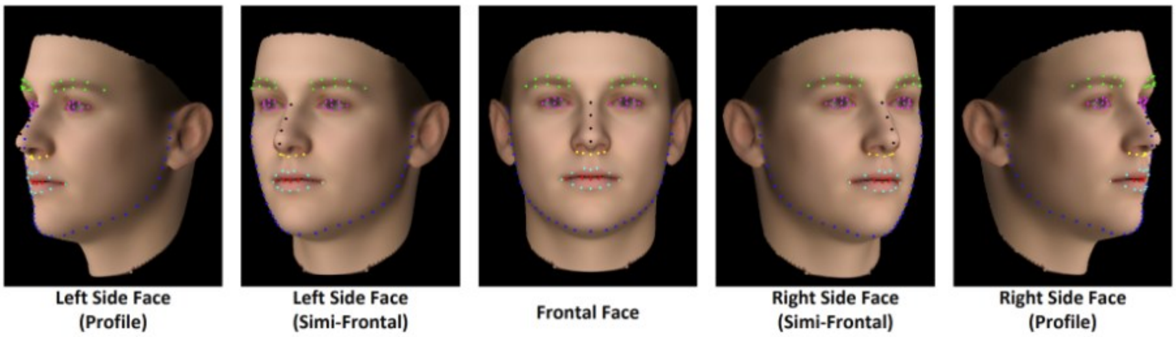

Figure 2 Multi-view illustration for landmark definition [35]

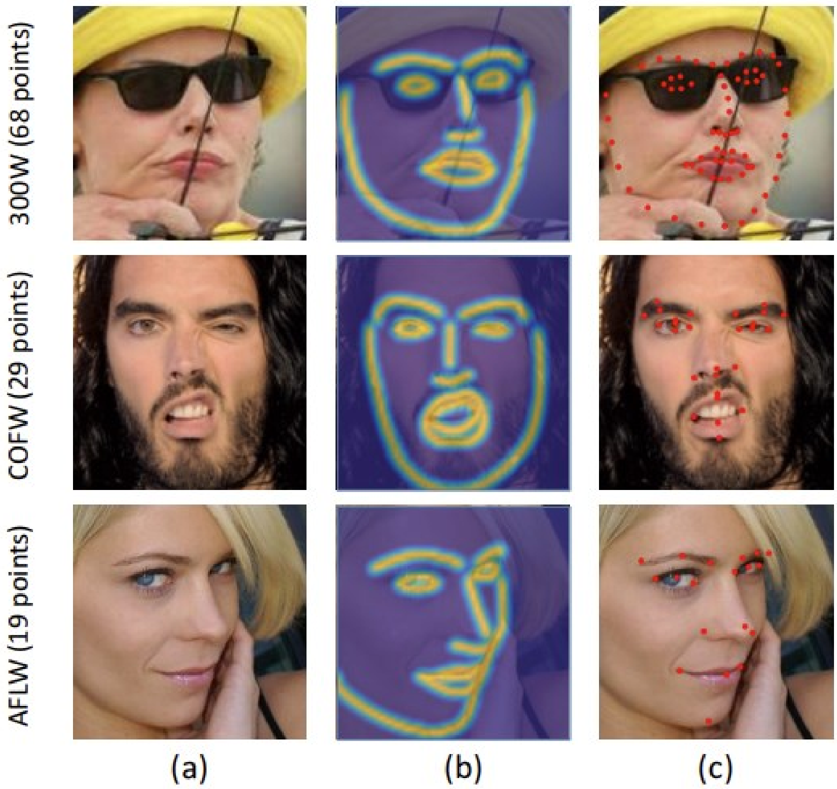

Figure 3 Column (a): the original photos from three different datasets, column (b) estimation of facial boundaries, column (c) landmark detection [35]

With the use of these methods, the face of a person could be detected not only when they are looking straight into the camera. Moreover, having the landmarks by points detecting fatigue can be done by calculating the distance between those points over time. By having a distance between a point on an upper and lower lid of an eye, blinking frequency and duration can be derived. A similar approach can be used to determine the frequency of yawning. However, yawning is similar to talking in terms of opening and closing the mouth. Therefore, constraints of the duration of the yawn, mouth dimension, and frequency of the mouth opening should be applied to ensure that talking is not being classified as yawning. And by a change in the head's orientation, the head tilt and the frequency can also be derived.

However, the system version described by Wu [45] does not operate in real-time. With using TITAN X GPU, the runtime of this algorithm averaged 60ms [45]. The GPU used is already a powerful one and most likely will not be implemented into the truck. Therefore, the runtime could extend the period by 300ms, making this method unsuitable for detecting fatigue symptoms related to the eyes. Nevertheless, as yawn lasts longer than the blink, yawning could still be detected.

2.4.4 OpenFace

OpenFace is an open-source framework that allows for facial features detection [46]. For facial landmark detection and tracking, it uses the Point Distribution Model (PDM) and patch experts for modelling the variations in the appearance of the landmarks.

"A Point Distribution Model (PDM) is a statistical parametric model of the shape of the deformable object, which is an essential part of many state-of-the-art deformable models.[47]*"*

In the case of OpenFace, the PDM is used to capture variations in the shapes of the landmarks. PDM and patch experts are the two main components of Conditional Local Neural Fields (CLNF), which can detect 68 facial landmarks – Figure 4.

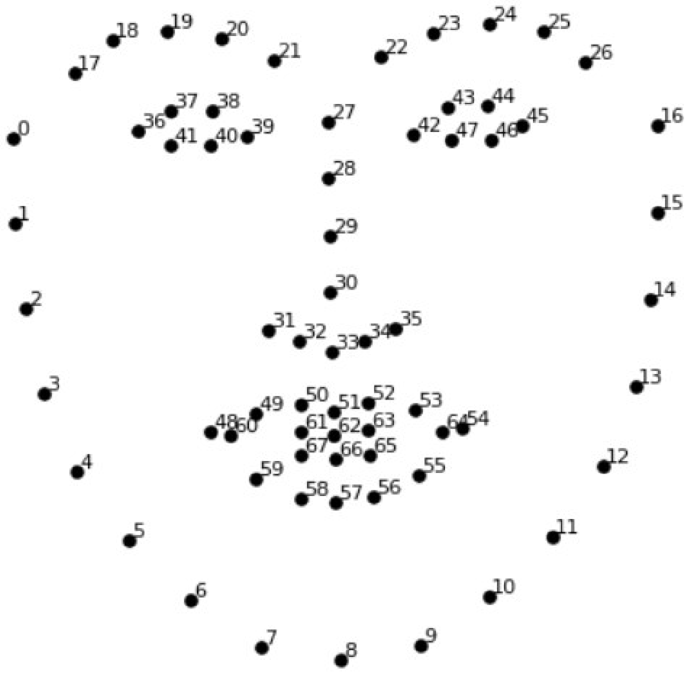

Figure 4 68 facial landmarks detected by CLNF

OpenFace also uses Convolutional Neural Network (CNN) to predict the expected facial landmark detection error due to drift. Drift can appear when a face is being tracked for an extended period of time [46]. The fatigue-related parameters can be derived by tracking the distance between the points, as described in the previous section. As for time constraints, Baltrusaitis [46] points out that OpenFace is capable of performing in real time.

2.4.5 OpenFace 2.0

There exists an improved version of OpenFace, namely OpenFace 2.0 [48]. One of the main upgrades is the speed of this tool. OpenFace 2.0 is 1.5 time faster than OpenFace and it is suitable to be applied to cases when data analysis in real-time is needed [48]. The Conditional Local Neural Fields (CLNF) was replaced by Convolutional Experts Constrained Local Model (CE-CLM) that consist of Convolutional Experts Network (CEN) and PDM. Those are the changes that allowed for increased speed with minimal loss of accuracy [48].

Zadeh [49] explains that CEN is the local detector used by CE-CLM. Its function is to determine alignment of each individual landmark. This is done by having the CEN take the region of interest (ROI) around current estimated position of the landmark on a given image and model the probability of the final alignment of that landmark. The modelling of the probability is possible by having a mixture of experts where each expert is trained on a subspace of the whole model. The variations between subspaces can include for example different lighting condition. The response from individual experts is combined to determine the final alignment probability. However, the output returned by CEN only takes into account individual landmarks without the context of the whole face. To regularize the shape of the face the PDM is used. PDM penalises irregular landmark locations thus controlling landmark detection. Figure 5 shows differences in accuracy between CLNF and CE-CLM in landmark position localization. CEN and LNF columns depict the probability of the landmark location in that region.

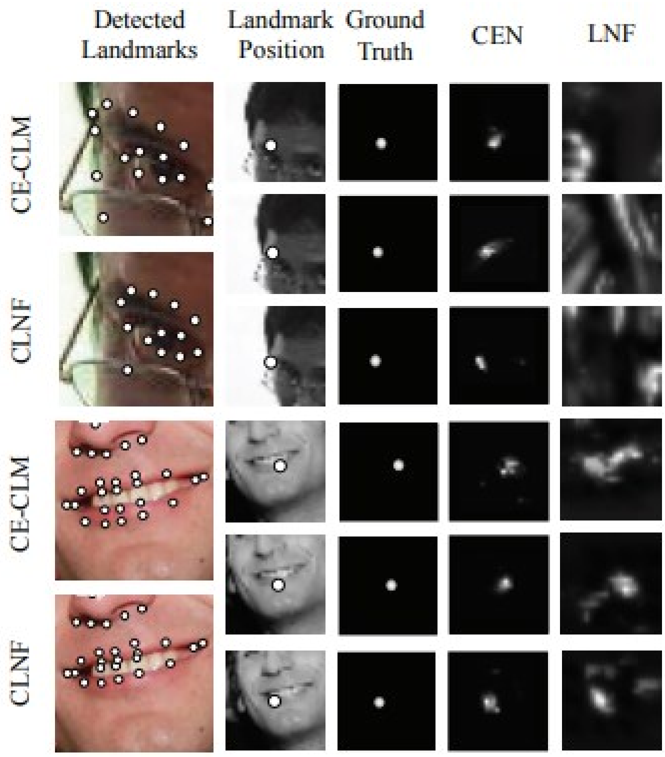

Figure 5 Comparison between CLNF and CE-CLM [45]

Another advantage of the CE-CLM using mixture of experts is the ability of extracting 3D facial landmarks without the need to be trained using the 3D data [49]. This allows the CE-CLM used in OpenFace 2.0 to deal with pose variation (when face is not frontally situated to the camera). Before the CE-CLM is applied to the image first the face detection takes place. OpenFace 2.0 uses Multi-task Cascaded Convolutional Network (MTCCN). The MTCCN provides a bounding box around the face [50]. Afterwards the 68 facial landmarks are extracted from the box. To speed up the landmark detection process in video or live feed input the CE-CLM is initialized based on the landmark location from the previous frame [48]. OpenFace 2.0 is capable of detecting head pose, eye gaze, facial expressions, and facial landmark location in real-time.

2.5 State of the art for fatigue detection systems

System for detecting fatigue detection have been already implemented in cars and trucks. Approaches to detect fatigue of the driver from various car manufactures are presented in this section

2.5.1 Mercedes’s Attention Assists

Within the first few minutes, the system analyses over 70 parameters to assess personal driving technique. Afterwards, the system assesses and identifies fatigue-related patterns derived from those 70 parameters. Finally, combining those patterns with information about external conditions such as road conditions, the system alerts the driver, suggesting them to take a break if the changes in steering behaviour show fatigue patterns.

2.5.2 Volvo’s Driver Alert Control

The Driver Alert Control (DAC) activates when the speed of the car is greater than 65 km/h and continues to monitor the driver’s behaviour as long as the speed is above 60 km/h. The parameter taken into account is the car position on the line. A camera detects how far the car is from both edges of the lane and compares those distances with steering wheel movements. When the car's behaviour on the line becomes significantly inconsistent, the system alerts the driver that it is time to take a break.

2.5.3 Nissan’s Intelligent Driver Alertness

Similarly, the DAC Intelligent Driver Alertness (I-DA) only activates after the car's speed reaches 60 km/h. This system monitors steering input patterns. By using steering angle sensors, the system can assess how smoothly or how roughly the driver is steering. If the difference between the baseline and detected steering pattern is significant, the I-DA alerts the driver prompting them to take a break.

2.6 Conclusion

It is possible to obtain information about one's fatigue by looking at that person's facial features. The best results in determining fatigue would be obtained when all fatigue-related areas (head, mouth, and eyes) and the information derived from the changes in them are combined.

There is a multitude of existing face detection methods. One set of them detects faces by the use of invariant facial features. The second one uses learned templates to detect faces in the given images. Similar, there is plenty of available toolkits that allow for facial feature recognition. However, the OpenFace 2.0, with the head pose, eye gaze, facial expression detection and facial landmark location, appears to be the most suitable tool for developing a fatigue detection system.

3 Requirements capture and ideation

Before starting the ideation process, this project's requirements must be listed and analysed. The chapter begins with requirements capture and is followed by an ethics discussion. Lastly, a listing of possible ideas and their advantages and disadvantages are provided.

3.1 Requirements and constraints capture

The client of this Graduation Project is the company Techspread. Techspread is a data-oriented start-up approached by several companies connected to the truck industry. Those companies inquired whether it would be possible to develop a system to detect the truck driver's fatigue by means of computer vision. The computer vison technique would be focused on observing and recognizing certain fatigue symptoms from the driver's face. This implies having a camera located inside the truck and pointing towards the driver. Therefore, to protect the driver's privacy, the paramount requirement named by Techspread was that the feed captured by the camera does not get stored. This means that the calculations performed by the system ideally need to happen in real-time. That puts an initial constraint on a type of facial-recognition technique that could be suitable for this system. The second named requirement has to do with the fact that sometimes truck drivers switch mid-journey. Therefore, there should be some feature to reset the system mid-drive so that the data collected on fatigue of the first driver does not influence what is detected for the second driver. Another requirement has to do with camera placing. The camera's position inside the truck should not be a hindrance in any way to the driver. Yet the device must be placed in a spot that allows the camera to capture the driver's face.

3.2 Stakeholders

Techspread, truck companies, truck drivers, supervisor and critical observer of this GP are the stakeholders in this project. Each one of them has different goals and different principles. Techspread, as a company, wants to deliver a good, reliable product that truck companies and truck drivers as individuals would be interested in. The truck companies have the safety of their employees as one of the essential aspects. The mentioned safety could be improved by introducing checked and reliable products. The truck company might also want to ensure that the product does exactly what it is supposed to do, as perhaps introducing it in a truck would require consent from the truck driver. The truck drivers might be interested in their safety, as well as a guarantee that the system only detects tiredness and not the fact that the truck driver is, for example, using a phone behind the wheel.

Furthermore, Techspread, as a company, does not employ any truck drivers. Therefore the system they aim to develop would not be used by people directly associated with the company. This means that the truck driver has a small power on the project but might be highly interested in it as it could be implemented into their trucks. Therefore, all the listed stakeholders have different interests to influence ratios illustrated in the Interest vs Power Map in Figure 6.

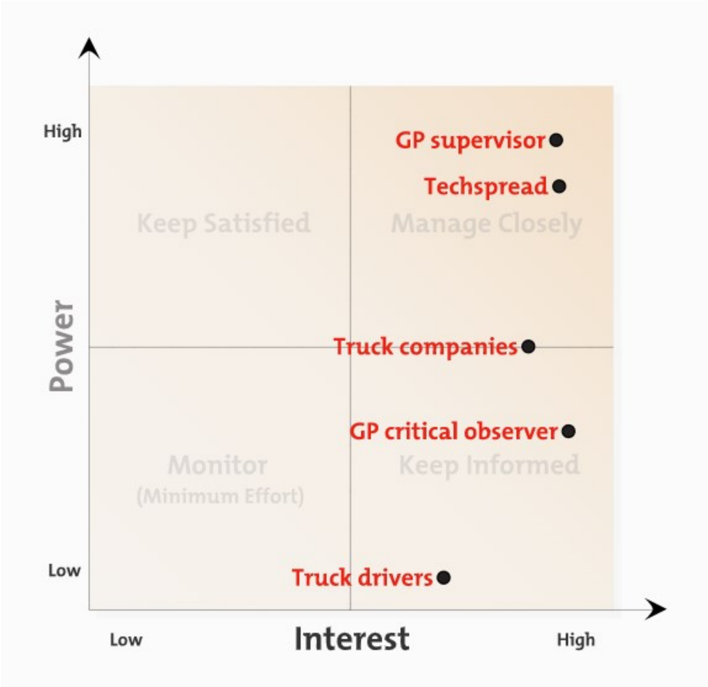

Figure 6 Interest vs Power stakeholder's map

3.3 Ethics

The scope of this Graduation project does not go beyond research. Therefore, when looking within the mentioned scope, there are not many ethical concerns. However, ethical dilemmas arise when taking into account possible usage, such as introducing this system into the industry. The following three sections describe the ethical dilemmas and provide possible mitigation of that risks.

3.3.1 Lack of knowledge

The system on which this research is being conducted on might end up being used by truck/transport companies to recognize tiredness in drivers. Although the facial recognition and facial feature recognition algorithms are well documented there is a high chance that nor the truck drivers nor their employees will read those papers. The same applies for the fatigue detection. That means that there could be some truck drivers who will have little to no idea on how the system works and why it makes certain decisions. From this also follows that some user of the system could have unrealistic expectations of the system. For example, if a truck driver will perform action similar to yawning (without actually yawning) frequently the system would then inform them that they are tired. With uninformed truck driver and unreal expectation of a system that person could assume that they are indeed tired (because the system has said so) and proceed to go for a break. This is the less harmful scenario, although it could result in truck driver not meeting his daily goal of kilometres travelled. However, there is a more harmful side to it. If a truck driver has an unrealistic expectation of the system and its not fully informed, they could assume that the system is always right. Even when the truck driver would feel tired, but the system would be saying that there are no signs of tiredness, the truck driver could continue his journey endangering for all road users. Therefore, in the future the acceptable use of this system is to have it only as an indication of one's state in terms of fatigue and not as a definitive truth. The unacceptable use of it would be for the employers and the truck drivers to use this system and only this system as a deterministic factor when it would come to make decisions regarding one's fatigue.

3.3.2 Extracting more data by the employer

Moreover, there is also a risk that the employers of the truck drivers will use this AI to evaluate how quickly their employees get tired and do some cuts in the team based on that. Of course, there is also the bright side that using the same evaluation the employer will adjust working hours/times of their employees. The AI could possibly be extended to recognize when the truck drivers are using their phone during the ride. That could also be used against the truck drivers, however there is also a chance that recognizing phone usage while a person is behind the wheel could improve the overall road safety. For the risk mitigation, when introducing the system into the industry the researcher would make sure that the data collected by the system about action performed by the truck driver and their fatigue will be deleted after each drive, so that the employer will not have access to it. And furthermore, having a disclaimer informing truck drivers that the communicates sent by the system are only an indication.

3.3.3 Forced consent

As the system could be introduced by employers to their truck driver the truck driver could have a choice of either quitting their job or consenting to the usage of this AI model. A possible mitigation of this risk would be for the system to always ask for a consent of the truck driver before each drive. However, it is not guaranteed that the employers will not "force" upon the truck drivers to always consent to this. Another risk is that if there is already a camera pointed towards truck drivers face the whole drive could be recorded, which could violate one's privacy. To mitigate it, the system working on the camera could not allow another system (for recording) to get the access to the camera. However, again nothing could stop truck drivers' employers from putting another camera in. In summary the acceptable use scenario for this system is to only have inform non-deterministically the truck drivers how fatigue they are (in accordance with the system). The unacceptable use scenarios involve truck drivers taking the information given by the system as the only truth, and employers accessing the data collected by the system and making decision involving the truck drivers based on that. As well as the employers installing other systems next to the one mention in this research to monitor the truck drivers and "forcing" the truck drivers to consent to that.

3.4 Specification

To prioritize requirements and therefore create the specification of the system the MoSCow Technique is used. MoSCoW uses for labels: *Must Have, Should Have, Could Have, Won't Have* to sort the features in terms of their priority. The meaning of the following labels is as follows

- *Must Have* – requirements that have to be included in the final product, highest priority
- *Should Have* – not critical requirements for the final product but if included, would add high value to the final product
- *Could Have* - requirements which could be nice to include (not crucial) and do not cost a lot of effort to be implemented into the final product
- *Won't Have* – requirements that have been requested but are outside the scope of the project but still can be included in the future iteration of the final product, lowest priority [51]

Based on the labels a list of requirements is constructed with the explanation as to why that priority was assigned.

Must Have

- The system must detect fatigue using facial feature – requirement stated explicitly by the client
- The system must not store video data – requirement stated explicitly by the client.

- The system must perform in real-time, close to it – requirement follows from the previous one. Since the no video data can be stored the system needs to perform calculation "on the spot."
- The system must only recognise fatigue – requirement follows from the ethic discussion
- The truck driver must not be identifiable based one data collected – requirements follow from the ethic discussion
- The system must not have any gender or racial biased

Should Have

- The system should not be hindering the driver's performance – requirement follows from requirements capture
- The system should work with different lighting conditions - requirement follows from requirements capture
- The system should take into account truck drivers switching mid drive - requirements stated explicitly by the client

Could Have

- The system could use input from another sources (such as sensors) to determine fatigue - combining input from the sensors and parameters used by the companies (section 2.4) could further improve the system
- The system could use multiple facial features to determine fatigue – combining more facial features could further improve the system

Won't Have

- The system won't have feedback to the driver – providing feedback to the driver in a best possible way lies outside the scope of this graduation project

The requirement present in this list will be used to evaluate the system.

3.5 Ideation

Before proceeding with the Realization phase of this project, an initial idea had to be created. This was done based on the requirement listed by the client and information gathered in section 2. Therefore, this section outlines the paramount requirements and ideated solutions matching them. From the concept of the project, it can be deduced that the system should work well during the day and the night. As mentioned in chapter two, lighting conditions can severely impact algorithms' performance for face detection. Moreover, using a standard camera during the night without any additional light source, as placing one in a truck cabin could be a distraction for the truck driver, will result in very dark, unreadable images. An infrared camera could potentially solve that problem, as it can be used in low light conditions [52].

Another constraint is related to the camera's frame rate and the speed at which the fatigue state analysis can be done. Firstly, the program analysing the face should work fast enough so as not to miss any signs of fatigue happening on the face. This is also connected with the frame rate of the camera. Slow acquisition of data from the camera might result in missing those signs. Moreover, recognising fatigue symptoms should also happen in real-time or close to it to provide the most accurate and updated information about the driver's state. Therefore, for the real-time performance, the Boundary-Aware Face Alignment Algorithm had to be excluded from the consideration as it was not suitable for real-time performance.

The remaining techniques from section 2 are the colour approach and OpenFace. For the OpenFace, since both OpenFace and OpenFace 2 are available, and OpenFace 2 is an improved version of OpenFace, OpenFace is also excluded from the consideration. Therefore, possible approaches are the colour-based approach or OpenFace 2.

Fatigue signs detected from the face can come from the eye region, mouth region or head movement. The literature stated the maximal duration of the blink in a non-fatigued person is 300 ms. Therefore, to successfully detect whether the blinking takes more than that, the camera rate should be at least four frames per second (fps). That way, the image is captured every 250 ms, which allows to at least determine whether the blink last longer than 300 ms or not. Unfortunately, the colour approach is not suitable for detecting the opening and closing of the eye, and there is not enough precise information about how fast the OpenFace 2 can operate when analysing eye-opening patterns. Therefore, the sign chosen for fatigue detection is yawning.

With the fatigue sign chosen, there were two approaches distinguished to consider. First yawn detection with colour approach and yawn detection with OpenFace 2. Initially, the first approach was selected to proceed with. After discussing this potential approach with the client, there was an agreement reached that this research could proceed with yawn detection as a means to detect fatigue.

4 Methodology and Realization

This chapter revolves around the development of a fatigue detecting system. Following deriving the initial idea in the previous section, more requirements are added accordingly to the MoSCoW method. Subsequently, a methodology section to describe the system's performance is examined. Finally, realization is described by means of iterations necessary to reach a final state of the system in this GP.

4.1 Requirements update

From the ideation phase it was determined to detect fatigue of a driver by the mean of yawning patterns. With this information the list of requirements was updated. The requirements listed in section 3.5 are still valid and following requirements are added:

Must Have

- The system *must* detect yawning
- The system *must* determine fatigue based on the yawning pattern information

Should Have

- The camera *should* not disturb the driver

Could Have

- The camera *could* be connected to a piece of hardware that can run a software used for fatigue detection

The requirement present in this list will be used to further evaluate the system.

4.2 Methodology

Assessing the performance of a system requires a predefined methodology or evaluation criteria. To ensure the quality of the final product and adherence to the derived requirements following evaluation criteria are used. For this GP, the evaluation criteria can be divided into statistical evaluation and user evaluation.

4.2.1 Statistical Evaluation

The accuracy of the system, in terms of detection, is determined by analysis of the confusion matrix [53]. When detecting occurrence of an event four cases can happen:

- True Positive TP: The event has occurred and was detected
- True Negative TN: The event has not occurred and was not detected
- False Positive FP: The even has not occurred but was detected
- False Negative FN: The even has occurred but was not detected

After assigning obtained results to the four cases other parameters can also be calculated and analysed:

- Recall: What fraction of True cases was detected correctly
$$Recall = \frac{TP}{TP + TN}$$
- Precision: What fraction of all True prediction was correct
$$Precision = \frac{TP}{TP + FP}$$
- Accuracy: How often the detection is correct
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
- F1 – score: Take both precision and recall into consideration
$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

4.2.2 User Data

The evaluation based on user data in this GP project is applied first to evaluate how well the system can detect fatigue symptoms and, secondly, how well the system can determine a fatigue state of a person. For the first part, the participants were asked to sit in front of a laptop with a webcam on. The first user evaluation lasted approximately 20 minutes, during which the participants were asked to move their heads around and yawn (or pretend to yawn to the best of their abilities). The researcher was present for the full duration of the study and assessed the system's output. By comparing the output with what was the participant doing, the researcher could assess how well the system performs and under which conditions it falls short.

The second phase required the participants to sit in front of their laptops while they performed tasks such as studying, working or reading on their laptops. A webcam was usually placed above the participant's laptop and connected to another laptop on which OpenFace 2 was running. The participants were asked to assess their fatigue on a scale of 1 – 10 every 20 to 30 minutes. As self-reporting fatigue is very subjective to a participant, and different people tend to scale their fatigue differently (for one person, fatigue on a level of 5 would be on a level of 7 for somebody else), the numbers obtained from participants were only used as an indication whether the system had correctly increased or decreasing fatigue of a person. During this phase, the researcher was absent, not looking at the participant while participating in the study.

4.3 Realization

The development of the fatigue detection system can be divided into two major approaches. Approach one focuses on yawn detection based on colour differences, and approach two implements OpenFace 2 to detect fatigue. Moreover, the realization of the system was done in the scope of several iterations. Each iteration consists of either testing an idea, evaluating certain features or introducing relevant changes to the system, as well as combinations of all three possibilities. The iteration followed the principles of the iterative design described by Schell [54].

4.3.1 Approach 1: Detecting yawn through colour

The first approach aims to detect yawning by the use of the colour approach described in section 2.3.1. The main idea of this method is to detect, using OpenCV, the open mouth shape by the difference in colour between the inside of the open mouth and the skin around it.

4.3.1.1 Iteration 1.1: Colour method

The colour of the inside of the mouth was assigned by analysing a few pictures of people yawning and determining the most fitting colour value – "*colour 1"*. The same procedure was used to determine the colour of the skin, *"colour 2"*. After determining the most fitting colour values, the yawn detector was set to assign a yawn to a particular frame when both *colour 1* and *colour 2* were detected, and *colour 1* was surrounded by *colour 2*. This method was outputting a lot of false positives as the inside of the nostrils and inside of the mouth have, in some cases, similar colours. If a face was titled slightly backwards so that the nostrils were visible, the system categorised frames obtained at that moment as frames depicting yawning.

4.3.1.2 Iteration 1.2: Adjusting the shape

The second iteration for the colour method included assigning shape constrains to the detected colours. For the system to categorize a given frame as a frame containing a yawning face, the detected *colour 1* should approximate oval by its shape and be of a certain minimal size. These constraints eliminated the issue of the system outputting many false positives. Nevertheless, the colour method proved to be unreliable with different lighting conditions; the assignment of *colours 1* and *2* should constantly be changed to recognize the yawning. Furthermore, extending the colour assignment thresholds (so that more of the similar colour could be categorized as either *colour 1* or *colour 2*) was not successful. On the contrary, it caused even more chaos as items visible in the frame with cluttered background were also taken into account, resulting in multiple yawns detected in one frame with only one face visible. Therefore, this method was not further pursued.

4.3.2 Approach 2: OpenFace 2

After determining that the colour method is not suitable for this project the OpenFace 2.0 was examined.

4.3.2.1 Iteration 2.1: Understanding and tweaking OpenFace 2.0

OpenFace 2.0 is a complex tool that can detect head pose, eye gaze, facial expressions, and facial landmark location as described in section 2.3.4.2. It is available, provided a proper citation of resources used (Appendix A). The current (as of June 2022) version of OpenFace present on GitHub is version 2.2.0. That version includes an application through which the user can input images, videos or webcam feed for the program to analyse. After input is created, OpenFace 2 starts writing results to a CSV file; the results are:

- Gaze angle (x, y)
- Gaze location (x, y, z) - location of the pupil for both right and left eye
- Head Pose (x, y, z)
- Head Orientation (x, y, z) as an angle
- Location of 2D facial landmarks (x, y)
- Location of 3D facial landmarks (x, y, z)
- Facial action unit presence
- Facial action unit intensity

The Facial Action Coding System (FACS) is used to classify facial movements of people by how the movements appear on the face. Each facial expression can be deconstructed into Action Units (AU) that created the expression. In the original paper published in 1978, there were 64 AUS derived. The OpenFace 2 recognises 18 of them (Figure 7).

AU	Full name	Illustration
AU1	Inner brow raiser	
AU2	Outer brow raiser	
AU4	Brow lowerer	
AU5	Upper lid raiser	
AU6	Cheek raiser	
AU7	Lid tightener	
AU9	Nose wrinkler	
AU10	Upper lip raiser	
AU12	Lip corner puller	
AU14	Dimpler	
AU15	Lip corner depressor	
AU17	Chin raiser	
AU20	Lip stretched	
AU23	Lip tightener	
AU25	Lips part	
AU26	Jaw drop	
AU28	Lip suck	
AU45	Blink	

Figure 7 AUs recognized by OpenFace2

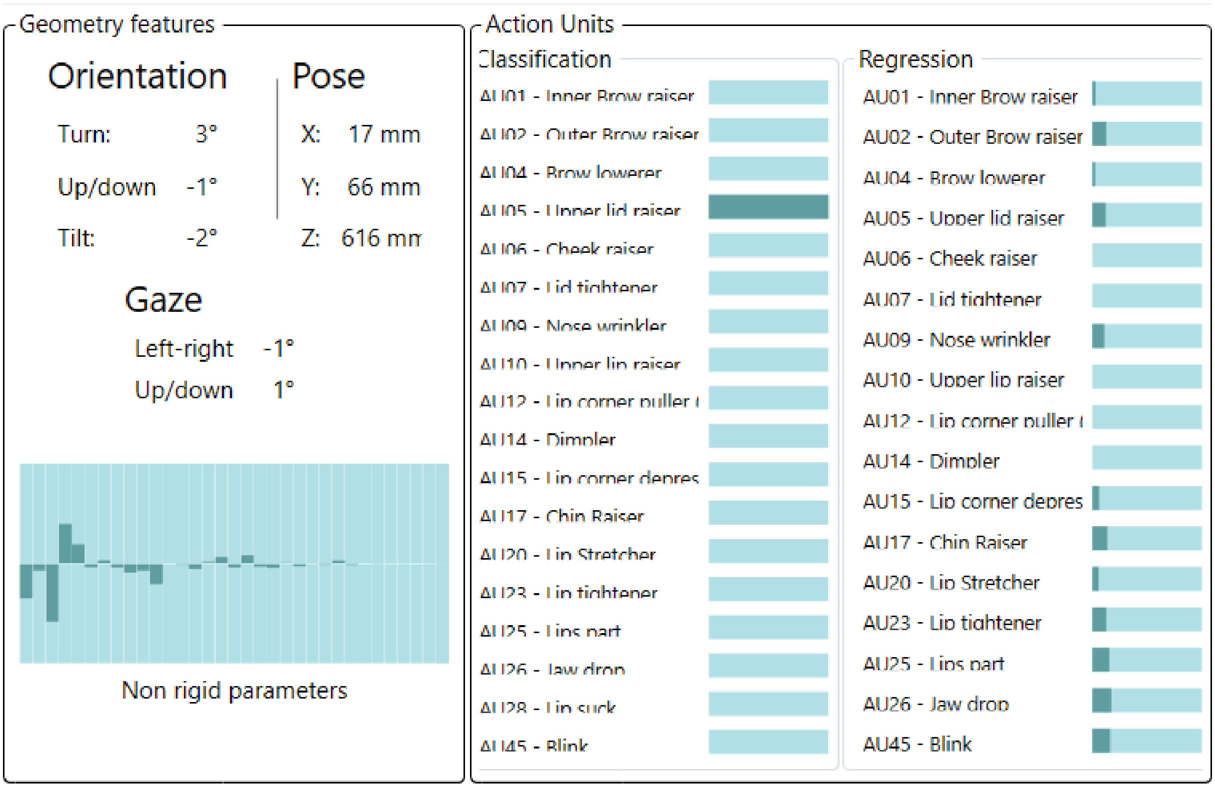

Figure 8 OpenFace2 demo app interface

Installing all the necessary dependencies for the OpenFace 2 to work also contributed to the total workload of this project. The GitHub repository of OpenFace 2 suggested using the 2017 version of Visual Studio Code (VSC) to avoid compatibility issues. I have determined as of June 2022, the OpenFace 2 is not compatible with VSC version 1.68, released in May 2022. However, the GitHub repository suggests that it is possible to achieve compatibility with newer versions of VSC after creating proper DLL and LIB files. Nevertheless, in this Graduation Project, the 2017 version of VSC is used. Furthermore, after editing the code of the OpenFace 2, the compiled program was significantly slower than the provided app version of it. Finding correct settings for the compiler and the correct version of Windows SDK to ensure that the frame rate goes back to around 30 analysed frames per second (the frame rate dropped to around 2 analysed frames per second) also required time investment.

4.3.2.2 Iteration 2.2: Threshold estimation and data labelling

Following an understanding of what is in the output of the OpenFace2 sample, pictures with yawing and not-yawing faces were analysed. The occurrence and intensity of particular AUs were taken into account. After calculating the average score of occurrence and intensity for both yawning and now yawning faces, a threshold was determined. If the values from the most prominent AUs detected when yawning (17, 25, 10, 7) felt into the threshold, the frame from which they were obtained was considered as a frame containing yawning. This method worked well when the input image was one of the images used to create the threshold; however, when additional pictures not included in threshold creation were introduced, the accuracy of the detection drastically plummeted.

With that finding, another approach to detect yawning was needed, and more annotated data was needed. The YawDD dataset provides videos of people sitting in the car, talking, moving their heads, and yawning. The videos were of both people with and without glasses. But only videos with people without glasses were analysed. The videos were put through OpenFace2, and afterwards, a value of yawn (1) and no yawn (0) were assigned. In the annotated data, the beginning of yawn is only categorised when the person is yawning and has the mouth maximally wide open (when the person is in the acme phase) – Figure 9. The end of the yawn is categorised when the mouth of the person yawning is not fully wide open anymore. That classification does not follow any literature. I have decided to classify the yawn only when the mouth is fully open, as, before that, various people have various facial expressions. Classifying those “pre-yawn” expressions as yawn could lead to inaccuracies with yawn detection. The obtained data set has 36223 entries, from which 4485 are yawning and 31738 are not yawning. The data included in the dataset came from both men and women with ethnically different backgrounds. However, within the data set, there were no videos present of people with dark skin.

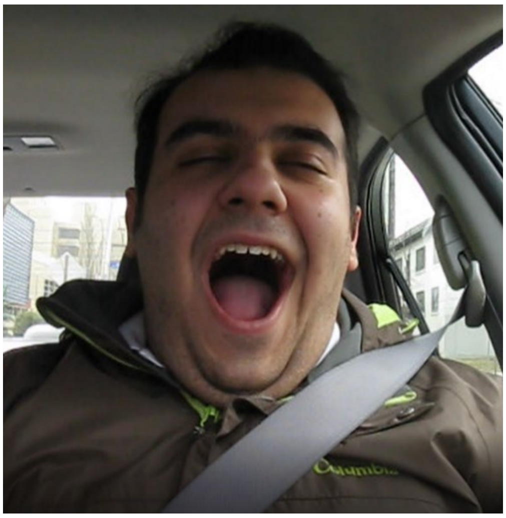

Figure 9 Face labelled as yawn

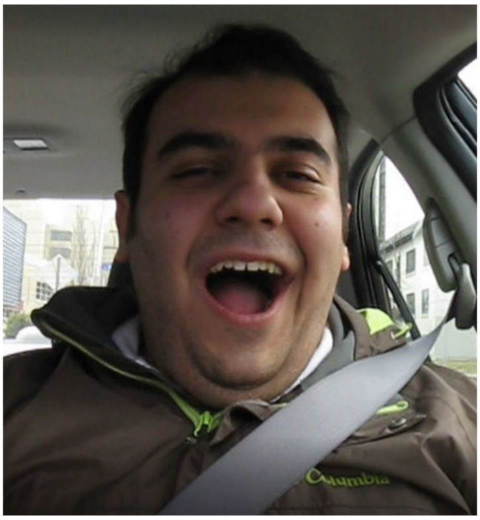

Figure 10 Face labelled as no yawn

4.3.2.3 Iteration 2.3: Machine Learning

Yawn detection should return only two values: true and false – yawn detected and no yawn. Support Vector Machine (SVM) - supervised learning model was analysed. Supervised meaning that the model required pre-labelled data for paired input-output examples to be able to perform classification on any other given input as soon as the input consists of the same data types as the input used for training the model. To classify data, the SVM creates a plane, or in the case of two-dimensional data, a line that would best separate the instances that fall into different categories. The best possible plane or line is determined by the distance it has to the closest member of each category. The furthest possible distance is desirable [55]. When the data cannot be separated within the dimension it was originally entered, the SVM does a kernel trick, which makes the data scattered over higher-dimensional space and thus possible to categorise. Figure 11 illustrates the kernel trick.

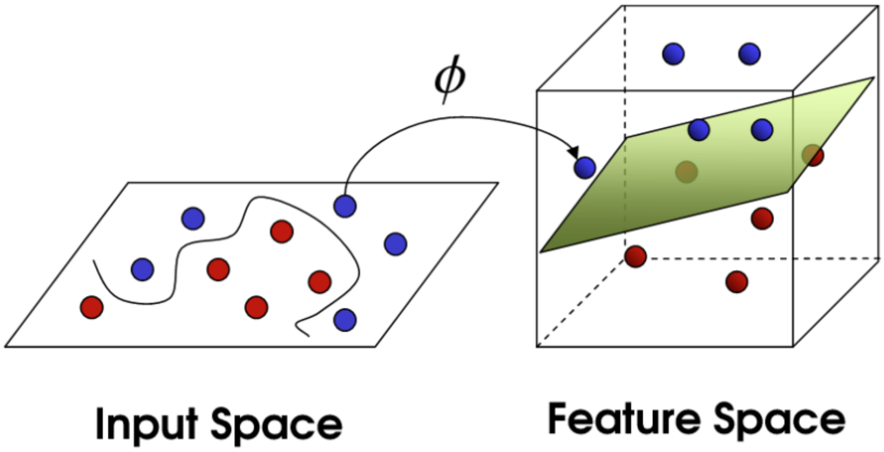

Figure 11 SVM kernel trick [56]

Scikit-learn library was used to create and evaluate the SVM. To analyse whether the speed of the SVM algorithm was satisfactory for this problem, a time was obtained in which it takes the model to predict values of randomly selected test data out of the data obtained from the previous iteration. The dataset videos and pictures of people in the acme phase were split into test and train data in the following manner. In the initial approach, the data was split into train and test sets randomly. Random rows from the whole data set were assigned to the training set, and the remaining ones created the testing set. As the data rows were created after video frames, randomly split rows are not independent of each, thus resulting in high accuracy, precision and recall. The confusion matrix for SVM for this wrongly divided data is shown in Figure 13, and the statistical criteria in Table 5. The time it took SVM to predict values of around 12000 entries was 1.11 s. The conclusion was that the accuracy and the time it takes the SVM to evaluate given input makes it an appropriate algorithm to use for yawn detection. Unfortunately, this mistake was only discovered after all the iterations had been completed. However, the model used for the remaining development of this system remained the same, as the contents of the dataset did not change. Therefore, further iterations and conclusions from them remain valid, but the first assumption about the model's accuracy, precision and recall is incorrect.

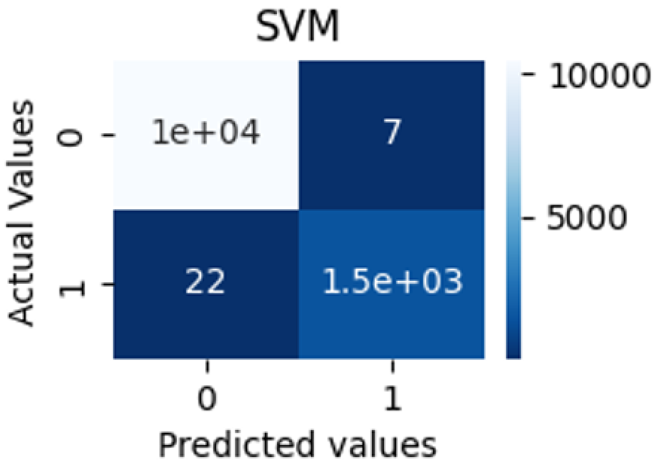

Figure 12 Confusion matrix for SVM

Accuracy	0.998
Precision	0.995
Recall	0.985
F1 - score	0.990

Table 5 SVM evaluation incorrect data split

The correct data division is as follows: the total dataset contains videos of 9 different people and pictures of random people during the acme phase. 2/3 of all videos and all the pictures of people in the acme phase were used to create the training set, and the remaining 1/3 of videos created the testing set. The confusion matrix for SVM for the correctly divided data is shown in Figure 13, and the statistical criteria in Table 6.

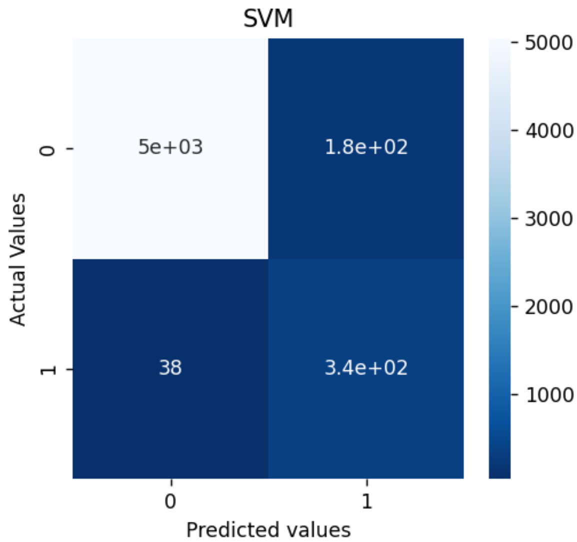

Figure 13 Confusion matrix for SVM

Accuracy	0.961
Precision	0.651
Recall	0.899
F1 - score	0.755

Table 6 SVM evaluation correct data split

4.3.2.4 Iteration 2.4: Real-time calculation

One of the must-have requirements of this project is for the system to be able to detect and determine fatigue states in real-time. The OpenFace 2 can analyse incoming images up to 30 frames per second. With webcam feed as input, the speed with which OpenFace 2 can analyse incoming frames oscillates between 22 – 30 frames per second. This is satisfactory speed when it comes to real-life calculations. As explained in section 4.2.3, SVM is also fast enough to proceed with real-time fatigue detection.

OpenFace 2, written in C++ outputs to the CSV file. That file is opened by another program, written in python and read line by line. To prevent the reading of an incomplete line, the python programs only read a line if it has a certain length and ends with an "end" string. If even one of those conditions is not true, the program waits till there is a complete line to be read.

Next to that, there was an attempt to implement a port communication between OpenFace 2 and the python program, which would eliminate the incomplete line problem. However, with the CSV file, it was more feasible to analyse the data multiple times without running both programs simultaneously. As the OpenFace 2 would save the file after it has been terminated, and the python program then is able to access and analyse the file at any given time. With port communication, that would not be possible as the OpenFace 2 needs to send the data to the python code at least once so that the python code can save it. Nevertheless, the port idea was not implemented, but it might be worth considering for future applications to avoid writing and reading the same file simultaneously.

4.3.2.5 Iteration 3.1: Yawn classification improvement

SVM algorithm has very high accuracy regarding yawn detection (around 0.94). However, it would also classify a wide-open mouth as a yawn. That state of mouth sometimes occurs when a person is speaking. To minimise the wrong yawn detection, the yawn is only classified if it has lasted more than 2 seconds (minimal yawn duration). After a series of frames is detected as NO YAWN, a YAWN frame occurs, and then the program stores the time of the YAWN frames and checks what frame will come next. If the following frame is also a YAWN frame, nothing happens. However, if the following frame is NO YAWN, the program stores the information about the time of the NO YAWN frame and waits for another. If another frame is YAWN and the time difference between NO YAWN and the newest YAWN frame is smaller than 0.1 s, the NO YAWN frame is discarded. Otherwise, if the difference is bigger than 0.1s, the program checks whether the time difference between the time of the initial YAWN frame and the time of the NO YAWN frame is greater than 2 seconds. If yes, there is a yawn detected of duration NO YAWN time – initial YAWN time, if no yawn not classified. That way of detection and classification reduces the amount of wrongly categorised yawns. Figure 14 shows initial yawn classification without the minimal duration implemented, and Figure 15 shows the same data but with the minimum duration detection applied. 0 means NO YAWN, and 1 means YAWN. The three peaks appear between 10 and 13 seconds in the mouth opening. The first approach would classify that action as a yawn. The second approach with the minimal duration did not classify those actions as a yawn, however, it failed to recognise the second actual yawn as a yawn.

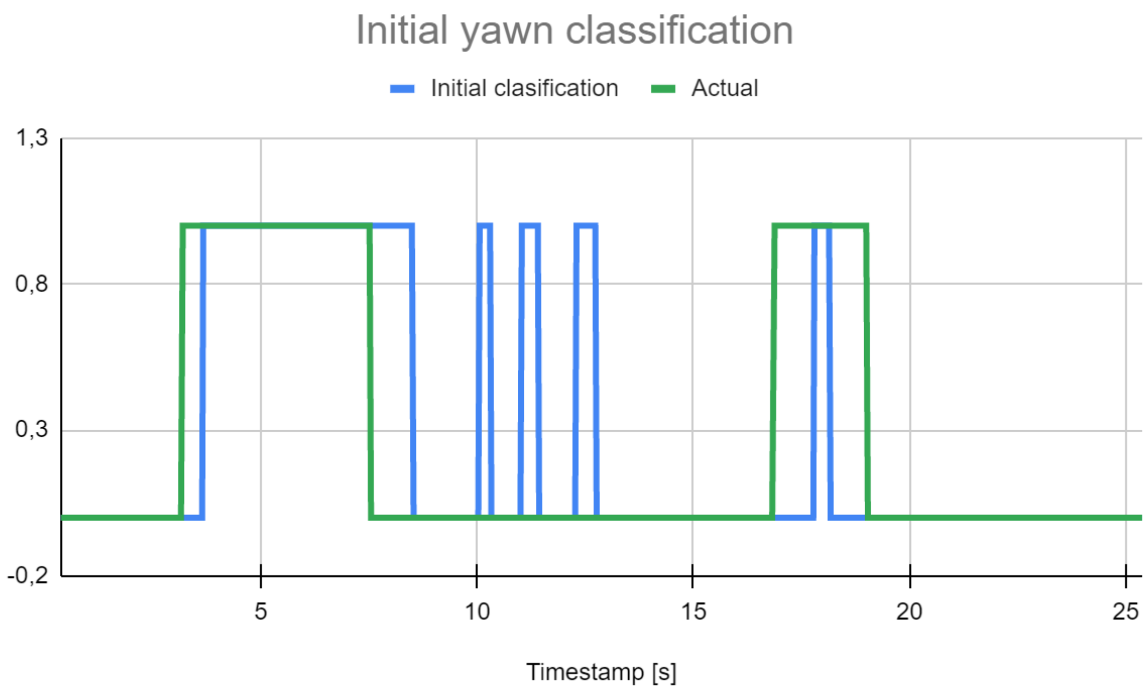

Figure 14 Initial yawn classification graph

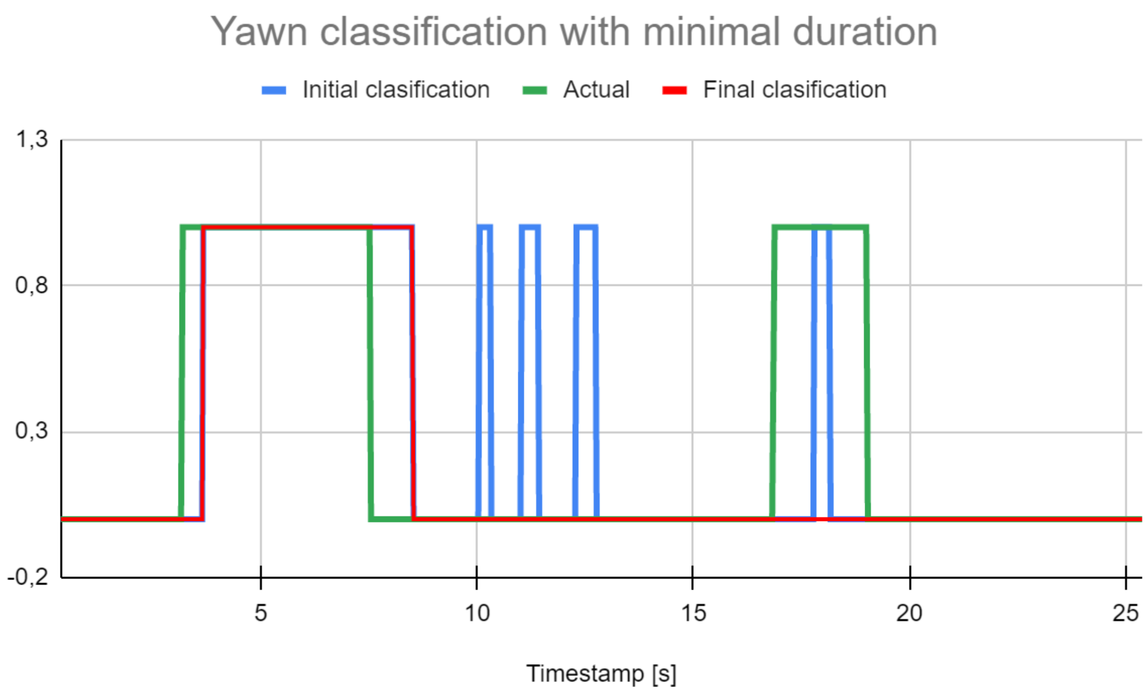

Figure 15 Yawn classification with minimal duration graph

4.3.2.6 Iteration 3.2: Accuracy improvement

To further improve the accuracy of the system, another parameter is examined. The OpenFace two also outputs a parameter called confidence. Confidence on a scale of 0 to 1 indicated the certainty with which the system has detected a face. For example, if the certainty is 0.1, that might indicate that if a face is present in OpenFace 2, the input is either rotated or heavily distorted. Analysing frames with low confidence also leads to incorrect classification. That is because parameters do not have values that are in accordance with real life. Therefore, the python program only considers frames with faces detected with 0.8 or higher confidence. The value of 0.8 was determined as the most suitable one as the yawn detection works well with that value, and yet the value is low enough so that the system can take plenty of frames into account. If the confidence is lower than this, the program takes the YAWN or NO YAWN value from the last properly detected frame and sets it a YAWN or NO YAWN value for the frame with confidence smaller than 0.8. If the last properly detected frame were a YAWN frame, the program would only continue to assign incoming frames as yawn for 2 seconds and then switch back to NO YAWN. This is implemented to prevent the system from recognising very long yawns only because a person decided to turn while yawning. Figure 16 shows the graph with yawn prediction without considering the Confidence value, and Figure 17 shows the same data but with the Confidence value in consideration.

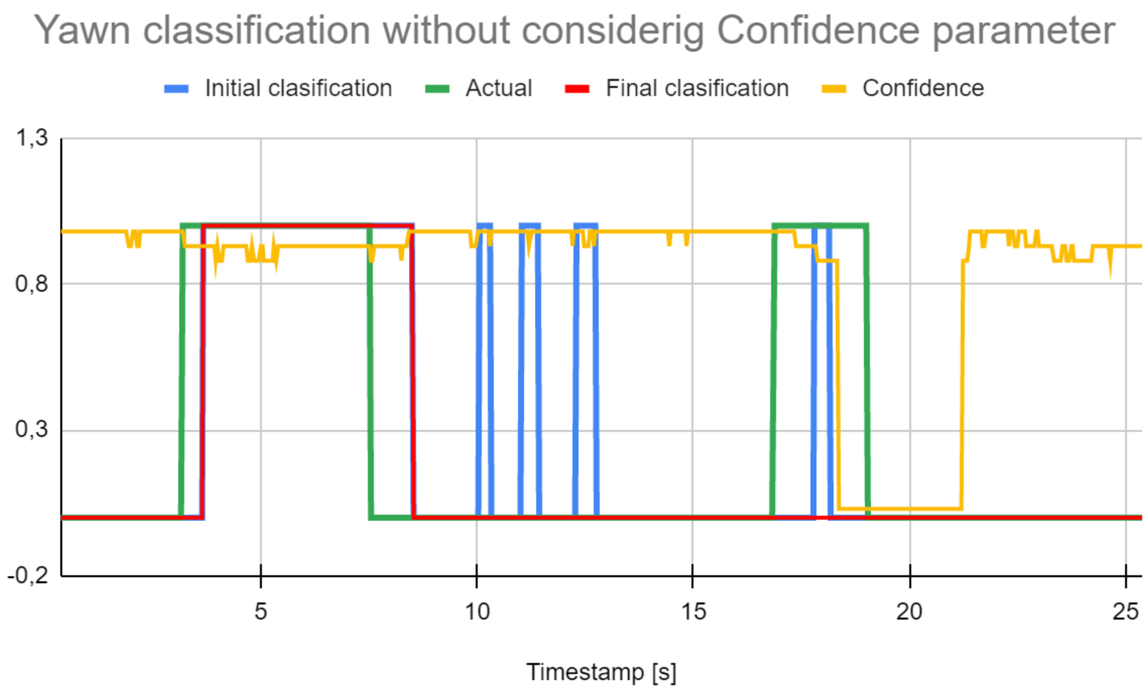

Figure 16 Yawn classification without considering Confidence parameter

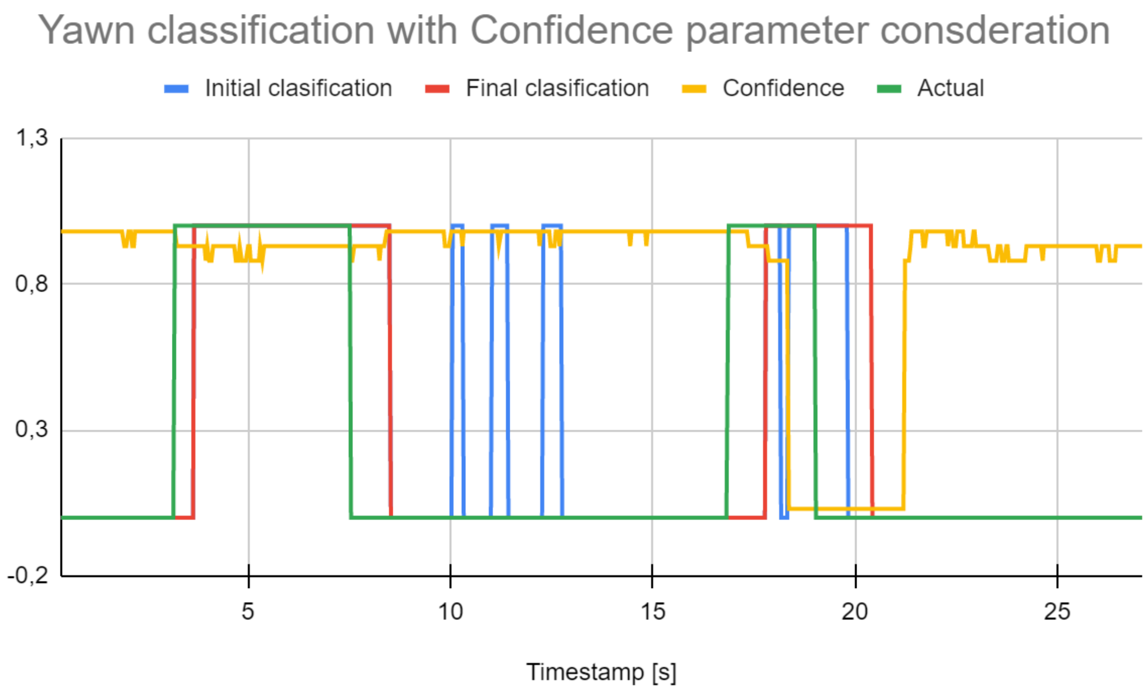

Figure 17 Yawn classification with Confidence parameter consideration

4.3.2.7 Iteration 3.3: Yawn calculations and yawn detection improvements

Iteration 9 mainly revolved around testing and reacting to test results. Nevertheless, firstly a few calculations for yawn parameters were introduced: yawn duration, overall yawn frequency, yawn frequency in the past 30 minutes and yawn number in the same way as yawn frequency. The first part of the user evaluation was conducted based on those parameters. Based on the findings from the initial part of the user evaluation, the main conclusion stated that the system sometimes detects false positives, which are later categorized as a yawn. Therefore, another machine learning method was introduced called Random Forest.

Random Forest is another Machine Learning model that can be used for classification. Random Forest consists of multiple classifiers. Those classifiers are Decision Tree models. A Decision Tree model starts from one node ("root") and then is split into more nodes – branches ("as the tree grows") and branches split into more branches etc. When a branch cannot be further split, that's the leaf of a tree, which also carries information about the classification made. The Random Forest model consists of multiple decision trees. Figure 18 shows the confusion matrix for both Decision Tree and Random Forest performed on the same data set, which was split into test and train data in the exact same way for both methods. Moreover, the time it took for Random forest to categorize around 12000 results was 0.1 s which also makes it suitable to be used within this fatigue detection system considering the real-time requirement.

With that result, the Random Forest is used to double-check the output of SVM if SVM detects yawning. Moreover, another check was introduced. Before checking what the output of SVM and Random Forest is, first, a value of the presence of AU number 25 (mouth part) is checked. After implementing the aforementioned checks, the number of false-positive detection was greatly reduced.

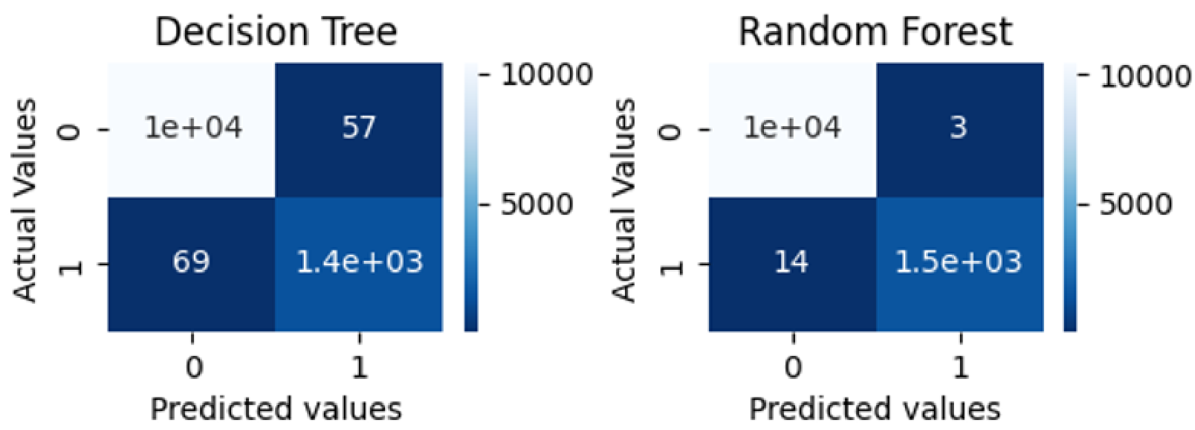

Figure 18 Confusion matrix for Decision Tree and Random Forest

	Decision Tree	Random forest
Accuracy	0.990	0.998
Precision	0.967	0.998
Recall	0.956	0.987
F1 - score	0.962	0.992

Table 7 Decision Tree and Random Forest Evaluation

This scores were obtained by using wrongly divided data into train and test set. Obtained values after correct division are presented in Figure 19 and Table 8.

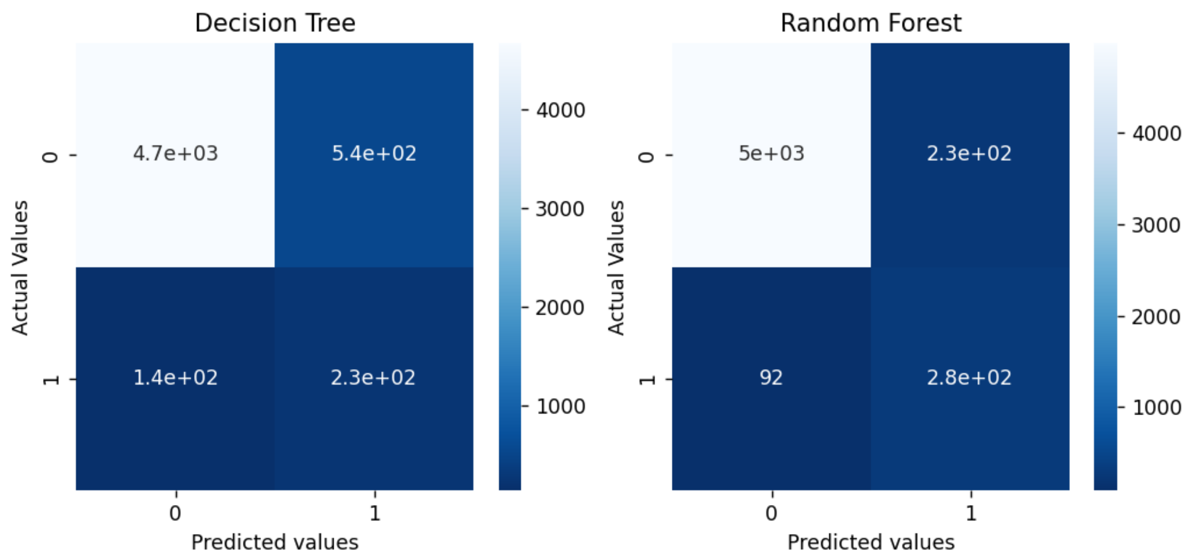

Figure 19 Confusion matrix for Decision Tree and Random Forest

	Decision Tree	Random forest
Accuracy	0.871	0.950
Precision	0.274	0.595
Recall	0.553	0.798
F1 - score	0.366	0.682

Table 8 Decision Tree and Random Forest Evaluation

4.3.2.8 Iteration 4.1: Head angle

Initial user evaluation showcased yet another issue. Even when the confidence score was above the earlier determined threshold, the yawning would not be detected correctly. This would happen as the general shape of the face expected by the OpenFace would be visible on the image. However, some facial landmarks would be positioned inaccurately. Another check for head angle with respect to the camera was added to prevent this. As the user evaluation was done with participants who were using their laptops, a threshold for the head angle was also determined by means of the laptop and webcam position. A “window” at which the participant could look so the system could detect the yawn was essentially the laptop screen with a bit of a margin on the top of the screen. OpenFace 2, besides AUs, can also detect head position with respect to the camera in three dimensions: x-axis – positive to the left, y-axis – positive down and z-axis – positive towards the person. Next to the head position, a head angle along the same three-axis is also detected.

To determine what are the maximal angles in the x and y direction with which the person’s face will still be suitable to perform yawing detection, the angles from the current head position to both x and y direction boundaries are determined. Subsequently, when the y angle is either higher or lower than the just obtained threshold, that frame is no longer suitable for detection. Figure 20 illustrates the process. The black line is the height of the “window”. Blue lines are the maximal angles at which a person can have their head oriented for the program to still classify the yawn correctly. The red line shows the height of the head with respect to the height of the “window”. The green line is the angle with which the person looks at the “window”. Even though the angle stays the same for both cases in the right image, the angle no longer falls behind the blue line as the head moves downward.

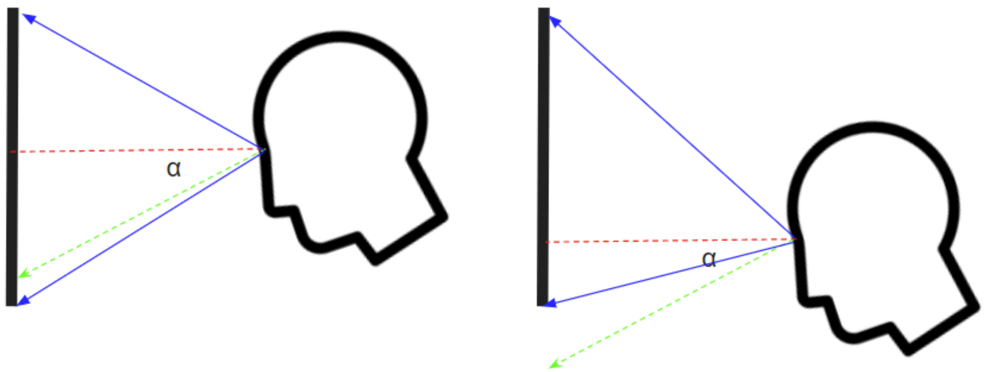

Figure 20 Angle threshold detection

Figure 21 shows the threshold and detected angle for the x-axis and Figure 22 for the y-axis. The data comes from the same measurement as the graphs present in the previous iteration. The angle threshold going to 0 is the moment when the Confidence value is below 0.8.

Figure 21 X axis angle threshold

Figure 22 Y axis angle threshold

The min angle y corrected parameter represents the extended upper boundary for the laptop test. If the angle does not fall withing the given threshold the yawn detection procedure is identical then to the one when confidence drops below 0.8.

4.3.2.9 Iteration 4.2: Head tilt

Analysing values of angle y of the head over time gives information about head tilt. Firstly, if the value of the angle goes above the threshold (in the case when the head is tilting down – positive direction of the y-axis) and stays in that position, there is the possibility that the person has dozed off or is experiencing microsleep. Therefore, the system also stores information about the frequency and duration of head nodes that go above the upper y threshold and last longer than a second. Subsequently, a peak detection function is applied to detect the frequency with which the smaller head nodding occurs. One node is when the head changes the direction of a y angle with which its moving, a peak is created which can be detected.

4.4 System

All 11 iterations describe the process of arriving at a final (for the scope of this graduation project) system. This section provides a brief overview of what the system can detect and a conclusion of section 4. The system can detect yawning and head movement. In the yawning section, duration, frequency and point in time when the yawning occurred are recognized. For the head movement section, the head-nodding frequency is calculated, as well as duration, frequency and point in time for head nodes that go above the y threshold and last longer than one second. It is important to note that the program, with its standard functionality, does not give any visual feedback, only numerical. The graphs present in the iterations were created from the numerical output of a program in a separate program as a visual aid to better understand the program functionality.

4.5 Fatigue detection

A fuzzy logic model is constructed to detect fatigue based on the data from the system described in sections 4.3 and 4.4 [57]. Fuzzy logic is a logic that can output a degree of truth. On the contrary, a Boolean logic can either output 1 or 0 and nothing in between. The Fuzzy Inference System maps the values from an input vector to an output vector. This is done by defining a set of rules based on which the final output value is derived [58]. To create the fuzzy inference system, inputs need to be defined. The inputs in the case of this detector are the number of yawns in the last 30 minutes from the point the measurement for fatigue detection is taken, the total number of yawns from the beginning of the system's session and overall nodding frequency. All the inputs take values ranging from Low, Medium and High. The values assigned to Low, Medium and High vary across the inputs. The final values are based on the information described in section 2 and findings from results obtained from initial participants from the second phase of the methodology. Table 9 showcases the input values present in the final model. Figure 23 shows both the input and output values on graphs Following sections explain the three values in each section and why the particular numbers are assigned to them.

	Low			Medium			High			Range		
Yawn	0	0	2	0	2	4	3	5	5	0	5	5
Nodding	0	0	7	5	10, 15	20	15	35	35	0	35	35
Total Yawn	0	0	15	10	20	30	25	40	40	0	40	40

Table 9 Fuzzy set values

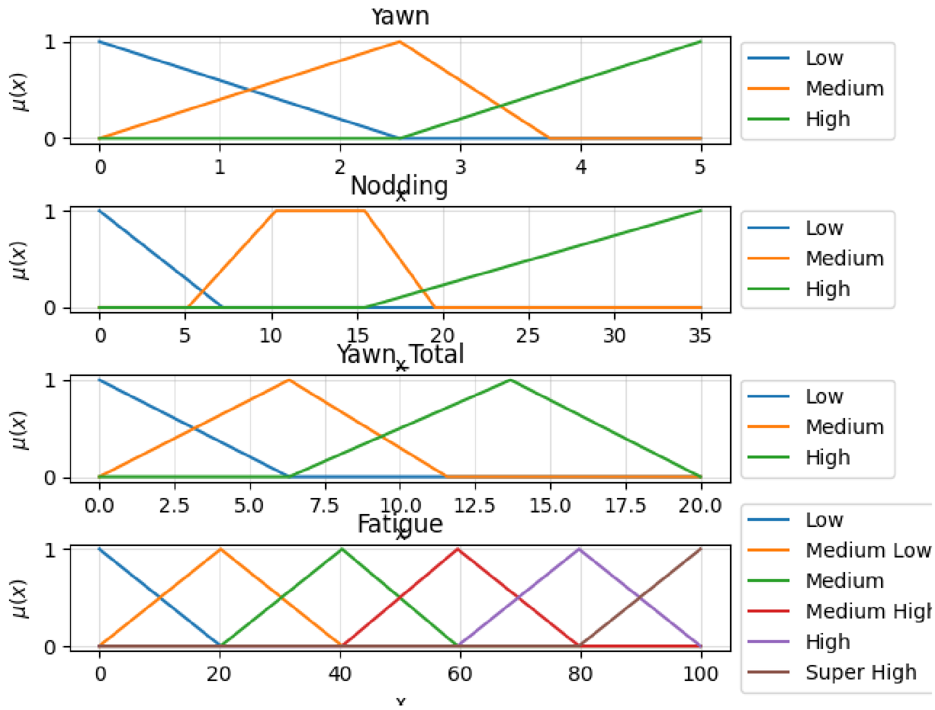

Figure 23 Graphs of input and output values

4.5.1 Parameter explanation

Each of the inputs is a fuzzy set. A fuzzy set determines the possible range of values. The first two columns in the 5th column define the beginning and end of the range (both included in the range), and the third number defines resolution in a number of steps between the beginning and the end of the range. As the Fuzzy Inference System implemented in this project is of Mamdani type, the output is also a fuzzy set.

4.5.1.1 Yawn Total

For the Yawn Total values the number were strongly based on the findings of study conducted by Zilli [20] where he determined that within two groups of people (people waking up early and people waking up late) two average number of yawns per day are present. For people who were waking up early the average was 11.08 yawns between 10:00 and 21:00 and for the second group the average was 22.643 yawns between 10:00 and 21:00. First group 1 yawns once per hour per average while the second group yawns twice as much. Therefore the Low ending boundary and Medium beginning boundary were adjusted to the second group. Two yawns per hour give a frequency of 0,0005Hz; every yawn total frequency value is multiplied by 10 000 for the fuzzy set. The value of 0,0005Hz results in 5. However after examining input from the participant who yawned during the study the yawn frequency on a level of 12 (after multiplication) still felt into category of not tired. Therefore the Low ending for total yawn is set to 15 and the Medium beginning to 10. As the value of a total yawn has a high influence on the total fatigue score the Medium ending was set to 25 and High beginning to 25 to enlarge the distance between the Low and High values for Yawn Total.

4.5.1.2 Yawn

The numbers for Low, Medium and High values for Yawn were obtained as a combination of observed values in participants and values deduced from the previous section. Instead of the total yawn frequency (measured from the start of the fatigue detection) the numbers represented amount of yawns in the last 30 minutes from when the measurement was taken. The end of Low section has a value of 2 (2 yawns in a span of last 30 minutes) which corresponds to average frequency of 0,001Hz. However, as any sign of yawning within 30 minutes corresponded with enlarged fatigue value for some participants the Medium begging was set to 0 and Medium end to 4. As having multiple instances of yawning within the span of 30 minutes is a sign of progressing fatigue a difference in numbers between Medium end and High begging is minimal. The High beginning was set to 3 and High end to 5.

4.5.1.3 Nodding frequency

Numbers for this parameter were assigning on the basis of findings of section 2.2.3.1. The frequency of head nods indicating alertness of a person falls between 0,05Hz and 0,2Hz. Frequencies below that range indicate that a person might be fatigued. Therefore in a set of rules explained in the following section the Medium is treated as a normal, non-fatigue value while Low and High as treated as the values indicating state of fatigue. The frequency value of nodding obtained by the system is multiplied by 100 to be implemented in the fuzzy set.

4.5.1.4 Fatigue

The fatigue fuzzy set has 6 values instead of 3 to facilitate creating IF – THEN ruleset. Input from three parameters with combination of values Low, Medium and High was more convenient to be denoted into 6 fatigue values instead of three. This due to discrepancies between input values on how do they contribute to the overall fatigue score. The range of fatigue is from 0 to 100 with resolution of 100. Values and numbers assigned to them are presented in the Table 10.

Name	Beginning	Peak	End
Low	0	0	20
Medium Low	0	20	40
Medium	20	40	60
Medium High	40	60	80
High	60	80	100
Super High	80	100	100

Table 10 Output values

4.5.2 Rules

In order for the fuzzy logic model to determine fatigue a set of IF – THEN rules is implemented. As there are three inputs with three values each there is a total of 27 rules to describe each possibility of inputs' values combination. All 27 rules are in appendix A. The rules are based on the findings of section 2. The main notion in rules creation is that both Low and High nodding frequency rules treated as fatigue increments. Subsequently, input of Yawn total has bigger influence on the overall fatigue than of Yawn in case of Yawn decreasing. If the Yawn Total values is high while the Yawn value for the last 30 minutes is low the system fatigue indication is oriented more towards Yawn Total input thus is higher. This done because a periodical break in yawning does not mean that the person is no longer fatigued. Only when the overall yawn frequency drops the detected fatigue will also be reduced. However when the Yawn Total frequency has a Low value but Yawn has a High value the system's fatigued indication is more oriented towards Yawn value. Low Yawn Total and High Yawn eventually results is also higher Yawn Total, therefore in the mentioned case system already anticipates increasing fatigue.

4.5.3 Final results

In order to create a final model for fatigue detection, results from 4 participants were taken into account. The participants were students from the University of Twente, aged 20 to 28, representing both genders. The results included the data from OpenFace 2 over the period of the research for a particular participant, system output (yawning frequency, number of yawns, head-nodding frequency) and participant's self-reported fatigue over time. The participants were asked to rank their fatigue on a scale of 1 to 10 every 20 to 30 minutes. That ranked fatigue is marked with the colour red on the graph in Figure 24. The yellow line is the detected fatigue. During the first 3600 seconds, both reported fatigue of the participant and the fatigue detected by the system are rising. Then after the 3600 seconds mark, the reported fatigued of the participant drops but the detected fatigue continuous growing. This is due to the fact explained in the rules section that the total yawn frequency has a bigger influence on the total fatigue than just the number of Yawns in the last 30 minutes when the total yawn frequency is high. Afterwards, at the 4800-second mark, the reported fatigue remains at the same level remains the same and the detected fatigue drops as the total yawn frequency is also lower at this point. Subsequently, the reported fatigue at 6000s increases which is followed by the same response from the system.

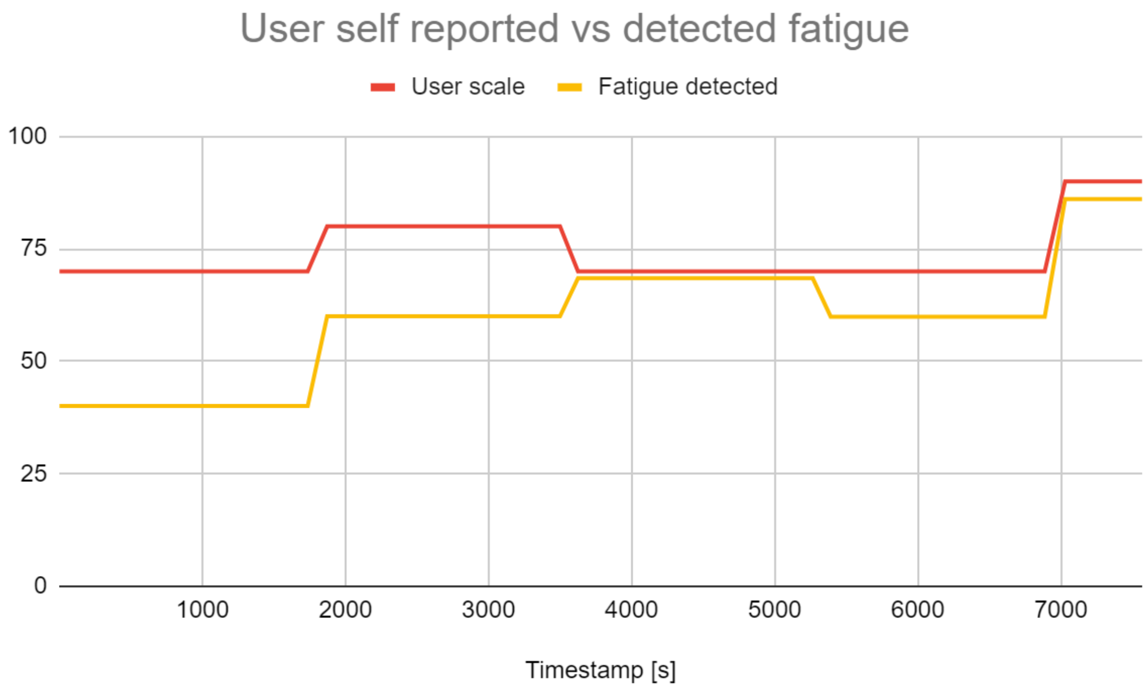

Figure 24 User self-reported vs detected fatigue

5 Evaluation

The system described in the section 4 was evaluated by comparing the final performance of the system to the listed requirements as well as the evaluation described in the methodology section. Furthermore, additional evaluation such as system performance with low light conditions and system performance with partially covered face are performed.

5.1 Requirements Evaluation

Evaluation for some requirement is straightforward while some other require more detailed description. A table is presented with the requirements that can be evaluated in a simple manner (Table 11). More complexed requirements evaluation is described in the subsection following the table below.

Requirement	**Evaluation**
The system *must* detect fatigue using facial feature	The system is taking into account yawning and head movements to detect fatigue.
The system *must* not store video data	The feed captured by the camera is not saved
The system *must* only recognise fatigue	System only detects yawning and head movement. No other behaviours are recognized
The truck driver *must* not be identifiable based one data collected	The data collected consist of only values assigned by OpenFace 2 (intensity and presence of AUs, head pose and angle, timestamp of detected frame and confidence) from which person identification is impossible
The system *must* perform in real-time, close to it	Both image analysis and fatigue detection are done in real time
The system must detect yawning	The system can detect yawning
The system *must* determine fatigue based on the yawning pattern information	The system is taking into account yawning and head movements to detect fatigue.
The system *should* distinguish between different truck drivers	Restarting the system resets all the values used to detect fatigue thus the other driver can have a fatigue prediction tailored to the after switching places with other driver.
The system *should* not be hindering the driver's performance	The system was not tested in real life scenario including truck drivers – impossible to determine
The camera *should* not disturb the driver	The system was not tested in real life scenario including truck drivers – impossible to determine
The system *should* work with different lighting conditions	There is a small section in this evaluation chapter regarding this requirement, however not enough testing was done to confirm that the requirement has been met.
The system *could* use input from another sources (such as sensors) to determine fatigue	Not implemented
The system *could* use multiple facial features to determine fatigue	System uses both yawning and head movement to determine fatigue

The camera *could* be connected to a piece of hardware that can run a software used for fatigue detection	Camera can be connected to the laptop that can run the software, however no other combination was tested
The system *won't* have feedback to the driver	Not implemented

Table 11 Requirements evaluation

5.1.1 The system must not have any gender or racial biased

Chen and Joo [59] in their study analysed the performance of OpenFace, the predecessor of OpenFace 2 used in this project. The performance was analysed in terms of bias. For the AUs and facial expressions that can be derived from them, the study concluded that the differences between the accuracy of detections for men and women are insignificant.

Racial bias occurs when the system is favourable towards one particular social group. Favourable in this case means that the system assumes a baseline related to that social group as a general baseline that is later applied to the other social groups that may use the same system, thus creating discrimination. However, the variety of databases and social groups portrayed in them, which were used to train OpenFace 2, allows us to assume that the program is not racially biased..

5.2 Performance evaluation

Performance evaluation of the system is split into several sections. Firstly, yawn detection itself is assessed and with secondly, the fatigue detection. Lastly system behaviour under bad lighting conditions and with occluded face is evaluated.

5.2.1 Yawning and fatigue detection

The metrics for the correctly divided data set for SVM, Decision Tree and Random Forest Classifier are shown in Figure 25.

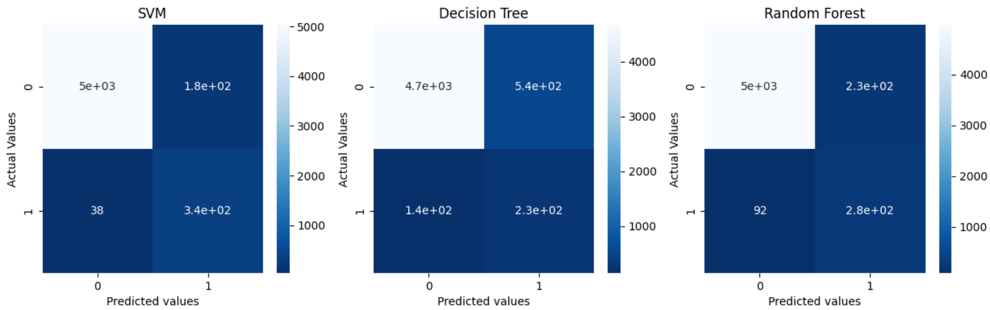

Figure 25 Confusion Matrices for SVM, Decision Tree and Random Forrest Classifier

As observed in the initial state of iteration, the system would detect a yawn when in reality observed person did not yawn. This can be explained by the high amount of False Positives in the models. The values for False Negatives are low in both SVM and Random Forrest. Therefore when an observed person yawns or open their mouth widely (in a manner of yawning), there is a high chance the system will classify this behaviour as a yawn.

Even though the Random Forest model returned a False Positive that SVM, the inputs for which it was returned partially differ. This means that having Random Forest as the second model to ensure that the categorised yawn is an actual yawn is not pointless, although the initial assumptions behind that idea were incorrect. Subsequently, an additional check for AU 25 indicating mouth part also further improves the yawn detection. However, as shown on the graph in Figure 14, the yawn is also detected when the mouth is widely open (the person is speaking). The additional check for yawn duration (yawn lasting longer than 2 hours) was implemented to prevent miscategorisation of mouth wide open as yawning. Nevertheless, when a person is signing a song which involves long vowel pronunciation (mouth open), that action is very likely to also be categorised as yawning. As yawning has a big input in fatigue detection fuzzy logic model that miscategorisation results in overstating the fatigue value. On the other hand, when a person has a very short yawn (the total duration might be longer than 2 seconds, but the acme phase is shortened than that), the yawn will not be recognised.

With the final evaluation, a participant was asked to report their fatigue every 20 minutes and note the number of yawns that happened during the duration of the research. Subsequently, the data were analysed by the system to output the number of yawns detected and detected fatigue state of the participants. The number of yawns detected by the system was equal to 18, while the participant reported 16. This yields that at least two results were incorrectly assigned as yawning. Figure 26 shows the graph of reported fatigue state by the participant and fatigue detected by the system. The fatigue detection was done every 20 minutes to match the frequency of participant reporting.

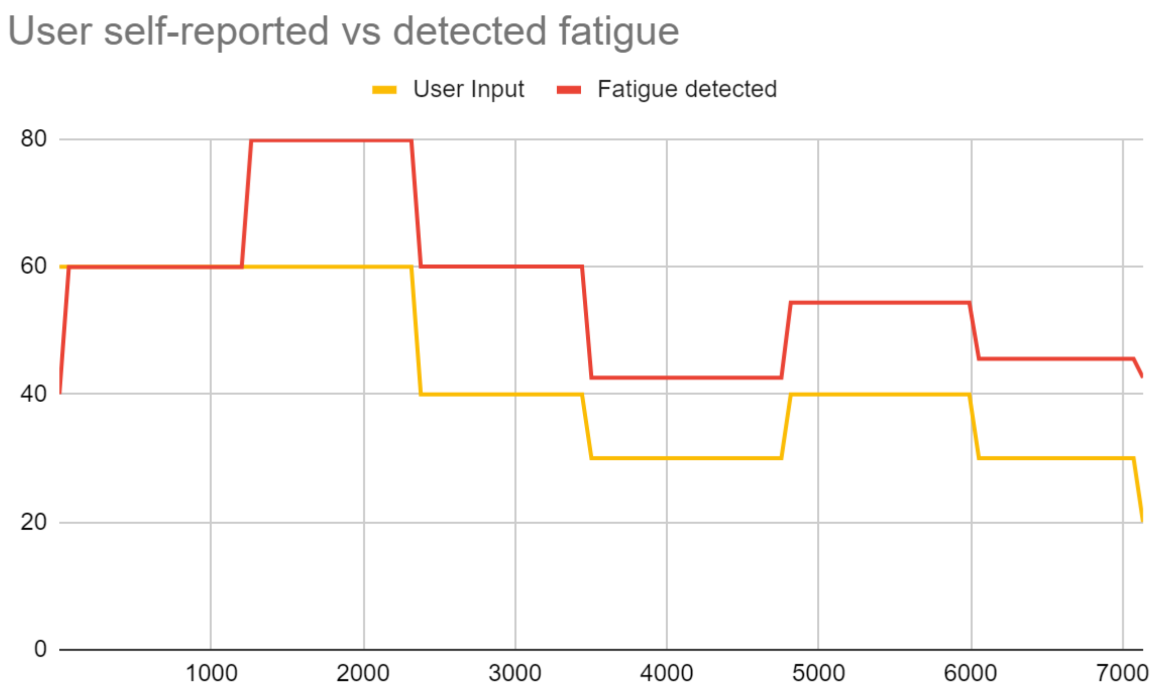

Figure 26 User self-reported vs detected fatigue, 20 minutes periods

At the 1200-second mark, the fatigue reported by the participant remains the same while the detected fatigue increases. This is caused by the advantage of total yawn frequency over periodical yawn frequency when the total yawn frequency is higher in value. Subsequently, the detected fatigue increases and decreases with the same pattern as the reported fatigue. Figure 27 show the same parameters with the addition of fatigue detection model output returned with the frequency of one output every 5 minutes instead of 20.

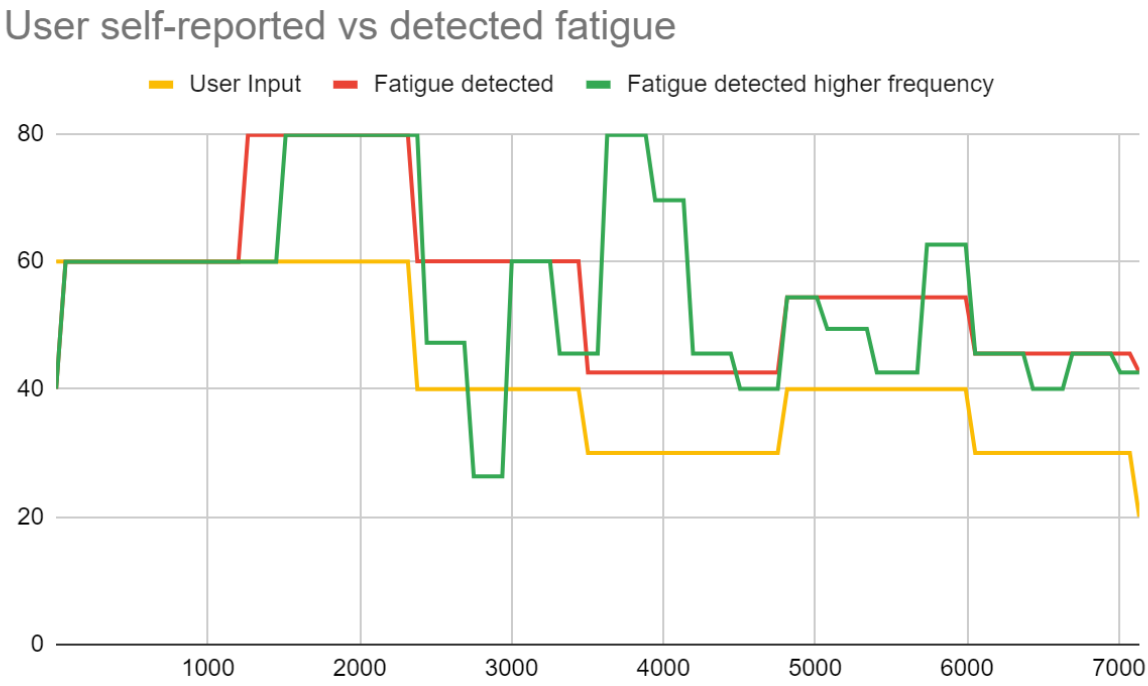

Figure 27 User self-reported vs detected fatigue, 5 minutes periods

With increased frequency, discrepancies between the reported and detected fatigue arise. At the point of 1200 seconds, the detected fatigue increases, although the reported one remains the same. This is again a result of the relationship between Yawn Total and Yawn, and it is to be expected. At the point of 2400 seconds, the reported fatigue drops so does the detected one. The drop in value of detected fatigue continues to occur up to the point of 3000 s when the fatigue increases again. Then when the value of reported fatigue drops, the value of detected fatigue drops as well a few seconds beforehand to increase again at 3600 seconds. This is the most unexpected increase in value as the reached detected fatigue value is very high compared to the user input. Having in mind at least two incorrectly detected yawns, this increase could be attributed to them. However, with every measurement, the model checks the number of yawns within the last 30 minutes. As the measurements are taken every 5 minutes, it can happen that the number of yawns differs by one or two or even more yawns than the previous input. These differences result in abrupt peaks in either direction in the detected fatigue graph. After 3800 seconds, the value of fatigue detected drops again to increase at the same time as the value of reported fatigue at the 4800 seconds mark. Furthermore, the detected fatigue values fluctuate slightly. However, as the measurements are taken more often in this case, that is to be expected. Finally, Figure 28 shows a graph with a measurement of fatigue detected taken every 1 minute.

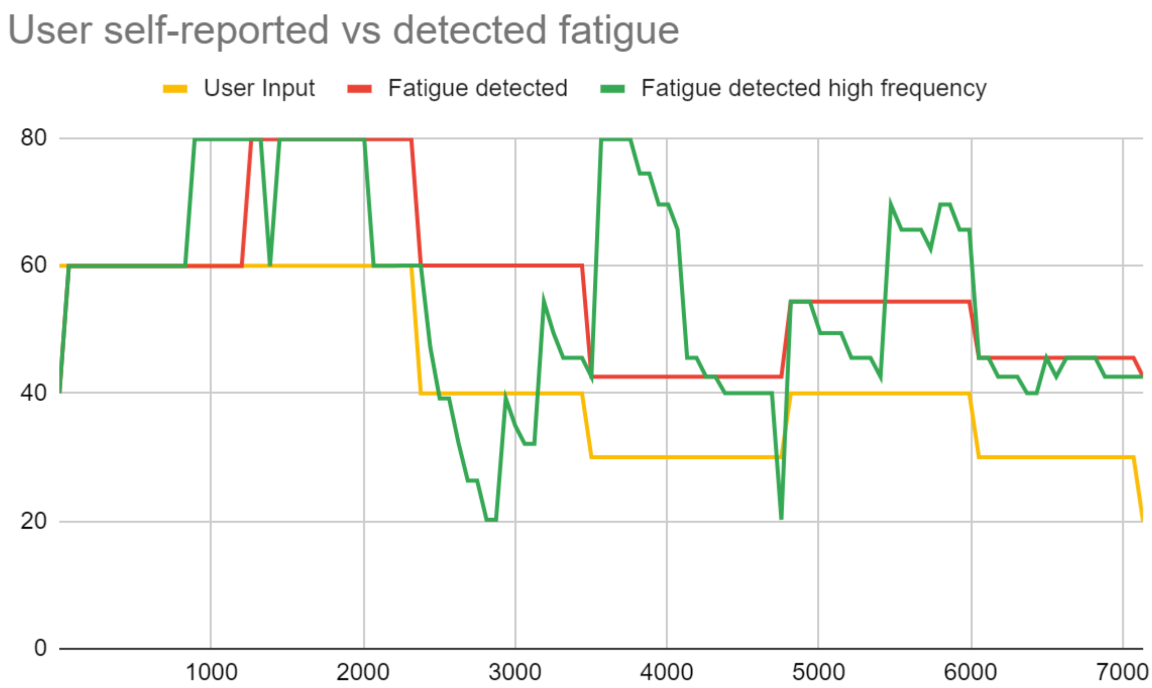

Figure 28 User self-reported vs detected fatigue, 1minute periods

In the case shown in Figure 28, the values returned by the fatigue detecting system correlate strongly with the ones outputted with measurements every 5 minutes. The discrepancies between measurements taken at different frequencies than user-reported fatigue show that the model is accurate with regards to trends in user fatigue when the measurements are taken with a long time between them. When the measurement is taken more often, there is more space for various parameters to change significantly in relation to each other. This results in big changes in values between consecutive measurements. Perhaps a better categorization and choice of parameters would help to create smoother output in fatigue values by the system. Moreover, with the last check, it was determined that the yawning detection has at least categorized two not-yawns as yawns which indicates that further improvements of the system in terms of reducing false positives are necessary for more reliable performance.

5.2.3 Unfavourable detection conditions

A brief evaluation of the system performance, mostly OpenFace 2 two with remarks from the yawning detection side, in not ideal conditions was done. Firstly the system was tested with limited lighting conditions, and secondly, with part of the face occluded.

5.2.3.1 Poor lighting condition

As part of the drives done by truck drivers takes place during the night, the system's capabilities of working with limited light without the use of any special camera were tested. The face detection, thus, head rotation used for detecting nodding off and nodding frequency was still possible. The yawning detection, however, was sometimes failing as the facial landmarks used to detect AUs were not updating based on the movement of the face. The changes in facial expression, regardless of high confidence value, were sometimes not possible to detect in partial darkness.

Therefore the location of facial landmarks could not be updated. Figure 29 shows a frame from the OpenFace 2 obtained with poor lighting conditions.

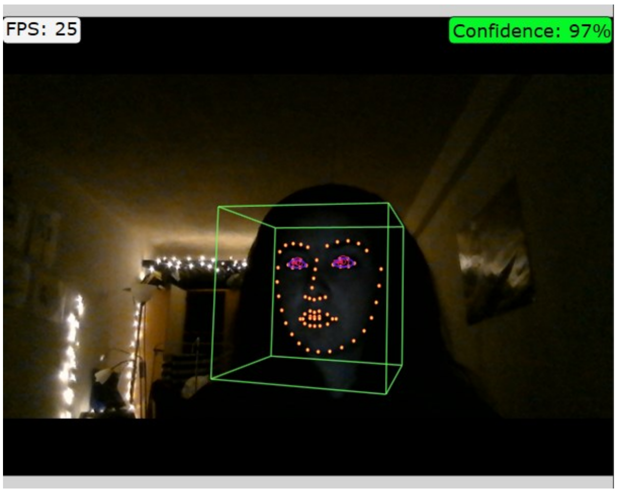

Figure 29 OpenFace 2 detection with poor lighting conditions

5.2.3.2 Face occlusion

Firstly a face with glasses on as an occluding element was evaluated. Two types of glasses were tested—one covering the eyebrows and the second where the eyebrows were partially visible. Similarly to the previous instance, the face detection and head angle measurements worked correctly. The yawing detection had about a 50% success rate – yawns were detected with the first type of glasses. The second type of glasses allowed for a bit higher success rate. This is because the model used to determine yawing also considers the values of AUs associated with eyebrows. Occluding that region of a face prevents the model from receiving all the necessary information.

Moreover, OpenFace 2 performance was tested with the lower part of the face occluded. Naturally, the yawn was impossible to detect at a point when the whole mouth was covered. However, the head pose and angle remained returning valid output as long as the confidence value remained high. When the lower part of the nose, altogether with the mouth region, was covered, the program was not able to detect the face anymore. Furthermore, when the face is no longer detectable, the frame rate of the whole program also drops. This is due to the fact that OpenFace 2 starts looking for a probable position of the facial landmarks based on the one detected in the previous frame [48]. Therefore if no landmarks were detected in the previous frame, the program needs to look at the whole image instead of just its selected Regions Of Interest.

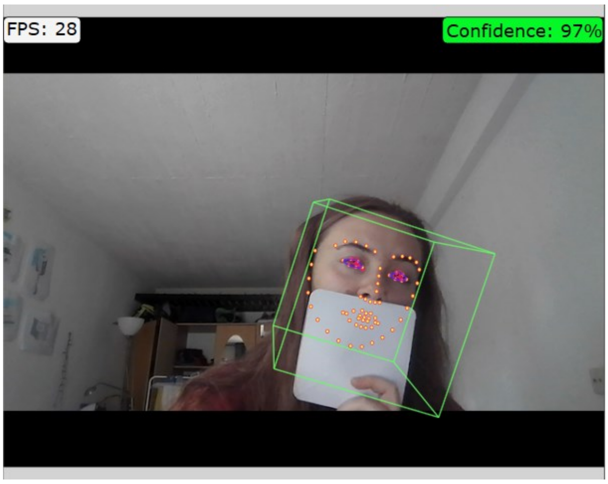

Figure 30 Face occlusion allowing for face detection

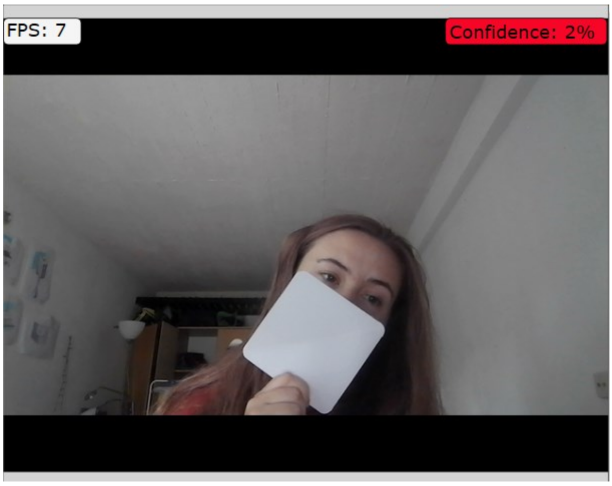

Figure 31 Face occlusion not allowing for face detection

6 Discussion and Recommendations

Following developing and evaluating the system for detecting fatigue, recommendations and fields worth discussing arise. Moreover, there were certain limitations to the system, which are also mentioned in this section.

6.1 Discussion

One of the user measurements showed that a person could experience and report progressing fatigue without yawning. Therefore, having a fatigue detection system with yawning as the primary criterium could yield little to no results in the above-described cases. Other parameters that can be derived from the face: namely blinking and PERCLOS, are more reliable. The reason is that a person has to blink just like they have to breathe. With yawning, that is not the case.

Nevertheless, the detection of face orientation to the camera was working in both cases, as long as enough face was not covered – the system could still detect the face in the image. Therefore, even when yawning detection is not possible, the detection of head nodding and nodding off should be feasible.

Referring again to the fact that the proposed system would work better when more parameters are added, adding more factors to fatigue detection could further improve the result this system has in terms of fatigue detection. Currently, there are only there parameters taken into account: yawning amount over the span of the 30 minutes preceding the measurement, frequency of head-nodding measured from the start of the whole measurement and frequency of yawning measured from the start of the whole measurement. However, the system is able also to output yawning duration, timestep of head nod forward lasting longer than 1 second, which can signalize nodding off of the driver and duration of those nod offs. All the listed parameters could also be included in the fuzzy logic model with IF-THEN rules in order to improve the final predictions of the model.

Next to the parameters available to implement from the system presented in this graduation project, parameters not detectable by this system could be introduced. Those parameters could include already mentioned information from the eyes but also information from other sensors, such as truck position between the line or drivers behaviour described in section two. Additionally, information about the time of the day, hours of drive, number of drives in the past days etc., could also be considered. From all the data, a more accurate and precise fuzzy logic model for determining fatigue could be created.

6.2 Limitations

The major limitation that affected this graduation project is connected to the form of recognizing participants' fatigue. The participants in the second evaluation phase were asked to rate their fatigue on a scale from 1 to 10 every 30 to 20 minutes. There are a few drawbacks to this method. First of all, fatigue perceived as 5 by one participant could be perceived as 7 by another, which in consequence, prevents creating a uniform scale for all the participants. Another major drawback in assessing the fatigue detection algorithm was the lack of deterministic data to compare the output to. The reporting fatigue by participants method only indicates whether fatigue of a person is increasing or decreasing and, if so, at what pace it is happening. Subsequently, the rating was only done in a selected period of time, and there was no information on how the participant felt in between the measurements. It is possible that within the period of 30 minutes, the fatigue of the participant increased and decreased afterwards. Therefore, in order to obtain more accurate data about the fatigue state of a person, more readings, time-wise as well as sensor, wise is necessary. The reading could be obtained by the use of driving behaviour sensors as well as sensor measuring physiological parameters.

Moreover, the system was only tested on university students while they were sitting in front of their laptops. A test with truck drivers is crucial to be fully confident on whether such a system can be reliably implemented in a truck.

6.3 Recommendations

Another worth considering aspect of yawning is that sometimes people cover their mouth with their hand while doing so. As described in the evaluation, this system does not handle that situation well. For future work, it could be profitable to combine hand detection with yawn detection. So that when a hand is detected in the frame moving towards the mouth, a yawn can be assumed. Furthermore, the performance of this system while people were wearing glasses, even though not thoroughly tested, was lower than in the case of people without glasses. Therefore, creating a new model based on people in glasses yawning could enhance the total accuracy of the system, regardless of whether the person is wearing glasses.

Currently, the system is only capable of predicting fatigue at a given moment considering three parameters. That information is relevant to the driver considering their safety. However, it leaves little to no time for the reaction. When the system reports a fatigue state of the driver, the driver is already in that state, and there is nothing they can do to prevent it. However, if the system would be able to predict how the fatigue state of the driver would progress given current data and inform the driver early on, it would leave the driver with more time to take action, such as finding a gas station to take a rest at.

7 Conclusion

This graduation project aimed to assess computer vision-related techniques to develop a system that could detect the fatigue of truck drivers. As fatigue is an everlasting element in the reality of truck drivers, such a system could tremendously improve their safety and wellbeing. This thesis focuses on yawning and head nodding movements as parameters to detect fatigue. OpenFace 2 was chosen as a suitable software to derive desired parameters from a face. The machine learning model used to detect yawning faces from values of Facial Action Units has an accuracy of 0.961, recall of 0.899 and precision of 0.651. Additional checks were implemented in the system to mitigate the consequences of low precision value. However, the mitigations do not entirely eliminate the problem. Nevertheless, the head-nodding detection is reliable even under poor lighting or partial face occlusion conditions.

Furthermore, a fatigue detection model was proposed based on fuzzy logic. The model output corresponded well to the reported fatigue by participants over periods from 20 minutes onward. During this graduation project, another important conclusion emerged: yawning is not the most reliable fatigue detection parameter. Although the proposed system allows for yawning detection, estimation of a person's state in terms of fatigue is more reliable when additional parameters are considered.

Bibliography

[1] K. Vitols, E. Voss, and L. Mackay, 'DRIVER FATIGUE IN EUROPEAN ROAD TRANSPORT', p. 60.

[2] G. Sikander and S. Anwar, 'Driver Fatigue Detection Systems: A Review', *IEEE Trans. Intell. Transport. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019, doi: 10.1109/TITS.2018.2868499.

[3] 'Volvo Trucks Safety Report 2017', p. 34.

[4] R. O. Phillips, 'Countermeasures for use in fatigue risk management', p. 76.

[5] E. Lupova, 'ATA-Fatigue-Hackathon-Research-Paper.pdf'.

[6] J. F. May and C. L. Baldwin, 'Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies', *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 3, pp. 218–224, May 2009, doi: 10.1016/j.trf.2008.11.005.

[7] J. A. Caldwell, M. M. Mallis, J. L. Caldwell, M. A. Paul, J. C. Miller, and D. F. Neri, 'Fatigue Countermeasures in Aviation', *aviat space environ med*, vol. 80, no. 1, pp. 29–59, Jan. 2009, doi: 10.3357/ASEM.2435.2009.

[8] S.-Y. Shi, W.-Z. Tang, and Y.-Y. Wang, 'A Review on Fatigue Driving Detection', *ITM Web Conf.*, vol. 12, p. 01019, 2017, doi: 10.1051/itmconf/20171201019.

[9] G. Borghini *et al.*, 'Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices', in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, Aug. 2012, pp. 6442–6445. doi: 10.1109/EMBC.2012.6347469.

[10] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, 'Real-Time System for Monitoring Driver Vigilance', *IEEE Trans. Intell. Transport. Syst.*, vol. 7, no. 1, pp. 63–77, Mar. 2006, doi: 10.1109/TITS.2006.869598.

[11] R. Ahmed, Kazi Emrul Kayes Emon, and Md. F. Hossain, 'Robust driver fatigue recognition using image processing', in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, Dhaka, Bangladesh, May 2014, pp. 1–6. doi: 10.1109/ICIEV.2014.6850713.

[12] J. Wongphanngam and S. Pumrin, 'Fatigue warning system for driver nodding off using depth image from Kinect', in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Mai, Thailand, Jun. 2016, pp. 1–6. doi: 10.1109/ECTICon.2016.7561274.

[13] H. H. Lic and H. Summala, 'Sleepiness at Work Among Commercial Truck Drivers', *Sleep*, vol. 23, no. 1, pp. 1–9, Jan. 2000, doi: 10.1093/sleep/23.1.1b.

[14] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, 'A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection', *IJAST*, vol. 64, pp. 73–100, Mar. 2014, doi: 10.14257/ijast.2014.64.07.

[15] 'IET Intelligent Trans Sys - 2016 - He - Evaluation of driver fatigue with multi-indicators based on artificial neural.pdf'.

[16] T. Sundelin, M. Lekander, G. Kecklund, E. J. W. Van Someren, A. Olsson, and J. Axelsson, 'Cues of Fatigue: Effects of Sleep Deprivation on Facial Appearance', *Sleep*, vol. 36, no. 9, pp. 1355–1360, Sep. 2013, doi: 10.5665/sleep.2964.

[17] S. Jin, S.-Y. Park, and J.-J. Lee, 'Driver fatigue detection using a genetic algorithm', *Artif Life Robotics*, vol. 11, no. 1, pp. 87–90, Jan. 2007, doi: 10.1007/s10015-006-0406-8.

[18] G. Daquin, J. Micallef, and O. Blin, 'Yawning', *Sleep Medicine Reviews*, vol. 5, no. 4, pp. 299–312, Aug. 2001, doi: 10.1053/smrv.2001.0175.

[19] G. Schino and F. Aureli, 'Do men yawn more than women?', *Ethology and Sociobiology*, vol. 10, no. 5, pp. 375–378, Jul. 1989, doi: 10.1016/0162-3095(89)90026-5.

[20] I. Zilli, F. Giganti, and P. Salzarulo, ‘Yawning in morning and evening types’, *Physiology & Behavior*, vol. 91, no. 2–3, pp. 218–222, Jun. 2007, doi: 10.1016/j.physbeh.2007.02.015.

[21] R. Schleicher, N. Galley, S. Briest, and L. Galley, ‘Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?’, *Ergonomics*, vol. 51, no. 7, pp. 982–1010, Jul. 2008, doi: 10.1080/00140130701817062.

[22] G. Zhang, K. K. W. Yau, X. Zhang, and Y. Li, ‘Traffic accidents involving fatigue driving and their extent of casualties’, *Accident Analysis & Prevention*, vol. 87, pp. 34–42, Feb. 2016, doi: 10.1016/j.aap.2015.10.033.

[23] M. C. Catalbas, T. Cegovnik, J. Sodnik, and A. Gulten, ‘Driver Fatigue Detection Based on Saccadic Eye Movements’, p. 5.

[24] Y. Wang, R. Huang, and L. Guo, ‘Eye gaze pattern analysis for fatigue detection based on GP-BCNN with ESM’, *Pattern Recognition Letters*, vol. 123, pp. 61–74, May 2019, doi: 10.1016/j.patrec.2019.03.013.

[25] S. Johansson, C. Ytterberg, B. Back, L. Holmqvist, and L. Koch, ‘The Swedish occupational fatigue inventory in people with multiple sclerosis’, *J Rehabil Med*, vol. 40, no. 9, pp. 737–743, 2008, doi: 10.2340/16501977-0236.

[26] E. Åhsberg, F. Garnberale, and A. Kjellberg, ‘Perceived quality of fatigue during different occupational tasks Development of a questionnaire’, *International Journal of Industrial Ergonomics*, vol. 20, no. 2, pp. 121–135, Aug. 1997, doi: 10.1016/S0169-8141(96)00044-3.

[27] E. Åhsberg, ‘Dimensions of fatigue in different working populations’, *Scandinavian Journal of Psychology*, vol. 41, no. 3, pp. 231–241, Sep. 2000, doi: 10.1111/1467-9450.00192.

[28] L. B. Krupp, ‘The Fatigue Severity Scale: Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus’, *Arch Neurol*, vol. 46, no. 10, p. 1121, Oct. 1989, doi: 10.1001/archneur.1989.00520460115022.

[29] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, ‘Karolinska Sleepiness Scale (KSS)’, in *STOP, THAT and One Hundred Other Sleep Scales*, A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, Eds. New York, NY: Springer New York, 2011, pp. 209–210. doi: 10.1007/978-1-4419-9893-4_47.

[30] K. Kaida *et al.*, ‘Validation of the Karolinska sleepiness scale against performance and EEG variables’, *Clinical Neurophysiology*, vol. 117, no. 7, pp. 1574–1581, Jul. 2006, doi: 10.1016/j.clinph.2006.03.011.

[31] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, ‘Stanford Sleepiness Scale (SSS)’, in *STOP, THAT and One Hundred Other Sleep Scales*, A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, Eds. New York, NY: Springer New York, 2011, pp. 369–370. doi: 10.1007/978-1-4419-9893-4_91.

[32] N. Williams, ‘The Borg Rating of Perceived Exertion (RPE) scale’, *Occupational Medicine*, vol. 67, no. 5, pp. 404–405, Jul. 2017, doi: 10.1093/occmed/kqx063.

[33] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, ‘Face Recognition Systems: A Survey’, *Sensors*, vol. 20, no. 2, p. 342, Jan. 2020, doi: 10.3390/s20020342.

[34] A. Kumar, A. Kaur, and M. Kumar, ‘Face detection techniques: a review’, *Artif Intell Rev*, vol. 52, no. 2, pp. 927–948, Aug. 2019, doi: 10.1007/s10462-018-9650-2.

[35] K. Dang and S. Sharma, ‘Review and comparison of face detection algorithms’, in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, India, Jan. 2017, pp. 629–633. doi: 10.1109/CONFLUENCE.2017.7943228.

[36] M. K. Dabhi and B. K. Pancholi, ‘Face Detection System Based on Viola - Jones Algorithm’, vol. 5, no. 4, p. 4, 2013.

[37] P. Viola and M. Jones, ‘Rapid object detection using a boosted cascade of simple features’, in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, vol. 1, p. I-511-I–518. doi: 10.1109/CVPR.2001.990517.

[38] E. Bagherian and R. W. O. K. Rahmat, 'Facial feature extraction for face recognition: a review', in *2008 International Symposium on Information Technology*, Kuala Lumpur, Malaysia, 2008, pp. 1–9. doi: 10.1109/ITSIM.2008.4631649.

[39] S. Abtahi, B. Hariri, and S. Shirmohammadi, 'Driver drowsiness monitoring based on yawning detection', in *2011 IEEE International Instrumentation and Measurement Technology Conference*, Hangzhou, China, May 2011, pp. 1–4. doi: 10.1109/IMTC.2011.5944101.

[40] C. Anitha, M. K. Venkatesha, and B. S. Adiga, 'A Two Fold Expert System for Yawning Detection', *Procedia Computer Science*, vol. 92, pp. 63–71, 2016, doi: 10.1016/j.procs.2016.07.324.

[41] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, and L. Skrypchuk, 'Analysis of Yawning Behaviour in Spontaneous Expressions of Drowsy Drivers', in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, May 2018, pp. 571–576. doi: 10.1109/FG.2018.00091.

[42] X. Deng *et al.*, 'Joint Hand Detection and Rotation Estimation Using CNN', *IEEE Trans. on Image Process.*, vol. 27, no. 4, pp. 1888–1900, Apr. 2018, doi: 10.1109/TIP.2017.2779600.

[43] T. H. N. Le, K. G. Quach, C. Zhu, C. N. Duong, K. Luu, and M. Savvides, 'Robust Hand Detection and Classification in Vehicles and in the Wild', in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1203–1210. doi: 10.1109/CVPRW.2017.159.

[44] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, 'Object Detection with Discriminatively Trained Part Based Models', p. 20.

[45] W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou, 'Look at Boundary: A Boundary-Aware Face Alignment Algorithm', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 2129–2138. doi: 10.1109/CVPR.2018.00227.

[46] T. Baltrusaitis, P. Robinson, and L.-P. Morency, 'OpenFace: An open source facial behavior analysis toolkit', in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1–10. doi: 10.1109/WACV.2016.7477553.

[47] 'Point Distribution Model · The Menpo Project'. https://www.menpo.org/menpofit/pdm.html (accessed Apr. 20, 2022).

[48] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, 'OpenFace 2.0: Facial Behavior Analysis Toolkit', in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, May 2018, pp. 59–66. doi: 10.1109/FG.2018.00019.

[49] A. Zadeh, Y. C. Lim, T. Baltrusaitis, and L.-P. Morency, 'Convolutional Experts Constrained Local Model for 3D Facial Landmark Detection', in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Oct. 2017, pp. 2519–2528. doi: 10.1109/ICCVW.2017.296.

[50] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, 'Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks', *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[51] A. Hudaib, R. Masadeh, M. H. Qasem, and A. Alzaqebah, 'Requirements Prioritization Techniques Comparison', *MAS*, vol. 12, no. 2, p. 62, Jan. 2018, doi: 10.5539/mas.v12n2p62.

[52] W. Tipprasert, 'A Method of Driver's Eyes Closure and Yawning Detection for Drowsiness Analysis by Infrared Camera', *Artificial Intelligence*, p. 4, 2019.

[53] A. Kulkarni, D. Chong, and F. A. Batarseh, 'Foundations of data imbalance and solutions for a data democracy', in *Data Democracy*, Elsevier, 2020, pp. 83–106. doi: 10.1016/B978-0-12-818366-3.00005-8.

[54] J. Schell, *The art of game design : a book of lenses*, 3rd edition., 1 online resource (xlii, 610 pages) : illustrations (some color) vols. Boca Raton, FL: CRC Press LLC, 2020. doi: 10.1201/b22101.

[55] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification*, vol. 36. Boston, MA: Springer US, 2016. doi: 10.1007/978-1-4899-7641-3.

[56] D. Wilimitis, ‘The Kernel Trick’, *Medium*, Feb. 21, 2019. https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f (accessed Jul. 07, 2022).
[57] C. Gafa, *ml_from_scratch*. 2022. Accessed: Jul. 08, 2022. [Online]. Available: https://github.com/carmelgafa/ml_from_scratch
[58] S. A. Kalogirou, ‘Designing and Modeling Solar Energy Systems’, p. 112.
[59] Y. Chen and J. Joo, ‘Understanding and Mitigating Annotation Bias in Facial Expression Recognition’. arXiv, Aug. 19, 2021. Accessed: Jul. 07, 2022. [Online]. Available: http://arxiv.org/abs/2108.08504

Appendix A - OpenFace 2.2.0 Citation

OpenFace 2.0: Facial Behavior Analysis Toolkit Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency, IEEE International Conference on Automatic Face and Gesture Recognition, 2018

Convolutional experts constrained local model for facial landmark detection A. Zadeh, T. Baltrušaitis, and Louis-Philippe Morency. Computer Vision and Pattern Recognition Workshops, 2017

Constrained Local Neural Fields for robust facial landmark detection in the wild Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. in IEEE Int. Conference on Computer Vision Workshops, 300 Faces in-the-Wild Challenge, 2013.

Rendering of Eyes for Eye-Shape Registration and Gaze Estimation Erroll Wood, Tadas Baltrušaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling in IEEE International Conference on Computer Vision (ICCV), 2015

Cross-dataset learning and person-specific normalisation for automatic Action Unit detection Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson in Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition, 2015

Appendix B – Fuzzy logic model rule set

	Yawn	Nodding	Yawn Total	Fatigue
1	Low	Low	Low	Medium Low
2	Low	Low	Medium	Medium
3	Low	Low	High	Medium High
4	Low	Medium	Low	Low
5	Low	High	Low	Medium Low
6	Medium	Low	Low	Medium High
7	High	Low	Low	High
8	High	Medium	Low	Medium High
9	High	High	Low	High
10	Medium	High	Low	Medium High
11	High	High	Medium	Super High
12	High	High	High	Super High
13	Low	Medium	Medium	Medium High
14	Medium	Medium	Low	Medium
15	Medium	Medium	Medium	Medium
16	Low	High	High	High
17	Low	Medium	High	Medium High
18	Medium	Low	High	High
19	High	Low	Medium	Super High
20	Medium	Medium	High	Medium High
21	Medium	High	Medium	Medium High

22	High	Medium	Medium	Medium High
23	Medium	High	High	High
24	High	Medium	High	High
25	Medium	Low	Medium	Medium High
26	Low	High	Medium	Medium
27	High	Low	High	Super High

Appendix C – Information brochure

Information Brochure for evaluating fatigue detecting system

Institution: University of Twente (053 489 9111)

Researcher information: Agata Sowa, a.m.sowa@student.utwente.nl (+48608298169)

Supervisor: Job Zwiers, j.zwiers@utwente.nl

Title: Development of a computer vision-based system for recognising fatigue of truck drivers

Purpose: Sixty per cent of truck drivers admitted to being fatigued every ninth or lesser drive. Furthermore, 30% of them stated that they had fallen asleep behind the wheel within 12 months before the study conducted by Vitols [1]. Current safety systems implemented in the trucks to recognise driver fatigue mostly take into account truck position between the lanes and steering patterns [2]. Of all road accidents involving heavy goods vehicles, 90% are caused by human factors. The human factors are split into more detailed categories. One of the most prominent ones is drivers' fatigue – 20% of all accidents caused by human factors result from driver's fatigue [3]. Those percentages indicate that there is still space and need for introducing new technology that could further improve fatigue detection in truck drivers. This research aims to develop a system that could recognize a truck driver's fatigue level in real-time or close to it. The system considers changes in the drivers' facial features and, based on that, evaluates their state.

Procedure: You will be asked to sit in front of your laptop and work, study or read something on your device. You will be free to choose how you want to spend the time. In front of you, a webcam will be placed. The webcam will only collect data about changes in your facial features – no video or audio will be recorded. Based on the data collected, the system will determine your fatigue level. Every 20 – 30 minutes, you will be asked to decide on a scale of 1 to 10 how fatigued you are (where one is not fatigued at all and ten is extremely fatigued). After you are finished, the researcher will use the obtained data to evaluate the system's performance. There will be no possibility to identify you via the collected data. While participating in this research, you can ask the researcher questions at any time. You are also free to choose whether you want the researcher to supervise/be present while participating in this study.

Duration: As long as you feel comfortable doing the research, you will be able to quit at any time or take a break.

Risks: Covid-19 regulations will be implemented, as well as additional precautions to ensure the health and safety of the participants.

Confidentiality: You will not be required to state any private information. The data and the findings will be anonymized. After participating in the research, you can ask the researcher to delete all your data. For this, you will have to state the day and time during the study, as the collected data will be anonymized. The contact information to do so has been given in this brochure and the consent form.

Bibliography:

[1] K. Vitols, E. Voss, and L. Mackay, 'DRIVER FATIGUE IN EUROPEAN ROAD TRANSPORT', p. 60.

[2] G. Sikander and S. Anwar, 'Driver Fatigue Detection Systems: A Review', *IEEE Trans. Intell. Transport. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019, doi: 10.1109/TITS.2018.2868499.

[3] 'Volvo Trucks Safety Report 2017', p. 34.

Information Brochure for testing fatigue detecting system

Institution: University of Twente (053 489 9111)

Researcher information: Agata Sowa, a.m.sowa@student.utwente.nl (+48608298169)

Supervisor: Job Zwiers, j.zwiers@utwente.nl

Title: Development of a computer vision-based system for recognising fatigue of truck drivers

Purpose: Sixty per cent of truck drivers admitted to being fatigued every ninth or lesser drive. Furthermore, 30% of them stated that they had fallen asleep behind the wheel within 12 months before the study conducted by Vitols [1]. Current safety systems implemented in the trucks to recognise driver fatigue mostly take into account truck position between the lanes and steering patterns [2]. Of all road accidents involving heavy goods vehicles, 90% are caused by human factors. The human factors are split into more detailed categories. One of the most prominent ones is drivers' fatigue – 20% of all accidents caused by human factors result from driver's fatigue [3]. Those percentages indicate that there is still space and need for introducing new technology that could further improve fatigue detection in truck drivers. This research aims to develop a system that could recognize a truck driver's fatigue level in real-time or close to it. The system considers changes in the drivers' facial features and, based on that, evaluates their state.

Procedure: You will be asked to sit in front of your laptop and work, study or read something on your device. During this, the researcher will ask you to yawn and move your head up and down in random time periods. In front of you, a webcam will be placed. The webcam will only collect data about changes in your facial features – no video or audio will be recorded. Based on the data collected, the system will predict whether a difference in your face could indicate a growth in the fatigue level. The researcher will also take notes about the system's performance in the meantime. After you are finished, the researcher will use the obtained data to evaluate the system's performance. There will be no possibility to identify you via the collected data. While participating in this research, you can ask the researcher questions at any time.

Duration: 20 – 30 minutes. Regardless, you will be able to quit at any time.

Risks: Covid-19 regulations will be implemented, as well as additional precautions to ensure the health and safety of the participants.

Confidentiality: You will not be required to state any private information. The data and the findings will be anonymized. After participating in the research, you can ask the researcher to delete all your data. For this, you will have to state the day and time during the study, as the collected data will be anonymized. The contact information to do so has been given in this brochure and the consent form.

Bibliography:

[1] K. Vitols, E. Voss, and L. Mackay, 'DRIVER FATIGUE IN EUROPEAN ROAD TRANSPORT', p. 60.

[2] G. Sikander and S. Anwar, 'Driver Fatigue Detection Systems: A Review', *IEEE Trans. Intell. Transport. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019, doi: 10.1109/TITS.2018.2868499.

[3] 'Volvo Trucks Safety Report 2017', p. 34.

Appendix D – Consent Form

Consent Form for testing fatigue detection system

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated [DD/MM/YYYY] (date yet to be set for each individual participant), or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves testing of the facial features recognition and fatigue detection system.	○	○
Use of the information in the study		
I understand that information I provide will be used for determining the accuracy of fatigue detection system developed by the researcher	○	○
I consent that obtained results can be described in the research outcomes with maintaining my anonymity as the study participant.	○	○

Signatures

______________________ ______________________ ________

Name of participant Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_________________________ ___________________ ________

Researcher name Signature Date

Study contact details for further information: Agata Sowa, **a.m.sowa@student.utwente.nl**
Supervisor: Job Zwiers, **j.zwiers@utwente.nl**

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-CIS@utwente.nl