Classification of Muscle Fatigue during Prolonged Driving

Noor Azlyn Ab Ghafar*, Nur Liyana Azmi, Khaerul Affendy Md Nor and Nor Hidayati Diyan Nordin

Department of Mechatronics Engineering, Kulliyah of Engineering, International Islamic University Malaysia.

*Corresponding author: liyanazmi@iium.edu.my

Abstract: Driving has become essential in transporting people from one place to another. However, prolonged driving could cause muscle fatigue, leading to drowsiness and microsleep. Electromyography (EMG) is an important type of electro-psychological signal that is used to measure electrical activity in muscles. The current study attempted to use machine learning algorithms to classify EMG signals recorded from the trapezius muscle of 10 healthy subjects in non-fatigue and fatigue conditions. The EMG signals and the time when muscle fatigue was experienced by the subjects were recorded. The mean frequency (MNF) and median frequency (MDF) of the EMG signals were extracted as dataset features. Six machine learning models were used for the classification: Logistic Regression, Support Vector Machine, Naïve Bayes, k-nearest Neighbour, Decision Tree and Random Forest. The results show that both the MNF and MDF are lower when fatigue conditions exist. In term of the classification performance, the Random Forest, Decision Tree and k-nearest Neighbour classifiers produced the accuracy levels of 85.00%, 83.75% and 81.25% respectively. Thus, the highest accuracy for classifying muscle fatigue due to prolonged driving was obtained by the Random Forest classifier, using both the MNF and MDF of the EMG signals. This method of using the MNF and MDF will be useful in classifying driver’s non-fatigue and fatigue conditions during prolonged driving.

Keywords: Classification, Electromyography, Median Frequency, Mean Frequency, Muscle fatigue

Article History: received 15 May 2022; accepted 17 October 2022; published 22 December 2022.

© 2022 Penerbit UTM Press. All rights reserved

1. INTRODUCTION
Road transportation is one of the major modes of transport used by Malaysians. Driving necessity because it is fast, cheap and practical way of moving people from one place to another [1]. According to a Ministry of Transport Malaysia (MOTM) report, in 2020, the number of registered vehicles recorded was 32.38 million. This number had increased to 33.57 million by 2021. In addition, cars have been recorded as the type of vehicle used most frequently by Malaysians with a rate of 47.10% followed by motorcycle at a rate of 46.19% in 2021 [2].

As a developing country, Malaysia gains income from greater productivity, which at the same time requires people to move faster and further [3]. Therefore, the transportation system of roads and highways should be greatly improved, which will enable Malaysians to experience better infrastructure, facilities and comfort [4]. However, while the increase in driving activity has offered major benefits, it has also had negative effects due to the increasing number of road accidents.

Malaysia has one of the highest rates of road accidents worldwide in relation to its population. Since 2012 to 2018, Malaysia has been ranked as the seventh-highest country in the world for the overall number of traffic accidents. Additionally, Malaysia has had the greatest global mortality per 100,000 people since 1996 [5]. In 2019, the Road Safety Department of Malaysia recorded 5764 cases of fatal accidents [6]. The three main causes of traffic accidents are human, environmental, and technical factors [7]. According to Malaysian Institute of Road Safety Research (MIROS) reported that the main contributor to road accident is human factor as much as 80% [8]. Mahat et al. (2020) categorized human factor into subfactor and according to their finding, the first ranking is drunk driving while drowsiness or microsleep rank as second factor contributing to road accident [6].

Microsleep or drowsiness is extremely dangerous when driving. It is defined as a sleep in a short period of time [9]. Fatigue is one of the factors leading to microsleep or drowsiness of the drivers besides prolonged driving, road condition, environment and health[10].

Thus, in this work, an experiment was conducted to classify the fatigue and non-fatigue condition of the driver during prolonged driving. Subjects were asked to drive a car on a monotonous highway, where the speed limit is restricted to 90km/h. Driving on the highway involves a monotonous driving environment because of the wide and flat pavement, fewer spatial references and high volume of traffic [11]. Prolonged driving in this type of environment requires drivers to sustain attention over long a period which decreases their alertness performance and lead to fatigue.

The car seat inclination angle was set to 10 degrees throughout the experiment. Majid et al. (2013) proposed...
that the optimal adjustment for a car seat was a seat inclination of 10° and a seat pan inclination from 0° to 5° [12]. Li et al. (2015) stated that a slight backward inclination angle of the backrest (approximately 10°) may reduce a driver’s muscle fatigue [13].

Electromyography (EMG) is an experimental technique concerned with the development, recording and analysis of myoelectric signals. Myoelectric signals are formed by physiological variations in the state of the muscle fibre membranes [14]. Two types of EMG muscle sensors are available in the market: intramuscular EMG and surface EMG [15]. The former is also called the invasive electrode approach as it uses needle electrodes that penetrate the skin. This type of EMG was not suitable for this study because only certified personnel can perform these tests while this type of EMG will also make the subject feel uncomfortable. Meanwhile, surface EMG, often called the non-invasive electrode approach, measures muscle activity on the surface of the skin. Surface EMG electrodes are fairly inexpensive and can be easily placed on various muscles, making them suitable for numerous purposes [16]. More than two electrodes are needed to measure the EMG signals because the sensor records the potential difference (voltage difference) between the two separate electrodes. In this work, surface EMG was used to measure EMG signals of the muscle.

The position and orientation of the EMG sensor’s electrodes have a vast effect on the signal strength. The electrodes should be placed in the middle of the muscle body and align with the orientation of the muscle fibres [17]. The placement of these electrodes should be based on the SENIAM standard. The European SENIAM (Surface EMG for the Non-Invasive Assessment of Muscles) project which aims to standardize the placement procedure of EMG sensor, processing the EMG signal and modeling the signal [18]. This work has followed this protocol.

Muscle fatigue has been defined as a reduction in the maximum capacity available to generate force or power output [19]. In previous research, the median frequency (MNF) and mean frequency (MDF), based on the Fourier Transform of the EMG signals have been used for muscle fatigue assessment [20][21]. When muscle fatigue occurs, the blood flow to the muscle decreases because the muscles contract intensely, reducing the blood flow and thus the availability of oxygen. Otherwise the muscle is simply working so intensely that there is literally not enough oxygen to meet the demand [4]. The energy reserves (sugar and phosphorous) are depleted, while lactic acid and carbon dioxide levels increase and the muscular tissue becomes acidic [22]. This results in the decreasing conduction velocity of the motor action potential on the muscle membrane. Thus, the power spectrum of the EMG signals recorded from the muscle shifts towards lower frequencies when muscles are in a fatigue condition. As a result, both the MNF and MDF values in non-fatigue conditions are higher than those obtained in fatigue condition [23].

The main objective of this work was to classify non-fatigue and fatigue conditions using the MNF and MDF of EMG signals recorded from the trapezius muscles of drivers during prolonged driving, based on subjective user reports using Machine Learning Classifier. Machine learning classifiers, namely Logistic Regression, Support Vector Machine, Naïve Bayes, k-nearest Neighbours, Decision Tree and Random Forest were used in this work. To date, Machine Learning classification of muscle fatigue using EMG has mainly focused on the areas of rehabilitation, sports science, human-computer interaction and medical research. However, no research has been conducted in the field of driving. Although this is an important topic as driving fatigue leads to accidents and the loss of life.

2. METHODOLOGY

2.1 Subjects

Five healthy male subjects and five healthy female subjects (mean age 30.8 ± 5.77 years, height 164.4 ± 6.06 cm and body mass 64.2 ± 12.7 kg) with no history of sleep related problem participated in this study. The nature of the study and the procedure of the experiment were fully explained to the subjects and consent form were obtained prior to the experiment. This study was approved by the International Islamic University Malaysia Research Ethics Committee (ID No: IREC 2020-069).

2.2 Experimental procedure

Two Ag-AgCl disc type disposable electrodes were placed on the left trapezius muscle according to the SENIAM standard [24]. The reference electrode was placed on the bony surface of the C7 vertebra as shown in Figure 1. In the figure below, the headrest was removed for ease of visualization. Before positioning the electrodes over the muscle, an alcohol swab (Isopropyl, approximately 70%) was applied to the skin surface in order to remove dirt and dried skin. The EMG signals were recorded using a BITalino biosignal acquisition board and acquired at a sampling rate of 1000Hz.

![Figure 1. Location of EMG electrodes](image)

The subjects needed to drive for two hours using a highway route on the East Coast Expressway phase 2, Malaysia. The driving duration was chosen for two hours based on the pilot study by El Falou et al. (2003) where the subjects reported experiencing muscle pain after two hours...
in sitting position [25]. The highway route was a monotonous environment with straight feature but also with some slanted ramps, unexpected downhills and bumpy features. The same car model, the Perodua Axia with automatic transmission was used as the test vehicle. The seat inclination angle was set to 10 degrees. All the subjects are needed to maintain a driving speed of 90km/h.

When driving, the researcher verbally asked questions from a questionnaire every 5 minutes. The subjects were asked if they felt drowsy or sleepy or were experiencing any muscle pain. Wang et al. (2017) asked a 10 minute interval questionnaire to the subject during the simulation driving to minimize the fluctuation and difference between subject [26]. In this work, it is assumed that 5 minutes interval is more accurate to detect muscle fatigue perceived by the subjects. The muscle fatigue perceived by the driver is known as the subjective measure and the time of muscle fatigue occurred will be recorded. The EMG signals before and after the time of the subjective muscle fatigue were considered as non-fatigue and fatigue conditions, respectively.

2.3 Data Processing

Having been collected using MATLAB software, the EMG signals were filtered using a fourth-order Butterworth band pass filter with a range of 20-500Hz to remove noise at the high-end cut-off and motion artefacts at the low-end cut-off. The MNF and MDF of the EMG signals were then computed using the sliding window technique, as simplified in Figure 2 [27]. The window size was 250 samples and the increments of 125 samples were used to segment the data. The window size was selected as suggested by Thongpanja et al. (2013) [28]. For each segment, the MNF and MDF were obtained.

MNF is defined as the average frequency and calculated as the sum of the product of the EMG power spectrum and the frequency divided by the total sum of the power spectrum [29]. The MNF equation is given as follows:

\[
MNF = \frac{\sum_{j=1}^{M} f_j P_j}{\sum_{j=1}^{M} P_j}
\]  

(1)

MDF is obtained by dividing the EMG total power spectrum into two equal halves. The MDF equation is given as follows:

\[
\sum_{j=1}^{MDF} P_j = \sum_{j=MDF}^{M} P_j = \frac{1}{2} \sum_{j=1}^{M} P_j
\]  

(2)

Where \( f_j \) is the frequency value of the EMG power spectrum at the frequency bin \( j \), \( P_j \) is the EMG power spectrum at the frequency bin \( j \), and \( M \) is the length of the frequency bin for both the MNF and MDF.

Based on the time of the subjective muscle fatigue for every subject, the MNF and MDF were extracted five times before and after the subjective fatigue time. The time gap between the subjective measure and the data taken occurred because, according to Sahayadhas et al. (2013), subjective measures do not fully coincide with physiological measures [30].

The MNF and MDF values before the time of the subjective muscle fatigue were considered non-fatigue conditions, whereas the values after the time of the subjective muscle fatigue were considered fatigue conditions. Hence, for every subject, five MNF values for non-fatigue data and another five values for fatigue data were collected. Same method applied to MDF values, All the dataset were normalised so that the differences in data range between the subjects would be eliminated [31].

2.4 Classification

The computed MNF and MDF values were used as the features dataset for classification. The data was divided: twenty-five percent used as the test set, while the remaining data became the training set. Six machine learning classifiers namely Logistic Regression, Support Vector Machine, Naïve Bayes, k-nearest Neighbours, Decision Tree and Random Forest were used to classify non-fatigue and fatigue conditions. A ten-fold cross

![Figure 2. A concept of the sliding window technique to compute the mean frequency and median frequency of EMG signals using consecutive overlapping FFTs with a window size (L) and a window increment (I).](image-url)
validation method was implemented to evaluate the performance (accuracy) of the classifiers.

3. RESULTS AND DISCUSSION

An EMG filtered signal is shown in Figure 3(a). The original EMG signal was filtered with the band pass filter to remove noise and motion artefacts. Next the median frequency (Figure 3(b)) and mean frequency (Figure 3(c)) were computed using the sliding window technique.

Figures 3 above show the EMG signal, MNF and MDF for a representative subject. According to the subjective muscle fatigue, which had been recorded by verbally questioning the subject, the time that muscle fatigue was experienced by the subject was in the 45th minutes of driving, that is after 2,700s. From Figure 3(a), it can be concluded that, after 2,700s, the MNF and MDF of the EMG signal decreased and fluctuated. This was because, as outlined in the literature, the association of the MNF and MDF values with fatigue decreases due to the reduction in the propagation velocity of the muscle’s action potential [32].

Based on the time of the subjective muscle fatigue, the MNF and MDF values were extracted five times before and another five times after the time of the subjective muscle fatigue. A total of 50 non-fatigue and fatigue datasets obtained from the 10 subjects overall were used as the features of the machine learning model. Figures 4 and 5 show the MDF and MNF value respectively.

Based on Figure 4 and 5, both the MNF and MDF values were higher in non-fatigue conditions. The characteristic of spectral shift towards a lower frequency region was used to evaluate the fatigued muscle. The results indicate that the MNF and MDF values could be used as the input or features in muscle fatigue classification.

Next, the MNF and MDF data were used to classify muscle fatigue using Machine Learning classifier. The performance result of the classification of non-fatigue and fatigue condition was evaluated using ten-fold cross-validation. The results are summarised in Table 1 below. The most accurate classifier for MDF was obtained using the Random Forest classifier, 81.96%. On the other hand, when only using the MNF dataset, the best accuracy was obtained by the Logistic Regression classifier, 77.68%. Lastly, when both the MDF and MNF were used as features in machine learning model, the Random Forest classifier was the most accurate, improving the accuracy to 85.00%.

Based on the result, the accuracy of the Random Forest classifier when using MDF dataset was only 81.96%, while using the MNF dataset only produced the rate of 73.57%. However, when both the MDF and MNF were used as the dataset of the model, the accuracy improved to 85.00%. Thus, it can be concluded that using more features in classification dataset yields higher classification accuracy in most of the classifier.
Thus, the most accurate form of muscle fatigue classification for prolonged driving was obtained by the Random Forest classifier using both the MDF and MNF values of EMG signals. This classifier is powerful and widely used because of the stability and robustness of the data, which features only slight variations. This classifier is constructed using multiple distinct decision trees and the final decision is predicted by most of the trees. Each decision tree is trained with different subsets of the training data using random sample from the original training set [32].

The second highest accuracy in this work was obtained using the Decision Tree classifier, which yielded 83.75%. This classifier performs well with an enormous volume of information, while unrelated features do not influence its results. However, the drawback is over-fitting as it is sensitive to information [33]. This is because when the result will extremely change to huge degree when the information changes. Lastly, the k-Nearest Neighbours classifier produced an accuracy rate of 81.25%, making it the third-best classifier. With training data, the k-Nearest Neighbours algorithm sets a group of k objects closest to the test object. It then assigns a class to the test object based on the neighbours. The three main stages of the k-Nearest Neighbours algorithm are initialising dataset and k-Nearest Neighbours, computing the distance between neighbours, and classifying the test data based on the majority of the neighbouring class data [31]. The value of k was iterated and set as five in this study based on the highest classification accuracy obtained when tested with MDF classification. The result is shown in Table 2 below where k=5 produces the highest classification accuracy. The usage of k=5 also been used in the research by Marri et al. (2016) in classifying muscle fatigue [34].

### Table 1. Cross validation accuracy result for MDF and MNF

<table>
<thead>
<tr>
<th>Classifier</th>
<th>MDF Accuracy (%)</th>
<th>MNF Accuracy (%)</th>
<th>MDF and MNF Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>80.36</td>
<td>76.25</td>
<td>77.86</td>
</tr>
<tr>
<td>Random Forest</td>
<td>81.96</td>
<td>73.57</td>
<td>85.00</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>79.29</td>
<td>76.25</td>
<td>76.43</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>79.11</td>
<td>77.68</td>
<td>79.11</td>
</tr>
<tr>
<td>k-Nearest Neighbours</td>
<td>79.46</td>
<td>73.21</td>
<td>81.25</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>79.46</td>
<td>74.64</td>
<td>83.75</td>
</tr>
</tbody>
</table>

Below summarized the result of different k-fold cross validation tested. The ten-fold cross validation was also used by previous researcher to classify muscle fatigue [35][33][36].

### Table 3. Selection of k value for k-fold cross validation

<table>
<thead>
<tr>
<th>k-fold cross validation value</th>
<th>k=3</th>
<th>k=5</th>
<th>k=10</th>
<th>k=15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier Accuracy (%)</td>
<td>73.33</td>
<td>77.33</td>
<td>79.46</td>
<td>78.67</td>
</tr>
</tbody>
</table>

This study has limitation. The number of samples is small. However, this pilot study managed to show that the classification of muscle fatigue for prolonged driving especially by the Random Forest classifier using both the MDF and MNF values of EMG signals.

### 4. CONCLUSION

Due to the increasing use of road transportation and accident rates in Malaysia, this study focused on the classification of non-fatigue and fatigue conditions in drivers during prolonged driving. EMG signals from the trapezius muscle were recorded and the MNF and MDF were computed. The MNF and MDF dataset were trained and tested using six machine learning models, the Logistic Regression, Support Vector Machine, Naïve Bayes, k-Nearest Neighbours, Decision Tree and Random Forest classifiers. The results show that both the MNF and MDF value were lower in fatigue conditions compared to non-fatigue conditions. In addition, the Random Forest, Decision Tree and k-nearest Neighbour classifiers performed with the accuracy levels of 85%, 83.75% and 81.25% respectively.

It is suggested that in the future, further research is conducted on detecting and classifying muscle fatigue during driving, while other psychophysical signals like electrocardiogram (ECG), electrocoulogram (EcG) and electroencephalogram (EEG) signals could be used to improve the classification accuracy. The outcomes of this work form important guidelines that could be used when studying driver’s muscle fatigue and to detect muscle fatigue when driving. Early detection of muscle fatigue of the driver is important to reduce fatigue, avoid musculoskeletal disorders, prevent accidents and loss of life.

### ACKNOWLEDGMENT

This work was supported and funded by Ministry of Education Malaysia and International Islamic University Malaysia (IIUM) under Research Acculturation Grant Scheme (IRAGS)(IRAGS18-023-0024).

### REFERENCES


