2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Impairment Screening Utilizing Biophysical Measurements and Machine Learning Algorithms

Saboora M. Roshan and Edward J. Park, *Senior Member, IEEE*

Abstract— **Drug recognition expert (DRE) officers employ a set of tests to investigate drivers who are under impairment and to determine the type of drug that they have used. Horizontal Gaze Nystagmus (HGN), Walk and Turn (WAT), and One Leg Stand (OLS) are the main three tests included in the Standardized Field Sobriety Tests (SFSTs), which lead the officers to evaluate the sobriety of drivers. Performing these tests requires trained officers, but the final decision may still be subjective. These tests along with a suite of comprehensive (yet manual) at-station testing are the basis of police decision making and are subjected to scrutiny by courts. Therefore, designing an automated system to detect impairment not only will help officers in making accurate decisions, but also will remove the subjectivity and can potentially serve as a court-admissible evidence. In this paper, a new method for automated impairment detection is introduced and implemented using data analysis and machine learning algorithms based on a comprehensive suite of tests performed on 34 participants.**

I. Introduction

Using drugs or alcohol may cause impairment that can make driving or operating heavy machinery dangerous. The percentage of fatality alcohol-impaired drivers in 2018 was 29% (10,511 people) out of all fatal vehicle crashes in the US [1]. In Canada, 55.4% (1,273 people) of all traffic deaths in 2014 was alcohol or drugs related crashes [2]. Therefore, it is critical to detect drivers who are under the influence of drugs or alcohol. Drug Recognition Expert (DRE) officers employ the Standardized Field Sobriety Tests (SFSTs) to investigate drivers who are under impairment [3]. While DRE officers perform the SFSTs on suspected drivers and decide if they are impaired or not; these tests are considered subjective and are dependent on DRE officers' qualification and opinion. One way to overcome this issue is to design an automated system to determine impairment using machine learning and data analysis methods. The SFSTs consist of an eye examination, divided attention psychophysical tests, and vital signs and biosignal measurements (blood pressure, temperature, and pulse) [4]. The most important subset of these tests is Horizontal Gaze Nystagmus (HGN), Walk and Turn (WAT), and One Leg Stand (OLS), specifically in detecting alcohol. Alcohol is one of the most popular central nervous system (CNS) depressants which reduces the CNS' activity and has a sedative and calming effect. Cannabis may cause depression, anxiety, poor judgment, short-term memory, and impaired motor skills. Cannabis affects the decision making of drivers and operators of heavy equipment which results in their control capabilities [5]. Table I illustrates drug indicators of the SFSTs which officers use for drug type classification [4].

TABLE I. Drug Indicators of SFSTs

Drug type	Drug Indicators				
	HGN	***Pulse***	***Blood Pressure***	***Temp-erature***	***Psycho-physical Impair-ment***
CNS Depressant	Present	Down	Down	Normal	Present
Cannabis	None	Up	Up	Normal	Present

To perform HGN, DRE officer positions a stimulus (e.g., the tip of a pen) 12 to 15 inches in front of the person's nose, moves it all the way to the subject's side, holds the eye at that position (the eye is turned as far to the side as possible) for at least 4 seconds, and examines the eye for evidence of distinct jerking. The HGN test contains three clues: lack of smooth pursuit, distinct and sustained nystagmus at maximum deviation, and angle of onset of nystagmus. A subject is considered to have failed the test when two clues from six clues in both eyes are detected by the officer. In WAT, the participant is instructed to walk for nine steps heel-to-toe back and forth while keeping balance. The officer investigates the following eight clues in this test: keeping balance while listening to the instruction, starting too soon, stopping while walking, having a one-half inch or more distance between toe and heel, stepping off the line, using arms to keep balance (raising arms 6 inches or more), performing improper turn, and having incorrect step counts. If the officer detects two or more clues from these eight clues, then the subject fails this test. The other psychophysical test is OLS, in which the subjects are first told to stand with heels together and arms down at their sides. When the test begins, the subjects should raise one leg about six inches from the ground and hold that position. At the same time, they should count 30 seconds starting from 1000, while watching their foot and maintaining their balance. This test has four clues: swaying while balancing, using arms to keep balance, hopping, and putting foot down. If DRE detects two clues during the test, the subject is considered to have failed [6].

Although these three tests are the primary ones to detect impaired drivers and are being widely used, their accuracy is not perfect: HGN (88%), WAT (79%), and OLS (83%) [6]. These tests, despite being gold standard, are not considered

* Research supported by Mitacs.
Saboora M. Roshan and Edward J. Park are with the School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, BC V3T 0A3, Canada (e-mail: sabooram@sfu.ca, ed_park@sfu.ca).

accurate enough, even when being performed by qualified professionals. Moreover, not all drugs show consistent symptoms on all the tests. For example, Porath-Waller *et al.* argued that only 1% of cannabis consumers can be detected with the HGN test [7], while Hartman *et al.* mentioned no difference in HGN test between cannabis impaired subjects and sober ones [8]. On the contrary, HGN is noticeable in alcohol consumers [9]. Moreover, Porath-Waller *et al.* predicted the type of drugs using drug evaluation dataset [10] and, similarly, the accuracy of drug evaluation tests is discussed in Beirness *et al.* [11]. Furthermore, Downey *et al.* presented the usefulness of the SFST as a screening tool to identify drug or alcohol impaired drivers by implementing point-biserial correlation method [12].

In this study, we used HGN, WAT, OLS, and biometric measurements to implement an automated impairment screening system. To our best knowledge, no previous studies have presented the method proposed in this paper. The rest of the paper is organized as follows. Section II explains the dataset, and Section III discusses the experimental methodology and data preparation method. Section IV discusses the results of the simulations and the comparison of different classifiers. The study's limitations are presented in Section V, and Section VI concludes the paper.

II. DATASET

We implemented the impairment detection on the data gathered from the automated impairment screening system that is developed by a partner organization, CannSight Technologies. These datasets were collected by CannSight via SFSTs under a registered IRB/REB trial and stored for the purpose of secondary use in our study. (The protocol numbers for alcohol and cannabis are 'Pro00037800' and 'Pro00037799', respectively.) The CannSight device consists of eye testing screen and camera system, body movement imaging system, and biometric measurement sensors. Our study protocol was approved by the Research Ethics Board of Simon Fraser University. (The ethics certificate number is '2020s0086'.)

The average age of 34 volunteers (12 females and 22 males) who participated in this study was 29.98. The number of sober, alcohol-impaired, and cannabis-impaired subjects were 17, 14, and 3, respectively. Both alcohol-impaired and cannabis-impaired were labeled "impaired" to have a more precise binary classification between impaired and sober subjects. The features that we used for impairment detection were: (i) HGN (right eye), (ii) HGN (left eye), (iii) WAT, (iv) OLS (right leg), (v) OLS (left leg), (vi) pulse rate, (vii) systolic blood pressure, (viii) diastolic blood pressure, and (ix) temperature. The pulse rate and blood pressure are the average of three values that were measured at three different times during the tests.

III. METHODS

A. Horizontal Gaze Nystagmus (HGN)

To implement the automated impairment detection from HGN tests, videos of eye movements were used. As discussed earlier, HGN is one of the SFSTs that is implemented to investigate impaired drivers and is defined as involuntary jerking of the eyes when the eyes gaze side to side. In this test, the participant is asked to stand in front of a visual stimulus at pre-determined distance and to look at the stimulus as it moves from side to side on a screen. The type of camera that is used in this test is ELP Video Surveillance Camera with resolution of 3840×2160 and CMOS IMX317 sensor, which is located close to the stimulus and captures eyes' movements. The stimulus moves three times from side to side for each eye and stays at a specific position where the eye is turned as far to the side as possible for at least 4 seconds. Eye movements are represented as signals in our study, and different signal processing methods can be used to analyze them as discussed in the previous studies. For instance, in Hindarto and Sumarno [13], a fast Fourier transform (FFT) was used to extract features for electroencephalography (EEG). FFT was also used by Azim *et al.* [14] for extracting features of human sleep EEG signal. Saxena *et al.* used FFT and machine learning for emotion detection [15].

The first steps of HGN data preparation are noise removal with median filter; histogram equalization using Contrast Limited Adaptive Histogram Equalization (CLAHE) method; conversion of RGB images to gray scale; and implementation of Canny edge detector. Then, some of the detected contours are removed from the images using a set of pre-defined conditions. These conditions are defined to find contours that have same size as pupils, and the defined values are measured on a trial-and-error basis. One set of conditions consists of removing contours that their lengths are less than 30 pixels (small contours), their ratio of the breadths over lengths of the rectangular around them are smaller than 0.5, and their diameter ratios are between 0.1 and 2. Needless to say, these numbers should be updated based on the dataset. After that, ellipses are fitted to the images, and the best ellipse is chosen as the pupil ellipse. To choose the best ellipse, the first condition is to choose ellipses that their major and minor axis are between 50 and 400 pixels. The second condition is to choose ellipses with their ratio of major axis over minor axis between 1 and 1.8, in order to remove very large ellipses. These parameters and ratios are derived from different experiments.

After finding the center of the ellipse, the bias due to the head and body movements should be removed. One way to remove this offset from the pupil's position is to subtract the center of the pupil from a fixed point on the face (e.g., a corner of the eye). In order to find the corner of the eyes, first, the eyes need to be detected. Dlib and OpenCV libraries are used to extract eye regions from the images. The dlib library is used to detect facial landmarks and map the coordinates to each facial region [16]. The position of the center is then subtracted from the corner of the eyes to remove head movements ($x_{removed-offset} = x_{center} - x_{corner}$). Then, the signals are normalized and filtered using Savitzky-Golay filtering method. After that, three features including RMS, Skewness, and Kurtosis of the signals are used for impairment detection from the HGN test. Kurtosis and Skewness of a signal are defined as follows:

$$Kurtosis[x] = \frac{E[(x-\mu)^3]}{E[(x-\mu)^2]^2} \quad (1)$$

$$Skew[x] = E[(\frac{x-\mu}{\sigma})^3] \quad (2)$$

where μ, σ, and E are the mean, standard deviation, and statistical expectation.

B. Walk and Turn (WAT)

WAT is another test in the SFST suite, in which the participant is instructed to walk for nine steps heel-to-toe back and forth while keeping balance. This test is recorded with the Intel RealSense Depth Camera D435 with Depth FOV of 85 ×58 degrees, which is located in front of the subject. WAT has four phases including instruction, walking, turning, and walking back. To classify the participants into two impaired and sober groups regarding their balance during each phase of WAT, first, the keypoints (i.e., indicator joints) of their body are extracted for each frame of the video using OpenPose model [17], which results in 14 keypoints. Figure 1 shows these keypoints on a sample picture. Noteworthy is that as the subject walks toward the camera, the size of skeleton changes because of the difference in the participant's distance to the camera. Hence, a scale needs to be utilized to have invariant skeleton in all frames as discussed in two previous studies, Li *et al.* [18] and Kumar and Babu [19]. The Euclidean distance between two keypoints (shoulder center and hip center) should be the same in all frames. The scale is the fraction of the Euclidean distance over 100 pixels [18]. Then, the subtraction between each position and the hip center (as a reference point) is scaled with respect to the scale and is defined as follows:

$$\hat{x}_i^n = \frac{x_i^n - x_c^n}{scale} \tag{3}$$

in which, x_i^n is the x position of i^{th} keypoint in n^{th} frame, x_c^n is the hip center, and $\hat{x}_i^n$ is the scaled position.

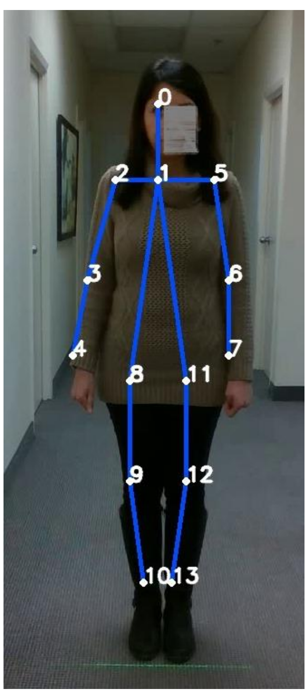

Figure 1. Keypoints of a sample picture using OpenPose model

As mentioned before, the WAT test can be divided into four sections. First, the keypoints' trajectories of each phase are transformed to frequency domain using Discrete Cosine Transform (DCT) to reduce noises and to remove high frequencies from the trajectory (50% of frequencies). After that, the trajectories are transformed back to time domain using inverse transform (IDCT) [17]. Next, the covariance matrix between all trajectories is computed to obtain relative movement of the skeletal points. Also, the difference between the keypoint trajectories and their mean in all frames is calculated for each keypoint, which hereby will be called the "difference value" for the remainder of this paper. Finally, a simple classifier is trained on each section based on the following features (refer to Fig. 1): relational movement between trajectories of the keypoints (3,4), (7,8), and (7,11), standard deviation of the difference value of the keypoints 4 and 7, the average of the difference value of the keypoints 4 and 7, and kurtosis of the difference value of the keypoints 4 and 7. In each section (instruction, walk, and walkback), we first computed the Spearman rank-order correlation between the above features to choose a single feature from each cluster and to reduce the correlation between the features, in order to produce acceptable performance in each phase. Then, we combined the selected features from all sections, and trained a Random Forest classifier to predict the impairment of the subject from the WAT test.

C. One Leg Stand (OLS)

Another divided attention psychophysical test is OLS, which contains raising one foot and counting 30 seconds while keeping balance. Videos of the OLS test is also recorded using the camera mentioned in Sec. III.B, which is placed in front of the subject. The procedure of preparing trajectories of the keypoints to remove the variance in the body movements is the same as that of Sec. III.B. The keypoints 4, 7, 10, and 13 (Fig. 1) are the most important keypoints for OLS impairment detection as they can show raising arms, putting foot down, hopping, and balancing clues which are all clues of this test. Therefore, the difference between each of these keypoints from their average in the whole video is calculated, and a value of 20 pixels is introduced as a threshold for detecting the clues by trial-and-error. If the difference between the keypoint trajectory and the mean is 20 pixels or higher in 5 consecutive frames, one clue is recorded. The result is the number of clues of all above keypoints in the video.

D. Biometric Measurements

Biometric measurement includes measurement of blood pressure, heart rate, and body temperature. Temperature and blood pressure are measured by Easy@Home Digital Thermometer and HOMIEE Blood Pressure Machine, respectively. Blood pressure and heart rate (pulse) are measured three times during the tests, and their average is used as biometric features.

E. Data Preparation and Classification

The first step of data preparation is to normalize the features. Also, the features of HGN, WAT, and OLS contain lots of zeros due to the nature of the impairment in the specific test and can potentially affect the performance of the classification algorithm. One way to solve this problem is to add a small value to all features. Therefore, we calculated the minimum value of all features (x) and added $log(x+(x/2))$ to all features to remove zeros in the dataset. Then, hierarchical clustering on the Spearman rank-order correlations is performed to keep single features from each cluster by choosing a threshold to select uncorrelated features. Because the dataset is relatively small, several classifiers, such as Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM) using k-fold cross validation, were implemented, and compared to determine the best solution for the available dataset.

Table II introduces the performance indices that we used for comparing the classifiers. In this table, TP, FP, TN, and FN represent True Positive, False Positive, True Negative and False Negative results of the classification algorithm, respectively.

TABLE II. PERFORMANCE INDICES

Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-Score	$\frac{2TP}{2TP+FP+FN}$

IV. RESULTS AND DISCUSSION

The dataset contains 34 subjects and nine features. Figures 2 and 3 show the Spearman linkage dendrogram of features and correlation heatmap, respectively. As shown in Fig. 2, the linkage distance between HGN-Right and HGN-Left is the lowest among all features. This shows the correlation between the two features, and one of them can be removed from the inputs. OLS of right foot (OLS-R) and pulse are also correlated to each other with a linkage distance of around 1.2. In addition, according to Fig. 3, some of the uncorrelated features are pulse and diastolic blood pressure, OLS-R and systolic blood pressure, and temperature (Temp) and both HGN. After removing the correlated features using the Spearman method, only five features are left, including HGN-Right, OLS-R, OLS-L, systolic blood pressure, and temperature. Therefore, this features set is used for the classifier training and performance evaluation.

Table III illustrates four well-known performance indices for each classifier before and after removing the correlated features (first row and second row, respectively). Note that these results are the average of scores of the cross-validation method (*k*-fold with 5 splits). Naïve Bayes before removing correlated features has the best results with an accuracy of 0.9 and a recall of 1.0. This table also shows that the performance of the classifiers does not significantly change after removing correlated features, which is probably due to the small size of the dataset. It seems safe to assume that the performance of the classifiers varies if the sample size is expanded. The optimized parameters for each classifier were chosen using grid search with cross-validation.

Figures 4 and 5 demonstrate the permutation feature importance for the Decision Tree and Random Forest classifiers, respectively. The important features in Decision Tree are WAT, pulse, and temperature. WAT has an importance of 0.6 in classification and, removing this feature (due to its correlation to the other features) decreases the accuracy from 0.87 to 0.85. Also, it can be seen from Fig. 5 that, again, WAT has the highest importance in the Random Forest classifier. Moreover, in this classifier, HGN-L and HGN-R have significant effect on the classifier's performance, while OLS-R and systolic blood pressure have the least effect. The other features also have noticeable effect on the Random Forest classifier, which was observed in the results of Table III.

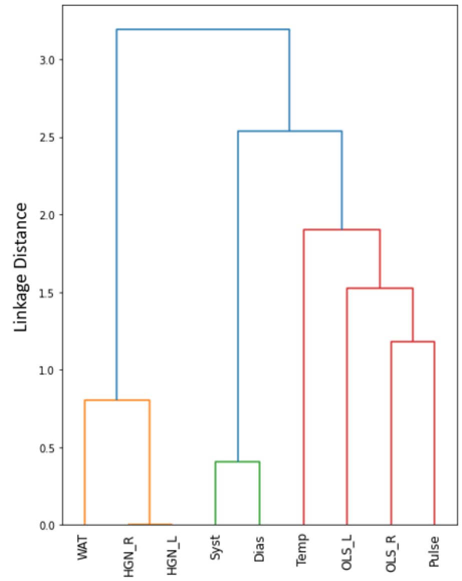

Figure 2. Linkage dendrogram of features

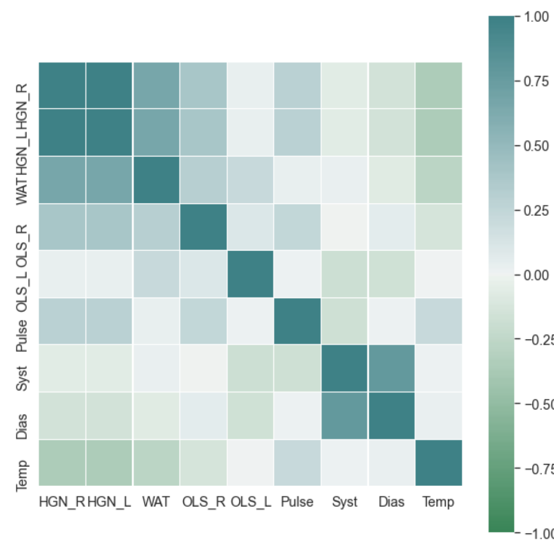

Figure 3. Correlation heatmap of features

TABLE III. CLASSIFIERS' PERFORMANCE

Classifier	Classification Performance			
	Accuracy	***Precision***	***Recall***	***F1-score***
Naïve Bayes	**0.9** 0.81	**0.82** 0.8	**1.0** 0.83	**0.89** 0.76
Decision Tree	**0.87** 0.85	**0.86** 0.85	**0.9** 0.86	**0.83** 0.81
Random Forest	0.85 0.81	0.9 0.86	0.79 0.79	0.80 0.78
SVM	0.79 0.85	0.82 0.9	0.79 0.79	0.74 0.80

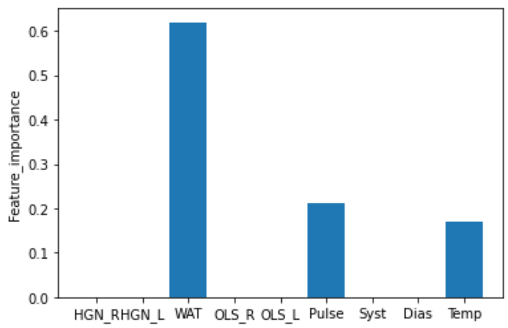

Figure 4. Permutation feature importance for Decision Tree

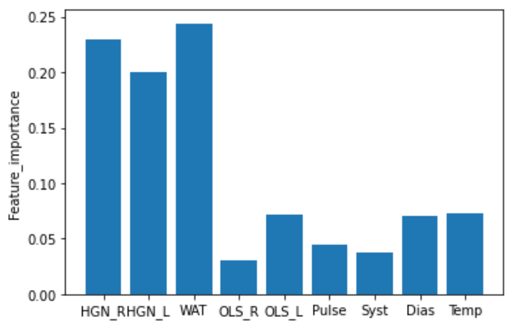

Figure 5. Permutation feature importance for Random Forest

V. LIMITATIONS AND FUTURE WORK

Due to the complexities in recruiting participants for the study and expanding the study size, the current research had limitations such as small dataset size, male-dominant participants, and limited/uneven coverage of consumed drugs. These limitations can potentially affect the ability to generalize the results of the proposed method. The next steps of the research will include obtaining an amended ethics approval for a more diverse and larger sample size to expand the research and to potentially detect the type of drug being used (another important factor for the law enforcement), as well as the degree of impairment.

VI. CONCLUSION

In this paper, an automated impairment screening system is introduced and implemented using a set of eye, psychomotor, and biometric tests. First, a set of features is selected by implementing Spearman correlation method. Then, the classifiers' performance is evaluated on both new set of features and set of all features. Based on the results, Naïve Bayes classifier showed the best performance, while Decision Tree using the same set of features had an acceptable performance. Future work concerns gathering a larger dataset, improving the performance of feature extraction, and optimizing the classifier parameters to improve the overall performance of the automated impairment screening system.

REFERENCES

[1] National Center for Statistics and Analysis, "2018 fatal motor vehicle crashes: Overview. (Traffic Safety Facts Research Note. Report No. DOT HS 812 826)," Washington, DC: National Highway Traffic Safety Administration, 2019. Accessed on: Dec. 20, 2021. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812826

[2] "*The Ultimate List of Canada Driving Statistics for 2020*", Accessed on: Dec. 20, 2021. [Online]. Available: https://tests.ca/driving-statistics/

[3] R. L. Hartman, J. E. Richman, C. E. Hayes, and M. A. Huestis, "Drug Recognition Expert (DRE) examination characteristics of cannabis impairment," *Accident Analysis and Prevention*, vol. 92, pp. 219–229, July 2016.

[4] "*DRE Training Manual 2018*", Accessed on: Dec. 20, 2021. [Online]. Available: https://www.njsp.org/division/investigations/pdf/adtu/2018_DRE_7-Day_Full_Participant_Manual.pdf.

[5] D. Schacter, D. Gilbert, M. Nock, and D. Wegner, *Psychology*, 5th ed, Chap. 5, Worth Publishers, 2020.

[6] "*DWI Detection and Standardized Field Sobriety Testing (SFST) Refresher*", Accessed on: Dec. 20, 2021. [Online]. Available: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/sfst_ig_refresher_manual.pdf.

[7] A. J. Porath-Waller and D. J. Beirness. "An examination of the validity of the standardized field sobriety test in detecting drug impairment using data from the drug evaluation and classification program," *Traffic Injury Prevention*, vol. 15, no. 2, pp. 125-131, 2014.

[8] R. L. Hartman, J. E. Richman, C. E. Hayes, M.A. Huestis, "Drug Recognition Expert (DRE) examination characteristics of cannabis impairment," *Accident Analysis and Prevention*, vol. 92, pp. 219–229, 2016. DOI: https://doi.org/10.1016/j.aap.2016.04.012.

[9] J. Stuster, "Validation of the standardized field sobriety test battery at 0.08% blood alcohol concentration," *Hum Factors*, vol. 48, no. 3, pp. 608-14, 2006. doi: 10.1518/001872006778606895. PMID: 17063973.

[10] A. J. Porath and D. J. Beirness, "Predicting categories of drugs used by suspected drug-impaired drivers using the Drug Evaluation and Classification Program tests," *Traffic Injury Prevention*, vol. 20, no. 3, pp. 255-263, 2019. DOI: 10.1080/15389588.2018.1562178.

[11] D. J. Beirness, E. Beasley, and J. Lecavalier, "The Accuracy of Evaluations by Drug Recognition Experts in Canada," *Canadian Society of Forensic Science Journal*, vol. 42, no. 1, pp. 75-79, 2009, DOI: 10.1080/00085030.2009.10757598.

[12] L. A. Downey, A. C. Hayley, A. J. Porath-Waller, M. Boorman, C. Stough, "The Standardized Field Sobriety Tests (SFST) and measures of cognitive functioning," *Accident Analysis & Prevention*, vol. 86, 2016, pp. 90-98.

[13] H. Hindarto and S. Sumarno, "Feature Extraction of Electroencephalography Signals Using Fast Fourier Transform," *CommIT (Communication and Information Technology) Journal*, vol. 10. No. 49, 2016, DOI: 10.21512/commit.v10i2.1548.

[14] M. R. Azim, M. S. Amin, S. A. Haque, M. N. Ambia and M. A. Shoeb, "Feature extraction of human sleep EEG signals using wavelet transform and Fourier transform," *2nd International Conference on Signal Processing Systems*, Dalian, 2010, pp. V3-701-V3-705, doi: 10.1109/ICSPS.2010.5555506.

[15] S. Anvita, et al. "Emotion Detection Through EEG Signals Using FFT and Machine Learning Techniques," *International Conference on Innovative Computing and Communications*. Singapore: Springer Singapore, pp. 543–550, 2020.

[16] "Facial landmarks extraction using dlib" Accessed on: Dec. 20, 2020. [Online]. Available:: https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/

[17] "*Human Pose Estimation*", Accessed on: June. 5, 2020. [Online]. Available: https://github.com/legolas123/cv-tricks.com/tree/master/OpenCV/Pose_Estimation

[18] Q. Li, et al. "Classification of gait anomalies from kinect," *Vis Comput*, vol. 34, pp. 229–241, 2018, DOI: https://doi.org/10.1007/s00371-016-1330-0.

[19] M. Kumar, M., R. V. Babu, "Human gait recognition using depth camera: a covariance based approach," *Proceedings of the 8th Indian Conference on Computer Vision, Graphics and Image Processing*, vol. 20, pp. 1-6, 2012, DOI: https://doi.org/10.1145/2425333.2425353.