Enhancing Road Safety with AI: A Secure System for Detecting Impairment through Horizontal Gaze Nystagmus

Chase Coleman*, James Coulthard†, Patrick Dodds‡, Ahmad Salman§, Rod MacDonald¶
College of Integrate Science and Engineering, James Madison University
Harrisonburg VA, 22801
Email: {*colem2cr, †coulthjw, ‡doddsph}@dukes.jmu.edu { §salmanaa,¶macdonrh}@jmu.edu

***Abstract*—Driving under the influence remains a pressing issue in the United States, contributing to approximately one-third of all fatal car crashes and claiming 11,000 lives annually. This research project aims to progress public safety by developing a technological solution aligned with the Department of Transportation's specifications. Our system administers an automated Horizontal Gaze Nystagmus (HGN) test that exposes and detects a driver's physiological impairment in order to prevent them from driving while intoxicated. The novelty in our system lies in the personalized machine-learning algorithm that ensures a precise and confidential assessment of a user's sobriety. Our algorithm is backed by significant amounts of data collected from controlled IRB testing that establishes a statistical baseline and is personalized through the performance of each user during their one-time calibration phase. Once calibrated, our system can determine if a user is unfit to operate their vehicle and will consequently immobilize their automobile's transmission. Throughout the entire process the user's data such as their identity, scores, or personalized scoring model, remains confidential through advanced encryption techniques. In all, these technological advancements within our system address the needs determined by the Department of Transportation in a personalized, robust, and confidential manner.**

***Index Terms*—Ignition interlock, Driving while Intoxicated (DWI), Artificial Intelligence (AI), Horizontal Gaze Nystagmus (HGN), Machine Learning (ML), Confidentiality & Privacy, Cryptography**

I. Introduction

As people have been returning to the streets after COVID, so too have the accidents involving them, both fatal and otherwise. According to the National Highway Traffic Safety Administration's most recent publication [1], total traffic accidents have risen by 10% in the time between 2020 and 2021, accounting for 42,939 deaths on the road. Of these deaths, alcohol use was involved in 13,384, nearly one third of the total in 2021 and a 14% rise from the previous year's findings. Of these alcohol impaired fatalities, closed-air vehicles account for the vast majority, with less than 11% of fatalities coming from motorcycles.

It is no surprise, then, that many attempts have been made to solve such a problem, though many have been reactionary. As stated in a journal from the National Library of Medicine [2], in 2011 all fifty states passed sanctions specific to those who have repeated driving while influenced counts on their record, and 38 states required sobriety checkpoints. On the other hand, interlock systems have been underutilized in legislation until recently, with only 9 states requiring their installment for DWI offenders. All of these penalties, however, rely on the metric of Blood Alcohol Content (BAC), a way in which to quantify and measure the level to which someone is alcohol impaired in use since 1938 that tests the amount of alcohol in a persons blood. More recently, legislation has been enacted using a different means of impairment testing, the Horizontal Gaze Nystagmus (HGN), a method by which a police officer may quickly test for indications of a person's impairment by examining the movement of their eyes when tracking an object [3]. In this paper, we outline the refinement and modularization of a proactive sobriety test based on current law enforcement's HGN test. Using this test, our system is capable of sensing, testing, and determining personalized conclusions on a user's level of intoxication based on the nystagmus (erratic eye movements) they may or may not exhibit.

The remainder of this paper is organized as follows. Section II discusses the background on the topic and related work and literature review. Section III details our methodology and the system components. Section IV represents our evaluation method for the system and its components. We present our results in Section V and we conclude the paper in Section VI.

II. Background and Related Work

Until the 1980s, legislation on drunk driving was lax, as we had little to no analogue to test the effects of drunk driving against. In 1982, however, President Reagan spearheaded a national commission to study drunk driving and create plans to deter it [4]. In this commission, they found that during the year the study took place, drunk driving was responsible for over half of all highway deaths, a staggering statistic that jump started modern driving legislation. The first real change happened in 1984 with the passing of the National Minimum Drinking Age Act, a law mandating that no citizens under the age of 21 would be allowed to drink. This law made it much more difficult for inexperienced drivers to acquire alcohol and is credited with preventing over 500 deaths every year. Drunk Driving laws continued to strengthen in the next twenty years due to the influence of groups like Mothers Against Drunk Driving (MADD), and in 1998 another massive leap was made [5]. In this year, the Transportation Equity Act

became federal law, requiring all states to pass laws restricting open containers in vehicles [4]. The next big change came about in the years following a grant enacted in 1998 designed to encourage states to lower the limit from the widely used 0.1 BAC to the modern standard. As of today, the BAC limit is held at 0.08, the point at which a driver is around four times more likely to be in an accident than with no alcohol in their system. Recent solutions being taken into consideration by law enforcement entities are pieces of technology such as the ignition interlock or the use of GPS to track convicted drunk drivers [6]. Though they are not mandated federally, states that require the use of ignition interlock devices by repeat offenders see a seventy-percent decrease in repeat DUIs, and it is estimated that 35,000 lives have been saved by this technology [4].

Many different approaches have been taken to solving the Driving While Intoxicated (DWI) problem. The primary focus has been on criminalizing the behavior to reduce the incidence of drinking and driving. A secondary approach has been to install ignition interlocks on the vehicles of repeat offenders. The majority of research teams take an approach similar to that of Srikar et al. [7], implementing a physical breathalyzer (an instrument used to measure BAC) inside a vehicle and connecting it directly to the ignition system. That interlock system would then halt the driver from operating the vehicle if their BAC was deemed too high by the system. While this approach is effective, it is quite intrusive, requiring the user to not only take out and use the breathalyzer on each ignition, but for sanitation's purpose, also clean it. Dubey et al. [8] resolved this intrusion issue by using a different BAC sensor than a breathalyzer, called an alcohol gas MQ3 sensor [9]. This sensor tracks the alcohol in a region of air and does not require the user to blow specifically into or near it. Unfortunately, since this method tests air across the whole vehicle, any passengers with a high BAC would trip the sensor as well, preventing the vehicle from starting even though the driver is sober. Abdullah et al. [10] use sensors in the steering wheel of a vehicle to test metrics such as the operator's pulse and grip strength. These metrics are then used to determine not just the driver's sobriety but can also be used to sense whether or not the driver is drowsy, or experiencing road rage. Being able to test so many criteria at once comes with a drawback, however, as pulse rate differs wildly between people, and grip strength can be altered by a number of different physiological factors. To remedy this, they use Artificial Neural Networks (ANN) to process data gathered from the sensors, allowing for the unit to be generalized across a variety of users. Kaur et al. [11] uses facial recognition software to interpret the driver's psychological state from a live video feed. With this method, the system can constantly monitor the driver, making decisions about their state while they drive. To take full advantage of this, the authors proposed system sends messages to the driver based on their state, telling them what the system believes their psychological state to be, as well as suggestions on how to improve. This method, however, gives the driver a multitude of distractions while driving, making for an overall more dangerous drive.

We separate ourselves from the above methods by making use of HGN, a method of testing for intoxication by examining the reactions of a subject's eyes while tracking an object, as well as by implementing machine learning algorithms such as Random Forest (RF) in data processing [12]. HGN is already used by law enforcement to determine probably cause for a breathalyzer after a suspect has been pulled over. A calibrated breathalyzer, after arrest, can be used to determine the BAC level of the driver and readings at or above 0.08 are considered per se evidence of intoxication [3]. In contrast to the court's use, we use the HGN test before the car is on the road, using the test as a prevention mechanism rather than a means to establish guilt. The Random Forest method involves the creation of many different and randomized decision trees, and then collecting their individual results into a single decision, hence the "forest". Researchers such as Callao et al. [13] have used RF in Driving While Intoxicated (DWI) prevention before as a way to analyze the effectiveness and placements of random BAC tests and has proven it to be an effective way to process data. Its largest advantage is its ability to not only compare results to a baseline, but also compare similarity between results. RF models also solve the issue of over fitting and noisy data present in features, as a single decision tree can be over tuned to training data, making it ineffective when put in practice.

III. System Components

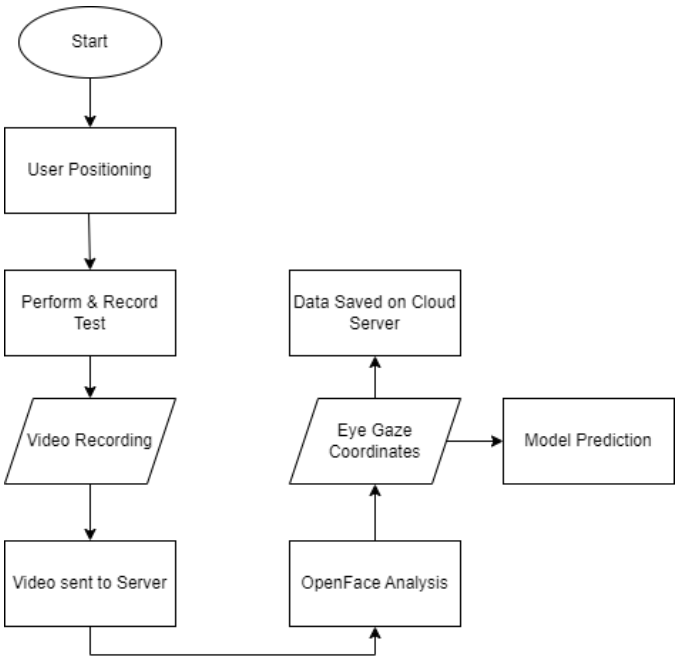

Fig. 1: HGN Test Flowchart

A. System Overview

The HGN Testing system we have developed is based on the work from [14]. However, we have integrated several components to ensure accurate and standardized testing for detecting

physiological impairment. Overall, it is an application that utilizes a display screen, camera, and an internet connection. At the forefront is our front end user interface, which employs the face-api JavaScript library for real-time face detection and positioning verification. Through visual feedback, participants are guided to align themselves correctly within the camera frame, ensuring consistent test conditions. Once positioned, the test begins with users tracking a moving square stimulus while their eye movements are analyzed. The recorded video is then securely transmitted to our cloud-based backend infrastructure, hosted on Amazon EC2 servers, where the OpenFace toolkit extracts facial features from the video for further analysis. Lastly, a Random Forest classifier processes the feature data to discern patterns indicative of impairment. A flowchart of the test is shown in Fig. 1

B. User Interface Frontend

1) Positioning: Before administering the HGN test, it is imperative to ensure that the participant's face is correctly positioned within the camera frame to obtain accurate eye gaze data [3]. To achieve this, we leverage the capabilities of the face-api JavaScript library, a powerful tool for face detection and recognition in real-time video streams. Using face-api, our system initiates a pre-test positioning verification process, wherein the participant's face is continuously monitored in real-time as they approach the camera [15]. Upon detecting the presence of a face within the camera frame, the face-api library assesses the positioning and orientation of the face relative to our predefined guidelines for optimal test conditions.

Once the participant's face is detected, the system prompts them to adjust their position if necessary, providing visual feedback in real-time to guide them towards the correct alignment. This feedback may include on-screen instructions or visual cues overlaid on the video feed, directing the participant to center their face within the frame and maintain a consistent distance from the camera.

The pre-test positioning verification process continues until the participant's face meets the specified criteria for proper alignment, as determined by the system's predefined thresholds and guidelines. Once the participant's face is correctly positioned, the system signals that the test can commence, automatically transitioning to the next phase of the HGN test. This seamless integration of face-api ensures that the HGN test is administered under standardized conditions, minimizing variability in eye gaze data and enhancing the reliability of the subsequent analysis performed by our machine learning model [15].

2) Testing: Our testing protocol begins with users removing any glasses they may be wearing to ensure optimal visibility and accuracy during the HGN test. Subsequently, users are prompted to initiate the test and position themselves correctly within the camera frame, guided by real-time visual feedback provided by our system. Once positioned correctly, the test commences with the presentation of a square stimulus that moves horizontally across the screen. The square starts at the left edge of the screen and gradually traverses to the right in a steady motion, pausing momentarily at the right edge before concluding the test as shown in Fig. 2. Throughout this process, users are instructed to maintain their gaze on the moving square, allowing our system to capture and analyze their eye movements for signs of physiological impairment. Upon completion of the test, the recorded video footage is automatically transmitted to our backend server for storage and further processing. This standardized testing protocol ensures consistency and accuracy in assessing users' sobriety levels, facilitating reliable detection of impairment through the HGN test.

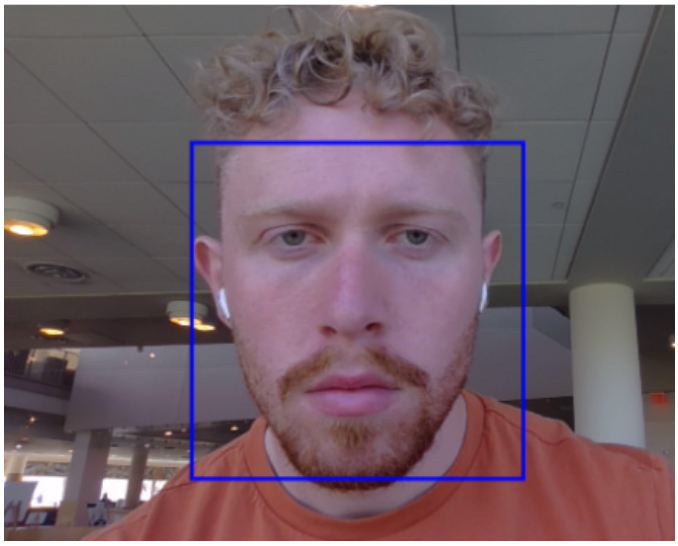

Fig. 2: HGN Test

C. Cloud Service Backend

For efficient and secure processing of user-generated videos and extraction of features for analysis, our HGN system leverages a cloud-based backend infrastructure deployed on Amazon EC2 (Elastic Compute Cloud) servers [16]. Upon completion of an HGN test, the system initiates the transfer of the recorded video footage to the backend EC2 server, where it undergoes processing using the OpenFace facial behavior analysis toolkit [17]. OpenFace is employed to extract relevant facial features and dynamics from the video, providing valuable data for subsequent analysis and classification of physiological impairment.

Following feature extraction, the resulting data, including the extracted features in CSV format and the original video file, are then securely stored in an encrypted Amazon Simple Storage Service (S3) bucket [18]. This encrypted storage ensures the confidentiality and integrity of the sensitive data, safeguarding it against unauthorized access or tampering. By utilizing Amazon S3's robust encryption mechanisms,

we uphold stringent data privacy standards and compliance requirements, thereby fostering trust and confidence in the security of our system.

D. Random Forest

Random Forest (RF) is a popular machine learning algorithm renowned for its versatility and robustness in handling classification and regression tasks [19]. It operates by constructing a multitude of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Each decision tree is built on a randomly sampled subset of the training data and a random subset of features, ensuring diversity among the trees. When making predictions, RF aggregates the outputs of these trees, which collectively offer more accurate and stable results compared to a single decision tree. This approach mitigates the risk of overfitting while maintaining high predictive performance [13]. In the context of our research, the Random Forest model plays a pivotal role in processing the data obtained from the HGN tests. By leveraging Random Forest, we can effectively analyze the HGN responses of individuals, compare them to established baselines, and discern patterns indicative of impairment with a high degree of accuracy.

E. OpenFace Library

OpenFace is an open-source facial behavior analysis toolkit that facilitates robust and efficient feature extraction from facial images [17]. Leveraging deep neural networks, OpenFace excels in capturing intricate facial dynamics and subtle cues that are indicative of physiological impairment, such as those observed during HGN tests as shown in Fig. 3. By employing OpenFace in our project, we can accurately extract facial features and dynamics from individuals undergoing HGN tests, enabling us to quantify and analyze their responses with precision. This feature extraction process serves as a crucial component in our system's ability to assess a user's sobriety effectively and confidentially, thereby enhancing its overall reliability and performance in promoting road safety.

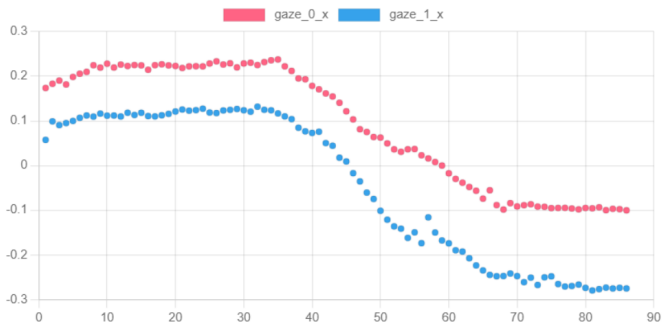

Fig. 3: OpenFace gaze coordinates for control test
Graphed x coordinates of the left eye's (blue dots) and right eye's (red dots) gaze coordinates for each frame of the video.

F. User Profiles

While the process of user profiling is not currently integrated within our system, we have conducted independent tests to demonstrate its potential for effectively managing user data and assessing participant conditions. Using the OpenCV library [20], we explored how user profiles could be created when a participant's face is detected within the camera frame. In optimal conditions, with proper lighting and facial expressions, our testing showed that capturing and analyzing facial features could be accurately accomplished in a single trial.

Although not implemented, our tests indicated that once established, these initial profiles could serve as a baseline for subsequent comparisons during theoretical operational use. As participants would undergo subsequent tests, their facial characteristics could be continually compared to their stored profile to detect any changes or deviations. This capability is envisaged to assess sobriety levels by comparing a participant's current facial features against their average and typical results within the profile. Significant deviations from this baseline, if observed in practice, might indicate potential intoxication, prompting further evaluation or intervention.

IV. Evaluation Method

A. Testing Protocol

During the Institutional Review Board (IRB) study (protocol ID: 23-3977), each of the 23 participants underwent a comprehensive testing regimen designed to evaluate the efficacy and accuracy of our HGN system under various conditions. Each participant recorded a total of thirteen HGN tests, comprising five sober control tests, five tests where they deliberately induced erratic eye movements while tracking the square's position, and three tests involving induced nystagmus. The five sober tests served as baseline measurements, allowing for the establishment of normal eye movement patterns and facilitating comparison with the next five failing test results. Furthermore, the final three tests involving induced nystagmus, achieved by spinning participants in a chair ten times, simulated conditions where nystagmus occurs naturally due to physiological factors such as alcohol intoxication or vestibular dysfunction [21]. Notably, we conducted only three spinning tests as participants may develop increased tolerance or resistance to spinning effects with repeated exposure. By subjecting participants to these diverse testing scenarios, our study aimed to comprehensively evaluate the robustness and reliability of the HGN system in detecting and quantifying impairment across different contexts and conditions.

B. PreProcessing

Raw eye gaze coordinate data extracted from the frames of the video presents challenges in its direct applicability for classification tasks within our Random Forest model [12]. Due to the inherent variability and noise present in the eye gaze coordinates across frames, direct utilization of this raw data can lead to suboptimal model performance. To address this issue, we implement a preprocessing step aimed at extracting

meaningful features from the raw eye gaze data. Specifically, we segment the data from each test sample into three distinct sections for each eye, reflecting different phases or intervals within the HGN test as shown in Fig.4. We then compute both the mean squared error for each section and the summed absolute difference between each value and it's neighbor, quantifying the deviation of the eye gaze coordinates within that section. These processed values effectively encapsulate the variability and patterns present in the eye gaze behavior throughout the test duration, thereby serving as informative features for our Random Forest model. By employing this preprocessing strategy, we transform the raw eye gaze data into a structured and interpretable format that enhances the discerning power of our model, ultimately enabling more accurate classification of physiological impairment based on HGN responses.

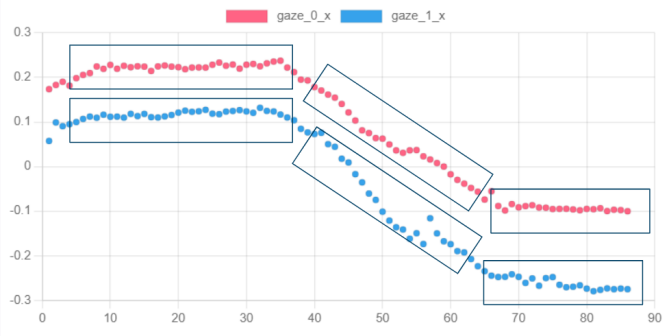

Fig. 4: Visualization of preprocessing sectioning

C. Machine Learning Model

To train and evaluate the Random Forest model for detecting impairment based on HGN responses, we transformed the testing data into a binary classification problem of pass and fail [12]. The first five sober tests recorded by each participant were assigned a label of 1, representing successful completion of the test under normal conditions. Conversely, the subsequent five tests involving deliberately induced erratic eye movements and the three tests involving induced nystagmus were labeled as 0, signifying impairment or abnormal eye behavior. This binary classification scheme enabled the Random Forest model to learn to distinguish between normal and impaired eye movement patterns, facilitating the accurate identification of impairment during the HGN test [19]. By training the model on this labeled dataset, we aimed to develop a robust and reliable classifier capable of effectively assessing a user's sobriety based on their HGN responses.

1) K Fold Validation: To properly assess the accuracy and generalization capability of our machine learning model, we employ k-fold cross-validation with a value of k set to 10 [22]. This technique involves partitioning the dataset into k equal-sized folds, where each fold serves as a validation set while the remaining k-1 folds are utilized for training. Subsequently, the model is trained and evaluated k times, with each fold taking turns as the validation set. By iteratively rotating the validation set across all folds, we ensure that every data point is used for validation exactly once, thereby minimizing the risk of bias and overfitting. Furthermore, the use of k-fold cross-validation provides a robust estimation of the model's performance metrics, including accuracy, precision, recall, and F1 score, across multiple iterations. The average performance metrics computed over the k iterations serve as reliable indicators of the model's effectiveness in discerning impairment based on the HGN responses. Through this rigorous validation approach, we can confidently ascertain the model's accuracy and its ability to generalize to unseen data, thereby bolstering the credibility and reliability of our proposed system for enhancing road safety.

2) N Estimator Selection: After evaluating the performance of our Random Forest model across various numbers of estimators, we found that the accuracy of the model stabilized at an estimator count of 75, as depicted in Table I.

TABLE I: Random Forest N Estimator Selection

Estimators	10-K Fold Accuracy
2	72.09%
5	77.87%
20	80.41%
35	80.45%
75	81.35%
100	81.05%

V. Results

A. Model Accuracy

The results presented in Table II highlight the model's effectiveness in distinguishing between various testing scenarios. Particularly, the model performs robustly in identifying impairment during control tests where the subjects were asked to track the square normally and as accurately as possible, as well as those with deliberately induced erratic eye movements that shift quickly from side to side of the square or have trouble following it across the screen, achieving a high level of accuracy. However, when faced with the added complexity of dizziness-induced nystagmus generated by spinning the subjects quickly in a chair, the model's accuracy diminishes, indicating some challenges in accurately identifying impairment under these conditions. Despite this, the model still demonstrates a notable capability to differentiate between impaired and non-impaired states across diverse testing conditions, emphasizing its potential value in enhancing road safety through early detection of impairment.

TABLE II: Model Testing Accuracy

Testing Scenario	(10) K Fold Accuracy
Control, Erratic, & Spinning	79.76%
Control & Erratic	90.04%
Control & Spinning	70.94%

B. Preventing False Positives

Our system aims to cut down the number of drunk driving fatalities, however our base model has a chance to impede

users who are sober. Achieving 100 percent accuracy in preventing false positives, where individuals who are sober do not fail the test, requires a nuanced approach that accounts for the intricacies of the HGN responses and the underlying features driving impairment detection. One effective strategy we employed involves refining the model's decision boundary to maximize specificity without sacrificing significant sensitivity. By adjusting the threshold for classifying responses as impaired, we can minimize the likelihood of misclassifying sober individuals while maintaining the ability to reasonably identify impairment when present. The Confusion matrix in Fig.5 depicts the performance of the tailored Random Forest model in preventing false positives. The tailored model was constructed using 90% of the available data samples and evaluated on the remaining 10% of unseen data samples.

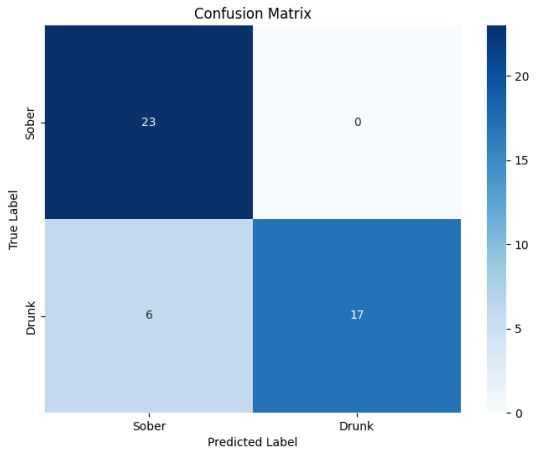

Fig. 5: Tailored Model Confusion Matrix

VI. Conclusion

In this paper, we introduced a passive transmission interlock system using the HGN test to determine driver impairment. Our model effectively distinguishes between users passing and failing the HGN test under both normal conditions and induced impairment scenarios, achieving high accuracy rates.

However, challenges arose in classifying impairment during tests with induced nystagmus, primarily due to individual differences in susceptibility to spinning-induced nystagmus. Some participants showed pronounced symptoms of impairment, while others had minimal effects, complicating accurate impairment classification.

This variability highlights the complexity of using physiological measures like HGN for impairment assessment, where factors such as vestibular sensitivity significantly influence outcomes. Our model's performance during spinning tests was affected by these individual differences.

Moving forward, further research and refinement of our model are necessary to account for individual variability and improve its accuracy across diverse populations. Strategies like incorporating additional features or adjusting decision boundaries based on individual characteristics could enhance the model's performance. These improvements will help increase the system's reliability in promoting road safety and preventing drunk driving incidents.

References

[1] T. Stewart, "Overview of motor vehicle traffic crashes in 2021 (report no. dot hs 813 435)," National Highway Traffic Safety Administration, Tech. Rep. 813435, APR 2023. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813435

[2] R. B. Voas and J. C. Fell, "Preventing impaired driving opportunities and problems," *National Library of Medicine*, 2011.

[3] N. T. L. Center, "Horizontal gaze nystagmus: The science and the law: A resource guide for judges, prosecutors and law enforcement," National Highway Traffic Safety Administration, Tech. Rep., 2021.

[4] J. Griffin, "Legislative history of drunk driving in america," *Center for Alcohol Policy*, 2021.

[5] "Home," Mar 2023. [Online]. Available: https://madd.org/

[6] R. Voas, "Enhancing the use of vehicle alcohol interlocks with emerging technology," *National Institute on Alcohol Abuse and Alcoholism*, 2014.

[7] N. S. K.A., H. V.C., H. B., and R. Hegde, "Low cost breath analyzed ignition interlock system," *2018 4th International Conference for Convergence in Technology (I2CT)*, OCT 2018.

[8] A. K. Dubey, D. K. Singh, R. Kumar, P. K. Das, and V. Yadav, "Smart automobile ignition locking system to avoid accidents due to alcohol consumption." *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*, NOV 2023.

[9] S. Hu, "Gas sensor (mq3)," Seeed Studio, Tech. Rep., 2023.

[10] C. Chengjun, "Vehicle safety sensory system," *UCP Journal of Engineering & Information Technology*, OCT 2023.

[11] J. Kaur and M. Elshakankiri, "Intelligent impaired driving detection and road emergency services system," *2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, DEC 2018.

[12] C. Vens and F. Costa, "Random forest based feature induction," in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 744–753.

[13] L. Ramon Callao, "Using tree-based models to predict drunk driving at police preventive checkpoints," Ph.D. dissertation, UPC, Facultat de Matemàtiques i Estadística, Jun 2016. [Online]. Available: http://hdl.handle.net/2117/88462

[14] C. Coleman, M. Jenkins, W. Roberts, C. Thomas, W. Westerkamp, R. MacDonald, and A. Salman, "Horizontal gaze nystagmus transmission interlock system," in *2023 Systems and Information Engineering Design Symposium (SIEDS)*, 2023, pp. 43–48.

[15] T. Severien, "Face detection on the web with face-api.js," JUN 2021, accessed: 2024-04-03. [Online]. Available: https://github.com/justadudewhohacks/face-api.js

[16] G. Juve, E. Deelman, K. Vahi, G. Mehta, B. Berriman, B. P. Berman, and P. Maechling, "Scientific workflow applications on amazon ec2," *2009 5th IEEE International Conference on E-Science Workshops*, DEC 2009.

[17] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[18] Z. Daher, , and H. Hajjdiab, "Cloud storage comparative analysis amazon simple storage vs. microsoft azure blob storage," *International Journal of Machine Learning and Computing*, vol. 8, pp. 85–89, FEB 2018.

[19] S. Hegelich, "Decision trees and random forests: Machine learning techniques to classify rare events," *European Policy Analysis*, vol. 2, pp. 98–120, MAR 2016.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] W. W. Krueger, "Method to mitigate nystagmus and motion sickness with head worn visual display during vestibular stimulation," *Journal of Otolaryngology-ENT Research*, vol. 7, JUN 2017.

[22] T.-T. Wong and P.-Y. Yeh, "Reliable accuracy estimates from ¡i¿k¡/i¿-fold cross validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, pp. 1586–1594, AUG 2020.