Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 10 (October) (2025)

Detection of Alcohol, Drug, and Sleepiness Conditions through Iris Behavior Curve Analysis

Arslan Ali (Corresponding author)
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan Email: arrsssal@gmail.com

Areej khan
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan Email: areejkhan762000@gmail.com

Naeem Aslam
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan Email: naeem.aslam@nfciet.edu.pk

Hira Saleem
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan Email: hira.saleem@nfciet.edu.pk

ABSTRACT
This paper provides an extensive analysis of iris behavior analysis as a means of determining fitness-for-duty by identifying impairment stemming out of alcohol, drugs and sleep deprivation. The system utilizes near-infrared image sequences to obtain biometric characteristics whereby the pupil and iris radii, their ratios and spatial positions are obtained and indicate substantial statistical variations between the control and impaired groups. The power of these features is confirmed by Kruskal-Wallis and Dunn tests, where the H-statistics are over 1200 of the important variables and p-values are below 0.001 in most the group comparisons, which prove that the selected biometric markers are discriminatory under different conditions. Models of classification such as the Random Forest, Gradient Boosting and the Multi-Layer Perceptron are reported to be having excellent accuracy up to 75.5 percent with the sensitivity and specificity of the model often being above 70 percent and 95 percent respectively. It is interesting to note that binary classification with fit and unfit states consistently has better performance and often has a sensitivity over 80% to detect unfit persons and therefore the practical use of the system in the occupational safety field. The non-invasive nature of the approach used by the system allows continuous monitoring without interrupting the work of operators: driving or piloting, which is why there are only fewer cases of sleep and drug impairment affecting the sensitivity of the classification in these groups. Segmentation can also be affected by image quality and limitations of NIR sensor. Future directions in work will aim to add to the data set with attention paid to underserved conditions, and the incorporation of deep sequential models like LSTMs to better model the temporal dynamics of biometric. Additional refinement of image processing, sensor strength, and ability to execute it in real-time will increase accuracy as well as application utility Such developments have scalable proactive safety evaluation potential, especially in safety-critical sectors that use biometric indicators to conduct real-time detectors of impairments.

Keywords: The Biometrics Of Iris, Fitness To Duty, Alcohol And Drug Testing, Sleep

Deprivation, Near-Infrared Imaging

Introduction

As a result of present-day twenty-four-hour society, much of the modern workforce is employed in the services of major sectors such as safety of its citizens, health care, and economy by working twenty-four-hour shifts. According to some recent studies, approximately 15-25 percent of the workforce performs in a shifted manner and with irregular schedule and some of them during the night. Night and rotational shifts have always been linked with the high risks of indulging in drug and alcohol, which are key antecedents of accidents in workplace in different countries like United States, Australia, United Kingdom and Japan [1]. It is estimated that nearly 15 percent of the workers in America have a substance use problem in the workplace, 10.9 percent of the workers claim that they have worked under the influence or hungover. Other Australian government studies have revealed that, 10 percent of regular schedule workers and 13 percent of shift workers consume alcohol in a manner that leads to short time harm. These are the ones that are generalized in the world regions. The inability of employees to be physically fit as a result of substance use and sleep disorders makes them less productive which poses the risk of accidents that lessen the economy with massive costs. Alcohol and drug abuse disorders increasingly affect physiological systems throughout the whole world, and the deterioration of the situation during the COVID-19 pandemic due to health and economic crises impacts the systems. According to the National Institute on Drug Abuse, the overall cost of alcohol, illegal substances and tobacco including lost productivity, criminal and medical care is between 600 billion and 740 billion in the US alone. The fatigue related expenses are also added to the list that adds to the loss of about 136 billion annually by the employer.

The effect of sleep deprivation on cognitive ability is very negative and particularly on sustained attention and memory which is linked to increased risks of industrial accidents and driving disasters. The OECD (2020) Road Safety Report suggests that fatigue and drowsiness cause 22 out of 100 injury crashes, and is projected to cause fatal crashes by 30 per cent, and WHO estimates suggest the loss of fatigue-related incidents is approximately 3 per cent of the GDP in most countries [2]. Given these critical impacts, there is need to implement preventive activities against drug addiction, alcoholism and work-related sleepiness. The existing interventions that are branded as Fitness for Duty (FFD) analysis systems are destined to assess the mental and physical readiness of the workers to perform job duties that are safe and effective. However, the prevailing FFD tools have a propensity of taking into consideration the individual impairment factors, but they do not incorporate the wide-ranged identification of overlapping causes, thus cannot properly eradicate the risks of the hazards in the workplace [3].

Research innovations revolve around the importance of the analyzed eye movements as bio-markers in order to get access to information regarding cognitive load and state of impairment. Specifically, Central Nervous System (CNS) regulates automatic reactions of the iris and pupil which are immediately altered by an exposure to light, consumption alcohol, drugs or even tiredness. It is also through iris recognition technology that the identification of a particular worker is done and therefore contributes to the fortification of impersonation and whereby the measurement of impairment is linked to the correct individual. On this fact, iris and pupil are valid objective data of work-fitness determination. There is therefore an imminent need to develop an automatic and dependable FFD model that will be based on the iris recognition framework. This model may assist in performing non-invasive, fast and multi-factorial impairments assessment,

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)
which will contribute to the workplace safety and efficiency [4].

Background

Based on the recent literature regarding iris-based alcohol detection system, the current technological advances are described by an impressive enhancement of the hardware and algorithms, and the principal extractors of features are Gabor filters. Navarro et al. (2016) proposed a system that captured the iris images with assistance of CCD cameras, run the image through Gabor filters and compared any change in the size of the pupil to the baseline values in order to establish the degree of intoxication [5],[6]. The method employs a software-hardware integrated system where the algorithms of MATLAB are employed to emulate real-time recognition to switch ignition control of the vehicles based on the alcohol detection. Tapia et al. (2022) establish this idea by capturing periocular NIR images and behavioral curves with reference to the alcohol influence, and are applying deep-learning computer neural networks like CNNs and LSTMs to classify the data, which comes after this communication. They indicate that they are more interested in the importance of multi-session data collection that can make them more resistant to variability as a result of pupil constriction or dilation that can compromise the detection accuracy [7]. Other biometric indicators like examining of pupillary light reflex (PLR) as alcohol sensors that are not invasive are researched and their detection rate is approximately 85. They include feature-based segmentation of images on the basis of maximum mydriasis, latency and contraction time and classify the responses in support vectors machines or k-nearest neighbor. Deep neural networks and optimization techniques have been employed as part of creative solutions to improve the selection of features and maximize the accuracy at which the alcohol effects are known. Wearable technologies used in alcohol detection are also on the rise with the adoption of non-invasive, pupil and iris-based biometric-based technologies to formulate real-time alcohol detectors and other physiological indicators of alcohol consumption, including heart rate variability are being integrated [8]. The systems are set to allow continuous measurements and outputs with few inconveniences and would be more trustworthy and applicable in actual life applications because of the recent research of iris recognition in alcohol detecting, which demonstrates high reliability and accuracy, and the fact that the new technology would not be invasive. These systems have a foundation on the basis of advanced image processing, deep learning, and biometric analysis methods within the current trends of research [5].

Commercial devices

Numerous Fitness for Duty (FFD) algorithms is developed and introduced into various commercial solutions used to assess the capability of an individual to perform his or her professional duties safely with particular consideration to the limitation posed by fatigue, alcohol and drugs. The primary examples of such devices are PMI FIT2000, Sobereye, and Optalert, and they are distinguished by the use of diverse technological applications and take into account numerous types of impairments. PMI FIT2000 is a system that utilizes application of physiological parameters involved in determining physiological conditions that are associated with exhaustion and substance impairments [9]. It is a fast test (1 minute) of variables (dilat of pupil, constriction of pupil, latency and saccadic velocity) and is compared to a baseline of that particular individual. Although it is an acceptable measure of the levels of impairment, it is not applicable in enabling biometric recognition to ascertain the identity of a subject under consideration.

Sober eye is a wearable product with the focus on Pupillary Light Reflex (PLR). The subject is made to close his eyes a minute or so by the system so that his pupil could be

dilated and then subject to a light stimulus. The smartphone is subsequently equipped with a camera that records high quality videos in order to gauge the dynamics of the response of the pupil. The instrument is based on the baseline established after a few days and can be considered to offer consistent detection of deviation that shows the possibility of impairment e.g. substance use or tiredness [10]. It is robust on the non-invasive monitoring of the real-time alterations of the PLR and supplying the iris recognition technology to verify whether the tests are accurate by confirming the individual. However, the requirement to gain user cooperation and the process of calibration of the baselines could limit its expediency and its flexibility in the workplace.
The Optalert is a special kind of FFD devices, which rely on the IR Reflectance (IR) oculography. It is able to gauge the rate and the time of the blinking eyelids by placing the infrared sensors on the spectacles frames used to gauge the drowsiness of drivers. The system generates a Johns Drowsiness Scale (JDS) score at an audio frequency after every minute and audio alerts in the event of the fatigue levels going to a critical level, which may lead to accidents. Its non-invasive and continuous technique of tracking is most effectively applied in transportation and industry. It should however have a calibration base and can only be applied to locate sleepiness and not necessarily assess other impairment areas like alcohol or drug intoxication [11],[12].
Cognitive alertness devices such as AlertMeter used as an addition to these hardware systems are a fast graphical measure of hand-eye coordination, reaction time and response time. Such computer-based tests provide a unique feedback information that is sensitive to circadian rhythms that can influence cognitive performance through comparison of daily test outcomes against individual baselines [7]. These techniques are good in mass thinking, but they are dynamic and therefore, could be restricted in their use in a non-monitoring environment. Moreover, the wearable health devices such as Fitbit, Jawbone, and Apple Watch have also made it possible to find physiological signs of fatigue at the workplace like heart rate variability and physical activity that are the primary indicators of fatigue. Continuous non-invasive data can be provided is based on the use of wearables with photoplethysmography and accelerators, which can facilitate fatigue monitoring, yet it often necessitates some other types of assessment in determining impairment [13],[4].
The method of FFD monitoring is introduced in the area of brainwave analysis with such innovations as the SmartCap LifeBand. The latter wearables can provide more precise data on fatigue levels as they will record the frequencies of brain EEG waves and other ratios between the state of wakefulness and the state of drowsiness. These devices leverage the power of signal processing and established algorithms to give fatigue scores to assist in the timely interventions. Nevertheless, certain drawbacks in popular use are associated with the price, comfort in the use of users, and the wide-range validation in other environments of work [14].
Despite the advantages, there are several major weaknesses to the existing FFD systems. Most of them are limited to identifying one factor of impairment at a time, fatigue, alcohol or drugs, and do not assess it in a multi-factorial fashion. Their reactive nature means that intervention occurs when the impairment is under progress, endangering their safety. In addition, majority of systems cannot be operated without the contribution of a user such as closing eyes or conducting tests not always available or valid. The reliance on the calibration of the baseline opens the baseline threats like manipulation or falsification of the outcome [15]. Moreover, absence of robust biometric validation of some systems introduces into consideration doubts on integrity of tests and impersonation of the subjects.
Our answer to the above problems is a new framework of FFD; this is founded on the study of behavioral curves of the near-infrared (NIR) iris pictures. Such a design enables

checking the alertness and fitness without any contacts, mobile and quick, and provides an intuitive interface with an inbuilt identity verification feature through iris recognition. Unlike the conventional systems that involve chemical or physical tests to determine the amount of alcohol mixed in the blood or breath, our system involves CNS response to the drug usage, alcohol consumption or fatigue by analyzing the dynamic changes in the iris and pupil [16]. The methodology further extends the FFD evaluation past the individual sources of impairment providing a multi-factorial approach of the individual readiness.

The main contributions of the study are the use of NIR iris-based contactless assessment as the first option to evaluate FFD and provide a rapid evaluation without patient discomfort or major behavioral disturbance. We identify behavioral trends that reliably indicate CNS status and impairment by examining the changes in the pupil and iris diameter throughout the time and using advanced machine learning models to identify their trends [17],[18]. Portability and anti-impersonation features of the system lead to increased practicality in a wide range of workplace environments, which facilitates active safety management. Moreover, we created a special annotated library of NIR periocular image data under different impairment conditions that will enable us to fully train and validate a model. Such developments all make our system a next-generation FFD tool, that will offer better detection accuracy, operational viability and full safety guarantees [19],[20].

Subsequent sections expound on the developed NIR iris image database, the finer algorithmic methodology contextualized within FFD detection, and prolific experimental evidences of efficiency. Finally, our work represents a massive leap in occupational safety technology with a combination of the state of the art in biometric analysis and the real-world practicality to protect health and performance in safety critical occupations.

Database

The development of the valid and complete database was a resource-demanding and difficult part of this research. We created the database of the near-infrared (NIR) iris images sequences database (FFD-NIR-Seq), which contains 10-second streams of near-infrared (NIR) iris images, and the protocol was approved by the ethical committee of the University of Chile. A variety of devices were used to record video to increase the richness of the data, including binocular set of periocular and monocular single-eye NIR, the Iritech MK2120UL, iCAM TD-100A, Iritech Gemini, and Gemini-Venus. In every image, one covered major part of the eye- the pupil, iris and sclera. When acquiring the data, the subjects were in front of the capture device, which automatically triggered eye capture and data recording. The database has four different subsets with different subject conditions such as Control (participants not under the influence of substances and well-rested), Alcohol (participants under the influence of alcohol) and Drugs (participants who mostly use substances such as cannabis or prescribed psychotropic drugs), and Sleep (participants that sleep deprived by work-related sleep disorders). This heterogeneous data base facilitates accurate behavioral analysis used to train and test the model that covers the impairment states that are important in Fitness for Duty assessment in an efficient and controlled way.

Alcohol consumption

Regarding the alcohol database, the participants underwent the subsequent protocol:

The initial NIR picture series acquisition took place at time 0 (before drinking).

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

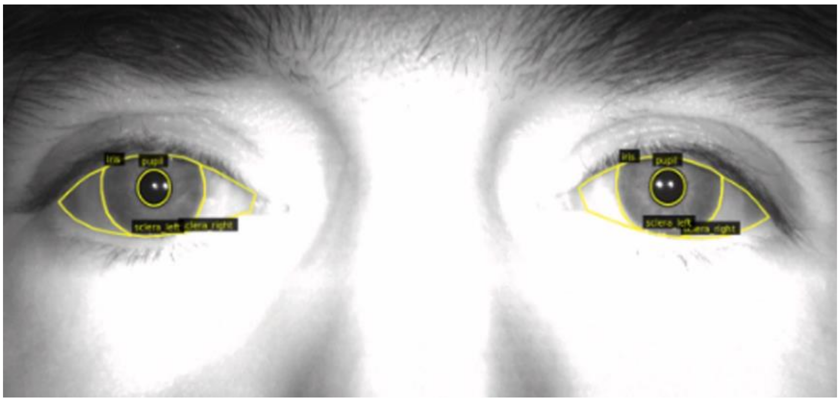
For up to 15 minutes, each volunteer consumed 200 milliliters of alcohol. The second acquisition was carried out 15 minutes after the first one, or right after the alcohol consumption was completed. Thirty minutes after time zero, the third acquisition was made. Forty-five minutes after time 0. 6, the fourth acquisition was made. Sixty minutes after time zero, the fifth purchase was finally made. As a result, one sequence of control photographs and five sequences of images showing the person while intoxicated were captured.

Drugs Consumption

The United Nations Office on Drugs and Crime's World Drug Report, 2021, states that cannabis8 is the most often consumed crop globally, with an annual prevalence of 15%. This is followed by pharmaceutical opioids and tranquillizers, which have annual prevalences of 5% and 2.5%, respectively. Because of this, data from our database indicate cannabis ingestion in almost 95% of cases. On the other hand, tranquilizers and more sophisticated narcotics (such as heroin and ecstasy) make up the remaining 5%. For the purpose of the drug database acquisitions, the volunteers were drug users, and the image recordings were made at least half an hour after the initial drug ingestion.

Sleep Conditions

For the sleep database, a specific image capture procedure was specified, where experiments were conducted in a controlled sleep deprivation environment. These recordings were made of a certain group of people who were exposed to varying degrees of sleep deprivation in order to assess the degree of weariness or drowsiness at various intervals. A smart band was used to track the volunteers' sleep patterns, both in terms of quantity and quality. Participants were arranged as follows:
Irregular sleep patterns
Sleep duration of less than 3 hours
Three to six hours of sleep each night
Sleeping more than six hours (normal sleep)
Volunteers conducted three daily image acquisitions during the recording season: (i) at the start of the working day, (ii) after lunch, and (iii) at the end of the working day. FFD-NIR-Seq contains fifteen hundred eye-disjoint pictures. 150 pictures are usually taken of each subject. That took ten seconds to complete. There were three categories for the image sequences: training, validation, and testing. The goal for the test set was to approximate 15%, which is the true percentage of unfit cases. Table 1 displays the



Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

Fig. 1: shows an example of a periocular NIR picture with labels. The picture displays both eyes together with the labels for the left and right sclera, pupils, and iris.

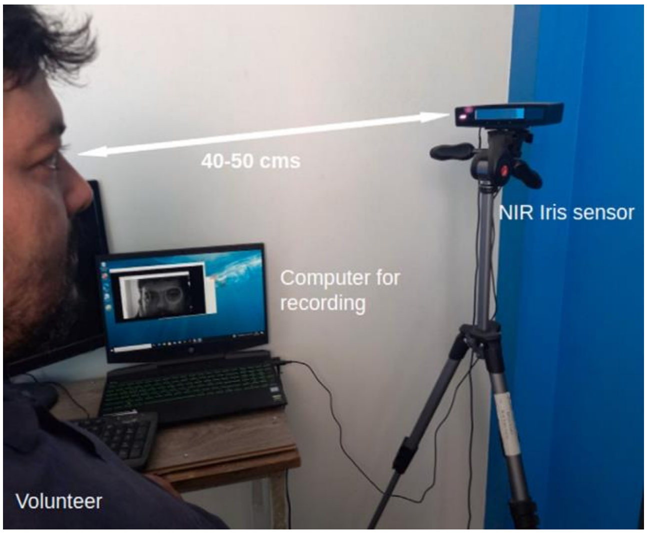


Fig. 2: Capture device description. The subject is placed 40–50 cm in front of the sensor. When the capture gadget notices eyes, it begins a 10-second recording.

Table 1: NIR image sequences by classes

Class	**Train set**	**Val. set**	**Test set**
Control	247	35	688
Alcohol	247	35	72
Drug	62	9	17
Sleep	69	9	20
Total	**625**	**88**	**797**

quantity of sequences. Over 150K photos across four classes were gathered in total. Examples of the obtained photos for each of the classes listed in the database are displayed in Fig. 4.
The object identification, segmentation, and final classification phases of the FFD model were all trained and validated using this database.

Methodology

The Fitness for Duty (FFD) model developed is an analytical cascade of four modules which are aimed at processing periocular near-infrared (NIR) images systematically. The eye detector module first detects and crops both eyes of input images producing sequences of single eyes to be analyzed further. The module utilizes the latest methods such as Cluster-Coordinated Net (CCNet) that performs semantic segmentation, classical tracking algorithms, and Eye-tiny-yolo that performs precise eye localization. Multiple Instance Learning (MIL) and Channel and Spatial Reliability Tracker (CSRT), are tracking methods that track the eye movement across the image sequence so that eye parts could be accurately cropped. After identifying the eyes, the iris and pupil segmentation module uses a semantic segmentation network that deploys a trained network using CCNet on aggressively augmented monocular NIR images to produce accurate masks as a representation of the pupil and iris. The pupil and iris centers and radii are determined using the localization algorithms which are fed with these masks and using the circles with optimal fit to the contours. Morphological erosion and logical XOR functionality enhance boundaries of segmentation, and finally mean square error criterion completes the most suitable circles. The obtained pupil and iris parameters in each frame are then fed into the feature extraction module which forms temporal behavior vectors. Lastly, the model categorizes the subjects into states of Fit/Unfit, control, drug, alcohol, and sleep-deprived, which allows performing full and automated FFD analysis.

Results

The Fitness for Duty (FFD) model was trained in a process of iterating on training and validation data sets with the final performance evaluated on a test set. The system also categorizes the subjects in 4 categories such as control, alcohol, drug and sleepiness. Statistical metrics show that the control and alcohol classes have the best performance, which is mainly because there were more images of the classes, whereas drug and sleep classes had fewer numbers, which affected the performance. The sensitivity and specificity were calculated using True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Control and alcohol categories had a sensitivity above 70% but reduced significantly to about 40% in drugs category and 25% in sleepiness. Control and alcohol groups had a specificity of between 75-80 percent, but specificity rose sharply in drug and sleep classes, perhaps because of imbalance in the classes. The uneven distribution of the dataset, which captures about 10-15 percent of subjects with impairment in real situations, is quite consistent with real-life scenarios, and thus, the model is suitable to be applied to practice despite the differences in the classes.

Two courses

The assembled bar chart shows how the three machine learning classifiers (Random Forest or RF, Gradient Boosting Machine or GBM and Multi-Layer Perceptron or MLP) performed in major evaluation measures (fit and unfit class prediction) of the three classifiers. It is worth noting that MLP has the best accuracy and sensitivity (75.3) with the closest RF and GBM with 70.8-73.1 and 70.1-73.1 accuracy and sensitivity respectively. GBM has the highest specificity (79.8%), but its accuracy (95.8%), as well as the accuracy of MLP and RF, is admirable (more than 94). This tendency supports the idea that ensemble and neural network models are quite powerful in the context of the robust classification task, but the differences in performance demonstrate the necessity to choose the suitable classifiers depending on the specifics of the dataset and the purpose of the task. Both high accuracy in all models and the gradual increases in specificity and sensitivity are good indications of effective reduction of class imbalance effects on model training and testing.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

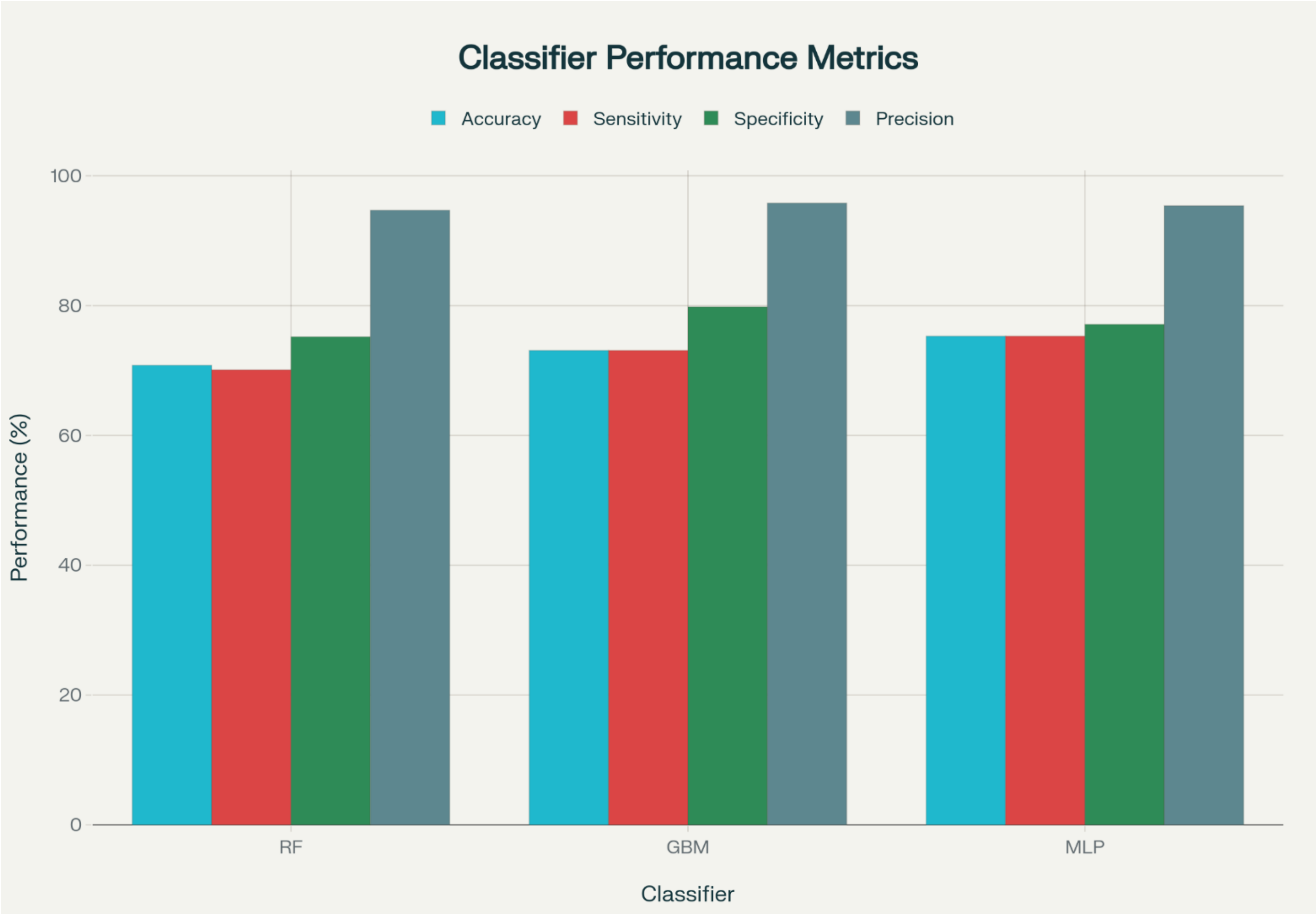


Figure 1: classifier performance Metrix

Discussion

As can be clearly seen in the table, all three classifiers, RF, GBM, and MLP, were adequately trained and the dataset had been constructed and divided to ensure a strong evaluation performance, which led to the similarity in the results in terms of performance measures. Sensitivity and accuracy represent the important indexes of trying to apply the system: they indicate the capacity of the system to acknowledge the relevant classes (fit/unfit) independently and give a general overview of the reliability of the modeling. RF gives control and alcohol classes sensitivity over 70% but its sensitivity and accuracy is low with drug and sleep classes given the misclassification of unfit cases that is prevalent in the analysis, though on the other hand, its performance can be enhanced when it is used to analyze fit/unfit groups. GBM is similar to the behavior of RF but with slightly higher scores of all metrics. MLP is better at control sensitivity and general performance, but not popular with decreased sensitivity in drug and alcohol groups. Such trends demonstrate the long-standing problem of discrepancy of classes, however, the successful determination of fit/unfit employees by increased sensitivity and accuracy suggest that the model is highly applicable to the practicable use of in the fitness-for-duty process with paying special attention to the security of the working environment.

Table 2: Performance metrics for RF, GBM, and MLP classifiers, showing sensitivity, accuracy, and class-wise results for Fit and Unfit states

Cla	Fit	Unfit	Overall	Control	Alcohol	Drug	Sleep

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

ssifier	**Sensitivity (%)**	**Sensitivity (%)**	**Accuracy (%)**	**Sensitivity (%)**	**Sensitivity (%)**	**Sensitivity (%)**	**Sensitivity (%)**
RF	70.1	75.2	70.8	>70.0	>70.0	Low	Low
GBM	73.1	79.8	74	>70.0	>70.0	Slightly better than RF	Slightly better than RF
MLP	75.3	77.1	75.5	75.3	70.8	29.4	~Same as others

Behavioral curves

The confusion matrices of the RF, GBM and MLP classifiers give us a scrupulous insight into the expertise of each model in differentiating between fit and unfit people. The three models all have high true positive rates of the fit class, with RF, GBM, and MLP having their results and predictions being accurate in most truly fit cases as indicated by the associated accuracy scores (0.726 in the case of RF, 0.765 in the case of GBM, and 0.762 in the case of MLP). The models are able to detect unfit cases with a success that is consistently high, but donning differences are observable in the number of false negatives and false positives among the classifiers. It is interesting to note that MLP has got the highest true positive rate whereas GBM has best balance of reducing misclassifications both in the fit and unfit subset. These findings support the high training and usefulness of all the three classifiers on the balanced subsets and the accuracy increase is explained by the careful feature engineering and hyperparameter optimization. Finally, the similarity in the performance underlines the effectiveness of the suggested FFD framework in the context of various machine-learning designs where the most important issue is to correctly classify both fit and unfit employees in real-life situational areas of operation.

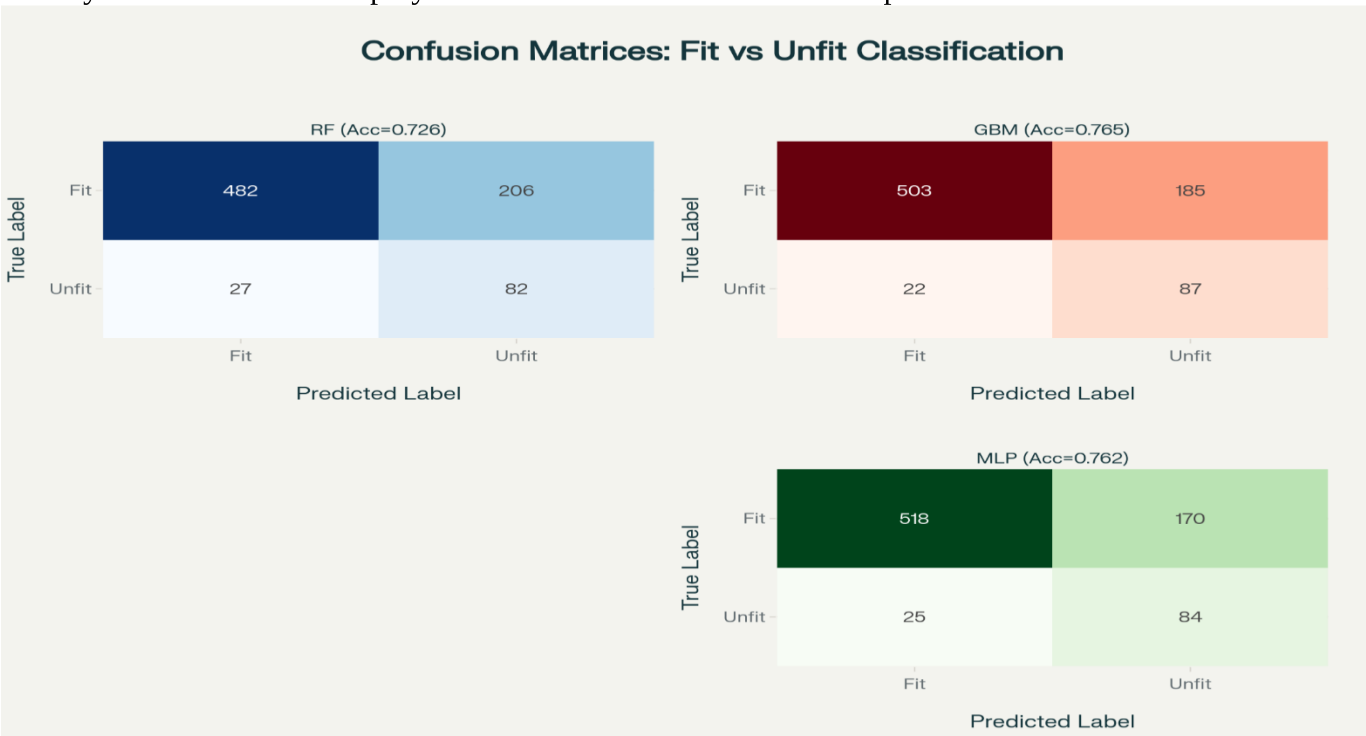


Figure 2: RF confusion matrix for Fit and Unfit classes

The plot group represents the time series dynamics of the pupil and iris radius of the four

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

different conditions Alcohol, Drug, Sleep and Control that provide detailed information about the dynamics of the class-specific changes in biometric variations through successive image acquisitions. It is also worth noting that the Alcohol group always has higher pupil radii (X and Y) than the other conditions, which is presumably due to the dilation caused by the substances. The Drug group, which is usually stable, depicts intermediate changes in radii, and however, the Sleep group exhibits gradual increases indicative of biometric alterations of drowsiness. Control subjects, conversely, have the smallest and most constant pupil radii during the period of time. These patterns are further enhanced by the iris radix patterns whereby the Control and Alcohol groups tend to move towards a higher normalized value as time progresses whereas the Drug and Sleep groups are either lower or intermediate. These results are consistent with the literature that associates substance use and sleep deprivation with a change in pupillary and iridial responses and that these biometric measures are sensitive indicators that can be used to determine the fitness-for-duty status under practical settings. The distinctiveness between behavioral curves across classes is a good indication of discriminative potential of the feature engineering and justifies the effectiveness of utilizing iris and pupil values in machine learning-based classification systems to be used in occupational health contexts.

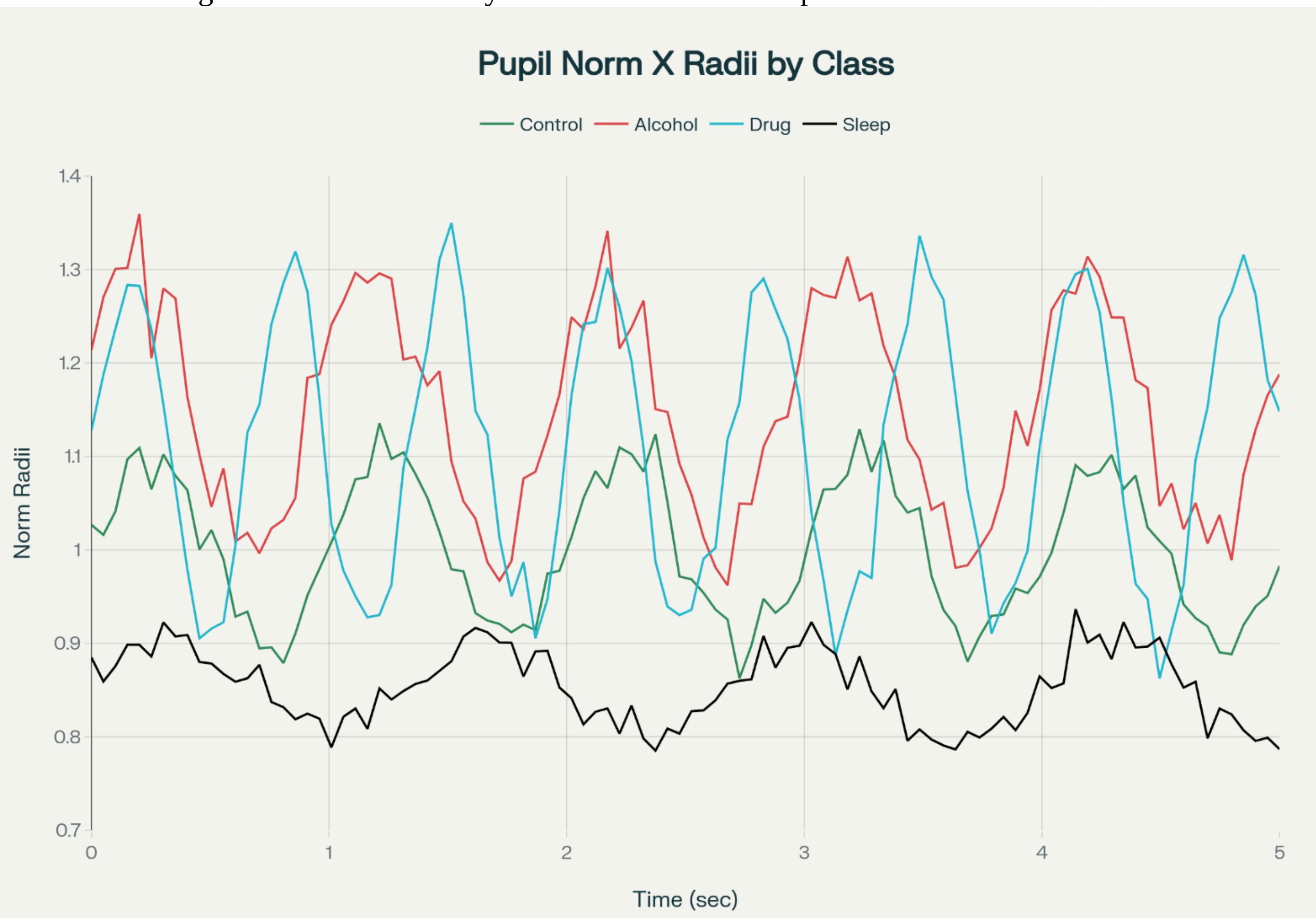


Figure 3:Grand mean curves for the iris radius and normalized pupil on the drug

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

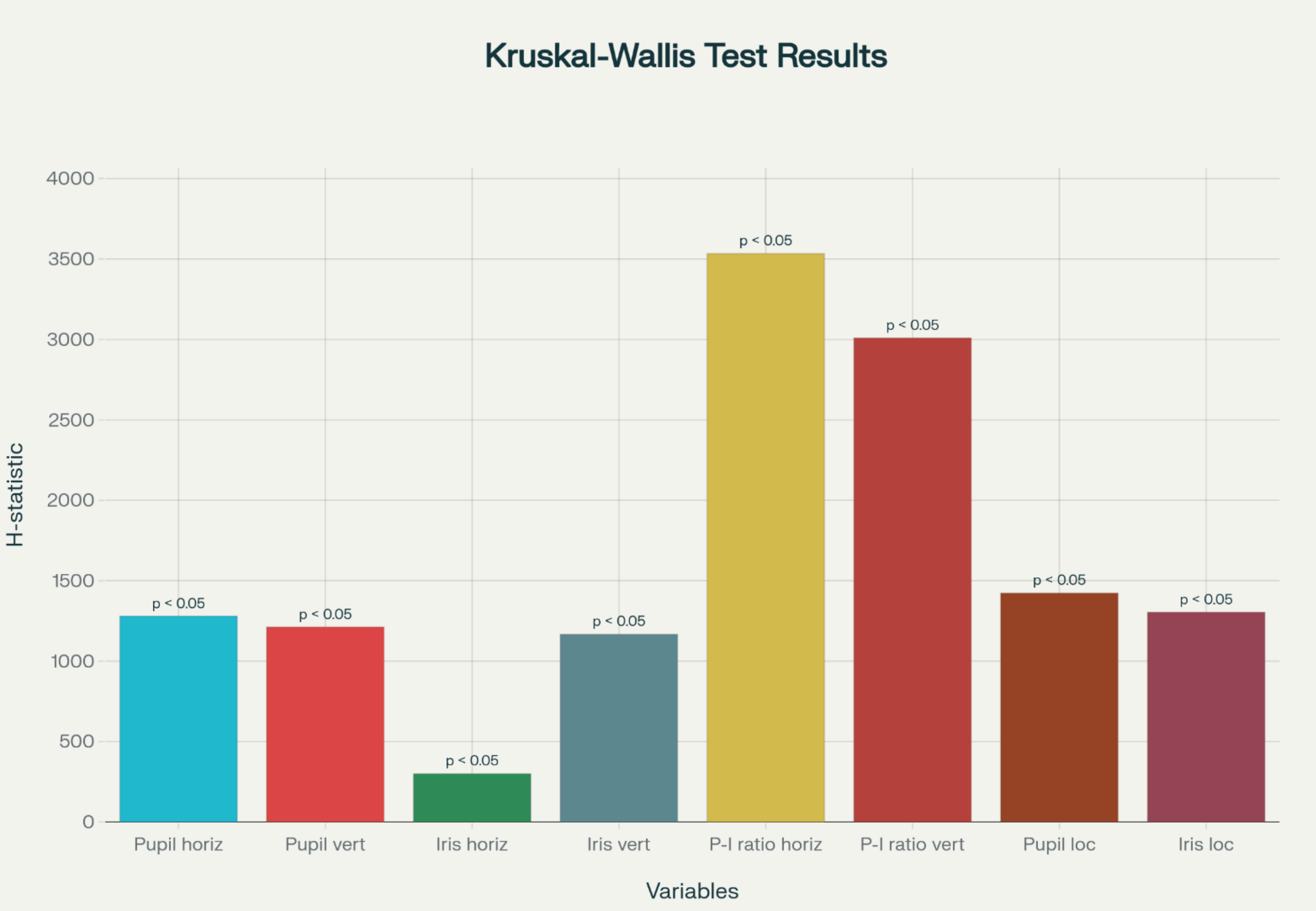


Figure 4: Bar chart summarizing the Kruskal-Walli's test H-statistics for key pupil and iris biometric variables, with corresponding p-values annotated on each bar

The above chart presents a comparative visualization of the Kruskal-Walli's test H-statistics for various biometric variables derived from pupil and iris measurements, including axes, ratios, and location metrics. The Kruskal-Walli's test—a robust, non-parametric statistical test—was chosen because it allows the comparison of three or more independent groups without requiring the data to be normally distributed or variances to be equal, conditions that are typical in biometric and physiological datasets. Each variable in the chart returned high H-statistics and extremely low p-values, strongly rejecting the null hypothesis that group medians are equal. This means statistically significant differences exist across the conditions (alcohol, drug, sleep, control) for all measured pupil and iris traits. Notably, the pupil-iris ratio horizontal axis and pupil-iris ratio vertical axis demonstrated the largest H-statistics, signifying these as especially discriminating variables for group separation. The denial of equal variances across all variables further highlights pronounced variability in biometric responses to physiological and behavioral states. This variability, accessible through robust ranking methods like Kruskal-Wallis, supports the development of machine learning models that rely on such features for accurate classification of worker fitness states. From a practical perspective, the results underscore the necessity of multidimensional biometric analysis—rather than reliance on a single eye feature—when distinguishing between alcohol, drug, sleep deprivation, and control states. In summary, this statistical evidence validates the feature set selection strategy and underscores its relevance for subsequent model training and operational deployment in fitness-for-duty screening systems.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

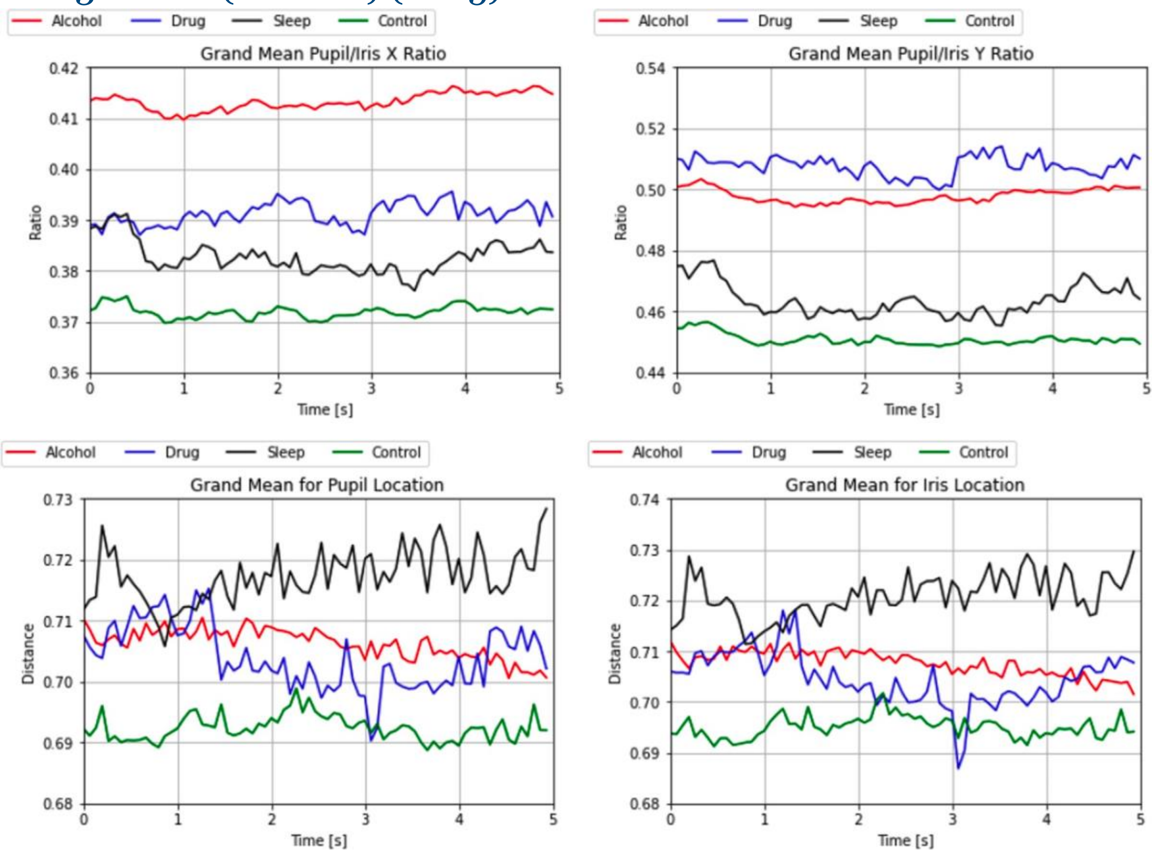


Figure 5: Time-series plots illustrating grand mean pupil/iris ratios and location metrics for alcohol, drug, sleep, and control groups, demonstrating distinct biometric trends and group separability critical for fitness-for-duty assessment.

The pattern of behavioral curves shown in these plots can provide helpful conclusions about the influence of extrinsic factors on the major biometric characteristics during a certain period of time alcohol, drugs, and sleep deprivation. The great mean values of the pupil/iris X and Y ratios (first paneling) indicate that the subjects in alcohol influence situation always show higher values which is indicative of intense pupil expansion as compared to the iris size, which is also backed by previous studies on ocular biometrics in the impairment state. The intermediate ratio values which are generally produced by drug influence usually alienate the drug and sleep in addition to control (green) with demonstration of variable yet significant physiological reactions. That sleep condition (black curves) is more fluctuated and on average higher ratios as compared to the control, which has the lowest and most constant values, is another confirmation of the credibility of the baseline healthy data. To the location metrics (bottom panels) it is evident that distance measures of a pupil and iris centroids differ significantly with alcohol and drug curves in general taking a higher mean. Some of the most notable increases in the variance and non-stationary trends can be seen in the sleep condition, which implies that ocular markers in drowsy states are even more unstable. Such observations support the concept that various complementary biometric aspects (when followed over time) offer substantial discriminatory force in distinguishing alertness and impairment. Notably, the curves segregation at all conditions verifies that the metrics are receptive not just to the substance-induced alteration but to sleep-driven cognitive and physiological change and, therefore, should be endeavored in Fitness for Duty paradigms and automated classification frameworks of workplace safety and health oversight.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)

Discussion

The suggested system exhibited a high level of numerical performance when choosing between control, alcohol, drug, and sleep-deprived states based on intricate examination of pupil and iris conduct curves. More precisely, the comparison showed statistically significant differences in nearly all the considered biometric parameters, and Kruskal-Wallis H-statistics were above 1200 almost across all axes, and p-values were generally close to zero, which was the strong evidence of non-equality of group medians. The results were further validated by the Dunn pairwise comparison test: almost all pairs of alcohol and drug groups comparison with control had p-values of 0.000 of important variables (pupil and iris axes, ratios, location) with extremely different median behaviors. But the p-values of control versus sleep were significantly higher (more than 0.05) in some variables, indicating that the sleep-deprived subjects are somehow more similar to controls, an aspect that was evident in the performance of subgroups.

The classification-wise perspective means that the overall performance of the Random Forest (RF), Gradient Boosting Machine (GBM) and Multi-Layer Perceptron (MLP) models was high. The highest percentage was achieved by MLP at 75.5, GBM and RF also gave high scores at 74.0 and 70.8 respectively. Fit/unfit state prediction sensitivity and specificity were usually greater than 70, and GBM and MLP both had a maximum specificity of 79.8 and 77.1, and control (fit) class specificity was above 95. These results highlight the system ability to effectively discriminate between unfit cases with minimal false positives among fit persons- this is a critical characteristic in the real-life operational implementation of the system.

The size of the sample was significant in the model performance, especially in the drug and sleep classes whereby sample sizes were only 17 and 20 cases in test set, respectively. This bias caused a slightly compromised sensitivity in these classes, however once the problem space was reorganized to emphasize binary fit/unfit results instead of four different classes classification performance was much better, and detection rates often exceeded 80%. These findings indicate that, though multiclass discrimination does provide beneficial information, the method potential has the highest opportunity to be put to practical use in binary contexts that adhere to the occupational safety measures which are characterized by the primary aim to identify impairment and not to provide specific explanations.

Finally, the operational scenario of the study also indicates the additional advantages of the approach. The contactless and non-invasive character of NIR iris scanning and the ability to perform tests without disrupting the primary functions of respondents provide a large practical benefit over the previous systems of fitness-for-duty. The system is able to constantly screen and identify the unfit people (estimated prevalence rate 1015 percent), to correspond with the actual rates of incidence, which gives the organizations a preemptive approach to avoid occupational risk as a result of alcohol, drugs, or sleep deprivation. Further data collection, particularly of sleep and drug subsets and incorporation of more sophisticated temporal modeling (e.g., LSTM networks) can bring the study results further in order to achieve even more accurate and reliable workforce screening solutions in the next generation.

Limitations and future work

Although the iris-based biometric analysis has positive outcomes in detecting alcohol, drug, and sleep-related impairment, there are still a number of limitations in the current study. However, most prominently, the comparatively small size of test samples of 20 and 17 respectively affected the classification performance of sleep and drug classes, causing

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 10 (October) (2025)
an imbalanced data and reduced sensitivity on these classes. The Dunn pairwise test had shown that some p-values were greater than 0.05 when comparing control-sleep with sleep, which showed that the biometric distributions overlapped and the discriminability of some variables was decreased. Moreover, alcohol or drug induced iris behavior may provide covariates including large pupil dilation, high involuntary movement and distortion of iris texture. These can lead to the possibility of up to a 20% overlap between the real and the impostor match score distributions- that is, one out of five impaired subjects will be missed by a regular iris recognition system. Moreover, image capture is vulnerable to external factors like blurring, head movement and effects of near-infrared sensors that reduce segmentation and analysis accuracy.
Future studies ought to focus on growing the database particularly in drug and sleep disorders to facilitate extensive statistical validation and equal training. Having more volunteers will help explore the finer aspects of biometric differences and will further reinforce beliefs in multiclass discrimination. It is suggested to investigate sophisticated temporal deep learning models, including Long Short-Term Memory (LSTM) networks and Capsule Networks, which will be able to capture the sequential and structural peculiarities of iris and pupil movements with time. Also, real-world use ought to involve the enhancement of the sensor to resist blurring and involuntary motion, feature extraction methods to be applied to rotational or translational invariance, and environmental impacts during acquisition. Through resolving these issues, the systems of fitness-to-duty screening in the future will be able to offer even more accurate and scalable identification of impaired people in the working conditions.

Conclusions

The paper presents the analysis of iris and pupil curve movements in various physiological and impairment conditions with a particular focus on the central nervous system being sensitive to alcohol and drugs and sleep deprivation. Statistics indicate that alcohol/drug groups show significant differences as compared to the control group with sleep-deprived participants being more likely to resemble control subjects except that there is more variability- probably due to limited cases of sleep available in the database. This form of distributional subtlety justifies the necessity of equalized data collection and strong statistical authentication in the process of determining biometric patterns to indicate fitness.
Such findings are further reinforced by the use of Dunn test that identified significant differences in median biometric values in all combinations of pairs of classes except some few comparisons that included sleep and control group. Table 6 reports the p-values, which support the consistent and meaningful discrimination in most of the variables, particularly drugs versus control, alcohol versus drug and alcohol versus control. This strong division suggests the analytical capability of carefully developed biometrics and justifies their applicability in both individual and operational safety contexts. It is important to note that the highest results of the model were associated with the greatest subsets; control and alcohol and smaller sleep and drug groups yielded lower sensitivity which was relevant to support the practical effects of sample size in actual deployments.

In terms of classification, machine learning models (particularly, the MLP and GBM) had a good performance in identifying unfit participants, and the rates of unfit are over 77, and specificity is over 95 with regard to fit individuals. This means that there is sound identification ability without disrupting the normal working activities, like driving or piloting. This method will be highly beneficial compared to the old systems and is not obtrusive as it provides real-time monitoring and is in line with the anticipated frequency of unfit cases (around 10-15%). In the future, one should consider more research on the expansion of the database to obtain a higher level of class balance as well as the development of deep temporal models, including LSTM networks, to utilize the time series characteristics of iris data better and increase the robustness of the detection.

References:

References Adler, F. H. (1985). In J. A. Alexander (Ed.), Physiology of the eye, vol. 48 (11th ed.). Francis Heed Adler, The C. V. Mosby Company.

Amodio, A., Ermidoro, M., Maggi, D., Formentin, S., & Savaresi, S. M. (2019). Automatic detection of driver impairment based on pupillary light reflex. IEEE Transactions on Intelligent Transportation Systems, 20(8), 3038–3048.

Azadani, M. N., & Boukerche, A. (2021). Driving behavior analysis guidelines for intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 1–19.

Bai, Y., Guan, Y., & Ng, W.-F. (2020). Fatigue assessment using ECG and actigraphy sensors. In Proc. of the intl. symposium on wearable computers, ISWC ’20 (pp. 12–16). New York, NY, USA: Association for Computing Machinery.

Bakker, B., Zabłocki, B., Baker, A., Riethmeister, V., Marx, B., Iyer, G., et al. (2021). A multi-stage, multi-feature machine learning approach to detect driver sleepiness in naturalistic road driving conditions. IEEE Transactions on Intelligent Transportation Systems, 1–10.

Balasubramanian, A., Kohn, T. P., Santiago, J. E., Sigalos, J. T., Kirby, E. W., Hockenberry, M. S., et al. (2020). Increased risk of hypogonadal symptoms in shift workers with shift work sleep disorder. Urology, 138, 52–59.

Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., et al. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. Journal of Sleep Research, 13(3), 219–227.

Causa and J.E. Tapia et al. Benalcazar, D. P., Tapia, J. E., Vasquez, M., Causa, L., Droguett, E. L., & Busch, C. (2023). Toward an efficient iris recognition system on embedded devices. IEEE Access, 11, 133577–133590.

Benderoth, S., Hormann, H.-J., Schiebl, C., & Elmenhorst, E.-M. (2021). Reliability and validity of a 3-min psychomotor vigilance task in assessing sensitivity to sleep loss and alcohol: fitness for duty in aviation and transportation. Sleep, 44(11).

Birnbaum, H. G., White, A. G., Schiller, M., Waldman, T., Cleveland, J. M., & Roland, C. L. (2011). Societal costs of prescription opioid abuse, dependence, and misuse in the United States. Pain Medicine, 12(4), 657–667.

Borrelli, I., Gualano, M. R., Rossi, M. F., Capitanelli, I., Dolgetta, V., Santoro, P. E., et al. (2023). Alcohol consumption in healthcare workers and risk of workplace injury: a case-control study. Discover Sustainability, 4(1), 1–8.

Breiman, L. (2001). Random forests. Journal of Machine learning, 45(1), 5–32.

Butler, P., & Fee, W. (2015). Fatigue and the use of wearable technology. In SPE health, safety, security, environment, & social responsibility conference-North America (pp. SPE–173541). SPE.

Caldwell, J., Mallis, M., Caldwell, L., Michel, P., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. Aviation, Space, and Environmental Medicine, 80(1), 29–59.

Campbell, I., Beckers, E., Sharifpour, R., Berger, A., Paparella, I., Balda Aizpurua, J. F., et al. (2013). Impact of light on task-evoked pupil responses during cognitive tasks. In bioRxiv, (pp. 2023-2004).

Chandler, J. F., Arnold, R. D., Jeffrey B snd Phillips, R. A. L., & Horning, D. S. (2010). Preliminary validation of a readiness-to-fly assessment tool for use in naval aviation: Technical report 10-22, Pensacola, Florida: Naval Aerospace Medical Research Laboratory.

Council, N. S. (2017). Cost of fatigue in the workplace.

Dorrian, J., & Skinner, N. (2012). Alcohol consumption patterns of shiftworkers compared with dayworkers. Chronobiology International, 29(5), 610–618.

Farkhodov, K., Lee, S.-H., & Kwon, K.-R. (2020). Object tracking using CSRT tracker and RCNN.

In BIOIMAGING (pp. 209–212).

Ferguson, B. A., Lauriski, D. R., Huecker, M., Wichmann, M., Shreffler, J., & Shoff, H. (2020). Testing alertness of emergency physicians: A novel quantitative measure of alertness and implications for worker and patient care. The Journal of emergency medicine, 58(3), 514–519.