Real-time detection of rider fatigue: a comparative study of black-box and glass-box artificial intelligence approaches

Cynthia Hayat[1], Iwan Aang Soenandi[2], Budi Harsono[3]
[1]Department of Information System, Center of Optimization System and Machine Learning, Krida Wacana Christian University, Jakarta, Indonesia
[2]Department of Industrial Engineering, Center of Optimization System and Machine Learning, Krida Wacana Christian University, Jakarta, Indonesia
[3]Department of Electrical Engineering, Center of Optimization System and Machine Learning, Krida Wacana Christian University, Jakarta, Indonesia

Article Info

Article history:

Received May 22, 2025
Revised Jan 5, 2026
Accepted Jan 22, 2026

Keywords:

Black-box models
Explainable artificial intelligence
Glass-box models
Human factors
Multimodal data
Rider fatigue detection

ABSTRACT

Rider fatigue poses a critical safety challenge in two-wheeled vehicle operation due to limited physical protection, increased balance demands, and prolonged exposure to environmental stressors. Effective real-time fatigue detection is essential to mitigate accident risks, particularly in high-traffic regions such as Indonesia. This study presents a comparative analysis of black-box and glass-box artificial intelligence (AI) models for real-time detection of rider fatigue, evaluated through a human factor's lens emphasizing interpretability, intrusiveness, and cognitive compatibility. Multimodal data comprising physiological signals, behavioral indicators, and environmental context were collected using wearable sensors and rider telemetry to train and assess the models. Experimental results reveal that black-box models, including convolutional neural network (CNN) + long short-term memory (LSTM), random forest (RF), and support vector machine (SVM), achieve superior predictive accuracy (94.3%, 91.5%, and 88.2%, respectively) but lack inherent transparency. Conversely, glass-box models such as decision tree (DT) and logistic regression (LR) offer greater interpretability, a critical factor in safety-sensitive applications, though with reduced accuracy (approximately 83–85%). These findings underscore the trade-off between predictive performance and explainability, highlighting the need to tailor model choice to specific operational requirements. This research advances the design of intelligent, human-centered rider support systems that balance accuracy, transparency, and user trust, fostering safer two-wheeled transportation.

Corresponding Author:

Cynthia Hayat
Department of Information System, Center of Optimization System and Machine Learning
Krida Wacana Christian University
Tanjung Duren Raya St. No. 4, Jakarta, Indonesia
Email: cynthia.hayat@ukrida.ac.id

1. INTRODUCTION

Fatigue has long been recognized as a critical factor influencing human performance and safety in transportation systems. In the context of two-wheeled vehicle operation, rider fatigue poses unique challenges due to reduced physical protection, greater demands on balance and situational awareness, and often prolonged exposure to environmental stressors [1]. These conditions make real-time fatigue monitoring not only a technical necessity but also pressing human factors concern. Accurately detecting and mitigating fatigue can greatly reduce the likelihood of accidents, injuries, and fatalities, particularly for motorcyclists who are more vulnerable than car drivers such as in Indonesia that has heavy traffic in many areas [2]. Despite its importance, the intersection of rider fatigue and human-machine systems remains relatively underexplored in applied human factors research [3].

Recent advancements in artificial intelligence (AI) present new opportunities for integrating intelligent monitoring systems into rider assistance frameworks. From a human factors perspective, such systems must not only be effective in recognizing states of fatigue but also be interpretable, trustworthy, and unobtrusive. While black-box AI models such as deep neural networks have shown considerable promise in pattern recognition and predictive analytics, their opaque decision-making processes raise concerns about system transparency, user acceptance, and accountability key issues within the human factors discipline [3]. In contrast, glass-box models such as decision tree (DT) and rule-based systems offer interpretability and align better with the principles of explainable AI, though often at the cost of reduced predictive performance.

Prior research in fatigue detection has largely focused on car drivers, with emphasis on facial recognition, eye tracking, heart rate (HR) variability, and vehicular behavior as indicators of drowsiness but for two wheeled riders it's very difficult to use facial recognition and eye tracking based sensors [4]. Studies employing machine learning and deep learning techniques have demonstrated success in offline detection settings, yet their real-world application remains limited by the lack of real-time adaptability and the cognitive load imposed on the user. Moreover, few studies directly compare the usability, interpretability, and contextual appropriateness of different AI model types in fatigue detection systems. In the realm of motorcycling, where riders may lack in-vehicle dashboards or passive monitoring infrastructure, a human-centered approach to fatigue detection is essential to ensure system effectiveness without compromising user experience or autonomy [5].

This study seeks to fill this gap by conducting a comparative analysis of black-box and glass-box AI approaches for real-time rider fatigue detection through a human factor's lens. We collect multimodal data including physiological signals, behavioral indicators, and environmental context using wearable sensors and rider telemetry. The models are evaluated not only on traditional metrics such as accuracy and latency but also on human centric criteria such as interpretability, intrusiveness, and cognitive compatibility. By foregrounding the human operator in both system design and evaluation, this work contributes to the development of AI-powered fatigue detection tools that are both effective and aligned with human capabilities, limitations, and expectations.

Ultimately, the findings aim to inform the design of intelligent rider support systems that prioritize safety, usability, and trust. In doing so, this study bridges a critical gap between AI system development and human factors engineering, emphasizing the importance of transparency, contextual relevance, and user-centered design in deploying real-time fatigue detection technologies for motorcycle riders. By integrating explainable and multimodal fatigue indicators into system design, the proposed approach supports more reliable decision making and enhances rider acceptance of AI-assisted safety technologies in real-world riding environments.

2. METHOD

2.1. Proposed architecture

This study aims to develop and rigorously evaluate a sensor-based fatigue detection system for motorcycle riders by employing two distinct types of AI models: black-box and glass-box approaches [6]. The proposed system integrates real-time physiological data such as HR and sweats measurement [7]. The overall architecture of this approach is illustrated in Figure 1, which presents the proposed architecture for data acquisition, processing, and model-based fatigue detection. By comparing the performance, interpretability, and reliability of both AI model types, the study seeks to provide comprehensive insights into effective fatigue detection mechanisms, ultimately enhancing rider safety and reducing accident risks on the road.

2.1.1. Step 1: data acquisition and pre-processing

The system's initial component is data, which involves two main processes:

i) Physiological data were collected using wearable sensors, including:
   - HR: measured in beats per minute, HR reflects physical and mental states. Abnormal HR patterns may signal driver fatigue, making it a crucial indicator in real-time fatigue detection systems.
   - Sweat measurement: a non-invasive method to monitor fatigue or stress by analyzing sweat rate or skin conductivity, typically using galvanic skin response/electrodermal activity (GSR/EDA) sensors.

ii) Data pre-processing:
   Raw sensor data undergo normalization, signal filtering, feature extraction, and time segmentation to transform it into structured inputs suitable for machine learning models.

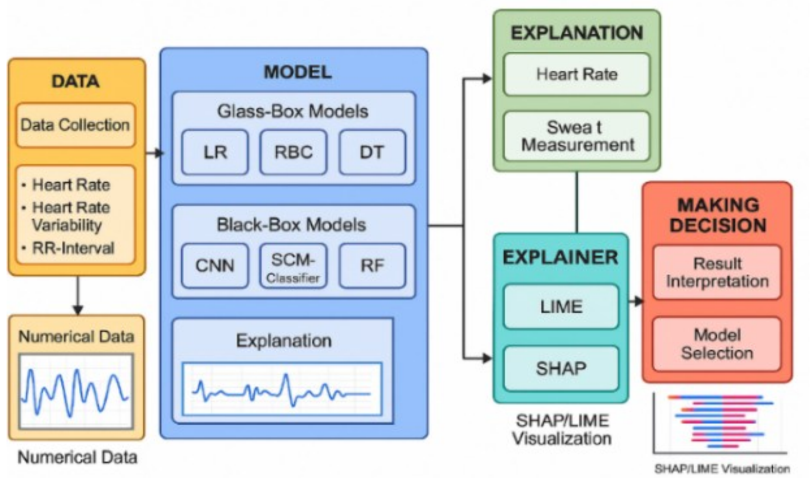

Figure 1. Proposed architecture

2.1.2. Step 2: glass-box vs black-box models

The modeling component is categorized based on interpretability [6], [8]:

i) Glass-box models (high transparency):
   - Logistic regression (LR): a simple linear model offering clear insight into feature contributions [9].
   - Rule-based classifier (RBC): uses human-readable rules for decision making [10].
   - DT: provides intuitive decision paths via tree structures [11].

ii) Black-box models (high performance and low interpretability):
   - Convolutional neural network (CNN): captures spatial and temporal features from physiological signals [12].
   - Set covering machine (SCM) classifier: marginal or combined classifiers with complex decision boundaries [13].
   - Random forest (RF): an ensemble of DT delivering high accuracy but limited transparency [14]. This comparison aims to evaluate the trade-off between model accuracy and interpretability in real-time fatigue detection [15], [16].

2.1.3. Step 3: XAI explainer layer SHAP and LIME

To interpret black-box models, this layer introduces explainable AI techniques [17], [18]. These techniques are: i) Shapley additive explanation (SHAP) quantifies each feature's contribution to predictions using game theory and ii) local interpretable model-agnostic explanations (LIME) builds interpretable surrogate models around individual predictions. These methods enable the system to explain why a driver is classified as "fatigued" or "not fatigued", even when using complex models.

2.1.4. Step 4: explanation layer

This layer translates explainer outputs into domain-specific information accessible to users and experts [19], [20]. Contributions from features like HR, eye blink rate, and head nodding are quantified and visualized. Re-processing is applied to ensure the presented information is clear, relevant, and interpretable by non-technical stakeholders.

2.1.5. Step 5: decision making

The final stage integrates classification results and their interpretations for informed decision-making: i) result interpretation: assesses whether the driver shows early signs of fatigue and ii) model selection: based on evaluation results, users can choose models prioritizing either accuracy (black-box) or transparency (glass-box). This architecture outlines a comprehensive framework for evaluating real-time driver fatigue detection models. It integrates interpretable and high-performance explainable artificial intelligence (XAI) techniques, such as SHAP and LIME, to provide transparent insights into model decision-making. By combining predictive accuracy with interpretability, the system ensures that decisions regarding driver fatigue are not only reliable but also explainable, which is an essential requirement for deployment in safety-critical applications. Furthermore, this framework supports continuous monitoring and evaluation, enabling the identification of potential failure points and the improvement of model robustness in real-world conditions [21], [22].

2.2. Physiological data collection and participant profile

This study used heart rate variability (HRV) and GSR to assess fatigue in 30 male motorcycle riders (M =32.05 years, SD =5.04; BMI 26–36; ≥2 years riding experience). Participants rode in controlled on-road sessions wearing non-invasive sensors, and those with medical conditions affecting HRV were excluded. Ethical approval was obtained, and all participants provided written informed consent. HRV, derived from electrocardiogram (ECG), reflects autonomic nervous system activity, while GSR measures skin conductance changes associated with fatigue. Data were continuously recorded using Polar H10 and Polar Ignite 2 devices to monitor real-time physiological dynamics. Examples of the recorded physiological data during non-fatigued (fit) and fatigued conditions are presented in Tables 1 and 2, respectively.

Table 1. Sample fit data

HR [bpm]	HRV [ms]	RR-interval [ms]	GSR signal	Condition
110	3.4	513	1.267	Fit
109	3.4	520	1.294	Fit
108	3.7	530	1.288	Fit
109	3.9	300	1.294	Fit
109	3.5	774	1.257	Fit
109	2.7	548	1.255	Fit
109	3	541	1.271	Fit

Table 2. Sample fatigue data

HR [bpm]	HRV [ms]	RR-interval	GSR Signal	Condition
99	2.6	565	1.423	Fatigue
98	2.6	564	1.421	Fatigue
98	2.6	566	1.409	Fatigue
98	2.6	563	1.415	Fatigue
98	2.6	566	1.44	Fatigue
98	2.6	570	1.408	Fatigue
99	2.6	585	1.415	Fatigue

The raw HRV and GSR signals are subjected to a series of preprocessing procedures to enhance data quality, including noise filtering, artefact correction, and signal normalization [23]. From the HRV data, both time-domain features such as root mean square of successive differences (RMSSD) and standard deviation of normal-to-normal intervals (SDNN) and frequency-domain metrics notably the low frequency to high frequency (LF/HF) ratio are extracted, as these parameters are well-established indicators of fatigue-related autonomic modulation. For GSR, the signals are processed to derive key features such as skin conductance level (SCL) and skin conductance response (SCR), which are known to reflect variations in sympathetic nervous activity linked to emotional arousal and fatigue states [24]–[26].

Following the evaluation, a comparative analysis was conducted to explore the trade-offs between black-box models and glass-box models. The analysis was structured around the following dimensions:

i) Accuracy: black-box models typically achieved higher predictive performance due to their ability to capture complex, non-linear patterns in physiological signals. Metrics showed that models like CNN and long short-term memory (LSTM) yielded superior F1-scores and reduced misclassification rates compared to their interpretable counterparts [27], [28].
ii) Interpretability: glass-box models demonstrated clear advantages in transparency. DT and LR provided understandable decision paths and feature weightings, making them preferable in settings where explainability and user trust are critical, such as for regulatory compliance or driver feedback systems [29]–[31].

This comparative framework provides a nuanced understanding of the accuracy–transparency trade-off in sensor-based fatigue detection systems. It supports informed model selection tailored to specific use-case requirements—whether prioritizing predictive precision or interpretability for user trust [32]–[34].

3. RESULTS AND DISCUSSION

3.1. Interpretation of feature contributions using SHAP summary plot and LIME

To investigate the contributions of physiological features to stress and fatigue prediction, we used SHAP and LIME techniques, as shown in Figures 2 and 3. Key features include HR, HRV, RR-interval, and GSR, which reflect autonomic nervous system activity. LIME visualizations show that higher HRV, lower HR, and longer RR-intervals reduce fatigue probability, while elevated GSR and certain HR values increase it. These explanations enhance model transparency and provide actionable insights for designing personalized fatigue monitoring systems.

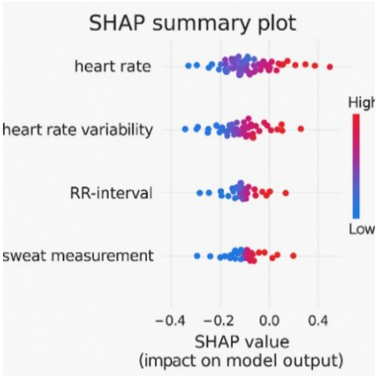

Figure 2. SHAP summary plot

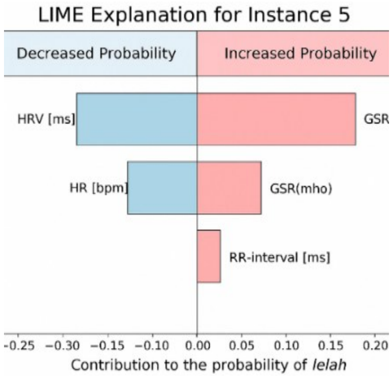

Figure 3. LIME explanation

3.2. Evaluation and comparison of models

This section presents a comprehensive evaluation of the implemented models with respect to two critical aspects: detection accuracy and interpretability. Furthermore, a comparative analysis is conducted between black-box and glass-box models to highlight the trade-offs involved in model selection. The performance of each machine learning model was assessed using standard evaluation metrics, including accuracy, precision, recall, and the area under the receiver operating characteristic (ROC) curve area under the receiver operating characteristic (AUROC). The results for the training dataset are summarized in Table 3, while the performance metrics obtained from the testing dataset are presented in Table 4. These tables provide a comparative overview of how black-box and glass-box models perform in terms of predictive capability, allowing further analysis of the trade-offs between accuracy and interpretability.

Table 3. Performance value for train dataset

Methods	No	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUROC-mean	Interpretability
Glass box-model	1	DT	83.3	80.4	84.2	82.2	0.88	High
	2	LR	81.1	78.2	80.5	79.3	0.85	High
	3	RBC	75.6	74.8	76.2	75.5	0.81	Very high
Black box-model	1	CNN+LSTM	89.4	88.7	90.1	89.4	0.94	Low
	2	SVM classifier	92.5	91.3	93.7	92.5	0.96	Low
	3	RF	93.2	92.5	94.2	93.2	0.95	Low

Table 4. Performance value for testing dataset

Methods	No	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUROC-mean	Interpretability
Glass box-model	1	DT	85.10	80.40	84.30	82.50	0.87	High
	2	LR	83.60	79.10	80.50	79.10	0.85	High
	3	RBC	84.20	75.70	74.60	75.90	0.8	Very high
Black box-model	1	CNN+LSTM	94.30	89.80	91.70	90.10	0.95	Low
	2	SVM classifier	88.20	92.50	94.20	92.10	0.95	Low
	3	RF	91.50	94.30	94.90	93.70	0.96	Low

The visualization presented highlights a clear trade-off between model accuracy and interpretability within the context of real-time motorcycle rider fatigue detection. This trade-off is central to the selection of appropriate machine learning models, especially when balancing predictive performance with system transparency.

i) High-accuracy, low-interpretability models: models such as CNN, LSTM, support vector machine (SVM), and RF achieve high accuracy by capturing complex, nonlinear patterns in physiological data. However, their decision-making processes are opaque, classifying them as "black-box" models. Specialized XAI techniques like SHAP or LIME are needed to interpret their predictions.

ii) High-interpretability, lower-accuracy models: interpretable or "glass-box" models, such as DT, LR, and RBC, provide transparent decision-making through readable rules or feature weights. They show how variables like HRV or sweat conductivity influence predictions. However, their predictive performance is generally lower, especially with high-dimensional or nonlinear data.

To illustrate the balance between predictive performance and model interpretability, a trade-off comparison among the evaluated machine learning models is visualized in Figure 4. This figure highlights how black-box models generally achieve higher accuracy at the expense of transparency, while glass-box models offer greater interpretability with moderate reductions in performance, emphasizing the need to align model selection with application-specific requirements. In the development of machine learning-based fatigue detection systems, selecting an appropriate model requires carefully balancing two crucial factors: interpretability and accuracy. This section presents a detailed comparative analysis of six representative models across these dimensions, as depicted in the accompanying trade-off plot. The plot maps six models on a two-dimensional space with interpretability on the x-axis and classification accuracy on the y-axis. LR and DT show high interpretability with moderate accuracy (~0.82–0.85), suitable for contexts requiring clear rationale. The RBC offers very high interpretability (~1.0) with slightly lower accuracy (~0.81), ideal for safety-critical environments. SVM and RF balance interpretability and accuracy, while CNN+LSTM achieves the highest accuracy (~0.94) but lowest interpretability, making it valuable where prediction performance is prioritized over transparency.

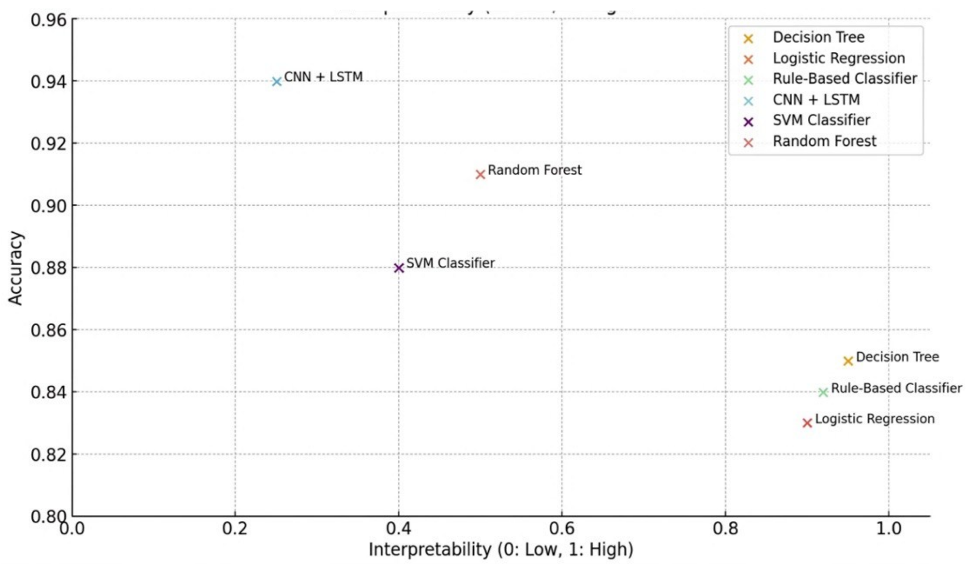

Figure 4. Trade-off of ML models

The trade-off curve emphasizes that model selection must align with the operational goals of fatigue detection systems. In safety-critical applications, transparency is preferred for human-in-the-loop validation, while high-performance models may be chosen for autonomous scenarios if their predictions can be interpreted via SHAP or LIME. This analysis highlights that accuracy alone should not dictate model choice; interpretability requirements specific to the application are equally important.

4. CONCLUSION

The experimental results demonstrate a significant trade-off between predictive performance and interpretability in fatigue detection systems based on physiological signals. Black-box models, particularly CNN+LTSM (94.3% accuracy), RF (91.5% accuracy) and SVM (88.2% accuracy) on the testing dataset, consistently outperformed glass-box models in terms of accuracy, precision, recall, and AUROC. These models effectively capture complex, nonlinear patterns within the data but lack intrinsic transparency, necessitating post-hoc interpretability methods such as SHAP or LIME. Conversely, glass-box models such as DT, rule-based classifier and LR provide more interpretable outputs critical in safety-sensitive applications like rider monitoring yet deliver relatively lower performance (e.g., 85.1%, 84.2% and 83.6% accuracy, respectively). These findings underscore the importance of aligning model selection with application-specific requirements. In contexts where real-time transparency and explainability are essential for user trust and system accountability, glass-box models may be preferred despite moderate reductions in accuracy. However, when predictive performance is paramount, especially in high-risk scenarios, black-box models augmented with explainability layers offer a viable solution. Future work should focus on optimizing this trade-off through hybrid approaches and the integration of interpretable machine learning techniques without compromising accuracy.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Research, and Technology of the Republic of Indonesia for the support provided through the Applied Research Model program. The authors also thank all parties who contributed to this research, including respondents, for their valuable input and constructive feedback.

FUNDING INFORMATION

This research was funded by the Ministry of Higher Education, Research, and Technology of the Republic of Indonesia through the Applied Research Model program, under contract number 124/C3/DT.05.00/PL/2025.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Cynthia Hayat	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		
Iwan Aang Soenandi	✓			✓		✓	✓	✓		✓	✓	✓	✓	✓
Budi Harsono	✓		✓	✓			✓			✓			✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest regarding the publication of this paper.

INFORMED CONSENT

All participants in this study provided informed consent prior to participation. They were fully informed about the objectives, procedures, potential risks, and benefits of the research, and they voluntarily agreed to participate.

ETHICAL APPROVAL

This study was approved by the Ethics Committee of Universitas Kristen Krida Wacana (UKRIDA) with approval number UKKW-EC-2025-12. All procedures involving human participants were conducted in accordance with relevant guidelines and regulations.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, upon reasonable request. The data include information obtained directly from research participants and contain sensitive personal information; therefore, they are not publicly available due to privacy and ethical restrictions.

REFERENCES

[1] A.-B. A. Al-Mekhlafi, A. S. N. Isha, N. Chileshe, M. Abdulrab, A. A. H. Saeed, and A. F. Kineber, "Modelling the relationship between the nature of work factors and driving performance mediating by role of fatigue," *International Journal of Environmental Research and Public Health*, vol. 18, no. 13, Jun. 2021, doi: 10.3390/ijerph18136752.

[2] S. Ayas, B. Donmez, and X. Tang, "Drowsiness mitigation through driver state monitoring systems: a scoping review," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 66, no. 9, pp. 2218–2243, Sep. 2024, doi: 10.1177/00187208231208523.

[3] A. Biró, A. I. C.-Vargas, and L. Szilágyi, "AI-assisted fatigue and stamina control for performance sports on IMU-generated multivariate times series datasets," *Sensors*, vol. 24, no. 1, Dec. 2023, doi: 10.3390/s24010132.

[4] J. Xu, J. Min, and J. Hu, "Real-time eye tracking for the assessment of driver fatigue," *Healthcare Technology Letters*, vol. 5, no. 2, 2018, doi: 10.1049/htl.2017.0020.

[5] Z. Dworakowski, K. Dragan, and T. Stepinski, "Artificial neural network ensembles for fatigue damage detection in aircraft," *Journal of Intelligent Material Systems and Structures*, vol. 28, no. 7, pp. 851–861, Apr. 2017, doi: 10.1177/1045389X16657428.

[6] T. Lee, J. Natalwala, V. Chapple, and Y. Liu, "A brief history of artificial intelligence embryo selection: from black-box to glass-box," *Human Reproduction*, vol. 39, no. 2, pp. 285–292, Feb. 2024, doi: 10.1093/humrep/dead254.

[7] R. Billones, J. K. Liwang, K. Butler, L. Graves, and L. N. Saligan, "Dissecting the fatigue experience: a scoping review of fatigue definitions, dimensions, and measures in non-oncologic medical conditions," *Brain, Behavior, & Immunity - Health*, vol. 15, Aug. 2021, doi: 10.1016/j.bbih.2021.100266.

[8] T. Dobbrick, J. Jakob, C.-H. Chan, and H. Wessler, "Enhancing theory-informed dictionary approaches with 'glass-box' machine learning: the case of integrative complexity in social media comments," *Communication Methods and Measures*, vol. 16, no. 4, pp. 303–320, Oct. 2022, doi: 10.1080/19312458.2021.1999913.

[9] Y. Xia, X. Li, F. Chen, and H. Liang, "Goodness of fit checking for stukel generalized logistic regression models," *Statistica Sinica*, 2026, doi: 10.5705/ss.202023.0231.

[10] M. M. Alam *et al.*, "An energy and leakage current monitoring system for abnormality detection in electrical appliances," *Scientific Reports*, vol. 12, no. 1, Nov. 2022, doi: 10.1038/s41598-022-22508-2.

[11] K. Maurya, S. Mahajan, and N. Chaube, "Decision tree (DT) and stacked vegetation indices based mangrove and non-mangrove discrimination using AVIRIS-NG hyperspectral data: a study at marine national park (MNP) Jamnagar, Gulf of Kutch," *Wetlands Ecology and Management*, vol. 31, no. 6, pp. 805–823, Dec. 2023, doi: 10.1007/s11273-023-09952-1.

[12] S. N, T. Shankar, and K. D. M. Sundaram, "Lemon fruit classification by transfer learning technique: experimental investigation of convolutional neural network (CNN)," *International Journal of Information and Decision Sciences*, vol. 17, no. 4, 2025, doi: 10.1504/IJIDS.2025.10061181.

[13] P. Trajdos and M. Kurzynski, "Soft confusion matrix classifier for stream classification," in *Computational Science (ICCS)*, Springer, Cham, 2021, doi: 10.1007/978-3-030-77967-2_1.

[14] X. Chen and H. Ishwaran, "Random forests for genomic data analysis," *Genomics*, vol. 99, no. 6. 2012, doi: 10.1016/j.ygeno.2012.04.003.

[15] A. Holzinger, M. Plass, K. Holzinger, G. C. Cris¸an, C.-M. Pintea, and V. Palade, "A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop," *Creative Mathematics and Informatics*, vol. 28, no. 2, 2019, doi: 10.37193/cmi.2019.02.04.

[16] L. M. Thimoteo, M. M. Vellasco, J. Amaral, K. Figueiredo, C. L. Yokoyama, and E. Marques, "Explainable artificial intelligence for COVID-19 diagnosis through blood test variables," *Journal of Control, Automation and Electrical Systems*, vol. 33, no. 2, 2022, doi: 10.1007/s40313-021-00858-y.

[17] R. Dwivedi *et al.*, "Explainable AI (XAI): core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, 2023, doi: 10.1145/3561048.

[18] W. Saeed and C. Omlin, "Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities," *Knowledge-Based Systems*, vol. 263, 2023, doi: 10.1016/j.knosys.2023.110273.

[19] A. M. Antoniadi *et al.*, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review," *Applied Sciences*, vol. 11, no. 11, 2021, doi: 10.3390/app11115088.

[20] L. Longo *et al.*, "Explainable artificial intelligence (XAI) 2.0: a manifesto of open challenges and interdisciplinary research directions," *Information Fusion*, vol. 106. 2024, doi: 10.1016/j.inffus.2024.102301.

[21] R. K. Sheu and M. S. Pardeshi, "A survey on medical explainable AI (XAI): recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20. 2022, doi: 10.3390/s22208068.

[22] S. Ali *et al.*, "Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence," *Information Fusion*, vol. 99, 2023, doi: 10.1016/j.inffus.2023.101805.

[23] S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biological Psychology*, vol. 55, no. 3. 2001, doi: 10.1016/S0301-0511(00)00085-5.

[24] S. Jaydarifard, K. Behara, D. Baker, and A. Paz, "Driver fatigue in taxi, ride-hailing, and ridesharing services: a systematic review," *Transport Reviews*, vol. 44, no. 3, 2024, doi: 10.1080/01441647.2023.2278446.

[25] B. K. Savaş and Y. Becerikli, "Real time driver fatigue detection system based on multi-task ConNN," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2963960.

[26] J. He, Z. Li, Y. Ma, L. Sun, and K. H. Ma, "Physiological and behavioral changes of passive fatigue on drivers during on-road driving," *Applied Sciences*, vol. 13, no. 2, 2023, doi: 10.3390/app13021200.

[27] A. Gramegna and P. Giudici, "SHAP and LIME: an evaluation of discriminative power in credit risk," *Frontiers in Artificial Intelligence*, vol. 4, Sep. 2021, doi: 10.3389/frai.2021.752558.

[28] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP," *Diagnostics*, vol. 13, no. 11, Jun. 2023, doi: 10.3390/diagnostics13111932.

[29] N. A. M. Ahmed and A. Alpkoçak, "A quantitative evaluation of explainable AI methods using the depth of decision tree," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 6, pp. 2054–2072, Sep. 2022, doi: 10.55730/1300-0632.3924.

[30] P. Thisovithan, H. Aththanayake, D. P. P. Meddage, I. U. Ekanayake, and U. Rathnayake, "A novel explainable AI-based approach to estimate the natural period of vibration of masonry infill reinforced concrete frame structures using different machine learning techniques," *Results in Engineering*, vol. 19, Sep. 2023, doi: 10.1016/j.rineng.2023.101388.

[31] A. S. Sunge, E. Abdurachman, Y. Heryadi, and I. H. Kartowisatro, "Interpretable using feature importance, SHAP and LIME for caesar's prediction," *International Journal of Applied Engineering and Technology* , vol. 5, no. 2, pp. 83–89, 2023.

[32] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and S. Alfarhood, "Explainable predictive maintenance of rotating machines using LIME, SHAP, PDP, ICE," *IEEE Access*, vol. 12, pp. 29345–29361, 2024, doi: 10.1109/ACCESS.2024.3367110.
[33] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, and A. A. Mäkitie, "Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP," *Scientific Reports*, vol. 13, no. 1, Jun. 2023, doi: 10.1038/s41598-023-35795-0.
[34] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron," *IEEE Access*, vol. 12, pp. 30164–30175, 2024, doi: 10.1109/ACCESS.2024.3368377.

BIOGRAPHIES OF AUTHORS

Cynthia Hayat is working as associate professor in the Department of Information System at the Krida Wacana Christian University, Indonesia. She is also member of Center of Optimization System and Machine Learning dan SAKURA Exchange Program in Science administrated by Japan Science and Technology Agency. Her research area includes machine learning, data science. She has been teaching for more than 12 years and has several research publications to her credit. She can be contacted at email: cynthia.hayat@ukrida.ac.id.

Iwan Aang Soenandi is an associate professor in the Department of Industrial Engineering at Krida Wacana Christian University, Indonesia. He is also member of Asean Registered Engineer. His research area includes ergonomics, simulation system and machine learning. He has been teaching for more than 14 years and has several research publications in international and national journal. He can be contacted at email: iwan.as@ukrida.ac.id.

Budi Harsono is an assistant professor in the Department of Electrical Engineering at Krida Wacana Christian University, Indonesia. He is also member of Center of Optimization System and Machine Learning. His research area includes signal processing, computer vision and robotics. He has been teaching for more than 15 years and has several research publications in international and national journal. He can be contacted at email: budi.harsono@ukrida.ac.id.