International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://www.theaspd.com/ijes.php

An Explainable Artificial Intelligence (XAI) Based Driver Fatigue Detection System For Safe And Care Human Life (SCHL)

P.Joy Kiruba[1], Dr.A.Sathya Sofia[2]

[1]Assistant Professor, B.S.Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.
Mail ID: pjoykiruba@gmail.com
[2]Associate Professor, PSNA College of Engineering and Technology, Dindigul. Tamilnadu, India.
Mail ID: sathyasofia@psnacet.edu.in

Abstract:

Road safety experts confirm driver fatigue as a major factor that leads to traffic accidents which results in deaths because existing fatigue detection methods operate without clear explanations of their analysis process. This paper investigates the fusion of Explainable Artificial Intelligence (XAI) systems into Driver Fatigue Detection Systems (DFDS) for Safe and Care Human Life (SCHL) purposes to boost safety levels and protect human life value. The use of XAI models enables AI systems to generate transparent explanations regarding their fatigue identification processes and resulting predictions. Decision trees and rule-based models and attention mechanisms and deep learning techniques are examined in this paper regarding their capabilities in predicting model outcomes. This review analyzes the technical capabilities of applied methods regarding their fatigue detection effect while prioritizing transparency standards needed for field implementation. The paper examines how different sensor data streams which include biometrics and vehicles Dynamics and environmental elements enhance detection precision while making the system more resilient. Additionally the review discusses strategies for incorporating XAI into DFDS systems as it examines future approaches to bridge model precision requirements with human readable explanations. This research shows how effective explanation systems serve as a means to boost the reliability while ensuring trustworthiness and enhancing effectiveness of driver fatigue detection systems for safer transportation environments.

Keywords: *Driver Fatigue Detection Systems (DFDS), Decision Trees, Safe and Care Human Life (SCHL), Deep Learning, eXplainable Artificial Intelligence (XAI)*

1. INTRODUCTION

Road safety conditions have worsened because of rising transportation-related crashes sourced from fatigued drivers. When drivers spend too much time behind the wheel without sufficient breaks alongside monotonous driving situations their cognitive functions deteriorate together with reduced reaction speed which creates conditions for more accidents to happen [1]. The World Health Organization (WHO) reveals drowsiness or fatigue causes 20% of every road collision. The need for an immediate fatigue detection system that prevents accidents has become an essential requirement due to this dangerous situation. The successful solution to this issue needs a dependable driver fatigue detection system which operates in real-time. Current AI systems together with ML technologies allow developers to build fatigue detection solutions through combining multiple bodily information with behavioral indicators [2], [3]. Traditional driver fatigue detection AI systems face limitations when it comes to demonstrating their decision-making approval to the public. Although these predictive models have good accuracy in fatigue detection they work in an unintelligible manner because their prediction processes remain hidden. The inability to interpret their operation weakens public trust and hinders actual use because transparency becomes essential for reliable systems [4]. The solution to this issue has materialized through Explainable Artificial Intelligence (XAI). The XAI framework improves machine learning systems through the generation of predictive models which deliver accurate results alongside easy-understandable explanations about the decision-making process [5].

The system requires complete transparency most especially when used in safety-critical environments such as driver fatigue detection. **Figure 1** illustrates the overall effectiveness and limitations of the AI framework.

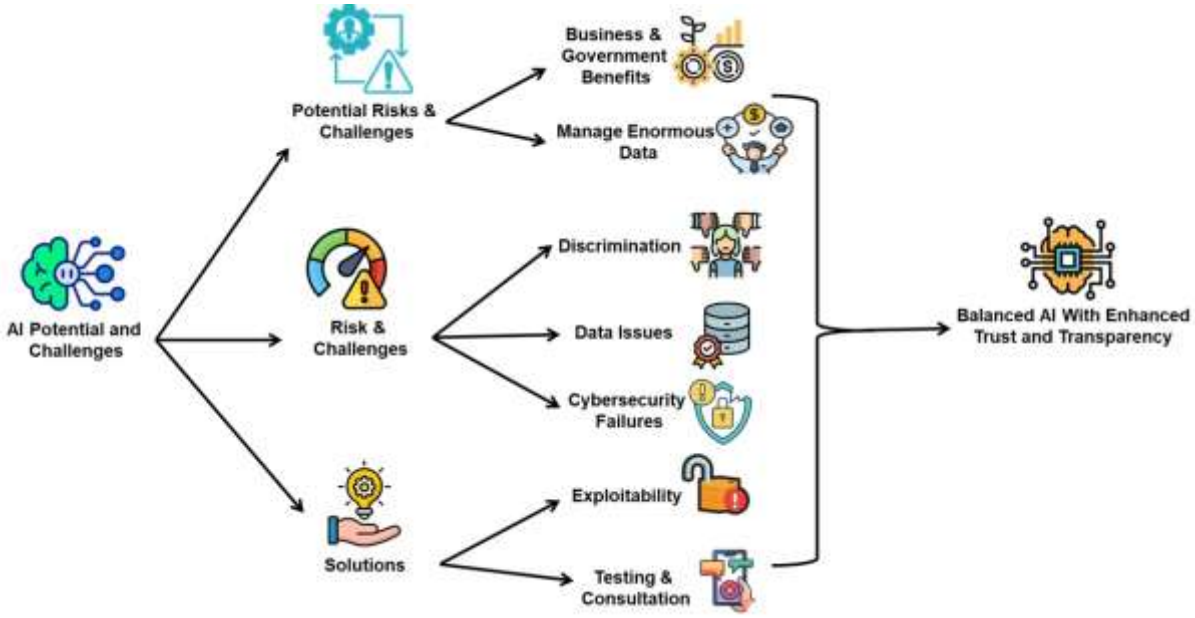

Figure 1: The potential and challenges of AI

This review analyzes how XAI systems can enhance driver fatigue detection methods for creating better safety conditions and improved well-being for drivers while exploring different approaches towards achieving this goal. These systems gain the ability to notify drivers about fatigue as well as disclose information regarding the motivation behind detecting particular actions indicating fatigue through XAI integration. Through these measures all stakeholders can depend on system alerts because they obtain transparent explanations from the system that support their decisions about interventions. The study assesses multiple XAI approaches along with their performance in aiding decision systems which results in Safe and Care Human Life (SCHL) advancement. The inclusion of XAI functions into fatigue detection systems would increase public confidence in AI safety technologies thus leading to their better adoption in real-world operational environments [6], [7].

The detection of driver fatigue before has employed driving behavioral patterns together with eye tracking measurements along with assessing physiological signals. These methods showed initial value but they cannot scale up for broad implementation because they have persistent precision and scalability drawbacks. Machine learning-based fatigue detection systems with their traditional approach fail to deliver explanation of their prediction procedures so users develop mistrust in the system. The stopping of important information from users about how these technologies work becomes an obstacle to their widespread implementation [8], [9].

XAI solves this problem by offering transparent descriptions about how decisions regarding fatigue detection are made. An XAI-based system demonstrates through explanations what features led to fatigue detection along with the reason why specific interventions are recommended. The immediate display of information is essential for users to build confidence especially when operating time-sensitive applications that need high reliability from users [10].

The implementation of XAI within driving fatigue systems will result in effective and trustworthy monitoring capabilities. Systems with XAI enhancements show users how fatigue alerts are processed so drivers together with safety authorities can perform the best preventive responses. The expansion of

autonomous vehicles as a market highlights the necessity to use explainable AI during monitoring of human and automated driving activities particularly in extended periods of monotonous driving tasks.

The study investigates XAI technology applications to develop a dependable fatigue detection system with clear explanations which meets the requirements of Safe and Care Human Life (SCHL) applications. XAI when partnered with advanced driver fatigue detection systems produces technology which performs accurate fatigue detection and reveals fatigue causes for improved decision-making processes. The implementation of XAI-based fatigue detection systems faces difficulties between maintenance of system complexity and interpretability as well as requirements for real-time processing.

The review evaluates XAI applications for transportation to offer a promising improvement in road safety and weariness accident prevention and traffic system-based driver and passenger protection. XAI serves as a necessary element to upgrade driver fatigue detection systems toward more dependable transparent operation which leads to enhanced road safety performance.

2. CURRENT STATE OF EXPLAINABLE AI (XAI)

For ethical, legal, and safety considerations, it is crucial to explain the outcomes of machine learning (ML) models when using AI algorithms in fields such as healthcare, credit scoring, loan approval, and others [11]. While there are several reasons why XAI is important, research indicates that the three main issues are as follows: 1) the dependability of AI algorithms; 2) the transparency of AI algorithms; and 3) the fairness and equity of AI algorithms. XAI techniques must address all three of these issues when working with highly nonlinear deep learning (DL) algorithms that have millions of parameters in ML pipelines [12], [13].

2.1 Rising Demand for Transparency

The demand for explainability in AI has become more pronounced as AI technologies are increasingly integrated into sensitive areas [14]. In healthcare, for instance, AI-driven diagnostic tools and decision support systems require clear explanations to ensure trust among doctors and patients. Similarly, in finance, credit scoring models and loan approval systems must be transparent to ensure fairness and compliance with anti-discrimination regulations. This widespread adoption of AI in high-stakes applications has created a pressing need for interpretable and accountable AI systems [15], [16].

2.2 Explainable AI Techniques

There are two main categories of XAI techniques: intrinsic explainability and post-hoc explainability.

Intrinsic Explainability: Strategies exist which enable the creation of models that generate built-in interpretability. Decision trees, linear models and rule-based systems show their decision-making processes through their built-in explanations which enables easy process understanding. Simple models along with clear interpretation tend to perform well although they can lack the ability to tackle intricate tasks including image recognition and natural language processing [17], [18].

Post-hoc Explainability: The approaches function on complex models such as deep neural networks following their training completion. LIME (Local Interpretable Model-agnostic Explanations) together with SHAP (Shapley Additive Explanations) and attention mechanisms aid predictive explanation by pinpointing significant features which influenced the model outcome. Such interpretation techniques deliver helpful findings yet their calculations demand extensive resources while their output might not present meaningful insights to users [19].

2.3 Model-Agnostic vs. Model-Specific Methods

XAI methods can also be divided into model-agnostic and model-specific approaches:

Model-agnostic methods: A model-agnostic method functions with any ML model regardless of its complexity. LIME and SHAP enable explanation of model behavior across any classification methods including decision trees and support vector machines as well as neural networks. The approximation techniques explain model predictions through simpler and interpretable models that substitute complex structures [20], [21].

Model-specific methods: The interpretation methods like Class Activation Mapping (CAM) and saliency maps describe how convolutional neural networks (CNNs) used for image recognition function by highlighting regions in pictures that influenced model outcomes most significantly. The methods have been optimized specifically for working with particular models that have unique structural characteristics and operational properties [22], [23].

2.4 Challenges in Achieving Explainability

Accuracy vs. Interpretability Trade-off: Complex deep learning networks prove most accurate but maintain difficult levels of explanation [24]. Decision trees provide easier interpretability than more accurate complex models although they might possess inferior capabilities to address challenging tasks. XAI faces a fundamental dilemma because experts need balanced accuracy and interpretability in their systems.

Lack of Standardization: XAI research remains new because the field lacks a single accepted framework for explanation standards. Different sectors and operational needs necessitate distinct explanation types thus there exists no universal solution to XAI applications. Standardization issues regarding explanation formats together with evaluation metrics prevent the assessment of XAI technique quality and effectiveness across different application domains.

Scalability Issues: When sophisticated AI models require increasingly complex and massive data processing the delivery of clear explanations for every decision becomes an increasingly hard task. It poses major scalability challenges to XAI methods because real-time applications demand instant explanations particularly in cases like autonomous vehicles.

Subjectivity of Explanations: Different explanations show subjectivity because they depend on three factors: the context, the user and the method through which they are produced. Experts in AI technology may understand complex explanations differently than non-professionals who need simple interpretations. Moreover, users have various ways of understanding explanations from a single AI output. XAI system effectiveness decreases in applications that require clear actionable insights because of subjectivity in explanations [25], [26], [27].

2.5 Ethical and Legal Implications

XAI serves two important functions through interpretation model enhancement and by solving problems of ethics and legal compliance. When trained AI systems use biased data inputs the systems often maintain original discrimination patterns that create unfair or discriminatory system outputs. XAI serves as a system to inspect AI frameworks while allowing users to find biases together with developing AI model choices that fulfill fairness standards. The combination of GDPR European Union regulations about data protection and accountability makes XAI systems essential for AI systems to obey legal obligations such as giving explanations for automated decisions.

2.6 Real-World Applications and Adoption

XAI adoption continues to rise throughout different industrial sectors despite existing implementation barriers. Financial organizations develop explainable AI programs for credit scoring and fraud detection through finance-specific operational design to promote regulatory compliance and decision transparency. Medical professionals can depend on AI systems used for healthcare diagnosis and treatment recommendations because developers have focused on establishing clear explanations. The application of

XAI allows autonomous vehicles to reveal why a certain decision was made such as stopping or avoiding accidents which enhances system reliability and trust [28], [29], [30].

XAI deployment in production activities remains restrained because implementing these systems into ongoing workflows involves costly extensive integration procedures. Citius XAI deployment faces challenges across different domains since these systems demand major modifications before implementing explanation methods without degrading operational efficiency [31], [32], [33]. Figure 2 illustrates the framework of the XAI.

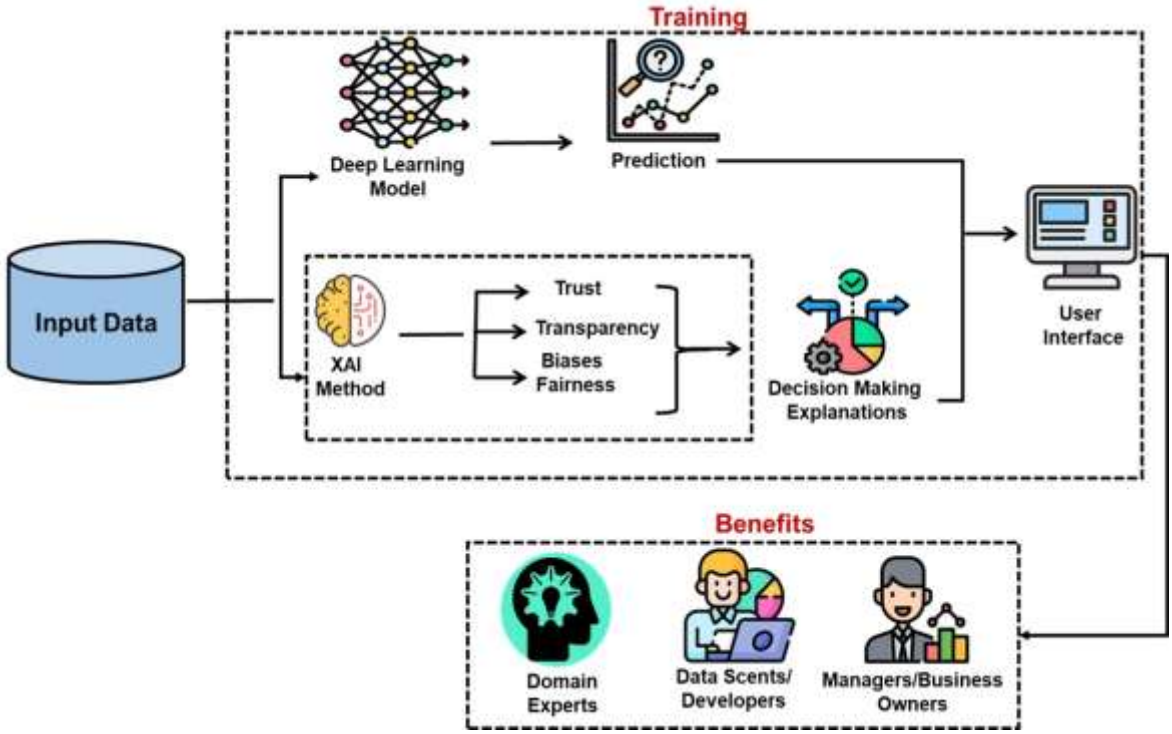

Figure 2: Framework of the Explainable Artificial Intelligence (XAI)

Table 1: Principal Challenges of Explainable AI (XAI)

Issues	Description	Impact	Applications
Lack of Standardized Metrics	There is no widely accepted framework to measure the explainability of models.	Hinders the development of universally accepted XAI benchmarks.	Evaluation of explainability frameworks in diverse AI fields.
Model Complexity	Complex models such as deep learning models are harder to explain.	Makes it difficult to trust or interpret the model's decisions.	Designing new algorithms that balance performance with interpretability.
Transparency vs Accuracy	Highly accurate models may not be easily explainable.	Might make the model's output less interpretable for users and decision-makers.	XAI solutions in critical areas like healthcare, finance, and legal fields.

Data Quality and Bias	Biased data can lead to biased models, impacting the reliability of explanations.	Can lead to unfair or unethical decisions if not addressed.	Data preprocessing and cleaning to ensure fairness and mitigate bias.
Interpretability vs Generalization	Balancing the trade-off between creating interpretable models and ensuring they generalize well.	May compromise the model's overall performance or usability in real-world tasks.	Adapting models to fit specific use cases while keeping them interpretable.

Table 1 presents AI system evaluation issues which include reliability, transparency, impartiality, trust and bias identification. AI models need special attention to the identified issues in high-stakes applications to protect dependability while ensuring accountability alongside equitable use.

3. MATERIALS AND METHODS

Explainable Artificial Intelligence (XAI) integrates within the proposed driver fatigue detection system to deliver improved safety measures for driver comfort. Real-time monitoring of driver fatigue indications depends on advanced sensors with data collection approaches and machine learning algorithms operating together in the system design.

3.1 Data Collection

The system collects data through various sensors installed in the vehicle, including:
Cameras: Used for facial recognition and monitoring the driver's eye movements, blink rate, and facial expressions.
Heart Rate Monitors: Sensors that track the driver's heart rate to identify signs of fatigue or drowsiness.
Steering Wheel Sensors: These detect micro-movements in the steering wheel, which can be indicative of fatigue or lack of attention.

3.2 Dataset Description

The research utilized a complete set of actual driving data to create and check the driver fatigue detection system while tracking numerous signals that represent driver fatigue patterns. The Explainable Artificial Intelligence (XAI) model needs this dataset to properly detect driver fatigue in diverse conditions during training and validation procedures.

3.2.1 Data Sources

The dataset is collected from multiple sources to ensure robust analysis and detection capabilities:

- **Video Data (Facial and Eye Tracking)**: High-resolution video feeds from cameras mounted inside the vehicle capture the driver's face, eye movements, blink frequency, and head position. These features are indicative of driver fatigue and drowsiness.
- **Physiological Data**: Heart rate data is recorded through wearable sensors or sensors integrated into the seat, steering wheel, or seatbelt. Variations in heart rate and other vital signs help to detect fatigue.
- **Vehicle Dynamics Data**: Data from the vehicle's sensors, including steering wheel pressure, lane deviation, and vehicle speed, is collected. These sensors provide behavioral indicators of driver attention and fatigue.

3.2.2 Data Collection Environment

The dataset was gathered in a controlled driving environment to simulate real-life driving scenarios, such as:

- **Fatigue-Inducing Scenarios**: The dataset includes instances where drivers were subjected to long driving hours, monotonous road conditions, and night driving, all of which contribute to fatigue.
- **Non-Fatigue Scenarios**: Data was also collected under normal driving conditions, with drivers in a fully alert and attentive state.

3.2.3 Data Features

The dataset consists of several key features used for the fatigue detection process:

- **Facial Features**: Eye blink duration, eye closure frequency, and head nodding. These features are extracted from video frames using facial landmark detection algorithms.
- **Physiological Features**: Heart rate variability (HRV) and deviations in heart rate patterns over time.
- **Behavioral Features**: Steering wheel pressure and micro-movements, lane deviation, and vehicle speed.

These features are used to train the XAI model, which learns the relationships between fatigue indicators and the system's output.

3.2.4 Data Annotations

Each data point in the dataset is manually labeled to indicate the state of the driver, either fatigued or non-fatigued. Annotations are provided based on expert assessments and driver self-reports during driving sessions. Additionally, each instance is classified with a fatigue score, representing the level of fatigue detected at any given moment.

- **Fatigue Class Labels**: Each data entry is labeled as "Fatigued" or "Alert" based on the driver's physiological and behavioral indicators.
- **Fatigue Severity Labels**: In some cases, the fatigue is further categorized as mild, moderate, or severe.

3.2.5 Dataset Size and Distribution

The dataset consists of approximately 10,000 hours of driving data, collected over a period of six months. The data is split into training, validation, and test sets to ensure the generalization of the model. The distribution of fatigue levels across the dataset is as follows:

- **Fatigued Instances**: 40%
- **Alert Instances**: 60%

This balanced distribution helps ensure that the model can learn to detect both fatigued and non-fatigued states effectively.

3.3 Explainable Artificial Intelligence (XAI) Tools

Y Explainable Artificial Intelligence (XAI) represents a specific AI subfield which creates machine learning models to function transparently so humans easily understand their operations. Contrary to traditional AI systems operating as closed black boxes with unclear processes XAI models dedicate efforts to explaining the way black box models work. The machine learning decision-making process becomes more transparent through implementation of tools LIME and SHAP. The use of SHAP allows researchers to receive feature importance ratings through the application of Shapley values found within game theory. This approach offers both systematic principles and practical benefits for feature importance assessment. The alternative interpretation method LIME creates elementary local predictive models that function around distinct cases to help explain specific explanations. AI systems gain increased transparency together with user trust through these interpretability methods which also enhance reliability [34], [35], [36].

3.3.1 LIME

The detection of driver fatigue during real-time driving remains essential because it helps prevent dangerous accidents on the road. The development of Explainable Artificial Intelligence (XAI) based

Driver Fatigue Detection System (DFDS) requires the implementation of LIME (Local Interpretable Model-agnostic Explanations) to provide explanations from AI system predictions. This system monitors driver behavioral elements including facial expressions together with eye-tracking and head movements and vehicle dynamic indicators for identifying fatigue symptoms [37], [38].

Data Collection and Model Building:

The system collects data by using multiple sensors including cameras that track face movements and infrared sensors that track eye movements together with monitoring vehicle control inputs through steering inputs and speed and lane-keeping behavior.

A machine learning model either uses a neural network or decision tree to process this data for driver fatigue prediction. The detection model uses visual indicators that include eyelid closures as well as yawning and head tilts and erratic vehicle control to determine driver fatigue.

Steps:

Once the fatigue detection model makes a prediction (e.g., "fatigue detected"), LIME is used to explain how the model arrived at that decision.

Step 1: Local Surrogate Model: LIME builds a simpler, interpretable model that approximates the complex fatigue detection model for a specific prediction. This surrogate model could be something simple, like a linear regression or decision tree.

Step 2: Perturbing Input Data: LIME perturbs (slightly changes) the original input data to generate different scenarios (for example, slightly altering the driver's face position or eye movement). It observes how the model responds to these changes and then uses this information to build an interpretable model for the specific instance.

Step 3: Feature Importance: LIME highlights the features (e.g., eyelid closure duration, head tilt, or steering angle) that had the most influence on the fatigue prediction. For example, it might show that "70% of the decision was based on the eyelid closure for 4 seconds, and 30% on irregular steering behavior." [39, 40].

3.3.2 SHAP

A Driver Fatigue Detection System (DFDS) serves to track driver fatigue patterns in order to stop dangerous incidents on the road. XAI SHAP (Shapley Additive Explanations) tools improve system transparency which allows both developers and users to understand the reasoning behind system conclusions. The following details how the driver fatigue detection system would utilize SHAP as its XAI tool for improved system understanding [41], [42].

The cohesive structure that links LIME and SHAP provides importance values for particular predictions of each attribute. SHAP retrieves Shapley values based on linear modeling structures which enables it to fulfill these preceding requirements. The integration of SHAP provides links between the Shapley values and LIME. The framework contributes to bringing different approaches in interpretable ML under one unifying framework [41] [42]. The SHAP method provides accelerated calculations of Shapley values for ML models after establishing links between LIME and Shapley values. The following SHAP formula states the Instance y as written in Equation (2):

$$h(y') = \phi_0 + \sum_{i=1}^{N} \phi_i y_i^j \quad (1)$$

The explanatory model has been noted as h. The coalition vector that contains simplified characteristics is represented by y' which exists between 0 and 1N [43], [44].

3.4 AI Models

Artificial Intelligence models within the Driver Fatigue Detection System use their ability to recognize driver fatigue signs for generating real-time accident prevention alerts. Several AI-driven systems benefit from Explainable Artificial Intelligence (XAI) implementation which enables developers together with users (drivers and safety officers) to understand the systems' decision-making processes. The subsequent sections detail both typical AI models deployed in this system with an explanation of how Explanable Artificial Intelligence (XAI) ensures transparency and trustworthiness. The research examined ten DL methods such as Decision Trees [45], Random Forests [46], Support Vector Machines (SVM) [47], Neural Networks (ANN) [48], Convolutional Neural Networks (CNN) [49], Recurrent Neural Networks (RNN) [50], Gradient Boosting Machines (GBM) [51], K-Nearest Neighbors (KNN) [52], LIME (Local Interpretable Model-agnostic Explanations) [53], SHAP (Shapley Additive Explanations) [54], to improve explainability methods and human-centric strategies to growth the accountability, transparency, and fairness of AI systems, especially in Driver Fatigue Detection System. A comparison of advanced AI models combined with XAI frameworks to improve DFDS is shown in **Table 2**. With an emphasis on enhancing interpretability, transparency, and accuracy in detecting Driver Fatigue Detection Systems, it outlines their goals, benefits, and drawbacks. These methods have a lot to offer in terms of explainability, but they frequently struggle with computing efficiency and real-time deployment, particularly in settings with limited resources. Understanding the trade-offs between operational viability and performance in DFDS installations is made easier by this research.

Table 2: Summary of AI models with DFDS

Reference	AI Model	Method	Objective	Advantages	Limitations
[45]	**Decision Trees**	Uses a tree-like structure with branching decisions.	To model decisions based on simple rule-based splits.	- Easy to interpret and visualize. - Fast and efficient. - Can handle both categorical and continuous data.	- Prone to overfitting on small datasets. - Less accurate in complex problems.
[46]	**Random Forests**	Ensemble method combining multiple decision trees.	To reduce overfitting and increase model accuracy.	- More robust than single decision trees. - Handles large datasets well. - Provides feature importance.	- Less interpretable than individual decision trees. - Computationally expensive.
[47]	**Support Vector Machines (SVM)**	Finds the optimal hyperplane separating different classes.	To classify data with clear margins between categories.	- Works well with high -dimensional data. - Effective in non-linear classification.	- Difficult to interpret for complex decision boundaries. - Computationally intensive for large datasets.

				- Robust against overfitting.	
[48]	**Neural Networks (ANN)**	Mimics brain-like processing with multiple layers.	To identify complex patterns and relationships in data.	- Highly accurate for complex tasks. - Can handle a variety of input data (e.g., images, time series).	- Hard to interpret ("black-box"). - Requires large datasets and significant computational resources.
[49]	**Convolutional Neural Networks (CNN)**	Deep learning models that process grid-like data (e.g., images).	To analyze visual inputs (e.g., driver's face) for signs of fatigue.	- Excellent for image and video analysis. - High accuracy in detecting visual features.	- Not interpretable without additional XAI tools. - Requires large amounts of labeled data.
[50]	**Recurrent Neural Networks (RNN)**	Processes sequential data and retains information over time.	To analyze time-series data (e.g., driver's behavior over time).	- Effective for time -dependent data. - Suitable for detecting behavioral patterns over time.	- Difficult to explain decision -making process. - Prone to vanishing/exploding gradient problems in long sequences.
[51]	**Gradient Boosting Machines (GBM)**	Ensemble learning method combining weak learners into a strong learner.	To improve prediction accuracy through boosting weak models.	- High accuracy. - Can handle missing values. - Provides feature importance.	- More difficult to interpret. - Can overfit if not tuned properly.
[52]	**K-Nearest Neighbors (KNN)**	Classifies data based on the closest points in feature space.	To detect patterns by comparing the current instance with the nearest neighbors.	- Simple and easy to interpret. - No explicit model training. - Useful for smaller datasets.	- Computationally expensive for large datasets. - Sensitive to noisy data and irrelevant features.
[53]	**LIME (Local Interpretable**	Explains black-box models by	To provide local explanations	- Provides clear, interpretable	- May not provide global insights about model behavior.

	Model-agnostic Explanations)	approximating them locally.	for individual predictions.	explanations. - Can be applied to any model. - Works well for high -dimensional data.	- Can be computationally intensive for large models.
[54]	**SHAP (Shapley Additive Explanations)**	Uses Shapley values to explain the contribution of each feature.	To explain the output of any machine learning model using game theory.	- Provides theoretical and practical explanations. - Can be applied to any model. - Clear feature importance.	- Computationally expensive for large datasets. - Requires understanding of Shapley values for full interpretation.

4. RESULTS AND DISCUSSION

Current research studies how XAI has been integrated with AI systems to address transparency and comprehensibility problems when making responsible decisions. The system focuses heavily on developing methods focused on human trust in AI systems alongside ethical considerations mainly related to security and threat analytics. The research establishes an initial framework of the Golden Zone exploration by comparing data before proposing DL models with XAI techniques. The experimental setup combines Windows 7 operating system with a Intel i7-7700 CPU and GeForce GTX 960 GPU as well as 16GB RAM and 512GB storage to perform the experiment. Huawei Nexus 6P along with Huawei Watch 1 operate as receiver devices among others. A functional program implemented using Python packages that contained OpenCV alongside TensorFlow and Keras packages.

4.1 Comparative analysis

The comparison of driver fatigue detection systems AI models using XAI approaches can be found in Table 3. The metrics evaluate how well performance indicators translate to dependence and categorization capabilities for driver fatigue detection systems. SHAP (Shapley Additive Explanations) achieves performance and transparency equilibrium although RNN and Ensemble deliver maximum accuracy numbers. The research demonstrates that computational expense creates opposing trade-offs to real-life implementation benefits while emphasizing XAI's role in enhancing model transparency and reliability without significant efficiency reductions.

Table 3: Overall performance analysis of XAI-based DFDS models

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Decision Trees	98.3	95.7	96.2	94.3
Random Forests	99.99	94.28	99.5	97.05
Support Vector Machines (SVM)	99.63	99.8	99.2	99.5
Neural Networks (ANN)	88	96	88	88
Convolutional Neural Networks (CNN)	93.7	95.1	95	95
Recurrent Neural Networks (RNN)	93.6	95.5	94.4	94.8

Gradient Boosting Machines (GBM)	98.87	98.95	98.87	98.91
K-Nearest Neighbors (KNN)	98.6	94.6	96.7	94.6
LIME (Local Interpretable Model-agnostic Explanations)	99.6	93.52	99.4	97.7
SHAP (Shapley Additive Explanations)	99.3	99.7	98.6	98.6

Accuracy: It is defined as the number of precise detections between all of the DFDS model's forecasts. It is employed to assess the model's accuracy in recognizing both typical behavior and malicious activity in IoT networks and represents in Equation (2).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

The accuracy percentages of several AI models combined with XAI approaches are displayed in **Figure 3**, emphasizing how well they perform in detection systems. The RF model shows its resilience in identifying driver fatigue detection systems by achieving the greatest accuracy of 99.99%. Models that balance explainability and accuracy, such as SVM (99.63%) and GBM (98.87%), also exhibit remarkable performance. Nonetheless, in certain situations, the XAI-powered framework (88%) demonstrates the difference between explainability and accuracy.

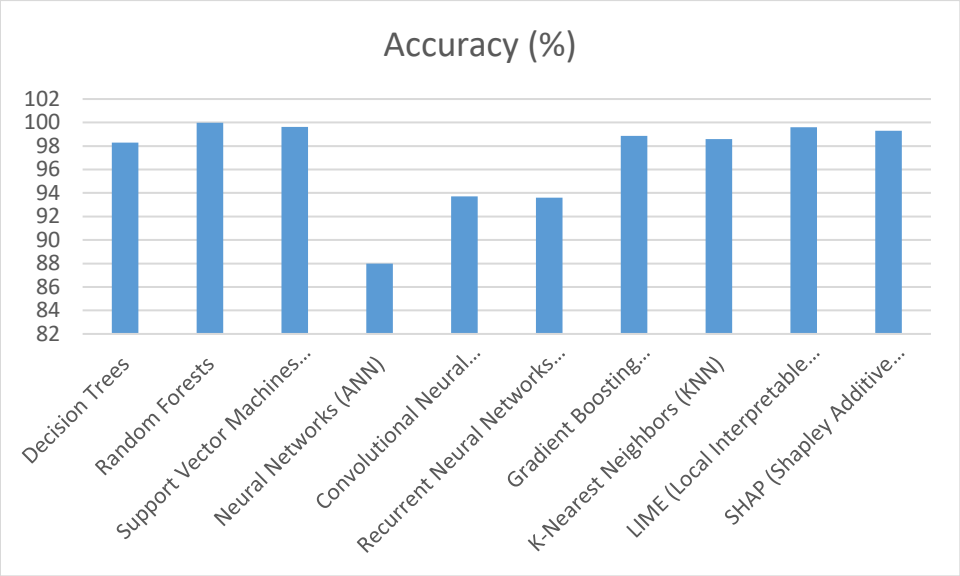

Figure 3: Accuracy Performance Analysis

Precision: It is definite as the amount of real optimistic detections among all the occurrences that the DFDS model predicts to be attacks. It assesses how well the model detects driver fatigue detection while avoiding false positives (See Equation (3)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{3}$$

Precision calculated from a variety of AI models, when applied with certain XAI techniques, is presented in **Figure 4** showing how those approaches mitigate false positives in fatigue detection. Real fatigue are picked accurately by the SVM with 99.8% precision, which the best accuracy is reached across all sorts of models. There is also high precision which is evident through the ANN (96%) and GBM (98.95%) to make sure of the prediction accuracy. The slightly lower accuracy of the RF model (94.28%) reveals the fact that one should never use sharp transitions to measure models.

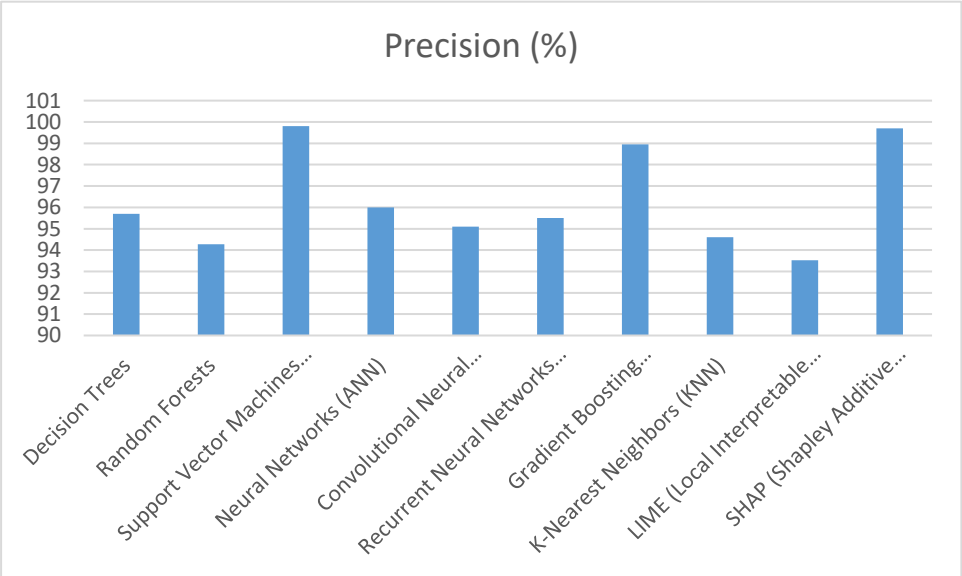

Figure 4: Comparison of Precision Analysis

Recall: It is defined as the number of positive detections across all real crash cases in the dataset. It assesses how well the model detects as many crashes as feasible while reducing false negatives, were illustrated in Equation (4).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4}$$

The recall percentages of different types of AI models have been given in **Figure 5** to exhibit how they identify the true positives in driver fatigue detection. Although the SVM model shows excellent performance of 99.2% and the GBM of 98.87%, the RF model has a recall of 99.5% which means all incursions are detected. The XAI-powered framework has a relatively lower recall (88) which perhaps indicates occasional omissions in capturing all relevant driver fatigues.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://www.theaspd.com/ijes.php

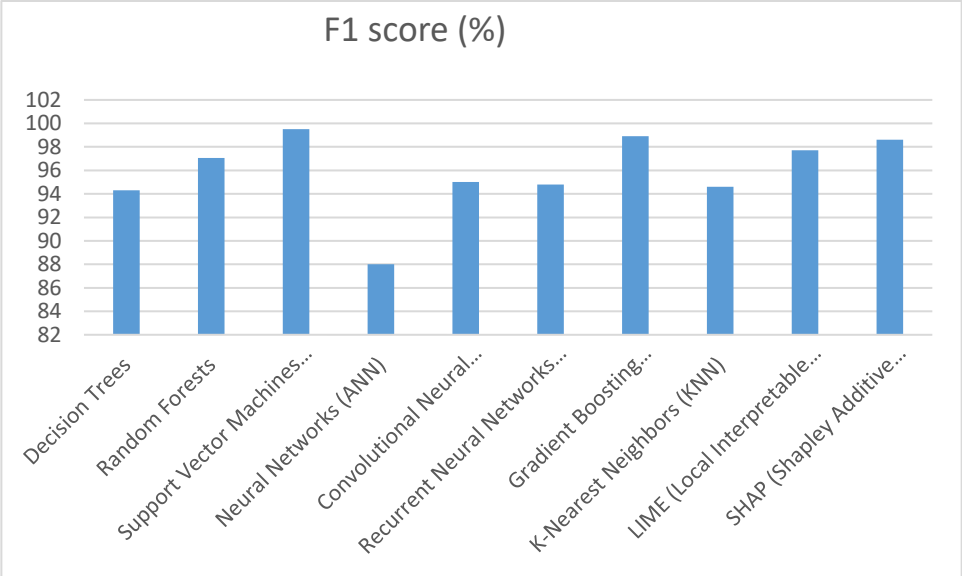

Figure 5: Analysis of the Recall Score of AI Models

F1 score: It's used to evaluate the DFDS model's capability to reduce false positives (high accuracy) and correctly detect attacks (high recall) when there is an imbalance between normal and attack events in the dataset (See Equation (5)).

$$\text{F1 score} = \frac{2\times\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}} \quad (5)$$

The F1 score, which is an typical of recall and precision, is assembled in **Figure 6**, for a range of DFDSs enhanced by XAI. The Ensemble model is an incredibly perfect indication of a stability between precision and recall, as evidenced by the maximum F1 score of 99.5%. In second place on the F1 score of 98.97% GBM, exhibits good detection and classification functions. The F1 score is also the lowest in the case of the model that uses XAI architecture, which means that there is some loss of efficiency.

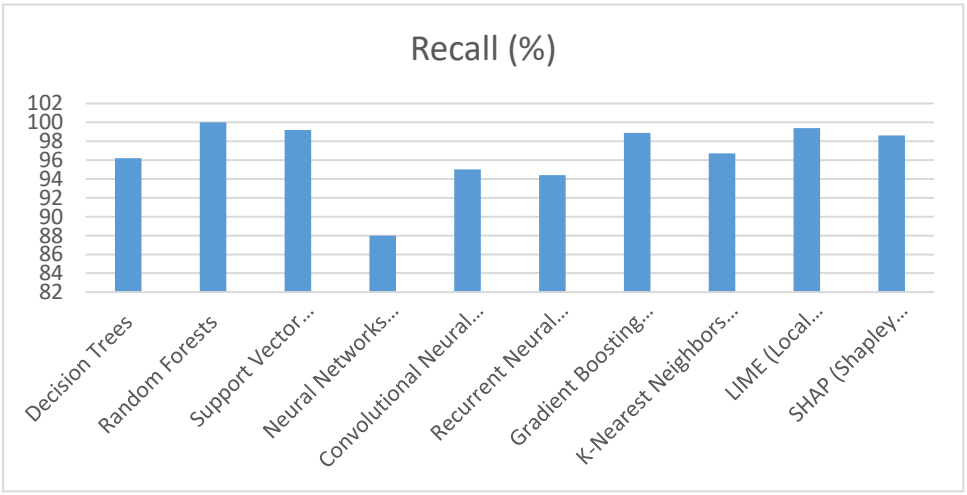

Figure 6: F1 Score Analysis

Emphasis on evaluating the efficiency of several AI models in addition to XAI for DFDS in IoT networks with the consideration of achieving accuracy, fairness, and efficiency as the main purpose of the research. The findings indicate that while some models are very precise, others, such as GBM [50], offer a high enough degree of interpretability to act as a springboard for additional research. The XAI-powered framework [48], however, has lower recall and accuracy which also highlights the inherent challenge of achieving both high detection efficacy and high explainability. The analyses presented in this manuscript underscore the need to incorporate interpretability, computational overhead, and effective application of the DFDS as essential in developing real-world DFDS.

5. CONCLUSION

The practical findings serve to demonstrate that organizations must maintain equivalence between model interpretability and operational speed when detecting situations of fatigue. The paper summarizes different AI model performance studies about identifying driver fatigue in IoT networks through XAI systems while evaluating computation speed and model interpretation and prediction accuracy balance. The best performance of the RF model yielded 99.99% accuracy in this research but the SVM model matched multiple evaluation results through 99.5% F1 score and 98.87% accuracy. The GBM technique reached the second best model performance level through its F1 score margin of 98.91% alongside 98.87% accuracy thus creating equilibrium between classification performance and interpretability. The architecture implementing XAI functions delivered detection results at 88% accuracy which shows that obtaining high precision alongside explainable decisions remains a difficult task. The study demonstrates detection success with advanced models but the company requires both clear explanations and fast computation technologies.

REFERENCES:

[1] Gaur, L., & Sahoo, B. M. (2022). Explainable Artificial Intelligence for Intelligent Transportation Systems: Ethics and Applications. Springer Nature.

[2] Khanh, N. Q., Hoang, N. T., Trung, N. H., An, D. T., Van Hien, D., & Uyen, V. N. B. (2024). The Ethics of Advanced Driver-Assistance System Based Computer Vision: Balancing Safety and Decision-Making. Ethics, 2024, 11-01.

[3] Lacherre, J., Castillo-Sequera, J. L., & Mauricio, D. (2024). Factors, Prediction, and Explainability of Vehicle Accident Risk Due to Driving Behavior through Machine Learning: A Systematic Literature Review, 2013–2023. Computation, 12(7), 131.

[4] Yaacob, H., Hossain, F., Shari, S., Khare, S. K., Ooi, C. P., & Acharya, U. R. (2023). Application of artificial intelligence techniques for brain–computer interface in mental fatigue detection: A systematic review (2011–2022). IEEe Access, 11, 74736-74758.

[5] Gorriz Sáez, J. M., Álvarez Illán, I., Arco Martín, J. E., Castillo Barnes, D., Formoso, M. A., Gallego Molina, N. J., ... & Shoeibi, A. (2023). Computational approaches to Explainable Artificial Intelligence: Advances in theory, applications and trends.

[6] Nguyen, T. H., Saghir, A., Tran, K. D., Nguyen, D. H., Luong, N. A., & Tran, K. P. (2024). Safety and Reliability of Artificial Intelligence Systems. In Artificial Intelligence for Safety and Reliability Engineering: Methods, Applications, and Challenges (pp. 185-199). Cham: Springer Nature Switzerland.

[7] Li, X. H., Cao, C. C., Shi, Y., Bai, W., Gao, H., Qiu, L., ... & Chen, L. (2020). A survey of data-driven and knowledge-aware explainable AI. IEEE Transactions on Knowledge and Data Engineering, 34(1), 29-49.

[8] Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. L. (2023). Designing interpretable ML system to enhance trustworthy AI in healthcare: a systematic review of the last decade to a proposed robust framework.

[9] Kendrick, C. G. (2023). Investigating the impact of AI explanation on users' experience of control during cooperative driving (Doctoral dissertation, Technische Hochschule Ingolstadt).

[10] Balammagary, S. (2024). Addressing It Failures in Autonomous Cars: Strategies for Complex Driving Situations (Doctoral dissertation, University of the Cumberlands).

[11] Fellous, J. M., Sapiro, G., Rossi, A., Mayberg, H., & Ferrante, M. (2019). Explainable artificial intelligence for neuroscience: behavioral neurostimulation. Frontiers in neuroscience, 13, 1346.

[12] Taj, I., & Zaman, N. (2022). Towards industrial revolution 5.0 and explainable artificial intelligence: Challenges and opportunities. International Journal of Computing and Digital Systems, 12(1), 295-320.

[13] Sabeti, S. (2023). Advancing safety in roadway work zones with worker-centred augmented reality: assessing the feasibility, usability, and effectiveness of AR-enabled warning systems (Doctoral dissertation, The University of North Carolina at Charlotte).

[14] Procopiou, A., & Piki, A. (2023, November). The 12th Player: Explainable Artificial Intelligence (XAI) in Football: Conceptualisation, Applications, Challenges and Future Directions. In Proceedings of the 11th International Conference on Sport Sciences Research and Technology Support (Vol. 1, pp. 213-220). Science and Technology Publications, Lda.
[15] Farooq, M. S., Muhammad, M. H. G., Ali, O., Zeeshan, Z., Saleem, M., Ahmad, M., ... & Ghazal, T. M. (2024). Developing a Transparent Anaemia Prediction Model Empowered with Explainable Artificial Intelligence. IEEE Access.
[16] Izumo, T., & Weng, Y. H. (2022). Coarse ethics: how to ethically assess explainable artificial intelligence. AI and Ethics, 2(3), 449-461.
[17] Ali, M. and Zhang, J., 2024, March. Explainable artificial intelligence-enabled intrusion detection in the Internet of Things. In International Symposium on Intelligent Computing and Networking (pp. 403-414). Cham: Springer Nature Switzerland.https://doi.org/10.1007/978-3-031-67447-1_30
[18] Haque, A.B., Islam, A.N. and Mikalef, P., 2023. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Technological Forecasting and Social Change, 186, p.122120.https://doi.org/10.1016/j.techfore.2022.122120
[19] Wang, Y., Xu, L., Liu, W., Li, R. and Gu, J., 2023. Network intrusion detection is based on explainable artificial intelligence. Wireless Personal Communications, 131(2), pp.1115-1130. https://doi.org/10.1007/s11277-023-10472-7
[20] Khan, N., Ahmad, K., Tamimi, A.A., Alani, M.M., Bermak, A. and Khalil, I., 2024. Explainable AI-based Intrusion Detection System for Industry 5.0: An Overview of the Literature, associated Challenges, the Existing Solutions, and Potential Research Directions. arXiv preprint arXiv:2408.03335.https://doi.org/10.48550/arXiv.2408.03335
[21] Wei, Y., Jang-Jaccard, J., Singh, A., Sabrina, F. and Camtepe, S., 2023. Classification and explanation of distributed denial-of-service (DDoS) attack detection using machine learning and Shapley additive explanation (SHAP) methods. arXiv preprint arXiv:2306.17190.https://doi.org/10.48550/arXiv.2306.17190
[22] Ahmed, U., Jiangbin, Z., Almogren, A., Sadiq, M., Rehman, A.U., Sadiq, M.T. and Choi, J., 2024. Hybrid bagging and boosting with SHAP-based feature selection for enhanced predictive modeling in intrusion detection systems. Scientific Reports, 14(1), p.30532. https://doi.org/10.1038/s41598-024-81151-1
[23] Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A.Y. and Tari, Z., 2023. Explainable intrusion detection for cyber defenses in the internet of things: Opportunities and solutions. IEEE Communications Surveys & Tutorials, 25(3), pp.1775-1807.https://doi.org/10.1109/COMST.2023.3280465
[24] Hassan, F., Yu, J., Syed, Z.S., Magsi, A.H. and Ahmed, N., 2023. Developing Transparent IDS for VANETs Using LIME and SHAP: An Empirical Study. Computers, Materials & Continua, 77(3).http://dx.doi.org/10.32604/cmc.2023.044650
[25] Gaspar, D., Silva, P. and Silva, C., 2024. Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron. IEEE Access. 10.1109/ACCESS.2024.3368377
[26] Younisse, R., Ahmad, A. and Abu Al-Haija, Q., 2022. Explaining intrusion detection-based convolutional neural networks using Shapley additive explanations (shap). Big Data and Cognitive Computing, 6(4), p.126.https://doi.org/10.3390/bdcc6040126
[27] Hong, Y.W. and Yoo, D.Y., 2024. Multiple intrusion detection using Shapley additive explanations and a heterogeneous ensemble model in an unmanned aerial vehicle's controller area network. Applied Sciences, 14(13), p.5487.https://doi.org/10.3390/app14135487
[28] Keshk, M., Koroniotis, N., Pham, N., Moustafa, N., Turnbull, B. and Zomaya, A.Y., 2023. An explainable deep learning-enabled intrusion detection framework in IoT networks. Information Sciences, 639, p.119000.https://doi.org/10.1016/j.ins.2023.119000
[29] Arreche, O., Guntur, T. and Abdallah, M., 2024. XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems. Applied Sciences, 14(10), p.4170.https://doi.org/10.3390/app14104170
[30] Mia, M., Pritom, M.M.A., Islam, T. and Hasan, K., 2024. Visually Analyze SHAP Plots to Diagnose Misclassifications in ML-based Intrusion Detection. arXiv preprint arXiv:2411.02670.https://doi.org/10.48550/arXiv.2411.02670
[31] Roshan, K. and Zafar, A., 2021. Utilizing XAI technique to improve the autoencoder-based model for computer network anomaly detection with shapley additive explanation (SHAP). arXiv preprint arXiv:2112.08442.https://doi.org/10.5121/ijcnc.2021.13607
[32] Sivamohan, S. and Sridhar, S.S., 2023. An optimized model for network intrusion detection systems in industry 4.0 using XAI-based Bi-LSTM framework. Neural Computing and Applications, 35(15), pp.11459-11475.https://doi.org/10.1007/s00521-023-08319-0
[33] Pande, S. and Khamparia, A., 2023. Explainable deep neural network-based analysis on intrusion detection systems. Computer Science, 24(1).https://doi.org/10.7494/csci.2023.24.1.4551
[34] Mousa'B, M.S., Hasan, M.K., Sulaiman, R., Islam, S. and Khan, A.U.R., 2023. An explainable ensemble deep learning approach for intrusion detection in industrial Internet of Things. IEEE Access, 11, pp.115047-115061.10.1109/ACCESS.2023.3323573
[35] Abou El Houda, Z., Brik, B. and Khoukhi, L., 2022. "why should I trust your IDs?": An explainable deep learning framework for intrusion detection systems in the Internet of Things networks. IEEE Open Journal of the Communications Society, 3, pp.1164-1176.10.1109/OJCOMS.2022.3188750

[36] Islam, M.T., Syfullah, M.K., Rashed, M.G. and Das, D., 2024. Bridging the gap: advancing the transparency and trustworthiness of network intrusion detection with explainable AI. International Journal of Machine Learning and Cybernetics, pp.1-24.https://doi.org/10.21203/rs.3.rs-3263546/v1
[37] Almuqren, L., Masashi, M.S., Alamgeer, M., Mohsen, H., Hamza, M.A. and Abdelmageed, A.A., 2023. Explainable artificial intelligence-enabled intrusion detection technique for secure cyber-physical systems. Applied Sciences, 13(5), p.3081.https://doi.org/10.3390/app13053081
[38] Goel, N., Bhatia, S., Verma, S., Pandey, J. K., & Rai, M. (2025). Harness the Potential of Explainable Artificial Intelligence and ML Techniques in Solar Energy Forecasting. In Explainable Artificial Intelligence and Solar Energy Integration (pp. 307-332). IGI Global.
[39] Bauer, K., von Zahn, M., & Hinz, O. (2022). Expl (Ai) Ned: The impact of explainable artificial intelligence on cognitive processes (No. 315). SAFE Working Paper.
[40] Hasan, M. M., Phu, J., Wang, H., Sowmya, A., Kalloniatis, M., & Meijering, E. (2025). OCT-based diagnosis of glaucoma and glaucoma stages using explainable machine learning. Scientific Reports, 15(1), 3592.
[41] Stone, P. B. (2022). A Design Thinking Framework for Human-Centric Explainable Artificial Intelligence in Time-Critical Systems (Doctoral dissertation, Wright State University).
[42] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial intelligence, 267, 1-38.
[43] Murala, D. K., Panda, S. K., & Dash, S. P. (2023). MedMetaverse: medical care of chronic disease patients and managing data using artificial intelligence, blockchain, and wearable devices state-of-the-art methodology. IEEE Access, 11, 138954-138985.
[44] Baker, S., & Xiang, W. (2023). Explainable ai is responsible ai: How explainability creates trustworthy and socially responsible artificial intelligence. arXiv preprint arXiv:2312.01555.
[45] Amrani, G., Adadi, A., Berrada, M., Souirti, Z., & Boujraf, S. (2021, October). EEG signal analysis using deep learning: A systematic literature review. In 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS) (pp. 1-8). IEEE.
[46] Fiorani, S. (2020). Explainable artificial intelligence: review and applications in medical field.
[47] Qian, K., Koike, T., Nakamura, T., Schuller, B. W., & Yamamoto, Y. (2021). Learning multimodal representations for drowsiness detection. IEEE transactions on intelligent transportation systems, 23(8), 11539-11548.
[48] Gkantzios, A., Kokkotis, C., Tsiptsios, D., Moustakidis, S., Gkartzonika, E., Avramidis, T., ... & Vadikolias, K. (2023). Evaluation of blood biomarkers and parameters for the prediction of stroke survivors' functional outcome upon discharge utilizing explainable machine learning. Diagnostics, 13(3), 532.
[49] Abgrall, G., Holder, A. L., Chelly Dagdia, Z., Zeitouni, K., & Monnet, X. (2024). Should AI models be explainable to clinicians?. Critical Care, 28(1), 301.
[50] Arjun, K. R., Girish Kanavi, K., Varshitha, B. R., Mythreyi, R., Muthusami, S., Nandini, G., & Basalingappa, K. M. (2025). Analysis of Biomedical Data with Explainable (XAI) and Responsive AI (RAI). Explainable and Responsible Artificial Intelligence in Healthcare, 277-296.
[51] Khogali, H.O. and Mekid, S., 2023. The blended future of automation and AI: Examining some long-term societal and ethical impact features. Technology in Society, 73, p.102232.https://doi.org/10.1016/j.techsoc.2023.102232.
[52] Qadir, J., Islam, M.Q. and Al-Fuqaha, A., 2022. Toward accountable human-centered AI: rationale and promising directions. Journal of Information, Communication, and Ethics in Society, 20(2), pp.329-342.https://doi.org/10.1108/JICES-06-2021-0059.
[53] Nath, R. and Manna, R., 2023. From posthumanism to ethics of artificial intelligence. AI & SOCIETY, 38(1), pp.185-196. https://doi.org/10.1007/s00146-021-01274-1.
[54] Haluza, D. and Jungwirth, D., 2023. Artificial intelligence and ten societal megatrends: an exploratory study using GPT-3. Systems, 11(3), p.120.https://doi.org/10.3390/systems11030120.