

Hybrid neuro-fuzzy and recurrent network for sequential modeling

Mouhcine Lakhroufi, El Mehdi Mellouli

Laboratory of Engineering, Systems and Applications, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Article Info

Article history:

Received Sep 23, 2025

Revised Jul 22, 2026

Accepted Aug 20, 2026

Keywords:

Adaptive neuro-fuzzy inference system

Driver drowsiness detection

Expectation maximization

Gaussian mixture model

Long short-term memory

Neuro-fuzzy systems

ABSTRACT

Driving drowsiness is a major cause of road accidents, underscoring the need for reliable detection systems to enhance road safety. This paper introduces a novel hybrid architecture, adaptive neuro-fuzzy inference system (ANFIS) (EM)-LSTM, which integrates the complete fuzzy inference process with recurrent modeling, combining interpretability and sequential learning. In this design, the premise parameters are initialized and refined using a Gaussian mixture model with expectation-maximization (EM-GMM), while the consequent parameters are optimized independently during training. Each fuzzy rule generates an explicit output at the frame level, which is then organized into temporal windows and processed by a long short-term memory (LSTM) to capture both short- and long-term behavioral dependencies. Applied to driver drowsiness detection on the public NTHU-DDD dataset, the proposed model achieves substantial improvements: a 37% relative gain in accuracy and 45% in F1-score compared to standalone ANFIS, as well as a 10% accuracy and 14% F1-score improvement over standalone LSTM. These results confirm the effectiveness and novelty of coupling fuzzy inference with sequential modeling, demonstrating its potential for deployment in safety-critical applications such as intelligent transportation systems.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Mouhcine Lakhroufi

Laboratory of Engineering, Systems and Applications, National School of Applied Sciences

Sidi Mohamed Ben Abdellah University

Fez, Morocco

Email: mouhcine.lakhroufi@usmba.ac.ma

1. INTRODUCTION

Driver drowsiness is one of the major causes of road crashes and a serious public safety concern. As a countermeasure, several drowsiness measurement methods have been proposed in the literature, and these tend to cluster into four broad classes: physiological measures electroencephalogram, electrooculogram, and electrocardiogram (EEG, EOG, and ECG), which provide highly precise data but remain intrusive [1]-[3], behavioral measures (eye blinks, yawns, and head movements), which are non-intrusive and based on visual observations [4], [5], subjective measures (questionnaires and self-ratings), and vehicle-based measures (lane departure, steering behaviors, and speed analysis), which rely on driving dynamics to estimate fatigue [6], [7]. Although physiological approaches provide high accuracy, they are costly, intrusive, and therefore unsuitable for real-world driving conditions. In contrast, behavioral measures are inexpensive, non-intrusive, and directly correlated with visible signs of fatigue. For these reasons, they are particularly well suited for integration into embedded drowsiness detection systems.

Recurrent neural networks (RNNs), introduced in [8], have become a reference model for sequential data processing thanks to their ability to model temporal dependencies in signals. Among their variants, long short-term memory (LSTM), proposed in [9], and bidirectional LSTM (BiLSTM), introduced in [10], stand out for their ability to capture both short- and long-range dependencies, making them particularly effective for complex sequence analysis. These architectures have been widely applied to drowsiness detection. For instance, [11] demonstrated the relevance of RNNs for video sequence analysis, while [7], [12] combined LSTMs with convolutional networks to jointly exploit spatial and temporal information. Despite these advances, several limitations remain: their internal mechanisms are often opaque, making interpretation difficult. They often depend on large datasets for optimal accuracy, and their complex structures are highly sensitive to hyperparameters, which complicates optimization [13]. To overcome these drawbacks, fuzzy systems provide a complementary approach.

Neuro-fuzzy systems, such as the adaptive neuro-fuzzy inference system (ANFIS) proposed by Jang [14], combine the adaptive learning capability of neural networks with the transparency of fuzzy rules derived from Zadeh's fuzzy logic [15]. They enable uncertainty in data to be modeled while providing an explicit decision-making process through interpretable if-then rules. This property gives ANFIS a significant advantage in sensitive applications where the justification of results is as important as their accuracy. In the field of drowsiness detection, several studies have highlighted the relevance of these approaches. ANFIS has been applied to estimate vigilance levels from entropic features extracted via wavelets [16], to steering-based detection through adaptive feature selection [17], and to wearable systems exploiting fuzzy logic [18]. More recently, methodological advances reported in recent studies [19], [20] have demonstrated the potential of hybrid neuro-fuzzy systems combined with recurrent networks (e.g., BiLSTM), both for drowsiness detection and for other critical tasks such as violence detection in videos. Nonetheless, when used in isolation, these models remain limited in their ability to capture complex temporal dependencies. This limitation naturally motivates to hybridization with sequential models, combining the transparency of fuzzy inference with the predictive power of RNNs.

In parallel, the emergence of neuro-symbolic and fuzzy-attention models has reinforced the trend toward integrating interpretable reasoning and uncertainty management within deep architectures. Recent studies, such as TSK-type fuzzy neural networks for fatigue estimation from EEG signals [21], [22], and fuzzy deep attention mechanisms applied to driver monitoring [23], have shown that embedding fuzzy logic within attention layers improves both robustness and interpretability. In addition, hybrid transformer-based architectures [24] have demonstrated strong capabilities for modeling complex spatio-temporal dependencies in visual and physiological signals, achieving state-of-the-art performance in drowsiness detection. These advances illustrate a growing convergence toward hybrid neuro-symbolic and attention-based paradigms, aiming to combine symbolic interpretability with deep temporal learning.

In this work, we introduce a hybrid ANFIS(EM)-LSTM architecture for driver drowsiness detection. This approach combines the interpretability of fuzzy systems with the ability of recurrent networks to model complex temporal dynamics. The ANFIS, whose membership functions are initialized using a Gaussian mixture model (GMM), transforms behavioral indicators (eye aspect ratio (EAR) and mouth aspect ratio (MAR)) into explicit fuzzy rules capable of representing data uncertainty. These rules, once weighted and normalized, are aggregated and passed to an LSTM network, which learns temporal dependencies to ensure reliable classification of the driver's vigilance state.

The main contributions of this work can be summarized as follows:

- The design of a hybrid ANFIS(EM)-LSTM architecture that combines fuzzy premises derived from EM-GMM with an LSTM to effectively capture temporal dependencies.
- The explicit decoupling of premise and consequent parameters during training, preserving the interpretability of the rules while ensuring stable optimization.
- Experimental validation on the NTHU-DDD dataset, demonstrating strong performance (accuracy $\approx$ 96% and F1-score $>$ 93%) and a comparative analysis showing that our approach achieves a balanced trade-off between interpretability and predictive performance.

The structure of this paper is as follows: section 2 introduces the theoretical foundations, section 3 explains the architecture of the proposed hybrid model, section 4 reports and discusses the experimental outcomes, and section 5 summarizes the main conclusions and future directions.

2. THEORETICAL BASIS

2.1. Adaptive neuro-fuzzy inference system

ANFIS is organized as a five-layer framework [25], blending neural networks and fuzzy logic, so that it may learn the data automatically and generate fuzzy rules. In the first layer, each input variable is processed through parameterized membership functions, and their outputs combine so that the rule's firing strength may be computed. These levels of activation become normalized, and each rule is of a linear

consequent of the Takagi–Sugeno type, and the overall network output is obtained by aggregating the individual rule outputs. These rules are generally expressed in the following form:

$$\text{If } (x_1 \text{ is } A_i) \text{ and } (x_2 \text{ is } B_j) \text{ then } f = p \cdot x_1 + q \cdot x_2 + r$$

The overall ANFIS structure is illustrated in Figure 1, where input features are fuzzified, combined through rule activation, normalized, and aggregated through linear consequents to produce the final decision.

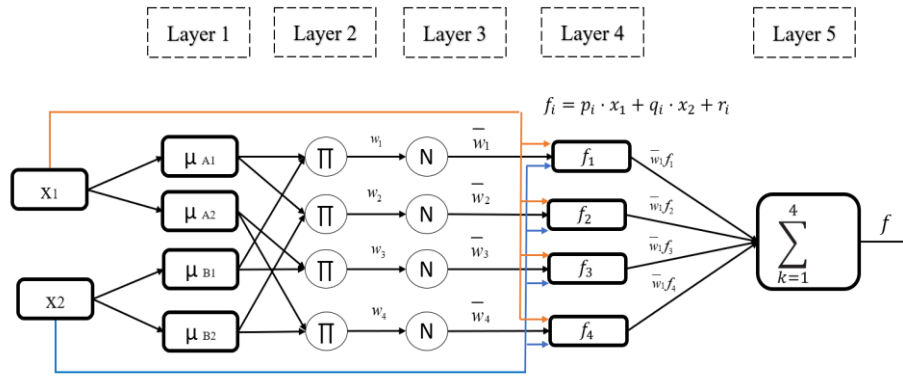


Figure 1. Architecture of the ANFIS

2.2. Expectation-maximization for Gaussian mixture models

GMMs are often applied to describe complex data distributions by expressing them as a sum of several Gaussian components. Each component represents a latent subgroup of the data, defined by its mean and covariance, while the mixture weights indicate its relative importance. This probabilistic framework provides a flexible way to partition the input space and is particularly well suited to capturing multimodal structures in the data.

The parameters of a GMM are typically estimated with the expectation–maximization (EM) algorithm [19], which iteratively by alternating between two steps:

- E-step (expectation): compute the posterior probability w_m^t that a data point x^t belongs to Gaussian component m . These probabilities are normalized such that $\sum_{m=1}^M w_m^t = 1$, yielding a soft assignment of each sample across the mixture components.
- M-step (maximization): update the parameters of each Gaussian component using the weighted statistics of the data. The mean and covariance of component m are computed as:

$$c_m^{new} = \frac{1}{T_m} \sum_{t=1}^T w_m^t \cdot x^t, T_m = \sum_{t=1}^T w_m^t,$$

$$\Sigma_m^{new} = \frac{1}{T_m} \sum_{t=1}^T w_m^t \cdot (x^t - c_m^{new})(x^t - c_m^{new})^T$$

These steps are repeated until convergence, ensuring that the estimated mixture maximizes the likelihood of the observed data.

In the proposed ANFIS(EM)–LSTM model, the Gaussian membership functions of the fuzzy system are initialized and optimized through this EM-GMM procedure. This ensures that the premise parameters $(c_{k,j}, \sigma_{k,j})$ are statistically aligned with the distribution of the input features EAR and MAR, improving interpretability and training stability.

2.3. Long short-term memory

The LSTM network is a type of recurrent neural architecture developed to overcome the vanishing gradient problem that affects traditional RNNs. By introducing a memory cell along with input, forget, and output gates, it can control the flow of information, making it possible to capture both short and long-term dependencies in sequential data and to maintain stable training.

At each time step t , the input vector $z(t)$ passes through these gates, which control the integration of new information, the retention of previous knowledge, and the update of the hidden state. The corresponding update equations are given in (1), where c_t denotes the memory cell state and h_t the hidden state at time t . The schematic representation of these operations is shown in Figure 2, illustrating the flow of information within the LSTM cell.

$$\begin{aligned}
 i_t &= \sigma(W_i z(t) + U_i h_{t-1} + b_i) \\
 f_t &= \sigma(W_f z(t) + U_f h_{t-1} + b_f) \\
 o_t &= \sigma(W_o z(t) + U_o h_{t-1} + b_o) \\
 \tilde{c}_t &= \tanh(W_c z(t) + U_c h_{t-1} + b_c) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned} \tag{1}$$

Through this mechanism, the LSTM retains relevant historical information while filtering out irrelevant patterns, making it well suited for sequential modeling of driver behavior. Within our hybrid framework, it is specifically used to capture the sequential dynamics of the fuzzy-rule-based features, allowing the network to identify both short-term fluctuations and long-term patterns associated with the onset of drowsiness (see section 3).

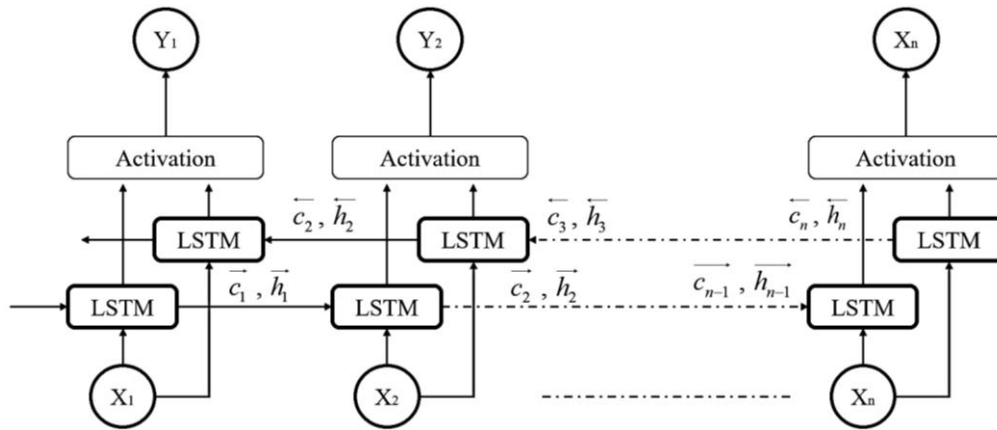


Figure 2. General structure of an LSTM network

3. METHOD

3.1. Overview of hybrid architecture

The proposed ANFIS(EM)–LSTM framework integrates the interpretability of fuzzy inference systems with the sequential modeling capabilities of RNNs. The architecture operates in two main stages. In the first stage, an ANFIS module processes frame-level input features. The raw inputs (x_1, x_2) , here represented by the EAR and MAR, are fuzzified using Gaussian membership functions whose premise parameters are estimated through the EM-GMM algorithm. Each fuzzy rule produces an activation weight, which is normalized and combined with a linear Takagi–Sugeno consequent. This results in extended feature vectors that preserve both the interpretability of fuzzy rules and their statistical consistency with the data distribution.

In the second stage, the vectors $z(t)$ are not used directly for classification but are concatenated into temporal windows and provided as input sequences to the LSTM. In this configuration, ANFIS serves as a frame-level feature extractor, while the LSTM is responsible for temporal modeling. The recurrent network updates its hidden and cell states at each time step, as described in Figure 2, which enables it to capture both short-term fluctuations and long-term dependencies in driver behavior. Finally, the temporal representation is mapped into a probability through a sigmoid output layer (see subsection 3.5), interpreted as the likelihood of drowsiness and ensuring a binary yet interpretable prediction suitable for safety-critical applications.

The overall workflow of the proposed model is illustrated in Figure 3, which highlights the integration of EM-GMM-based premise estimation, fuzzy rule consequents, and sequential modeling through LSTM. This hybrid design combines interpretability at the rule level with adaptability to temporal dynamics, making it particularly suitable for driver drowsiness detection.

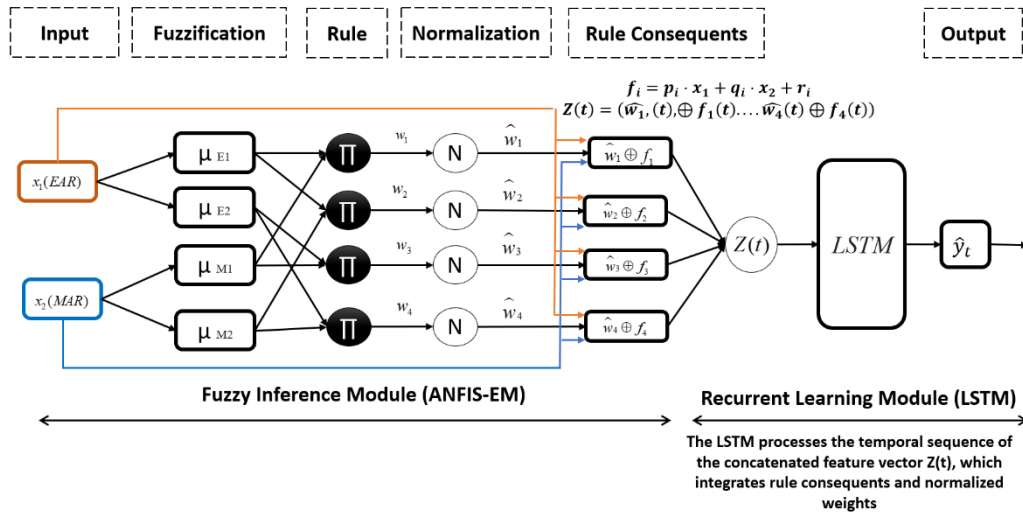


Figure 3. Architecture of the proposed 4-rule ANFIS learning via EM-GMM premises and LSTM sequential modeling

3.2. Fuzzification and premise parameter estimation

In the first stage, to avoid the computationally expensive backpropagation process, the premise parameters of ANFIS are estimated using an EM-GMM. EM-GMM automatically learns the parameters of the membership functions (centers and variances), providing a probabilistic fuzzification that is well adapted to the data without requiring gradient-based optimization. Given an input vector $x_t = [x_{t,1}, x_{t,2}, \dots, x_{t,N}]$ each Gaussian membership function for rule k and feature j is defined as (2):

$$\mu_{k,j}(x_{t,j}) = \exp\left(-\frac{1}{2}\left(\frac{x_{t,j}-c_{k,j}}{\sigma_{k,j}}\right)^2\right) \quad (2)$$

where $c_{k,j}$ and $\sigma_{k,j}$ are the mean and standard deviation estimated by EM-GMM for component k of feature j .

3.3. Rule activation and normalization

The firing strength $w_k(t)$ of each rule k at time t is computed by aggregating the MFs of all features using the product operator.

$$w_k(t) = \prod_{j=1}^N \mu_{k,j}(x_{t,j}) \quad (3)$$

The normalized firing strength $\hat{w}_k(t)$ is then:

$$\hat{w}_k(t) = \frac{w_k(t)}{\sum_{r=1}^K w_r(t)} \text{ with } \hat{v}_k(t) = 0, \sum_{k=1}^K \hat{w}_k(t) = 1 \quad (4)$$

where K is the total number of fuzzy rules.

3.4. Sugeno consequent layer

We adopt a linear consequent per rule (first-order Sugeno).

$$f_k(t) = \theta_k^T \bar{x}_t, \bar{x}_t = \begin{bmatrix} x_t \\ 1 \end{bmatrix} \in \mathbb{R}^{N+1}, \theta_k = (p_k, q_k, r_k) \in \mathbb{R}^{N+1} \quad (5)$$

Stacking outputs by rule.

$$F(t) = [f_1(t), \dots, f_R(t)]^T \in \mathbb{R}^R \quad (6)$$

The extended vector z is obtained by concatenating the per-rule consequent outputs with their corresponding normalized weights, preserving the interpretability of the fuzzy inference structure.

$$z(t) = [F(t) \oplus \hat{w}(t)] \in \mathbb{R}^{2R} \quad (7)$$

3.5. Temporal windowing and sequential modeling with long short-term memory

The extended fuzzy-transformed vectors $z(t)$, defined in (1) are grouped into temporal windows of length L to form the sequences.

$$z_t = [z(t - L + 1), \dots, z(t)] \in \mathbb{R}^{L \times 2R} \quad (8)$$

As defined in (8), these sequences are then fed into an LSTM, whose internal mechanism outlined in (1) updates the hidden and cell states (h_t, c_t) over time. In the classical formulation, the input would be raw feature vectors z_t , but in our hybrid architecture this role is replaced by the fuzzy-transformed vectors $z(t)$, which encapsulate the normalized rule activations and consequent outputs generated by ANFIS at each frame. Through this recurrent process, the network captures both short-term fluctuations and long-term dependencies in driver behavior. The resulting temporal representation is finally passed through a sigmoid output layer given in (9), which produces a probability interpreted as the likelihood of the drowsy state, while the normal state is implicitly represented by the complementary probability $1 - \hat{y}_t$. This ensures that the model output is always non-negative and restricted to the two classes only.

$$\hat{y}_t = \sigma(w_o^\top \hat{h}_t + b_o) \in [0, 1] \quad (9)$$

3.6. Learning objective and parameter optimization

The network is optimized using the binary cross-entropy (BCE) loss.

$$\mathcal{L} = -\frac{1}{M} \sum_{t \in \mathcal{S}} [y_t \log \hat{y}_t + (1 - y_t) \log(1 - \hat{y}_t)] \quad (10)$$

where $\mathcal{S}$ is the set of valid prediction indices and $M = |\mathcal{S}|$.

In the proposed ANFIS(EM)–LSTM model, two categories of parameters are clearly distinguished.

- Premise parameters ($c_{k,j}, \sigma_{k,j}$) which define the Gaussian membership functions, are estimated in an unsupervised manner using the EM algorithm applied to EM-GMM. This preprocessing step provides robust fuzzy partitions that are statistically consistent with the underlying data distribution. Once estimated, these parameters are kept fixed to preserve stability and interpretability of the fuzzy rule base.
- Consequent parameters θ_k , associated with the linear Takagi–Sugeno functions, are learned in a supervised fashion. Unlike the premises, these parameters are updated automatically through error backpropagation propagating through the LSTM. The Adam optimizer is employed to adjust both the consequent parameters and the recurrent network weights, ensuring dynamic adaptation to temporal sequences.

Formally, the update of each vector θ_k results from the gradient of the BCE loss expressed in (11) with respect to the local consequents $f_k(t)$ given in (5).

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_{t \in \mathcal{S}} \frac{\partial \mathcal{L}}{\partial f_k(t)} \bar{x}_t \quad (11)$$

where $\bar{x}_t = [x_t, 1]^\top$ denotes the extended input vector. Consequently, the consequent parameters are adjusted proportionally to the prediction error and to the corresponding input features.

This methodological separation guarantees a complementary role: the statistical robustness of EM-GMM for estimating the premise parameters and the flexibility of supervised learning for optimizing the consequents and modeling temporal dependencies. As a result, the proposed hybrid model benefits both from enhanced interpretability, provided by the fuzzy-rule structure, and from improved predictive performance, owing to the integration of temporal dynamics through the LSTM.

3.7. Experimental setup

All experiments were conducted on Google Colab under a Python 3 environment, which was installed with a NVIDIA Tesla T4 GPU. It relied on TensorFlow (v2.x) and Keras for modeling neural networks, scikit-learn for data pre-processing (normalization, parameter estimation of EM-GMM, and metric calculation), and Matplotlib/Seaborn for learning curve, confusion matrix, and ROC plotting.

Concerning the hyperparameters, the ANFIS(EM)–LSTM model was initialized with two Gaussian membership functions per input variable, thus obtaining four fuzzy rules overall. Temporal relations were captured through the formation of input sequences by sliding windows of ten contiguous frames, while the LSTM network was utilized with 32 hidden units. The learning process was conducted with the Adam learner

that utilized a learning rate of 0.005 fixed. The loss function was set to BCE and the learner was allowed to run to a maximum of 100 epochs. To avoid overfitting, an early stopping criterion with a patience of ten epochs was implemented. The batch size was utilized with a value of 32 for both the independent ANFIS and the coupled ANFIS(EM)–LSTM models. The choice succeeded in keeping the learner stable during the learning process, and this is further evident through the learning curves.

4. RESULTS AND DISCUSSION

4.1. Dataset and features

Although the original dataset consists of video sequences in AVI format (640×480 pixels, 30 frames per second, without audio), our study relies on a numerical representation derived from the NTHU-DDD dataset. Each frame was processed to extract facial landmarks, particularly around the eyes and mouth, using the CNN-based landmark detector from the dlib library [26]. These landmarks [27] were then used to compute behavioral indicators of drowsiness. The EAR, as defined in [28], measures the degree of eye openness and allows the detection of both prolonged closures and frequent blinking. Similarly, the MAR quantifies mouth opening allowing the identification of yawning episodes [28], a common indicator of driver drowsiness.

The extracted features were compiled into a CSV file, where each row represents a specific time instant and includes the corresponding visual variables together with a binary label indicating the driver's state (0=alert and 1=drowsy). In addition, the dataset provides valuable contextual information such as the driver's status (drowsy or non-drowsy), the recording condition (with glasses or without glasses), and the observed behavior, covering several drowsiness-related patterns like sleepy combination, slow blink with nodding, and yawning. This structured, time-aligned representation enables the construction of temporal sequences that are well-suited for sequential learning models and are used as inputs to the proposed ANFIS(EM)–LSTM architecture. By transforming raw video data into meaningful behavioral indicators, the proposed approach strikes a balance between descriptive richness and computational efficiency, reinforcing the suitability of the NTHU-DDD dataset for automatic driver drowsiness detection.

4.2. Evaluating model performance

To evaluate the ANFIS(EM)–LSTM model's ability to detect driver drowsiness, we looked at both regression and classification metrics. For regression, we used the mean squared error (MSE) defined in (12) to capture how much the predictions deviate from the actual values on average, giving a sense of the model's overall accuracy. For classification, we considered several complementary measures. Accuracy (13) tells us the proportion of correct predictions; precision (14) reflects how well the model avoids false alarms; recall (15) indicates how effectively it identifies all drowsy cases; and the F1-score (16), which combines precision and recall, offers a balanced view, particularly when the classes are unevenly distributed. By combining these metrics, the evaluation offers a comprehensive perspective: MSE captures the closeness of predictions to ground truth, while accuracy, precision, recall, and F1-score highlight the model's discriminative performance. Together, they ensure a robust and reliable assessment of the ANFIS(EM)–LSTM model's effectiveness for driver drowsiness detection.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{act}^{(i)} - y_{pred}^{(i)})^2 \quad (12)$$

$$accuracy = \frac{TN+TP}{TP+FP+TN+FN} \quad (13)$$

$$Precision = \frac{TP}{TP+FP} \quad (14)$$

$$recall = \frac{TP}{TP+FN} \quad (15)$$

$$F1 = 2 \frac{Precision * recall}{Precision + recall} \quad (16)$$

4.3. Presentation and discussions

The experimental results obtained from the public NTHU-DDD dataset are illustrated through a series of figures that visualize and compare the performance of different models. This was selected because it is a recognized benchmark to compare detection methods for drowsiness with a high number of varied situations (with/without glasses, yawn, slow blinking, and head nodding). Figures 4 and 5 show the

training/validation accuracy and loss functions for the standalone LSTM approach, while Figures 6 and 7 present those measures for the standalone ANFIS approach. Finally, Figures 8 and 9 present those curves for our hybrid approach based on ANFIS(EM)–LSTM network. These graphs allow us to obtain an assessment based on learning stability and potential signs of overfitting/fluctuations.

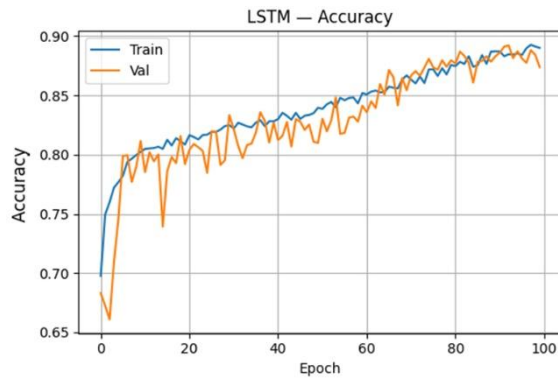


Figure 4. Evolution of training and validation accuracy for the LSTM model

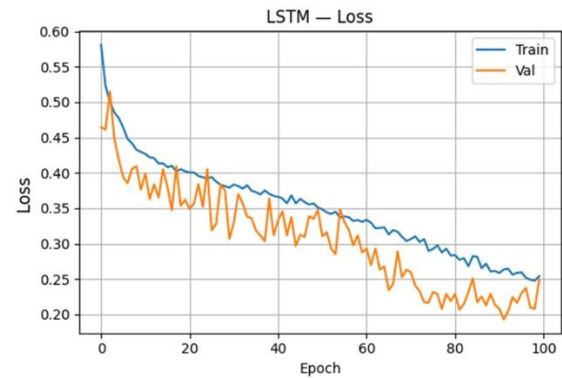


Figure 5. Evolution of training and validation loss for the LSTM model

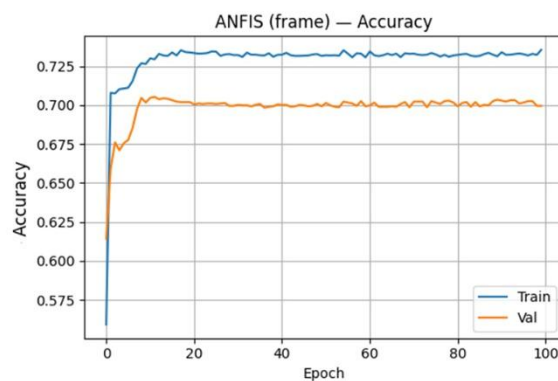


Figure 6. Evolution of training and validation accuracy for the ANFIS model

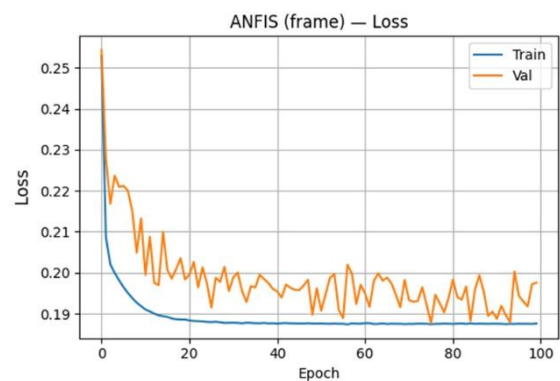


Figure 7. Evolution of training and validation loss for the ANFIS model

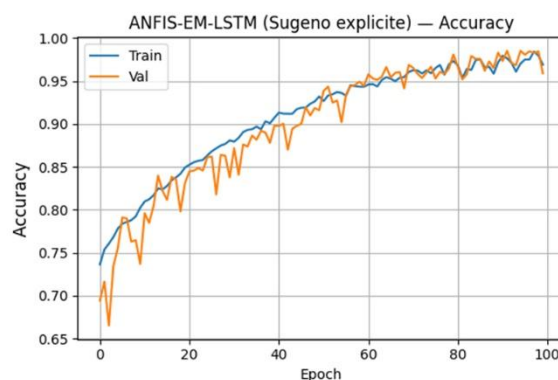


Figure 8. Training and validation accuracy curve of the proposed ANFIS(EM)–LSTM model

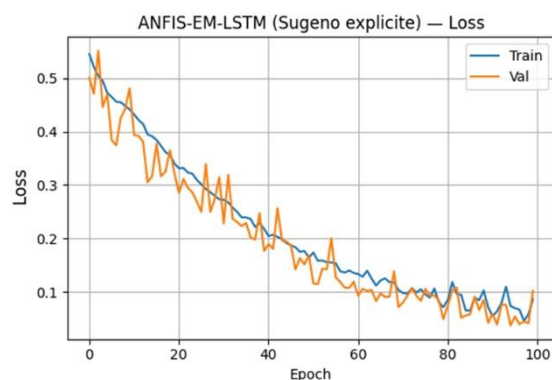


Figure 9. Training and validation loss curve of the proposed ANFIS(EM)–LSTM model

Comparing the classification performance plots in Figures 4, 6, and 8, we observe that the ANFIS(EM)–LSTM overwhelms the two baseline models. The standalone LSTM stabilizes around 87%, reflecting its ability to adequately capture temporal dependencies. The standalone ANFIS plateaus at around 71%, indicating limited performance typical of fuzzy modeling when used alone and without consideration of temporal behavior. The hybrid ANFIS(EM)–LSTM framework combines the advantages of both paradigms and achieves a superior stabilized accuracy of nearly 98%. The notable improvement shows added value by combining explicit fuzzy rules with sequential learning.

The loss curves in Figures 5, 7, and 9 indicate clear convergence behavior of the models. The LSTM alone exhibits a smooth decrease in loss before leveling off at 0.25 toward the end of training and validation. The standalone ANFIS shows less stability, with loss oscillations stabilizing around 0.19, indicating more difficult generalization. However, ANFIS(EM)–LSTM converges fast and stably with loss monotonically decreasing to reach 0.1 at steady state. These observations provide evidence to show that combining statistical premise estimation (by EM-GMM), supervised optimization of consequents, and sequential modeling with LSTM achieves steady convergence while controlling overfitting.

4.4. Result analysis

An examination of the loss curves in Figures 5 and 7 reveals a clear difference between the baseline models. The standalone ANFIS converges rapidly; within the first 20 epochs, its loss decreases to around 0.20, reflecting efficient parameter optimization. In contrast, the LSTM converges more slowly, reaching approximately 0.25 only after about 60 iterations. The hybrid ANFIS(EM)–LSTM model leverages the strengths of both approaches; it benefits from the fast convergence of ANFIS while exploiting the sequential memory of the LSTM. This complementarity results in stable and robust convergence around 0.10 (Figure 9), indicating a more balanced and well-controlled learning process.

The confusion matrix of the ANFIS(EM)–LSTM model in Figure 10 shows a correct classification rate exceeding 98% for both the “Normal” and “Drowsy” classes. By comparison, the standalone LSTM achieves about 87% overall accuracy but fails to detect nearly 13% of drowsy cases, whereas ANFIS alone reaches only around 71%, with a large proportion of misclassifications. The hybrid model therefore significantly reduces false negatives compared with the other two models, confirming its ability to more accurately distinguish between alert and drowsy states, an aspect that is crucial for road safety applications.

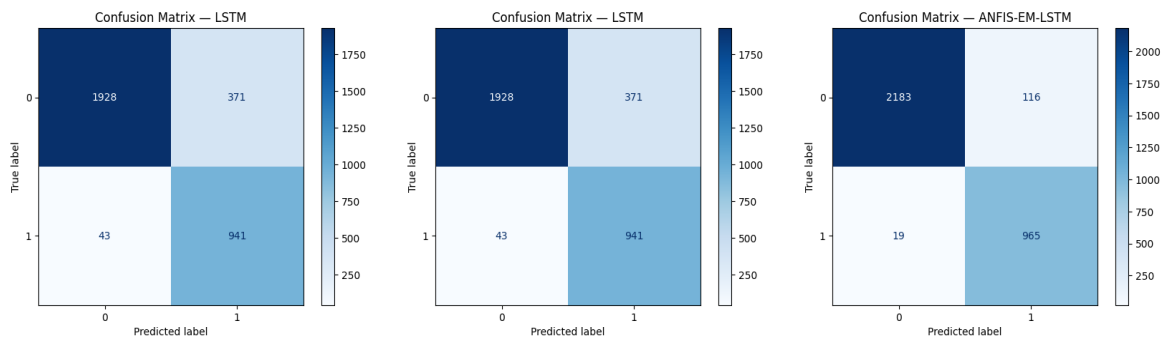


Figure 10. Confusion matrices of LSTM, ANFIS, and ANFIS(EM)–LSTM models

The ROC analysis in Figure 11 further confirms the superiority of the hybrid model ANFIS(EM)–LSTM achieves an AUC above 0.95, compared with 0.87 for LSTM and 0.71 for ANFIS. This demonstrates a stronger discriminative capacity between alert and drowsy states.

After presenting the overall performance metrics of the proposed model, an additional ablation analysis was conducted to better understand the specific contribution of the fuzzy and recurrent components. Three configurations were compared: the standalone ANFIS, the standalone LSTM, and the hybrid ANFIS(EM)–LSTM model. The results summarized in Table 1 show that the standalone ANFIS achieved an accuracy of 0.6996 and an F1-score of 0.6431, confirming its capacity to handle uncertainty but revealing its limitation in modeling temporal dependencies. The LSTM model achieved better results (accuracy=0.8739, F1-score=0.8197) due to its ability to capture sequential behavioral dynamics such as blinking and yawning patterns. When both components were integrated into the hybrid ANFIS(EM)–LSTM architecture, performance increased significantly, reaching an accuracy of 0.9589, an F1-score of 0.9346, and an AUC of 0.9961. These findings clearly demonstrate that the fuzzy inference layer enhances interpretability and

robustness to uncertainty, while the recurrent layer contributes to effective temporal sequence learning. Their joint integration provides complementary advantages, resulting in superior overall accuracy, generalization, and reliability compared with either component used independently.

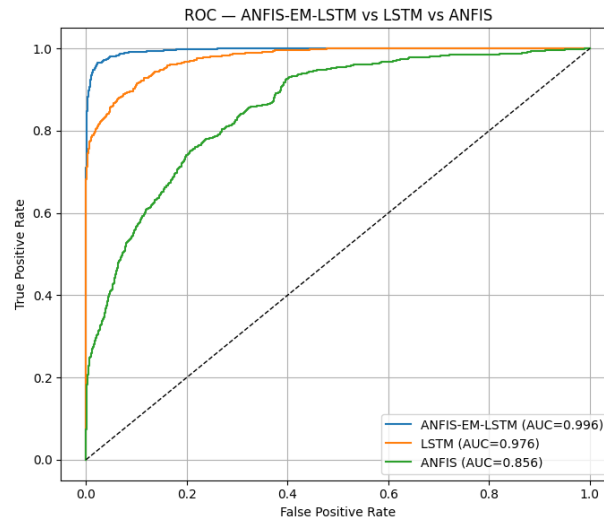


Figure 11. ROC curves of LSTM, ANFIS, and ANFIS(EM)–LSTM models

Table 1. Comparative performance metrics of LSTM, ANFIS, and ANFIS(EM)–LSTM models

Model	Accuracy	Precision	Recall	F1-score	MSE	AUC
ANFIS(EM)–LSTM	0.9589	0.8927	0.9807	0.9346	0.0307	0.9961
LSTM	0.8739	0.7172	0.9563	0.8197	0.0849	0.9760
ANFIS	0.6996	0.5011	0.8973	0.6431	0.1975	0.8561

4.5. Comparative analysis

The experimental results show that our ANFIS(EM)–LSTM model has nearly outperformed all the approaches reported in Table 2. With accuracy close to 96% and an F1-score above 93%, it clearly surpasses traditional methods based solely on LSTM or ANFIS, positioning itself among the best recent techniques by providing significant gains in robustness and reliability. While Ed-Doughmi *et al.* [11] achieved a slightly higher accuracy (97.3%), their F1-score remained limited to 85%, reflecting a less satisfactory balance between precision and recall. In contrast, our model maintains a more consistent equilibrium, reaching both 95.9% accuracy and a 93.4% F1-score, an essential property for reducing false negatives in safety-critical contexts.

Table 2. Comparative analysis of drowsiness detection methods

Authors	Classification methods	Drowsiness measures	Dataset	Performance (%)
Ed-Doughmi [11]	RNN	Driver image	NTHU-DDD	Accuracy=97.3 F1-score=85.0
Rajamohana <i>et al.</i> [29]	CNN, BiLSTM	Facial image, eye blink, and eye closure	Eye data, 1104 images	Accuracy=96.0
Bakheet and Al-Hamadi [5]	HOG + SVM	Driver image and eye pair region	NTHU-DDD	Accuracy=85.62 F1-score=87.84
Dua <i>et al.</i> [30]	Deep CNN	Hand gestures, facial expressions, and head movements	NTHU-DDD	Accuracy=85.0
Celecia <i>et al.</i> [18]	Mamdani FIS+RT	PERCLOS, eye closing duration, and average mouth opening time	300-W dataset	Accuracy=95.5 Precision=93.3
Chen <i>et al.</i> [12]	LSTM, CNN	Eye and face area	NTHU-DDD	Accuracy=93.30
Zhao <i>et al.</i> [31]	3D CNN	Driver face and facial landmarks	NTHU-DDD	Accuracy=88.60
Ghourabi <i>et al.</i> [32]	MLP, KNN	Eye closure and yawning	NTHU-DDD	Accuracy=94.31 F1-score=79.0
Qudus <i>et al.</i> [7]	LSTM, CNN	Blinking, eye closure, saccades, and fixation	Driving simulator (38 participants)	Accuracy=95–97
Our proposed model (2025)	ANFIS(EM)–LSTM	EAR and MAR	NTHU-DDD	Accuracy=95.9 F1-score=93.4

Indeed, in sensitive applications such as drowsiness detection, minimizing false negatives is more crucial than gaining a few points of accuracy, since failing to detect a truly drowsy driver poses a direct accident risk, whereas a false positive only triggers an additional alert, which remains acceptable in practice. Moreover, unlike purely deep architectures (CNN, RNN, or MLP), our approach relies on explicit fuzzy rules, providing greater interpretability and facilitating integration into real-time embedded systems for control, fault detection, and predictive maintenance. Its interpretable rule-based structure and lightweight computational requirements make it a practical and adaptable solution for real-world engineering applications, where reliability, transparency, and real-time operation are essential.

5. CONCLUSION

In this study, we proposed a hybrid ANFIS(EM)–LSTM architecture that integrates fuzzy inference with recurrent learning for driver drowsiness detection. The novelty of the approach lies in coupling a complete fuzzy inference system with an LSTM network, enabling the model to benefit from both the interpretability of fuzzy rules and the temporal modeling capability of recurrent networks. Furthermore, the consequent parameters are optimized independently, improving model stability, flexibility, and computational efficiency. The EM-GMM algorithm is employed to initialize the premise parameters and ensure coherent fuzzy partitions, while the integration with LSTM and the decoupled learning strategy distinguishes the proposed framework.

Experimental results on the public NTHU-DDD dataset show that our hybrid model achieves an accuracy of 95.89% and an F1-score of 93.46%, significantly outperforming both standalone ANFIS and LSTM models. The reduction in false negatives further supports its suitability for safety-critical applications.

Beyond drowsiness detection, the proposed framework can be applied to intelligent systems such as robotics and industrial automation, where interpretability, uncertainty modeling, and adaptive learning are essential for reliable decision-making.

Despite these promising results, this study is limited to a single dataset and the hybrid architecture involves higher computational cost compared to simpler models. Future work will focus on integrating fuzzy-attention mechanisms, exploring hybrid transformer-based extensions, validating the model on multi-source datasets, and investigating real-time deployment in intelligent transportation systems. Overall, the proposed ANFIS(EM)–LSTM framework provides a novel and interpretable hybrid modeling strategy that contributes to the development of reliable intelligent monitoring systems, particularly in safety-critical domains such as driver assistance and intelligent transportation.

ACKNOWLEDGMENTS

This manuscript was prepared without any research funding or grant support.

FUNDING INFORMATION

The authors report no financial support or grant involvement.

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
Mouhcine Lakhroufi	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
El Mehdi Mellouli		✓				✓		✓	✓	✓	✓	✓		

C : C onceptualization	I : I nterpretation	Vi : V isualization
M : M ethodology	R : R esources	Su : S upervision
So : S oftware	D : D ata Curation	P : P roject administration
Va : V alidation	O : Writing - O riginal Draft	Fu : F unding acquisition
Fo : F ormal analysis	E : Writing - Review & E ditng	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study was provided by the Computer Vision Laboratory at National Tsing Hua University and can be accessed online at <http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/>.

REFERENCES

- [1] S. Otmami, T. Pebayle, J. Roge, and A. Muzet, "Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers," *Physiol. Behav.*, vol. 84, no. 5, pp. 715–724, Apr. 2005, doi: 10.1016/j.physbeh.2005.02.021.
- [2] A. G. Correa, L. Orosco, and E. Laciari, "Automatic detection of drowsiness in EEG records based on multimodal analysis," *Med. Eng. Phys.*, vol. 36, no. 2, pp. 244–249, Feb. 2014, doi: 10.1016/j.medengphy.2013.07.011.
- [3] H. Zhang, C. Wu, X. Yan, and T. Z. Qiu, "The effect of fatigue driving on car following behavior," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 43, pp. 80–89, Nov. 2016, doi: 10.1016/j.trf.2016.06.017.
- [4] B. Cheng, W. Zhang, Y. Lin, R. Feng, and X. Zhang, "Driver drowsiness detection based on multisource information," *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 22, no. 5, pp. 450–467, Sep. 2012, doi: 10.1002/hfm.20395.
- [5] S. Bakheet and A. Al-Hamadi, "A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification," *Brain Sci.*, vol. 11, no. 2, pp. 1–15, Feb. 2021, doi: 10.3390/brainsci11020240.
- [6] X. Zhang, X. Wang, X. Yang, C. Xu, X. Zhu, and J. Wei, "Driver drowsiness detection using mixed-effect ordered logit model considering time cumulative effect," *Anal. Methods Accid. Res.*, vol. 26, p. 100114, Jun. 2020, doi: 10.1016/j.amar.2020.100114.
- [7] A. Qudus, A. S. Zandi, L. Prest, and F. J. E. Comeau, "Using long short term memory and convolutional neural networks for driver drowsiness detection," *Accid. Anal. Prev.*, vol. 156, p. 106107, Jun. 2021, doi: 10.1016/j.aap.2021.106107.
- [8] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: 10.1109/72.279181.
- [9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [10] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
- [11] Y. Ed-Doughmi, N. Idrissi, and Y. Hbali, "Real-Time System for Driver Fatigue Detection Based on a Recurrent Neuronal Network," *J. Imaging*, vol. 6, no. 3, pp. 1–15, Mar. 2020, doi: 10.3390/jimaging6030008.
- [12] S. Chen, Z. Wang, and W. Chen, "Driver Drowsiness Estimation Based on Factorized Bilinear Feature Fusion and a Long-Short-Term Recurrent Convolutional Network," *Information*, vol. 12, no. 1, pp. 1–15, Dec. 2020, doi: 10.3390/info12010003.
- [13] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [14] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, no. 3, pp. 665–685, 1993, doi: 10.1109/21.256541.
- [15] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.
- [16] A. Yildiz, M. Akin, M. Poyraz, and G. Kirbas, "Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7390–7399, May 2009, doi: 10.1016/j.eswa.2008.09.003.
- [17] S. Arefnezhad, S. Samiee, A. Eichberger, and A. Nahvi, "Driver Drowsiness Detection Based on Steering Wheel Data Applying Adaptive Neuro-Fuzzy Feature Selection," *Sensors*, vol. 19, no. 4, pp. 1–14, Feb. 2019, doi: 10.3390/s19040943.
- [18] A. Celecia, K. Figueiredo, M. Vellasco, and R. González, "A Portable Fuzzy Driver Drowsiness Estimation System," *Sensors*, vol. 20, no. 15, pp. 1–16, Jul. 2020, doi: 10.3390/s20154093.
- [19] S. Jabeen, M. Baig, and M. M. Awais, "ANFIS learning using expectation maximization based Gaussian mixture model and multilayer perceptron learning," *Appl. Soft Comput.*, vol. 149, p. 110958, Dec. 2023, doi: 10.1016/j.asoc.2023.110958.
- [20] S. Mishra, V. Jain, Y. A. Saraf, I. Kandasamy, and V. W.B., "Deep neuro-fuzzy system for violence detection," *Neurocomputing*, vol. 619, p. 129007, Feb. 2025, doi: 10.1016/j.neucom.2024.129007.
- [21] Z. Zhang, H. Wang, Q. You, L. Chen, and H. Ning, "A novel temporal adaptive fuzzy neural network for facial feature based fatigue assessment," *Expert Syst. Appl.*, vol. 252, p. 124124, Oct. 2024, doi: 10.1016/j.eswa.2024.124124.
- [22] H. Chen and J. Xie, "EEG-based TSK fuzzy graph neural network for driver drowsiness estimation," *Inf. Sci. (Nij.)*, vol. 679, p. 121101, Sep. 2024, doi: 10.1016/j.ins.2024.121101.
- [23] W. Xiao, G. Xie, H. Liu, W. Chen, and R. Li, "FDAN: Fuzzy deep attention networks for driver behavior recognition," *J. Syst. Archit.*, vol. 147, p. 103063, Feb. 2024, doi: 10.1016/j.sysarc.2023.103063.
- [24] S. A. Zendehbad, J. Ghasemi, and F. S. Khodadad, "FatigueNet: A hybrid graph neural network and transformer framework for real-time multimodal fatigue detection," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, Sep. 2025, doi: 10.1038/s41598-025-00640-z.
- [25] A. Pasha et al., "Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease," *Heliyon*, vol. 10, no. 9, May 2024, doi: 10.1016/j.heliyon.2024.e30241.
- [26] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC — CNN based efficient face recognition technique using Dlib," in *Proc. 2016 Int. Conf. Adv. Commun. Control Comput. Technol. (ICACCCT)*, IEEE, May 2016, pp. 192–195, doi: 10.1109/ICACCCT.2016.7831628.
- [27] T. S. Delwar et al., "AI- and Deep Learning-Powered Driver Drowsiness Detection Method Using Facial Analysis," *Applied Sciences*, vol. 15, no. 3, 2025, doi: 10.3390/app15031102.
- [28] Y. Albadawi, A. AlRedhaei, and M. Takruri, "Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features," *J. Imaging*, vol. 9, no. 5, 2023, doi: 10.3390/jimaging9050091.
- [29] S. P. Rajamohana, E. G. Radhika, S. Priya, and S. Sangeetha, "Driver drowsiness detection system using hybrid approach of convolutional neural network and bidirectional long short term memory (CNN_BILSTM)," *Mater. Today Proc.*, vol. 45 Part 2, pp. 2897–2901, 2021, doi: 10.1016/j.matpr.2020.11.898.
- [30] M. Dua, Shakshi, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3155–3168, Apr. 2021, doi: 10.1007/s00521-020-05209-7.
- [31] L. Zhao, Z. Wang, G. Zhang, and H. Gao, "Driver drowsiness recognition via transferred deep 3D convolutional network and state probability vector," *Multimed. Tools Appl.*, vol. 79, no. 35, pp. 26683–26701, Sep. 2020, doi: 10.1007/s11042-020-09259-w.
- [32] A. Ghourabi, H. Ghazouani, and W. Barhoumi, "Driver Drowsiness Detection Based on Joint Monitoring of Yawning, Blinking and Nodding," in *Proc. 2020 IEEE 16th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, IEEE, Sep. 2020, pp. 407–414, doi: 10.1109/ICCP51029.2020.9266160.

BIOGRAPHIES OF AUTHORS

Mouhcine Lakhroufi    received his master's degree in 2011 from the Faculty of Sciences, Sidi Mohammed Ben Abdellah University, Fez, Morocco. He is currently pursuing his Ph.D. at the National School of Applied Sciences, Sidi Mohammed Ben Abdellah University, Fez, Morocco, in the Laboratory of Engineering, Systems, and Applications. His research focuses on fuzzy logic and artificial neural networks. He can be contacted at email: mouhcine.lakhroufi@usmba.ac.ma.



El Mehdi Mellouli    obtained his doctorate in 2013 from the Faculty of Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco. His doctorate focused on the modeling and control of nonlinear and multivariable systems using fuzzy logic and various control techniques. Since 2017, he has been a professor at the National School of Applied Sciences, University of Fez, Morocco. He has numerous publications in reputable international journals. His research has focused on the modeling and generation of intelligent control law with artificial intelligence for autonomous systems, robotics, renewable energy systems, and unknown dynamic systems. He can be contacted at email: elmehdi.mellouli@usmba.ac.ma.