

Hybrid model for driver drowsiness prediction using neuro-fuzzy and random forest

Mouhcine Lakhroufi, El Mehdi Mellouli

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

Article Info

Article history:

Received Jan 21, 2025

Revised Jul 22, 2026

Accepted Aug 12, 2026

Keywords:

Adaptive neuro-fuzzy inference system

Drowsiness detection

Facial landmarks

Neuro-fuzzy systems

Percentage of eye closure

Random forest

ABSTRACT

In this research, a new hybrid model designed was introduced to identify driver drowsiness by integrating an adaptive neuro-fuzzy inference system (ANFIS) with the random forest (RF) method. Driver drowsiness is one of the major risk factors for road accidents, and timely, accurate identification of it is vital for improving traffic safety. The main emphasis of our work is on the combination of ANFIS with RF. ANFIS uses fuzzy logic to represent intricate systems with uncertainties and non-linear relationships, on the other hand, RF offers powerful classification by its ensemble learning technique, performing well in high-dimensional feature spaces. This combination capitalizes on each model's strengths: ANFIS for its dynamic responding to uncertain data and RF for its high classification accuracy, robustness, and resistance to overfitting. The results from the different experiments confirmed that the hybrid ANFIS-RF outperformed its components considerably. It was able to return an enhancement in accuracy by 2.70% compared to ANFIS, and 0.16% compared to RF, while at the same time, dropping the mean squared error (MSE) by 33.40% compared to ANFIS and 2.85% compared to RF. Knowing that RF is already better than other traditional machine learning methods in classification, these results emphasize its major role in the effectiveness of the hybrid model.

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



Corresponding Author:

Mouhcine Lakhroufi

Laboratory of Engineering, System and Application, National School of Applied Sciences

University of Sidi Mohamed Ben Abdellah

Fez, Morocco

Email: mouhcine.lakhroufi@usmba.ac.ma

1. INTRODUCTION

Driver drowsiness is the top contributing factor in fatal road accidents. It is therefore very risky for drivers as well as other road users. The World Health Organization has recognized that approximately 20% of traffic accidents result from sleepy driving and, therefore, they have made it a priority to develop drowsy state detection systems that warn drivers before they can lose control. Next generation in-vehicle drowsiness monitoring systems are capable of detecting subtle features of drowsiness to caution the driver in time.

Drowsiness detection methods are essentially those that measure physiological, vehicle-based, and behavioral changes. Physiological methods refer to monitoring the brain activity (electroencephalography (EEG)), heart rate (electrocardiography (ECG)), or skin conductance (galvanic skin response (GSR)) of a person which is a very accurate way of gathering data [1], [2]. However, such methods are usually intrusive, and therefore, it is not easy to incorporate them in cars, as an example, EEG sensors necessitate that the driver's head has electrodes on it which may divert his attention from the road [3], [4]. Vehicle-based

measures detect the signs of getting sleepy while driving by studying the driving patterns like lane deviations or steering corrections [5], [6]. Finally, behavioral measures focus on analyzing facial expressions, eye movements [7], blinks, and yawns to assess alertness. The techniques are based on visual perception and a camera, so there is no requirement for wearing a sensor, and it is easy to implement the methods in the real world [8].

This study primarily concentrates on behavioral measures because they provide a non-intrusive way of getting real-level insights into a driver's wakefulness through facial and eye movement recognition [9]. Compared to physiological methods that necessitate sensors to be attached, computer vision methods allow drivers to be monitored in real-time, thus they are the best fit for real driving situations. Some of the typical indicators are eye and mouth features that can show if a person is sleepy through signs like percentage of eye closure (PERCLOS) [10] and yawning.

Recent studies demonstrate the potential of machine learning and deep learning (DL)-based models for identifying sleepiness. To cite examples, a real-time application of a recurrent neural network (RNN) is introduced, which can constantly watch a driver and efficiently spot fatigue [11]. Apart from that, a 3D convolutional neural network (CNN) which identifies facial expressions and important points of the face has been documented as performing better in terms of accuracy for drowsiness detection [12]. Although less explored in this field adaptive neuro-fuzzy inference system (ANFIS) is well-suited for handling complex systems with uncertainties, a key aspect of driver state monitoring. Combining neural networks and fuzzy logic, ANFIS adapts effectively to driving data variations, as seen in [13] application to classify vigilance levels (alert, drowsy, and sleep) from EEG data using wavelet-entropy features. Additionally, ANFIS has shown promise in detecting drowsiness states from steering data for adaptive feature selection [14]. Similarly, random forest (RF) based methods have been effective in classifying fatigue, even with noisy or individual-specific data [15], as well as steering angle data for lane departure detection related to drowsiness [16].

The experimental results were evaluated using videos from the NTHU-DDD dataset [17], a commonly used reference in drowsiness detection research. The results obtained in this work were compared to classification reports and the findings of other studies that used the same dataset. The comparison between these works highlights the proposed model's capability to outperform the previous approaches in the literature. This paper mainly focuses on an innovative hybrid model which has been developed by combining the strengths of ANFIS with classification algorithms to tackle behavior data uncertainties for drowsiness detection. The method takes advantage of ANFIS's capability to represent nonlinear and uncertain relationships and at the same time, providing reliable decisions with the help of classification methods. Hence, our main objective is to minimize our prediction errors, enlarge our model's capability to be well generalized in diverse driving environments, and eventually contribute to the better prevention of drowsiness-related dangers on the road.

This paper has been structured in a way that: section 1 gives an account of the significance of driver drowsiness detection, and also it brings a brief review of the related research. Section 2 is concerned with the details of the methodology that was used in this study, and the dataset that was employed is described. Section 3 is devoted to the presentation of the experimental results and an extensive analysis is made highlighting the performance of the proposed hybrid model in comparison with the individual models and other related studies. Finally, section 4 sums up the paper through a discussion of the findings, the first three implications, and how they guide the future research directions.

2. METHOD

This study proposes a multi-step methodological approach, as shown in Figure 1, which details the complete pipeline for driver drowsiness detection. The framework starts with facial landmark extraction and the computation of behavioral features (eye aspect ratio (EAR), mouth aspect ratio (MAR), and PERCLOS), followed by independent predictions generated by the ANFIS and RF models. Finally, these predictions are combined through a weighted fusion strategy optimized using gradient descent.

Likewise, ANFIS and RF models can each be thought of as performing the first level of prediction independently. Here, ANFIS takes advantage of its strength in modelling uncertainties and non-linear relationships that are present in the drowsiness detection data. It goes through several stages such as fuzzification, normalization, and defuzzification to make predictions based on fuzzy rules. These stages allow the model to understand the slight changes in behavior that indicate drowsiness, and ANFIS' performance will be evaluated to give an initial prediction.

Meanwhile, the RF model builds up a collection of decision trees by applying the bootstrap sampling and feature selection methods. Each tree will give its prediction, and then these are combined using majority voting to formulate the final decision. Hence, the ensemble method (EM) not only increases the

model's stability but also makes it more efficient in recognizing drowsiness states. RF's prediction is evaluated as well, thus giving a complementary outcome to ANFIS's.

In the last phase, the ANFIS and RF model outputs are integrated through a weighted fusion model. This fusion takes advantage of the two models' fundamental differences and adjusts their weights accordingly by means of a gradient descent algorithm. This method determines the optimal combination of ANFIS and RF predictions that results in minimal error. Synergistically, the hybrid ANFIS-RF model results in increased prediction accuracy and less mean squared error (MSE) by effectively balancing ANFIS's flexibility dealing with the uncertain nature of data and RF's powerful classification features. Therefore, this fusion provides better results in driver drowsiness detection.

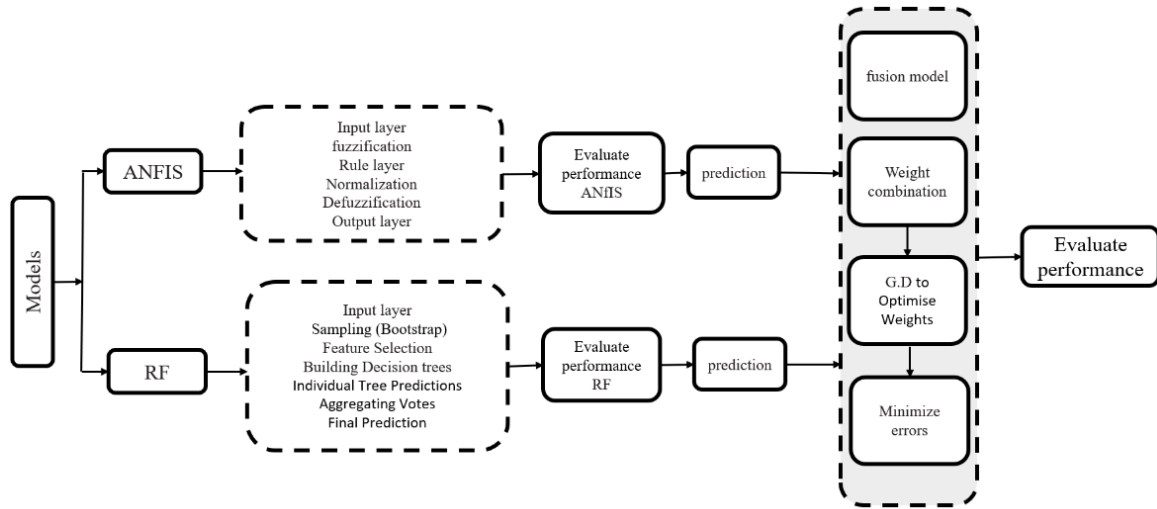


Figure 1. Overview of the proposed hybrid ANFIS–RF framework for driver drowsiness detection

2.1. Dataset description and feature extraction

The NTHU-DDD dataset aims to evaluate driver drowsiness and behavior (IDs 001–006) each recorded during 12 separate driving sessions under different conditions, including scenarios with and without glasses. Each recording is annotated with the driver's vigilance state (drowsy or non-drowsy) and associated behaviors such as sleepyCombination, slowBlinkWithNodding, and yawning, as illustrated in Table 1. In total, this dataset comprises 66,491 recordings from 6 individuals. It is divided so that 70% is used for training and the remaining 30% for testing. By analyzing this data collection, one can gain deep insight into the behavioral changes related to drowsiness, and it also forms a strong basis for the creation of predictive models that seek to improve road safety.

The facial landmark-based drowsiness detection system is an essential element of the research reported in this paper. To locate the face first, the authors used the Dlib library that utilizes a pretrained histogram of oriented gradients (HOG) detector along with a linear support vector machine (SVM). After recognizing the face region, 68 facial landmarks referring to the two eyes and mouth among other key facial components, were mapped for each frame.

From these landmarks, it is possible to calculate three behavioral characteristics: EAR, MAR, and PERCLOS. The great advantage of these features is that they give drivers' drowsiness non-intrusively and reliably. The EAR is a scalar value that explains how open or closed the eyes are [18], [19]. When a person is sleepy, eye EAR value slowly becomes zero. The EAR value for an eye is measured two times for both the right and left eye by using Euclidean distances between particular eye landmarks as per (1).

$$EAR_R = \frac{\|P38-P42\| + \|P39-P41\|}{2\|P37-P40\|} \text{ and } EAR_L = \frac{\|P44-P48\| + \|P45-P47\|}{2\|P43-P46\|} \quad (1)$$

The final EAR value is obtained by averaging the right and left eye ratios as defined in (2).

$$EAR = \frac{EAR_L + EAR_R}{2} \quad (2)$$

The MAR aims to assess mouth opening and is particularly effective for yawning detection [20]. MAR is computed using distances between selected mouth landmarks as (3).

$$MAR = \frac{\|P62-P68\| + \|P63-P67\| + \|P64-P66\|}{3\|P61-P65\|} \quad (3)$$

Table 1. Samples from the NTHU-DDD dataset illustrating drowsiness behaviors

ID	Status	Condition	Behavior	Behavioral samples
	Drowsy	Glasses	Sleepy combination Slow blink with nodding Yawning	
		No glasses	Sleepy combination Slow blink with nodding Yawning	
	No drowsy	Glasses	Sleepy combination Slow blink with nodding Yawning	
		No glasses	Sleepy combination Slow blink with nodding Yawning	

In addition, PERCLOS is obtained by calculating the ratio of frames in which the EAR remains below a predefined eye-closure threshold over a specified observation interval, indicating partial or full eye closure [21]. PERCLOS provides a robust temporal measure of prolonged eye closure and reduced alertness. Specifically, PERCLOS is obtained by applying a fixed threshold $T = 0.20$ to the frame-level EAR signal ($EAR < T$ indicating a closed eye), consistent with the EAR formulation of Soukupová [18] [19], and computing the proportion of closed-eye frames within a sliding window of $N = 30$ consecutive frames (approximately 1 second at the dataset's native frame rate), advanced frame by frame to produce a continuous signal.

The NTHU-DDD dataset used in this study is curated and annotated. Therefore, no missing values were observed in the extracted features. To mitigate noise introduced by facial landmark detection and inter-subject variability, all features (EAR, MAR, and PERCLOS) were normalized using min-max scaling to map their values into the range $[0,1]$. This normalization was applied globally across the pooled recordings of all subjects, and no additional temporal smoothing was applied to the frame-level EAR and MAR signals beyond this scaling, as PERCLOS already provides temporal aggregation of eye-closure information.

The normalized features were next examined frame by frame and directly used as inputs to the learning models. In the case of the ANFIS model, the characteristics were fuzzified by means of generalized bell-shaped membership functions for the representation of uncertainties and nonlinear relationships. On the other hand, for the RF classifier, the same normalized features were given as numerical inputs without any further transformation. This common preprocessing tactic guarantees numerical stability during training and allows for a fair comparison between models.

2.2. Model training procedure

The ANFIS model, based on Jang's work in 1993 [22] utilizes a neural network combined with fuzzy logic using Takagi-Sugeno-Kang (TSK) "if-then" rules [23]. The model's input layer receives the three behavioral descriptors, EAR, MAR, and PERCLOS, and it is followed by a fuzzification layer where each input is associated with three linguistic terms (low, medium, and high) using generalized bell-shaped membership functions (gbellmf), as depicted in Figure 2. The parameters of these membership functions are being optimized as training progresses using the Adam algorithm to enhance the ability of the model to adapt to data variability. The fuzzified inputs are then combined in the rule layer to produce 27 fuzzy rules that cover all possible combinations of the three linguistic states of the three input variables. The firing strength of every rule, which is computed by multiplying the membership degrees of the corresponding inputs, is presented in (4). Defuzzification outputs are computed as (5).

$$w_i = \mu_{E_i}(x_{ear}) \cdot \mu_{M_i}(x_{mar}) \cdot \mu_{P_i}(x_{perc}) \quad (4)$$

$$f_i = p_i(x_{ear}) + q_i(x_{mar}) + r_i(x_{perc}) + s_i \quad (5)$$

The RF model [24] is an ensemble learning method in machine learning that creates a model by aggregating several decision trees to improve the prediction, accuracy, and reliability of the model. Each decision tree in the forest is like a flowchart with a well-organized structure where the branches represent the possible decisions or actions that lead to the specific outcomes. RF is based on the collective power of many trees, with each tree being trained on a different subset of the data, which brings about some variation among them and thus reduces the chance of overfitting. This diversity enhances the model's ability to generalize well on new data. This work implements an RF classifier composed of 100 trees, providing a good trade-off between computational efficiency and prediction stability.

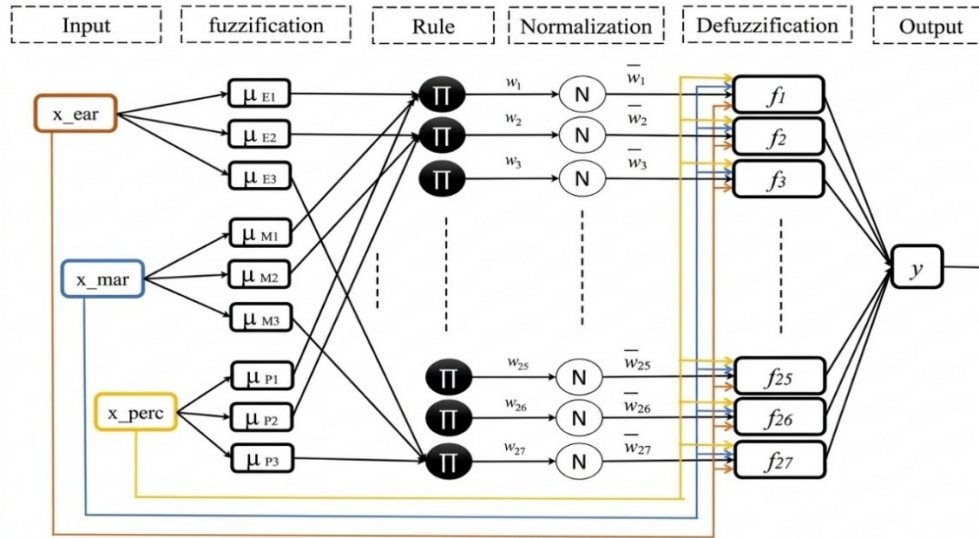


Figure 2. Structure of ANFIS with EAR, MAR, and PERCLOS inputs

2.3. Model fusion methodology

Gradient descent is employed to determine the optimal values of the weighting coefficients. The combined prediction model with the bias component is expressed in (6).

$$y_c = w_1 \cdot y_{anfis} + w_2 \cdot y_{rf} + b \quad (6)$$

The fused prediction y_c is computed by linearly combining the outputs of the ANFIS and RF models using the weighting coefficients w_1 , and w_2 , with b acting as the bias parameter. During training, these parameters are iteratively adjusted to minimize the MSE between the predicted and actual outputs, as formulated in (7).

$$L = \frac{1}{n} \sum_{i=1}^n (y_c^{(i)} - y^{(i)})^2 \quad (7)$$

The weights w_1 , w_2 , and the bias b are updated at each iteration of gradient descent as (8).

$$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1}; w_2 = w_2 - \alpha \frac{\partial L}{\partial w_2}; b = b - \alpha \frac{\partial L}{\partial b} \quad (8)$$

Where α is the learning rate.

3. RESULTS AND DISCUSSION

3.1. Evaluating model performance

In order to thoroughly assess how well our models can detect driver drowsiness, we employed several performance metrics. MSE reports the mean squared deviation of predictions from ground-truth

values [25], while root mean squared error (RMSE) gives a more relatable measure of this difference, and mean absolute error (MAE) measures the average absolute deviation. Recall and F1-score were evaluated to emphasize recall of drowsy states and the precision–recall balance, respectively, while area under the curve (AUC) assessed the models' ability to distinguish classes regardless of the decision threshold; accuracy quantified the overall proportion of correctly classified samples. These performance metrics should not be confused with the loss functions used for training, which depend on the model type: ANFIS was trained by minimizing MAE, whereas the DL models—artificial neural network (ANN), CNN, long short-term memory (LSTM), bidirectional long short-term memory (BLSTM), gated recurrent unit (GRU)—were trained through binary cross-entropy loss. All comparative results are based exclusively on performance metrics obtained on the test data, independent of the training process; MAE and MSE are reported only after thresholding the probabilistic outputs, quantifying decision-error size, while classification metrics indicate overall model efficiency.

Three model types were tested, namely fuzzy system (FS), DL, and EM, with hyperparameters carefully tuned for the best performance and least training fluctuation. Rather than relying on massively deep architectures, we favored simpler models to limit overfitting from excess parameters. All FS and DL models were optimized over 100 training epochs with Adam [26], chosen for its efficient convergence and stable learning rates. Binary cross-entropy was used for some models, particularly for binary classification, as it improved class separation (drowsy vs. non-drowsy) and reduced misclassification.

A batch size of 16 was used in some cases, along with dropout regularization to limit overfitting, while 'relu' and 'sigmoid' activations in dense layers improved the network's ability to capture nonlinearities. Table 2 shows the representative parameters of each model, reflecting tuning specialized to each algorithm's requirements. The hyperparameters summarized in Table 2 were determined empirically, through iterative manual experimentation guided by preliminary trials and values commonly reported in the literature for comparable tasks, rather than a formal grid-search or cross-validation procedure. For ANFIS, MAE served both as the training loss and as an evaluation metric, chosen for its robustness to outliers and stable gradient behavior; it was combined with MSE and RMSE as evaluation metrics for a fuller assessment of prediction errors and easier comparison across models.

Table 2. Hyperparameters of each model

Model type	Model	Hyperparameters
FS	ANFIS	n_input: 3, n_memb: 3, memb_func: gbellmf, optimizer: adam, loss: MeanAbsoluteError, n_epochs: 100
Deep neural networks	ANN	layers: [Dense(32, relu), Dense(16, relu), Dense(1, sigmoid)], loss: binary_crossentropy, optimizer: adam, metrics: accuracy, epochs: 100, batch_size: 16, validation_split: 0.2
	CNN	layers: [Conv1D(32, 2, relu), Flatten(), Dense(1, sigmoid)], loss: binary_crossentropy, optimizer: adam, metrics: accuracy, epochs: 100, batch_size: 16, validation_split: 0.2
	BLSTM	layers: [Bidirectional(LSTM(64, return_sequences=True)), Bidirectional(LSTM(64)), Dense(1, sigmoid)], loss: binary_crossentropy, optimizer: adam, metrics: accuracy, epochs: 100, batch_size: 16, validation_split: 0.2
	LSTM	layers: [LSTM(64, return_sequences=True), LSTM(64), Dense(1, sigmoid)], loss: binary_crossentropy, optimizer: adam, metrics: accuracy, epochs: 100, batch_size: 16, validation_split: 0.2
	GRU	layers: [GRU(64, return_sequences=True), GRU(64), Dense(1, sigmoid)], loss: binary_crossentropy, optimizer: adam, metrics: accuracy, epochs: 100, batch_size: 16, validation_split: 0.2
Traditional supervised learning methods	SVM	kernel: rbf, probability: True
	K-nearest neighbors (KNN)	n_neighbors: 7
Ensemble methods	Logistic regression	solver: lbfgs
	Random forest	n_estimators: 100, max_depth: 10
	Gradient boosting	n_estimators: 100, learning_rate: 0.05, max_depth: 5
	XGBoost	n_estimators: 100, learning_rate: 0.05, max_depth: 5
	AdaBoost	n_estimators: 100

The loss curves in Figure 3 are studied only to assess convergence and training stability, not for direct performance comparison, which relies on the evaluation metrics reported in Tables 3 and 4. Figure 3(a) stabilizes earliest and with the fewest oscillations, whereas the DL models in Figure 3(b) to 3(f) require more epochs and display comparatively less stable validation loss. As depicted in Figure 3(a), ANFIS converges much faster and more stable than the other models, rapidly reducing both training and validation loss within the first 30 epochs with only minor oscillations. ANN as in Figure 3(c) and CNN as in Figure 3(e), by contrast, are slower learners, converging only after 60–80 epochs. LSTM as in Figure 3(b), GRU as in

Figure 3(f), and BLSTM as in Figure 3(d) converge faster, stabilizing around 40 epochs, but their validation-loss curves are noticeably more unstable than ANFIS's, indicating higher sensitivity to temporal dynamics and lower generalization ability; BLSTM converges fastest among them, yet its converged loss remains higher than ANFIS's, confirming the hybrid neuro-FS greater optimization stability. This can be attributed to ANFIS's small number of trainable parameters (the membership-function parameters of three fuzzified inputs and linear consequents of 27 TSK rules), the smooth, differentiable bell-shaped membership functions, and the low-dimensional three-feature input space, all favoring stable gradient-based optimization. The DL architectures, designed to exploit larger, more heterogeneous datasets, are more prone to overfitting given NTHU-DDD's limited subject diversity, consistent with the validation-loss fluctuations noted above.

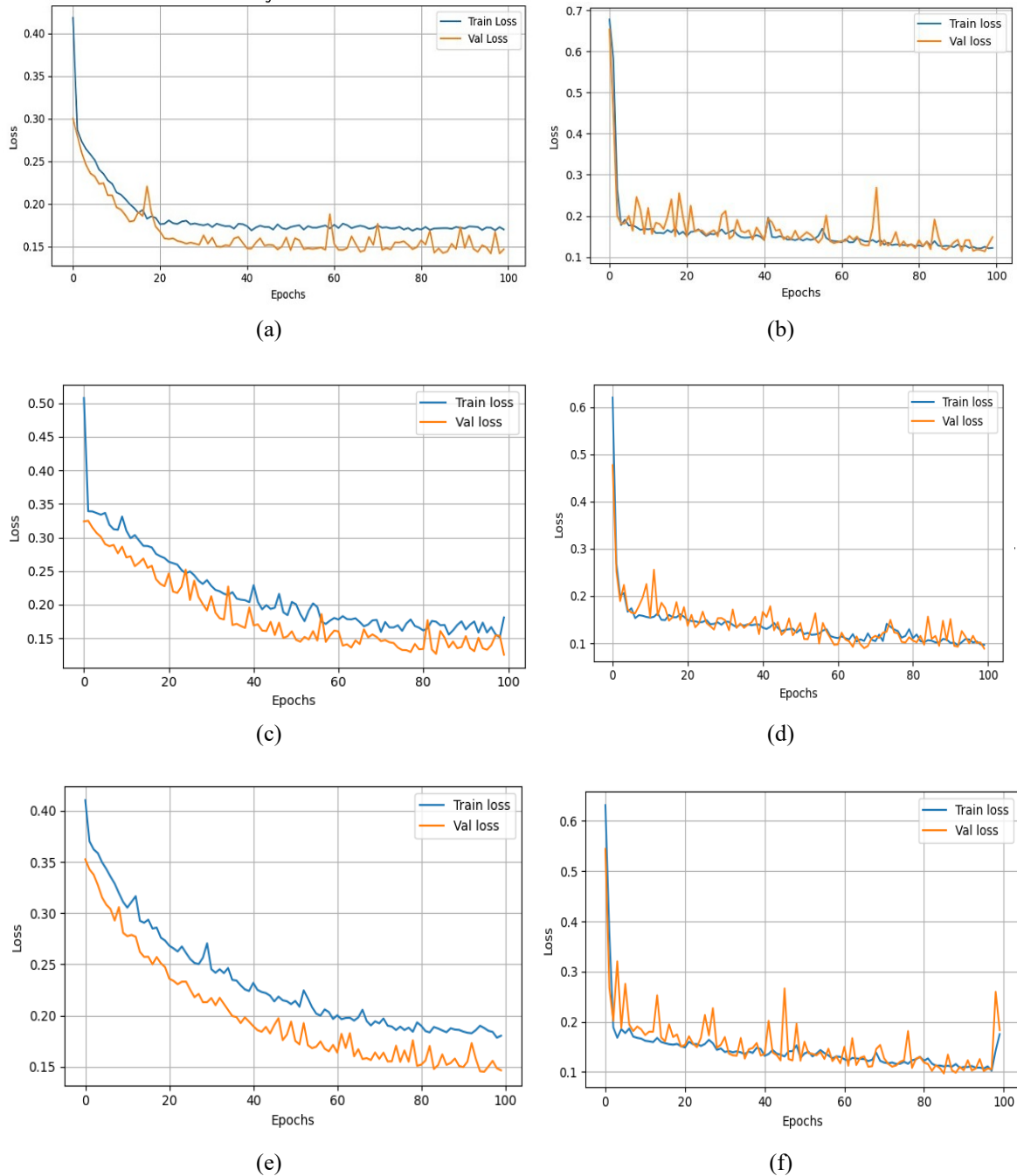


Figure 3. Training and validation loss curves over 100 epochs for (a) ANFIS, (b) LSTM, (c) ANN, (d) BLSTM, (e) CNN, and (f) GRU

3.2. Comparison of individual and hybrid model performance

To evaluate the effectiveness of the proposed approach, a comprehensive comparison between individual models and hybrid ANFIS–X configurations was conducted. The objective of this analysis is not only to assess the predictive capability of standalone classifiers, but also to examine the impact of integrating ANFIS as a fuzzy inference component within hybrid architectures. Rather than acting solely as an independent classifier, ANFIS is incorporated as a core module designed to handle uncertainty and gradual transitions between alert and drowsy states. Through its fuzzy rule-based reasoning, ANFIS provides stable probabilistic estimates and fast convergence behavior, which are particularly desirable in driver drowsiness detection. By hybridizing ANFIS with complementary classifiers, the proposed framework aims to exploit both the interpretability and stability of fuzzy inference and the strong generalization capabilities of data-driven models.

The performance of the standalone models is analyzed first. The individual model performances are reported in Table 3. Among the evaluated approaches, ensemble-based methods and RNNs demonstrate strong predictive capability. RF achieves the highest performance among standalone models, with an accuracy of 94.44%, an F1-score of the approximately 94.6%, and an AUC of 0.984, indicating robust generalization and effective discrimination between alert and drowsy states. Recurrent architectures such as LSTM, BLSTM, and GRU also yield competitive results, with accuracies exceeding 93% and AUC values above 0.97, confirming their ability to capture temporal dependencies in behavioral features. In contrast, simpler classifiers, including logistic regression, SVM, and KNN, exhibit comparatively lower performance, reflecting the limitations of linear or distance-based decision boundaries for this task. Although ANFIS alone does not achieve the highest accuracy (91.90%), it maintains a high recall (92.88%) and a strong AUC (0.975), demonstrating notable sensitivity to drowsy states and stable probabilistic behavior. These characteristics suggest that ANFIS constitutes a reliable and complementary component for hybrid integration.

Table 3. Comparison of valuation measures for individual models

Model	MSE	RMSE	MAE	Recall	F1-score	AUC	Accuracy
LSTM	0.060317	0.245596	0.060317	0.910979	0.941718	0.978216	0.939683
GradientBoosting	0.061905	0.248807	0.061905	0.916914	0.940639	0.979689	0.938095
XGBoost	0.063492	0.251976	0.063492	0.910979	0.938838	0.981264	0.936508
KNN	0.068254	0.261255	0.068254	0.926471	0.936107	0.96573	0.931746
GRU	0.068254	0.261255	0.068254	0.922849	0.935338	0.978114	0.931746
BLSTM	0.073016	0.270214	0.073016	0.905882	0.930514	0.9586	0.926984
ANN	0.07619	0.276026	0.07619	0.884273	0.925466	0.968352	0.92381
SVM	0.077778	0.278887	0.077778	0.869436	0.922835	0.963774	0.922222
AdaBoost	0.079365	0.281718	0.079365	0.890208	0.923077	0.97301	0.920635
CNN	0.080952	0.284521	0.080952	0.905045	0.922844	0.97141	0.919048
Random forest	0.055556	0.235702	0.055556	0.925816	0.946889	0.984161	0.944444
ANFIS	0.080952	0.284521	0.080952	0.928783	0.924668	0.975583	0.919048
Logistic regression	0.125397	0.354114	0.125397	0.896142	0.884334	0.941291	0.874603

The impact of hybridization is examined next. The hybrid ANFIS–X results are summarized in Table 4. A consistent improvement can be observed across all hybrid configurations when compared to their corresponding standalone counterparts, confirming the benefit of integrating ANFIS within the prediction framework. Among all hybrid models, ANFIS–RF achieves the best overall performance, with an accuracy of 94.60%, an F1-score of 94.82%, a recall of 92.28%, and a low MSE of 0.054, outperforming both standalone ANFIS and RF. Hybrid models based on recurrent architectures, such as ANFIS–BLSTM and ANFIS–LSTM, also demonstrate strong performance, particularly in terms of recall and AUC (both exceeding 0.98), highlighting the advantage of combining fuzzy inference with temporal modeling. Other hybrid configurations, including ANFIS–GB and ANFIS–XGBoost, show competitive but slightly lower performance, while hybrids involving simpler classifiers such as ANFIS–SVM and ANFIS–KNN yield moderate improvements without surpassing the leading hybrid models. Overall, these results indicate that the proposed ANFIS–X framework systematically enhances prediction performance by leveraging the complementary strengths of fuzzy inference and data-driven classifiers. Among all hybrid configurations, the ANFIS–RF model emerges as the most balanced and reliable solution for driver drowsiness detection.

As summarized in Table 5, the proposed ANFIS–RF hybrid model demonstrates superior balance in predictive performance compared to recent methods applied to the NTHU-DDD dataset. While certain approaches, such as CNN–LSTM of Chen *et al.* [27] (93.30% accuracy) and RNN of Ed-Doughmi *et al.* [11] (92.00% accuracy), achieved slightly higher accuracy, they either lack F1-score reporting or present lower F1-scores, indicating potential precision–recall imbalance. In contrast, our model attained a robust F1-score

of 90.6%, outperforming the deep CNN by Dua *et al.* [28] (84.09%) and the 3D CNN by Zhao *et al.* [12]. Furthermore, the combination of adaptive fuzzy reasoning and ensemble learning in ANFIS–RF enables effective generalization under facial variability. These results underline the relevance of neuro-fuzzy ensemble models as interpretable and competitive alternatives to deep architectures for real-time driver drowsiness detection. This should be interpreted in light of the dataset's scope: only six subjects are represented, limiting facial and demographic diversity, and drowsy-related behaviors are likely underrepresented relative to alert states, introducing some class imbalance. Both factors motivate our use of AUC alongside accuracy and F1-score.

Table 4. Comparison between individual and hybrid models

Model	MSE	RMSE	MAE	Accuracy	Precision	Recall	F1-score	AUC
ANFIS-RF	0.053968	0.232311	0.053968	0.946032	0.974922	0.922849	0.948171	0.978651
ANFIS-BLSTM	0.057143	0.239046	0.057143	0.942857	0.971787	0.919881	0.945122	0.982996
ANFIS-LSTM	0.05873	0.242343	0.05873	0.94127	0.977707	0.910979	0.943164	0.978312
ANFIS-GRU	0.061905	0.248807	0.061905	0.938095	0.965625	0.916914	0.940639	0.977218
ANFIS-GBoosting	0.063492	0.251976	0.063492	0.936508	0.971429	0.908012	0.93865	0.979639
ANFIS-XGBoost	0.066667	0.258199	0.066667	0.933333	0.971246	0.902077	0.935385	0.981396
ANFIS-AdaBoost	0.068254	0.261255	0.068254	0.931746	0.977273	0.893175	0.933333	0.962118
ANFIS-KNN	0.071429	0.267261	0.071429	0.928571	0.948328	0.917647	0.932735	0.9657
ANFIS-ANN	0.074603	0.273136	0.074603	0.925397	0.970779	0.88724	0.927132	0.967977
ANFIS-SVM	0.074603	0.273136	0.074603	0.925397	0.970779	0.88724	0.927132	0.970488
ANFIS-CNN	0.077778	0.278887	0.077778	0.922222	0.95	0.902077	0.925419	0.970929
ANFIS-LR	0.08254	0.287297	0.08254	0.91746	0.946708	0.896142	0.920732	0.965896

Table 5. Comparative studies on detecting driver drowsiness using the NTHU-DDD dataset

Study	Method	Used variables	Dataset	Accuracy (%)	F1-score (%)
[28]	Deep CNN	Hand gestures, facial expressions, head movements	NTHU-DDD	85.00	84.09
[11]	RNN	Driver image	NTHU-DDD	92.00	85.00
[12]	3D CNN	Driver face and facial landmarks	NTHU-DDD	88.60	
[29]	Deep CNN ensemble	Hand gestures, facial expressions, head movements	NTHU-DDD	85.00	84.09
[11]	RNN	Driver face and facial landmarks	NTHU-DDD	92.00	85.00
[12]	3D CNN	Driver face and facial landmarks	NTHU-DDD	88.60	
[27]	CNN + LSTM	Eye and face region	NTHU-DDD	93.30	
[29]	MLP + KNN	Eye closure, yawning	NTHU-DDD	94.31	79.00
This study	ANFIS-RF	Facial landmark (eye and mouth)	NTHU-DDD	94.60	94.81

4. CONCLUSION

This research proposed a hybrid model of ANFIS and RF to improve the precision and trustworthiness of driver drowsiness detection. Efficient ANFIS, handling uncertainty, coupled with RF's ensemble learning capability, enabled the hybrid model to outperform the standalone and several state-of-the-art models on the NTHU-DDD dataset. Specifically, the model reached an accuracy of 94.60% and an F1-score of 94.81%, which indicates its capability in handling a variety of facial expressions. Furthermore, the model, being both efficient and interpretable, can be a good candidate for real-time application on embedded systems. Explorations for future research will include combining more behavioral and physiological signals (e.g., heart rate, steering patterns) and diversifying the model's driving environment to improve generalization, robustness, and real-world applicability.

ACKNOWLEDGMENTS

The authors thank the faculty and administrative staff at the National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, for their support.

FUNDING INFORMATION

This research was not funded by any organization.

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 Lakhrouf	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
El Mehdi Mellouli		✓				✓		✓	✓	✓	✓	✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

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. Otmani, T. Pebayle, J. Roge, and A. Muzet, "Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers," *Physiology & Behavior*, vol. 84, no. 5, pp. 715–724, Apr. 2005, doi: 10.1016/j.physbeh.2005.02.021.
- [2] M. Awais, N. Badruddin, and M. Driberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, Aug. 2017, doi: 10.3390/s17091991.
- [3] A. G. Correa, L. Orosco, and E. Laciari, "Automatic detection of drowsiness in EEG records based on multimodal analysis," *Medical Engineering & Physics*, vol. 36, no. 2, pp. 244–249, Feb. 2014, doi: 10.1016/j.medengphy.2013.07.011.
- [4] H. Zhang, C. Wu, X. Yan, and T. Z. Qiu, "The effect of fatigue driving on car following behavior," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 43, pp. 80–89, Nov. 2016, doi: 10.1016/j.trf.2016.06.017.
- [5] 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," *Analytic Methods in Accident Research*, vol. 26, Jun. 2020, doi: 10.1016/j.amar.2020.100114.
- [6] 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," *Accident Analysis & Prevention*, vol. 156, Jun. 2021, doi: 10.1016/j.aap.2021.106107.
- [7] S. Bakheet and A. Al-Hamadi, "A framework for instantaneous driver drowsiness detection based on improved HOG features and Naïve Bayesian classification," *Brain Sciences*, vol. 11, no. 2, Feb. 2021, doi: 10.3390/brainsci11020240.
- [8] B. Cheng, W. Zhang, Y. Lin, R. Feng, and X. Zhang, "Driver drowsiness detection based on multisource information," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 22, no. 5, pp. 450–467, Sep. 2012, doi: 10.1002/hfm.20395.
- [9] S. E. Bekhouche, Y. Ruichek, and F. Dornaika, "Driver drowsiness detection in video sequences using hybrid selection of deep features," *Knowledge-Based Systems*, vol. 252, Sep. 2022, doi: 10.1016/j.knsys.2022.109436.
- [10] A. A. Bamidele *et al.*, "Non-intrusive driver drowsiness detection based on face and eye tracking," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 7, pp. 549–569, 2019, doi: 10.14569/IJACSA.2019.0100775.
- [11] Y. Ed-Doughmi, N. Idrissi, and Y. Hbali, "Real-time system for driver fatigue detection based on a recurrent neuronal network," *Journal of Imaging*, vol. 6, no. 3, Mar. 2020, doi: 10.3390/jimaging6030008.
- [12] L. Zhao, Z. Wang, G. Zhang, and H. Gao, "Driver drowsiness recognition via transferred deep 3D convolutional network and state probability vector," *Multimedia Tools and Applications*, vol. 79, no. 35–36, pp. 26683–26701, Sep. 2020, doi: 10.1007/s11042-020-09259-w.
- [13] 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 Systems with Applications*, vol. 36, no. 4, pp. 7390–7399, May 2009, doi: 10.1016/j.eswa.2008.09.003.
- [14] 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, Feb. 2019, doi: 10.3390/s19040943.
- [15] I. A. Fouad, "A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms," *Ain Shams Engineering Journal*, vol. 14, no. 3, Apr. 2023, doi: 10.1016/j.asej.2022.101895.
- [16] A. D. McDonald, J. D. Lee, C. Schwarz, and T. L. Brown, "Steering in a random forest," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 56, no. 5, pp. 986–998, Aug. 2014, doi: 10.1177/0018720813515272.
- [17] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Computer Vision – ACCV 2016 Workshops (ACCV 2016)*, Cham, Switzerland: Springer, 2017, pp. 117–133, doi: 10.1007/978-3-319-54526-4_9.
- [18] T. Soukupová and J. Cech, "Real-time eye-blink detection using facial landmarks," in *21st Computer Vision Winter Workshop*, 2016, pp. 1–8.
- [19] Q. Zhuang, Z. Kehua, J. Wang, and Q. Chen, "Driver fatigue detection method based on eye states with pupil and iris segmentation," *IEEE Access*, vol. 8, pp. 173440–173449, 2020, doi: 10.1109/ACCESS.2020.3025818.

- [20] A. G. Guggisberg, J. Mathis, U. S. Herrmann, and C. W. Hess, "The functional relationship between yawning and vigilance," *Behavioural Brain Research*, vol. 179, no. 1, pp. 159–166, Apr. 2007, doi: 10.1016/j.bbr.2007.01.027.
- [21] A. Celecia, K. Figueiredo, M. Vellasco, and R. González, "A portable fuzzy driver drowsiness estimation system," *Sensors*, vol. 20, no. 15, Jul. 2020, doi: 10.3390/s20154093.
- [22] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993, doi: 10.1109/21.256541.
- [23] E. M. Mellouli, M. Alfid, and I. Boumhidi, "Fuzzy sliding mode control for three-tank system based on linear matrix inequality," *International Journal of Automation and Control*, vol. 12, no. 2, pp. 237–250, 2018, doi: 10.1504/IJAAC.2018.090810.
- [24] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [25] J. Dessain, "Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric," *Expert Systems with Applications*, vol. 199, Aug. 2022, doi: 10.1016/j.eswa.2022.116970.
- [26] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015, pp. 1–15.
- [27] 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, Dec. 2020, doi: 10.3390/info12010003.
- [28] M. Dua, Shakshi, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3155–3168, Apr. 2021, doi: 10.1007/s00521-020-05209-7.
- [29] A. Ghourabi, H. Ghazouani, and W. Barhoumi, "Driver drowsiness detection based on joint monitoring of yawning, blinking and nodding," in *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sep. 2020, pp. 407–414, doi: 10.1109/ICCP51029.2020.9266160.

BIOGRAPHIES OF AUTHORS



Mouhcine Lakhroufi    received his master's degree in 2011 from Sidi Mohammed Ben Abdellah University, Faculty of Sciences, 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 Sidi Mohamed Ben Abdellah University, Faculty of Sciences, 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 international journals. His research has focused on the modeling and generation of intelligent control laws 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.