Precise Driver's Drowsiness Detection Using a Combination of Proven Methods with a single layer Neural Network

Ghazal Abdolbaghi and Alireza Yazdizadeh

Electrical Engineering Department, Shahid Beheshti University, Tehran, Iran

Introduction

Drowsiness reduces reaction time, leading to potentially fatal accidents. Most existing research focuses on only one symptom of drowsiness, which often results in false alarms. This paper introduces a novel approach for real-time drowsiness detection. The proposed method employs four deep learning architectures based on convolutional neural networks: AlexNet for environmental feature extraction, ResNet50V2 for hand gesture recognition, VGG-FaceNet for facial feature extraction, and FlowImageNet for behavioral feature analysis. To improve the voting system, the paper suggests using a layer with adjustable learning weights to reduce false results and enhance accuracy. Testing the static method on the NTHUDDD dataset and a custom dataset demonstrates that the proposed approach achieves higher accuracy (97.25% and 96.75% respectively) compared to existing methods.

The primary contributions of this paper are as follows:

1-This paper employs four deep learning structures based on convolutional neural net- works: AlexNet for environmental feature extraction, ResNet50V2 for hand gesture recognition, VGG-FaceNet for facial feature extraction, and FlowImageNet for behavioral feature analysis. Each structure is pretrained using transfer learning, with additional layers added to enhance results, thus leveraging the combined advantages of an integrated framework.

2- Instead of using a simple average voting mechanism to combine the four structures, this paper achieves better results by adding a layer to the outputs of these structures and adjusting the corresponding weights. This approach efficiently minimizes error and enhances accuracy.

To evaluate driver drowsiness, we adopt the levels outlined in [8], which are based on [9] with some difference's scales. The authors defined drowsiness on a scale with five levels, as shown in Table 1. The levels of drowsiness are shown (see Fig.1).

Drowsiness Levels	Features
1-Not Drowsy	Line of sight moves fast and frequently. Facial movements are active, accompanied by body movements.
2-Slightly Drowsy	Line of sight moves slowly. Lips are open.
3-Moderately Drowsy	Blinks are slow and frequent. There are mouth movements.
4-Significantly Drowsy	There are blinks that seem conscious. Frequent yawning.
5-Extremely Drowsy	Eyelids close. Head tilts forward or falls backward.

Table 1. Drowsiness levels

Methodology

This paper centers on Convolutional Neural Networks (CNNs), which are renowned for their capability to extract features directly from raw data and have shown fantastic performance. These networks are effective at building models that remain robust to transformations in the input data. In this study, rather than relying on a simple voting method, we implement a smart weighting approach by adding a layer at the output of the four models. Initially, transfer learning is used to enhance the models' accuracy by adding additional layers. Instead of using a basic voting mechanism, a single-neuron layer is introduced, where weights for each model are adjusted according to static rules.

Next, we examine the layers added to each structure through the transfer learning method and present the weight adjustment rules for each branch in the static mode of the ensemble model. We choose Exponential Linear Unit (ELU) as the activation function as we get better result with it.

It is important to note that for all models, we utilize the sparse categorical entropy loss function, which is well-suited for multi-class classification and yields superior results compared to other loss functions.

The layers illustrated, are incorporated into the pre-trained ResNet50V2, VGGFaceNet, AlexNet, and FlowImageNet using the transfer learning method to achieve improved results (see Fig.1, Fig.2, Fig3, and Fig.4). We calculate the weighting rules in static mode to create the smart weighting model. Ultimately, the proposed framework can be seen in Fig5.

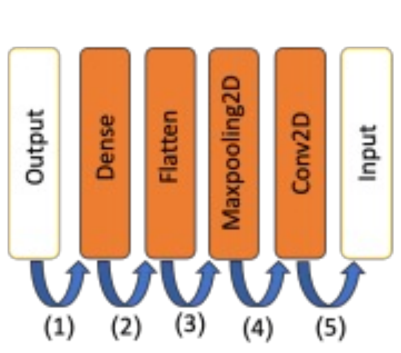

Fig.1. The layers added to a pre-trained ResNet50V2 structure

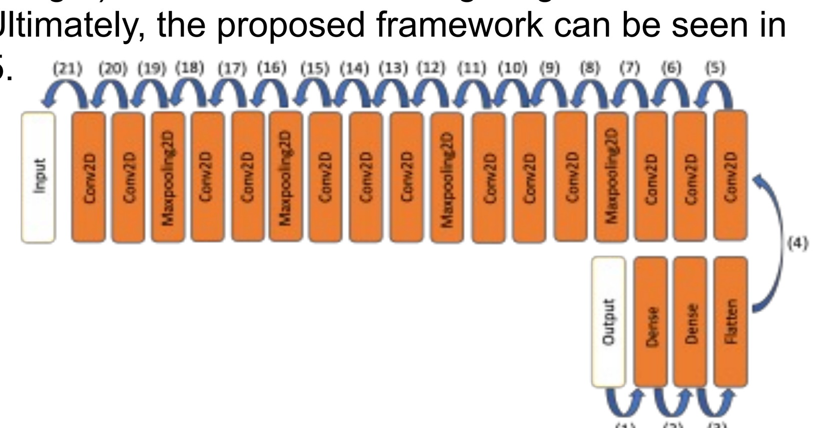

Fig.2. The layers added to a pre-trained VGG-FaceNet structure

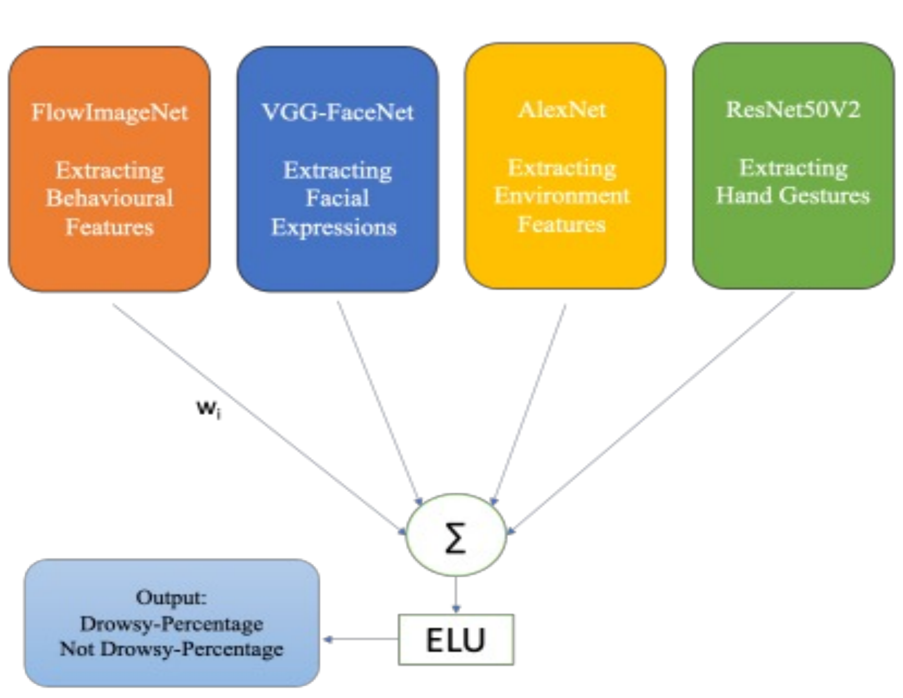

Fig.5. Proposed Framework

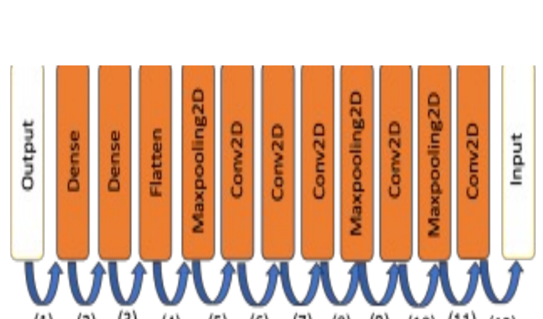

Fig.3. The layers added to a pre-trained AlexNet structure

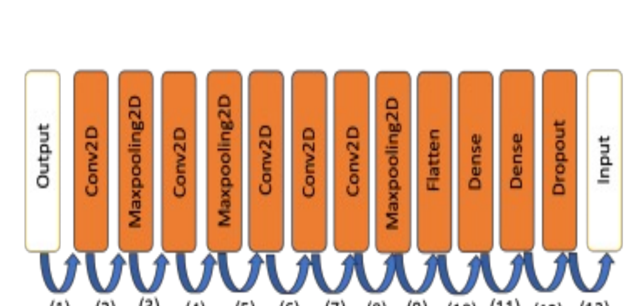

Fig.4. The layers added to a pre-trained FlowImageNet structure

Results

We have evaluated our proposed method using various datasets, first NTHUDDD which is a well-known dataset in drowsiness and a custom dataset that we compiled (see Fig.6).

Fig.6. Frames of our dataset

Given that our task involves classification, we utilized a confusion matrix to evaluate the performance of the proposed model. Specifically, we considered four metrics: ac- curacy, precision, recall, and F1-score.

Based on the calculations using the specified formulas, the results are summarized in Table 2. It is presented the confusion matrix, which includes four key parameters: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), il- lustrating two different scenarios (see Fig.7). It is displayed the accuracy and loss metrics throughout the training process for these scenarios (see Fig.8and Fig.9). The primary goal during network training is to minimize the loss (whether in terms of error or cost) observed in the output when the training data is processed. As evidenced by the training process in both scenarios, the accuracy for both training and validation steadily increases, while the loss for training and validation consistently decreases, indicating that the proposed method is functioning effectively. As we see, the values of TP and TN in scenario 1 for predicting drowsy and not_drowsy states are 96% and 98%, respectively (see Fig.8).

Num	Scenario	Label	Precision	Recall	F1-Score	Support	Accuracy
1	Static Structure on NTHUDDD	Drowsy -------------- Not-Drowsy	0.964 --------- 0.981	0.9803 --------- 0.9649	0.971 --------- 0.974	1010 ---------- 1153	**0.9725**
2	Static Structure on our Dataset	Drowsy -------------- Not-Drowsy	0.976 --------- 0.959	0.9595 --------- 0.976	0.9655 --------- 0.968	1010 ---------- 1153	0.9675

Table 2. The results of evaluation parameters on the proposed method with two datasets

Tables 3 and 4 highlight the differences in accuracy between two approaches: simple voting and the static model. The results clearly show that the static model yields significantly better accuracy across both datasets.

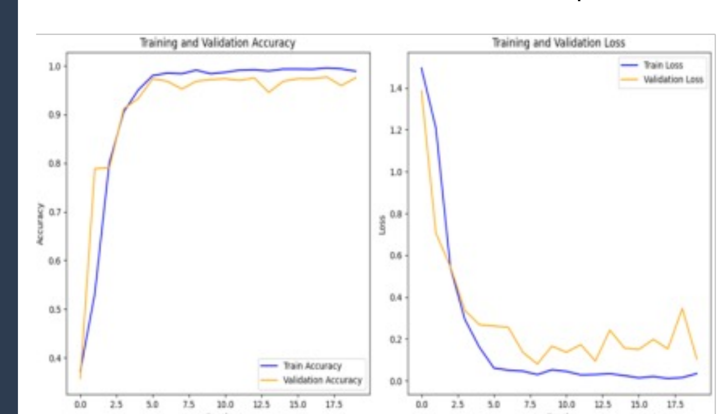

Fig.8. Metrics for scenario1:Accuracy and Loss

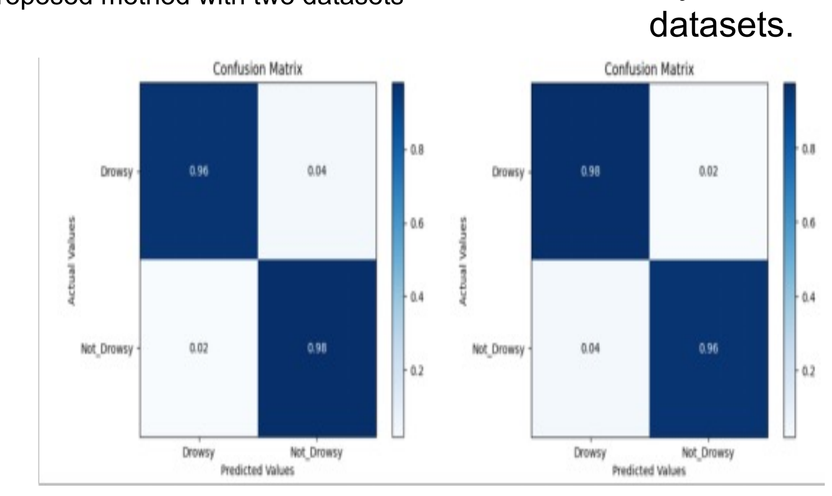

Fig.7. Confusion matrix for two scenarios

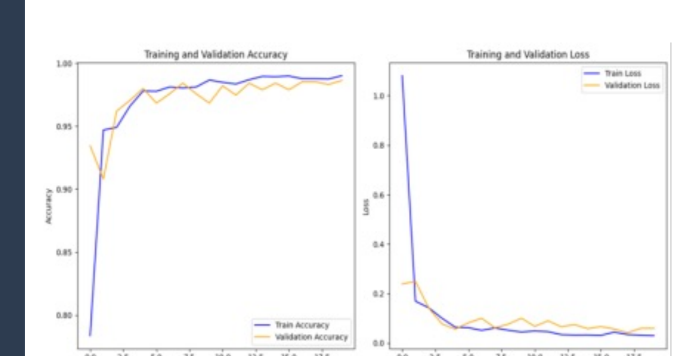

Fig.9. Metrics for scenario2:Accuracy and Loss

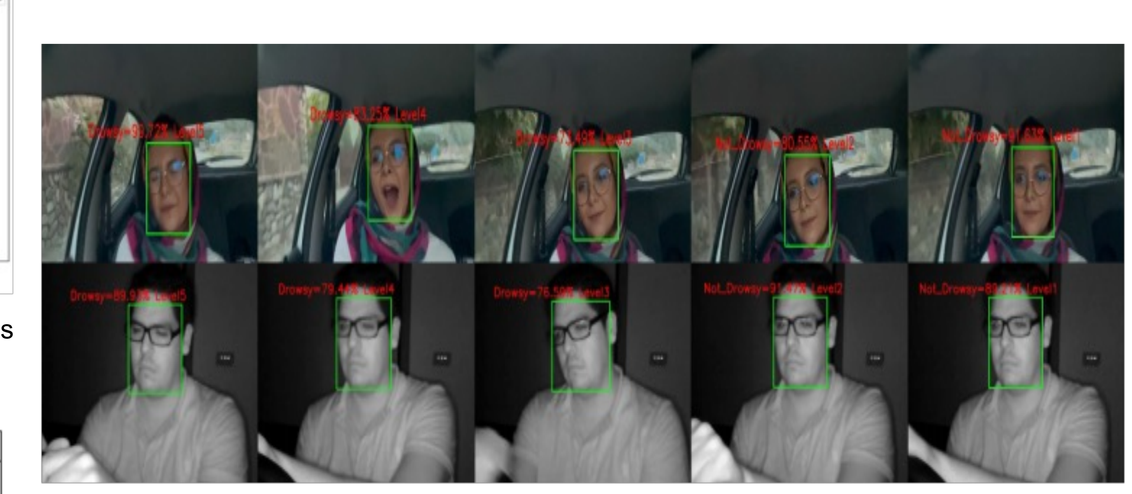

Fig.10. Evaluation of proposed structure on two datasets

Structure	Accuracy	Voting	Static Structure
AlexNet	92.73	87.47	**97.25**
VGG-FaceNet	83.03		
ResNet50v2	84.12		
FlowImageNet	90		

Table 3. Comparison of accuracy between two states on NTHUDDD dataset

Structure	Accuracy	Voting	Static Structure
AlexNet	92	86.49	**96.75**
VGG-FaceNet	81.93		
ResNet50v2	82.68		
FlowImageNet	89.35		

Table 4. Comparison of accuracy between two states on our dataset

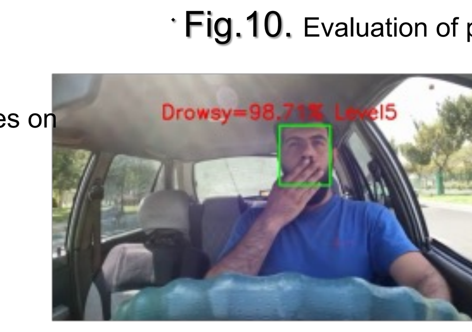

Fig.11. Detecting drowsiness even yawning with hand

As we see, the implementation results of the proposed structure on these datasets (see Fig.10). A notable advantage of this structure is its ability to accurately detect drowsiness even when yawning occurs with hands covering the face, as reflected by the impressive accuracy (see Fig.11).

Conclusion

This study utilized four convolutional neural network architectures AlexNet, Res- Net50V2, FlowImageNet, and VGG-FaceNet. The results from these networks were combined using static neural network approaches, leading to the final outcome. These frameworks were initially tested on the NTHUDDD dataset, which is widely recognized as a leading resource for sleepiness detection research. The results demonstrated that the proposed approach outperforms the individual methods. Following this, the method was evaluated on a dataset created specifically for this study, confirming its robustness across different datasets. **Notably, the static technique delivered significantly better results compared to earlier methods, such as averaging. Finally, the static architecture achieved accuracies of 97.25% and 96.75% on the two datasets, respectively.** In future work, the goal is to adapt this system for signal data, which differs from image data.

Contact

g.abdolbaghi@mail.sbu.ac.ir; a_yazdizadeh@sbu.ac.ir