IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 June 2024, accepted 26 July 2024, date of publication 5 August 2024, date of current version 18 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3438617

APPLIED RESEARCH

A Novel Hybrid Approach for Driver Drowsiness Detection Using a Custom Deep Learning Model

MUHAMMAD RAMZAN[1,2], ADNAN ABID[2,3], (Senior Member, IEEE), MUHAMMAD FAYYAZ[4], TAHANI JASER ALAHMADI[5], HAITHAM NOBANEE[6], (Fellow, IEEE), AND AMJAD REHMAN[7], (Senior Member, IEEE)

[1]Department of Software Engineering, Faculty of Computing and Information Technology, University of Sargodha, Sargodha 40100, Pakistan
[2]Department of Computer Science, University of Management and Technology, Lahore 54770, Pakistan
[3]Department of Data Science, Faculty of Computing and Information Technology, University of the Punjab, Lahore 54000, Pakistan
[4]Department of Computer Science, FAST–National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan
[5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[6]College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates
[7]Artificial Intelligence and Data Analytics Laboratory (AIDA), CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding author: Tahani Jaser Alahmadi (tjalahmadi@pnu.edu.sa)

The Funding of research is provided by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R513), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Driver Drowsiness Detection (D3) is a challenging task as it requires analysis based on various behavioral and physiological signs such as health issues, mental stress, and exhaustion. Data analytics reveals that driver drowsiness is the reason for one-fifth of all traffic accidents worldwide. Thus, safety devices are valuable for alerting sleepy drivers regarding more danger that may occur. Constant real-time drowsiness detection in complex conditions and denoting is still an open issue. However, facing these challenges, this article proposed a technique called Driver Drowsiness Detection using Custom Deep Learning Model (D3-CDLM). This approach contains four different modules: In the given procedure the investigation and exploration include, 1) feature extraction and selection; 2) machine learning and ensemble methods; 3) deep learning; and 4) the combination of the two, Hybrid. The first operation computes HOG, which stands for Histogram of Oriented Gradient, which is rotation and illumination invariant and resistant to the information in the local areas. Then Principal Component Analysis or PCA is used to obtain the best or top HOG features that are used as inputs to machine learning and ensemble methods-based modules. For hard to learn facial features, transfer learning is also carried out, and a new 30-layer CNNs structure is proposed called CDLM. Finally, the hybrid module's top features are investigated using the PCA control of the architecture in coordination with the proposed CDLM for detecting drowsiness. Empirical analysis that encompasses all districts were applied on Yawning Detection Dataset. The results reveal that the developed and designed deep learning and hybrid modules acquire better accuracy than the proposed and utilized machine learning-based module along with the compared existing approaches in the pertinent literature.

INDEX TERMS Unusual behavior, drowsiness detection, HOG features, PCA, deep CNN.

I. INTRODUCTION

Recognition from surveillance videos is trending and progressively domain emerged as an area of interest [1]. Such activities have been identified with the use of computer vision and deep learning techniques. Thus, it is required to differentiate between normal, suspect, and pathological motions of humans [2]. The video's media content relies on human action, which leads to human behavior analysis. Human behavior and actions are understandable based on different video features, which helps in categorizing activities as normal or abnormal, called suspicious. These actions are

The associate editor coordinating the review of this manuscript and approving it for publication was Joao Neves.

unusual in one pattern, but they may be reflected normally in another public place. For instance, running in the playground is normal, but running in a bank or marketplace is abnormal or suspicious [3]. The importance of recognizing human activity has grown due to the numerous challenges that must be overcome when recognizing action and human behavior. Drowsiness in a driver is also considered an unusual human activity [4]. Drowsiness or exhaustion on the road is one of the leading causes of serious injuries, fatalities, and economic losses. Driving performance suffers because of drowsiness. Road accidents are induced by a lack of awareness caused by the unintentional transition from awake to sleep [5]. Drowsiness of drivers is a physical condition caused by various factors such as health problems, mental stress, fatigue, and others [6]. These factors exhibit several symptoms that reflect the driver's condition, such as facial expressions and eye-opening or closing. According to international data on accident causes, these are one of the leading causes of traffic accidents, accounting for approximately 20% of all accidents. National Highway Traffic Safety Administration reported that sleepy driving causes around 100,000 traffic accidents and over 1,500 deaths in the United States each year [7]. Moreover, according to the World Health Organization and an American department named the National Highway Traffic Safety Administration (NHTSA), 100,000 mishaps happen in a year. Such mishaps lead to the loss of precious lives. Moreover, the costs of car accidents in Europe are estimated to be close to 160 billion Euros due to driver drowsiness. Lack of sleep, a lengthy journey, restlessness, alcohol intake, and mental stress can all contribute to driver sleepiness. Each of these scenarios carries the risk of being disastrous. As a result, the current transportation infrastructure is insufficient to handle these road risks. Some tragic incidents can be avoided by integrating autonomous fatigue detection technology into autos. The tiredness detection system constantly monitors the driver's attention level, which informs the driver before a severe threat to road safety occurs. Researchers have proposed numerous approaches for detecting driver drowsiness due to driver weariness's dangers on the road, each with its pros and cons. Drowsiness must be detected and treated as early as possible [8]. Keeping in view of the above discussion, this study proposes an automated system for detecting drowsiness that employs machine-learning and deep-learning modules individually as well as combined. In this regard, four modules are introduced in the proposed method, each of which is trained independently and used to make predictions. In this study, there are two types of output categories such as binary and ternary are used for predictions: (1) binary output consists of two classes such as normal/not yawning and yawning, and (2) ternary output consists of three classes such as normal/not yawning, yawning, and talking. The following are the main research contributions:

1. A robust pipeline named Driver Drowsiness Detection using Custom Deep Learning Model (D3-CDLM) is proposed to identify drowsiness from video frames.
2. HOG features are computed from video frames and PCA is applied for dimensionality reduction. Then, the PCA-based feature set is supplied to the classifier to evaluate the outcomes in terms of accuracy, precision, recall, and F1_measure related to D3 under the machine-learning module.
3. A novel 30-layer deep CNN architecture is also designed for binary (yawning and normal/not yawning) and ternary (yawning, not yawning, and talking) classification under a deep-learning module.
4. In a hybrid module, the use of HOG+PCA based features set with the proposed 30-layer architecture improves the overall accuracy when tested on a challenging Yawning Detection Dataset (YawDD) video dataset.

The rest of the paper is organized as follows: Section II examines the literature review on driver drowsiness detection. Section III goes over the proposed methodology in depth. Section IV discusses the dataset, experimental setup, and performance evaluation measures. Section V contains the empirical findings and discussion. Section 6 presents the proposed research study's conclusion and future work.

II. LITERATURE REVIEW

In recent years, several approaches for Human Activity Recognition (HAR) from video have been introduced. Many researchers have developed fully automated HAR for drowsiness detection using machine learning, ensemble, and deep learning-based methods. According to [9], this section divides existing literature into three categories based on drowsiness level: behavioral, physiological, and vehicular measures.

A. BEHAVIORAL MEASURES

Non-intrusive methods for detecting drowsiness rely on behavioral measures. These methods use behavioral cues such as head posture, eye closure ratio, blinking, yawning, and facial expressions to determine how tired the driver is. Current methods concentrate on the driver rather than the vehicle to acquire such measurements. They keep an eye on the driver's eye, facial, head, and yawning patterns using the camera [10]. Behavioral measures are provided by the driver's eye detection, eye blink rate, yawning, and facial expression analysis. With a single camera-based system, few issues are raised. For instance, computations are increased due to dealing with multiple images simultaneously. Also, it does not capture head movement and produces poor results. To decrease the computations, Park et al. [11] proposed a camera-based driver drowsiness system in which the shallow convolutional neural network (CNN) is used for the driver's face images obtained from two cameras to adaptively select camera images more suitable for detecting eye position. They employed a faster R-CNN for the driver images that were chosen, and once the driver's eyes are identified, the eye positions of the camera image on the opposite side are mapped using a geometric transformation matrix. The Columbia Gaze Data Set (CAVE-DB) open database and the self-built Dongguk Dual Camera-based Driver Database (DDCD-DB1), which included photos of 26 individuals taken from inside a car, were used for the experiments. The outcomes demonstrated that the suggested strategy performs better under these settings. Also, D3 approaches have a significant problem with darker-skinned drivers. In this regard, Ngxande et al. [12] proposed an innovative visualization technique that can assist in detecting groups of people where discrimination may exist. They employed Principal Component Analysis (PCA) to generate a grid of faces sorted by similarity and combined this with a model accuracy overlay. Three deep neural architectures such as ResNet50, VGG16, and InceptionV3 are fine-tuned by [13] to learn facial highlights comprised of 68 RGB video contribution properties. The VGG network has 16 layers, 13 of which are convolution layers and the remaining three are fully connected. Because the network is excessively deep, VGG has a model size of 533MB for VGG16. Whereas ResNet50 is 102MB in size and has 50 weight layers. ResNet50 uses global average pooling rather than fully connected layers, resulting in a smaller model size but a deeper model than VGG16. As a result, the ResNet50 model size has been reduced. The InceptionV3 network has 22 layers, 20 convolution layers, and 2 fully connected layers. Inception V3 weighs 96 MB, significantly less than the other two networks. The SoftMax classifier in these networks classifies the driver's sleepiness based on facial movements such as eye squinting, yawning, and head swaying. According to the findings of the experiments, InceptionV3 outperforms and reaches 78% accuracy. A method named two-stream network models with multi-facial features is presented by [14] for driver fatigue detection. This method comprises four sections. (1) Locating mouth and eye using Multi-Task Cascaded CNNs (MTCNNs) to perform multiple convolutional neural organization tasks. (2) Removing static structures from the fragmented facial picture. (3) Removing dynamic highlights from the inadequate facial optical stream. (4) Bonding both static and dynamic highlights with the assistance of two-stream neural organization for grouping reasons. Whereas partial face images used as network inputs can concentrate on information linked to weariness, two-stream networks can integrate static and dynamic image information, leading to improved performance. Additionally, they used gamma correction to improve image contrast, which improves the effectiveness of this technique. Another technique towards robust drowsiness detection proposed by [15] consists of two steps: the joint face detection and alignment step, followed by the drowsiness detection model step. They also employed the fastest, and most accurate face detectors named MTCNN for the face identification and alignment task. They compress the baseline model into a lightweight model deployable on an embedded board. Furthermore, a reduced network structure was built based on facial landmark input to determine whether the driver is drowsy or not. To effectively assess drowsiness levels, developing a detection system mostly depends on identifying and interpreting important facial cues. The face serves as a rich source of information, displaying a variety of cues like yawning, head movements, eye flashing, and a wide range of facial traits. However, there is a big issue in accurately and safely computing these requirements. The complexity arises from the need to measure multiple dynamic facial expressions and features consistently and precisely. Deep learning research is currently boosting D3 systems, notably in the context of exploiting facial features. For instance, Ngxande et al. [16] discussed different machine learning techniques such as support vector machines, CNN, and Hidden Markov models that are applied to detect the drowsiness level. These techniques are used to train models for drowsiness prediction by using facial features and labeled outputs.

B. PHYSIOLOGICAL MEASURES

Electroencephalograms (EEG), electrocardiograms (ECG), photoplethysmography (PPG), Source Ground Return Protection (SGR), pulse, and other electronic devices are attached to the human body to acquire physiological measurements [8]. Such measurements involve the application of physiological approximations to determine the level of fatigue in drivers. The precise ratio of physiological indicators to driver fatigue is taken into consideration when analyzing them. One of the most reliable methods for determining weariness is electroencephalography (EEG). For instance, using a CNN architecture, the authors proposed a new DD system based on EEG signals. In this study, they utilized Emotiv EPOC+ headset as a signal collection tool. Additionally, tiredness has been identified in our EEG data by the examination of alpha and theta waves from the temporal and occipital regions [5]. Despite EEG, the electrooculogram (EOG) based CNNs employed as an unsupervised learning approach for driver drowsiness detection [17]. To avoid utilizing manual features, a CNN with a linear regression layer is used for EOG signals. Also, a post-processing step named Linear Dynamic Systems (LDS) is implemented to capture the physiological status shifting. Finally, the correlation coefficients between the final outputs and the individuals' local error rates are used to assess how well the suggested model performs. Normally, visual information is required to classify drowsy driving and also make decisions based on the wavelet shift of Heart Rate Variability (HRV) signals over short time intervals [18]. Authors categorized driving events into alert and sleepy categories using the wavelet transform of HRV signals across brief time intervals. The performance of this approach in classification is compared with the traditional method, which employs characteristics based on the Fast Fourier Transform (FFT). According to the collected results, the averaged leave-one-out (LOO) classification accuracy utilizing the wavelet-based feature is 95%, whereas the FFT-based accuracy is 68.8%. Frike et al. [19] proposed a reliable method of estimating driver drowsiness which begins with the physiological signals and video signals. Specifically, the aim of this research was to propose a system for the automatic detection of drowsiness by using the driver's video and EEG signals wavelets, which exhibit brain signals. Moreover, this technique is handled, based on the comprehensive understanding of physicians who controlled the driver drowsiness that mainly tend to utilize ocular and cerebral data for detection. The first stage's key responsibility was to control the biosensor linked to a driver and examine its output to exhibit which features are linked to drowsiness. In the second stage, they created a sleepiness detection algorithm and a mobile app to alert drowsy drivers. Warwick et al. [20] developed a strategy that involved using a biosensor to collect physiological data from the driver and then analyzing the findings to discover the main elements driving drowsiness. To warn sleepy drivers, they created a mobile app and established an algorithm for sleepiness detection in the second step. In [21], the authors considered three groups of measures such as heart rate, brain activity, and alertness monitoring, and proved the relationship to (mental) work conditions in a simulator environment. EEG and EOG are examples of physiological signals that are useful non-invasive instruments to determine an individual's drowsiness [22]. EEG signals are non-fixed and have distinct characteristics, in which simple strategies effectively detect drowsiness levels. Also, prior techniques cannot achieve results without taking fundamental estimates hidden from the new signals. To solve these shortcomings, Chen et al. [22] proposed a technique for detecting drowsiness using physiological cues, which has four advantages: (1) decomposing EEG signals into wavelet sub-bands to extract more obvious information beyond raw signals; (2) extraction and fusion of nonlinear features from EEG sub-bands; (3) fusion of information from EEGs and eyelid movements; and (4) use of an efficient extremely learning machine for status classification. The outcomes of the proposed method not only achieve a high detection accuracy but also a very rapid computing speed. In [23] authors provided deep learning-based solutions for the prediction of drowsy or alert states of drivers using EEG information. In this technique, they discussed the new channel-wise convolutional neural network (CCNN) and its variant, CCNN-R, which substitutes the restricted Boltzmann machine for the convolutional filter. In addition, they examined bagging classifiers that utilize DL hidden units as a substitute for traditional DL solutions.

C. VEHICLE-BASED MEASURES

Vehicle-based measures such as Steering Wheel Movement (SWM), acceleration pedal, break patterns, and lane position are used in vehicle-based assessments [2], [24]. In vehicle-based techniques, sensors are placed on various parts (e.g., acceleration pedal and steering wheel) of the vehicle to detect driver drowsiness. In [25] authors utilized vehicle data based on deep neural networks for multi-level categorization of driver sleepiness. There are three different degrees of tiredness taken into consideration: alert, somewhat drowsy, and excessively drowsy. Five vehicle-based signals are investigated as input signals: steering wheel angle, steering wheel angular velocity, lateral deviation, yaw rate, and lateral acceleration. To automatically extract features from these input data, convolutional layers are used. These layers used the geometry of the input signals to identify important patterns. More complex features may be extracted by deeper convolution layers than by shallower levels.

For the classification of driver drowsiness, three distinct deep network structures are suggested in this study. The output of a convolutional network (CNN) is regarded as the final feature in the first network and is fed into a SoftMax classifier. As an additional hidden layer, a CNN has been merged with a Gated Recurrent Unit (GRU) layer in the second structure. The same strategy is used in the third network, which includes a Long-Short Term Memory (LSTM) layer on top of the CNN. GRU and LSTM are two forms of Recurrent Neural Networks (RNN) that interpret the extracted CNN characteristics to understand their temporal dynamics. Yang et al. [26] presented a vehicle-based D3 approach in which they recorded and analyzed the driver responses (left-lane drifting). This study used vehicles to detect driver drowsiness through the vehicle's active probe action. Six indicators of drowsiness are extracted from drivers' responses and then used to detect driver drowsiness using three recognition methods, namely, support vector machine, Gaussian kernel density estimation, and back-propagation neural networks, in comparison to traditional monitoring features such as steering-wheel movement. Experiment results show that the suggested active probe strategy beats standard monitor approaches for detecting driver tiredness. In [27] authors investigated a machine learning-based driver sleepiness monitoring and early warning system using car telemetry data. Through the real-time monitoring of driving patterns, the suggested system can guarantee safe driving. Since it doesn't require costly sensors, the proposed methodology turns out to be much more affordable than biometric and camera-based methods. In this work, the identification of a driver's drowsiness is formulated as a binary classification problem. As discussed above, behavioral, physiological, and vehicle-based measures effectively characterize the information for the detection of driver drowsiness. These measures passively monitor driver states, but they still are vulnerable to the uncertainties of different driving environments and individual differences. By considering behavioral measures, this manuscript proposes a robust D3 classification method using a blend of machine and deep learning approaches individually and jointly.

The main objective is to purposefully exploit driver behavioral characteristics (as described in section II-A) instead of physiological and vehicle-based measures to identify drowsy driving. Table 1 provides a comparison between existing methods that have been utilized by the researchers for driver drowsiness detection/classification.

III. PROPOSED METHODOLOGY FOR DRIVER DROWSINESS DETECTION

This section discusses the proposed method D3-CDLM for drowsiness detection, as shown in Figure 1. In this study, the

TABLE 1. Summary of the existing methods in terms of applied models, features, datasets, classes, and results.

Measure	Ref	Year	Model	Feature	Dataset	Classes	Results Acc (%)
Behavioral Measures	[11]	2019	Shallow CNN, Faster R-CNN	Driver's eyes	DDCD-DB1, CAVE-DB	2	99
	[12]	2020	CNN, PCA, VGGFace, VGG, and ResNet	Facial	NHTU-drowsy and DROZY CEW	-	85
	[13]	2019	ResNet50, VGG16 and InceptionV3	blinking, yawning, and head swaying	Drozy	3	78
	[14]	2019	MTCNNs	Mouth and eyes	NTHU-DDD	5	97.3
	[15]	2017	faster-RCNN, MTCNN, DNNN	Full Face	Drozy	3	90.5
Physiological Measures	[5]	2021	CNN-EEG	Partial and Full-face mask	Self-created	2	90
	[17]	2014	CNN-EOG, EEG	Eyes, partial face	Self-created	2	98
	[18]	2013	SVM-FFT, HRV	Heart rate wavelet-based features	Self-created	4	95
	[28]	2022	MTCNN, MSP, RESNET, Adaboost	Eye and mouth left or right	CSW, Yaw DD, Self-created	4	98
	[29]	2016	HTDBN	Eyes, Mouth, Head – Pairwise movement	Self-created	4	85.3
Vehicle-Based Measures	[25]	2020	CNN-LSTM, RNN, GRU, LSTM	face or eyes rubbing, restless posture, yawning, normal facial tone	drozy	3	96
	[26]	2019	SVM, SWA, GDKE, and BPNN	Steering, wheel, left or right lane movement	-	2	97
	[27]	2017	LR, SVM, MLP, RF	Signal processing feature	-	2	98
	[30]	2023	CNN-YOLO	Eyes and mouth	YawDD	2	96
	[31]	2021	CNN, MQ13, MAX30105, and L298N sensors	Face and eyes	Self-created	2	-

proposed work consists of four modules: machine learning and ensemble methods-based, deep learning-based, and a hybrid including feature extraction and selection. The proposed method uses the benchmark dataset to classify the presence of drowsiness. Initially, frames with unusual behavior are extracted from videos and then resized according to model input such as input with dimension $64 \times 64 \times 3$ for a deep-learning module. In subsequent sections, feature extraction and dimensionality reduction are carried out using HOG descriptor and PCA, respectively. The details regarding feature extraction and selection, and machine-learning/ensemble methods are provided in subsequent sections.

A. FEATURE EXTRACTION AND SELECTION MODULE

Histograms of an image represent a visualization of the predominant intensities of an image and are widely used as handcrafted shape-based feature descriptors [32], [33]. The image shows the frequency of pixels' intensity values [34] Considering the localized area of an image, horizontal and vertical gradients are calculated. The feature vector is made up of orientation displays gradient angles of 9 bins in the range [0,180], and then many image pixels that belong to each cell are selected. In the proposed module, there are 8×8 pixels per cell; hence cells per block (2×2) are set for contrast normalization across the blocks. Finally, the normalization parameters are set to true to make the descriptor independent of illumination invariance and achieve reasonable accuracy [35].

In this work, to compute the histogram features, the magnitude and orientation of gradients of each pixel are calculated using equations (1)-(2). Computed features can well describe local area information using gradient and orientation density distribution of the edge [36].

$$G = \sqrt{g_x^2 + g_y^2} \tag{1}$$

$$\theta = arctan\frac{g_x}{g_y} \tag{2}$$

where G and θ represent gradient magnitude and orientation of gradients or angles, respectively, however, g_x and g_y represent an axis to localize a specific pixel. For best feature extraction, PCA brings out strong patterns of input images. PCA is a statistical technique for identifying essential factors that explain variation in observations in a high-dimensional vector space. HOG feature extraction and selection is shown in Figure 2.

A few samples with HOG features representation are shown in Figure 3. PCA is a well-known linear feature extractor based on a few mathematical concepts such as Eigen Vectors and Eigen Values, Variance, and Covariance. It is widely applied as a common linear dimensionality reduction technique (PCA). PCA is an unsupervised technique for mapping or embedding data points from a high-dimensional space to a low-dimensional space while preserving significant linear structures. So, this is a feature extraction technique meant to reduce the dimensions of our input data. The data must be pre-processed to improve the efficiency and accuracy of data mining tasks on high-dimensional space, using an efficient dimensionality reduction method [32]. In this module, local features of the input image/video frame are obtained

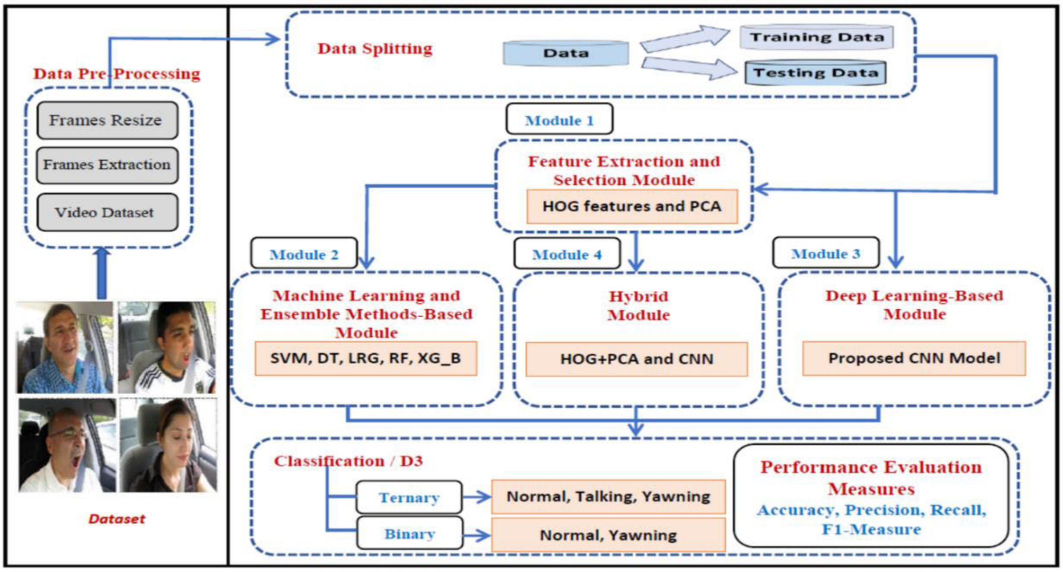


FIGURE 1. Overview of the proposed methodology consists of four modules (1) Feature Extraction and Selection Module, (2) Machine Learning and Ensemble Methods-based Module, (3) Deep Learning-based Module, and (4) Hybrid Module for Driver Drowsiness Detection.

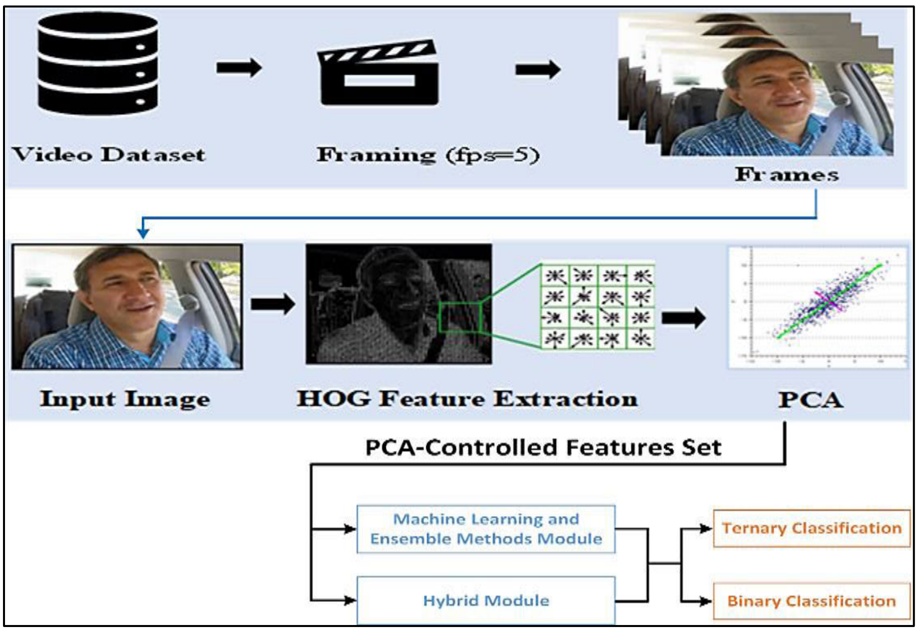


FIGURE 2. HOG feature extraction and dimensionality reduction using PCA.

using the HOG descriptor. Later, features are extracted by applying the PCA algorithm. The selected features subset is supplied to machine-learning and ensemble methods, and a hybrid module for drowsiness detection for both binary (normal/not yawning and yawning) and ternary (normal/not yawning, yawning, and talking) classification.

B. MACHINE LEARNING AND ENSEMBLE METHODS-BASED MODULE (MLEM-M)

This section discusses applied machine learning and ensemble methods such as support vector machine (SVM), decision tree (DT), linear regression (LRG), random forest (RF), and Extreme Gradient boosting (XG_B) which are widely used for classification tasks. Figure 4 covers these methods in the pipeline of the proposed MLEM module for the classification of driver drowsiness. The details related to these methods are provided in subsequent sections.

1) SUPPORT VECTOR MACHINE (SVM)

SVM is one of the extensively used supervised learning models/classifiers in machine learning. It can assign new data points to different categories when given a series of training

FIGURE 3. Results after applying the HOG descriptor on sample frames of YawDD video dataset.

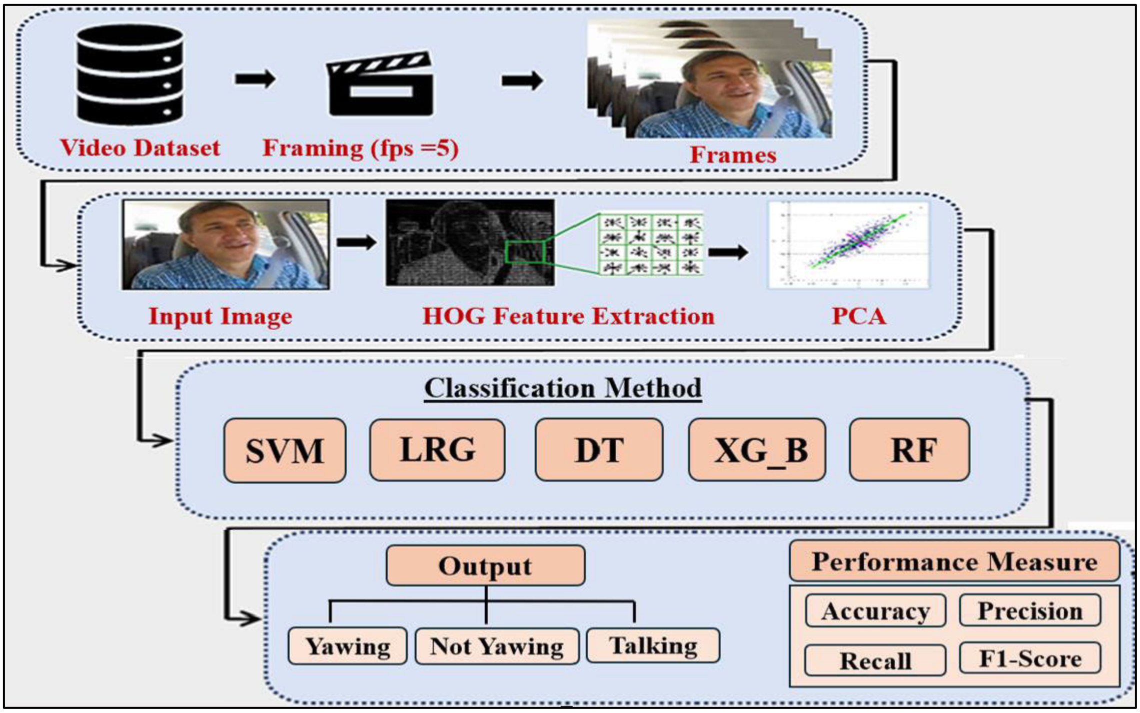


FIGURE 4. A general pipeline of the proposed MLEM module for D3.

samples [37]. This classifier is also a training algorithm designed to optimize linear and nonlinear separable problems by defining the maximum margin between data points [38]. The main advantages of this classifier is that it addresses the issue of over-fitting during training, and improves the generalization ability of learning [39]. The SVM classifier takes a long time to train on large datasets but produces very accurate classification results when modeling complex non-linear decision boundaries. Therefore, the proposed MLEM module utilizes SVM for the classification of driver drowsiness. The SVM formulates a set of hyperplanes in an infinite dimension based on labeled training data. The hyperplane with larger distance adjacent training points naturally causes good separation [12]. The hyperplane point division$H$ can be calculated using the following equation.

$$H = w^T (x) + b \tag{3}$$

where b and x are represent the intercept base term and feature vector, respectively and w is the direction of the vector.

2) DECISION TREE

The decision tree is shaped like a tree and is made up of a series of choices that have the potential to produce rules for classifying the dataset. Every distinct leaf node is devoted to a record that, following the splitting criterion, continuously advances toward a child node from the root. With the input records, the splitting criteria assess a branching condition on the present node [40]. Decision trees are simple to analyze and understand, providing valuable insights and allowing for the insertion of additional scenarios. They may calculate the best, average, and worst values for various circumstances and explain the results using Boolean logic as a white-box model. Additionally, decision trees can be integrated with other decision-making strategies. Therefore, in this study, the proposed MLEM module exploits the decision tree for the classification of driver drowsiness. The image understanding system is trained using a decision tree approach to perform supervised machine learning. The many attributes of the decision tree are formed by the image's various low-level characteristic features (for example, color, shape, texture, and so on) [31]. How these attributes are identified for the root node at each level is a major concern in decision trees. In this regard, uncertainty/impurity in a node is calculated using the entropy measure, which is associated with data. Entropy controls how a decision tree decides to split the data. It affects how a decision tree draws its boundaries. Entropy values range from 0 to 1, the less the value of entropy is trustworthy. These splits and boundaries are helpful in classification.

$$E = -\sum_{i=1}^{n} p_i \log(p_i) \tag{4}$$

where E represents entropy and p_i is the probability of an arbitrary tuple i associated with a particular class.

3) LINEAR REGRESSION

Linear regression analysis is a statistical technique that predicts the value of one variable based on another. The dependent variable is the one you wish to predict. The independent variable predicts the value of the other variable. A linear regression model is employed to evaluate this association and extract textural elements that distinguish the samples [41]. Linear regression is easy to use and understand, providing straightforward insights into feature influences. It is computationally efficient, making it ideal for large datasets and rapid forecasting. It also serves as a useful baseline for comparing performance to more complicated models. Equation 5 illustrates the linear regression model.

$$\hat{P}_i = l_0 + l_1 x_i \tag{5}$$

where $\hat{P}_i$ is the predicted value for i observation, and x_i is used to represent the value of x for i observation. Moreover, l_0 and l_1 represent the intercept of the line and slope of the regression line, respectively.

4) RANDOM FOREST

Breiman [42] created the ensemble learning technique known as random forest to address issues with regression and classification. Ensemble learning improves accuracy by combining different models to tackle a problem. Ensemble classifiers produce more accurate results than single classifiers. Using numerous classifiers reduces variation, particularly for unstable classifiers, and can lead to more dependable findings [43]. Random forest [10] is a classification ensemble-building method that uses decision trees that grow in randomly selected data subspaces. Random forest gives preference to hyper-parameters. According to experimental results, random forest classifiers can correctly classify the data in domains with multiple classes. Random forests have recently attracted image classification [31] and bioinformatics tasks.

5) XG BOOST

High-performance boosting is achieved with XGBoost, which reduces the loss function and is tuned through several configurations. It is a gradient-boosting technique that repeatedly adds models to a community by going through loops. The fundamental idea underlying boosting is to concentrate on difficult or unpredictable cases that the model is unable to accurately forecast. To make such measures look probable in a sample, the distribution of observations is skewed to emphasize these cases more. As a result, the subsequent weak student will concentrate more on accurately estimating difficult cases [44]. XGBoost, a potent predictor, is created by integrating all the straightforward prediction principles into a single general model [45]. Therefore, in this study, the proposed MLEM module exploits the XG boost for the classification of driver drowsiness. XG Boost is also popular among data scientists' scalable machine-learning approach for tree boost that prevents overfitting. It is a machine-learning technique widely used for classification and regression tasks. For classification and regression, XG Boost employs a K-tree ensemble. It is self-sufficient and has demonstrated success in various machine-learning challenges [46]. XG boost is a better option because it gives importance to functional space.

6) DEEP LEARNING-BASED MODULE (DLM)

Deep Learning (DL) is a subset of machine learning and is widely used in different domains such as digital image processing [39], [47], video recognition tasks [48], [49], and Natural Language Processing (NLP) tasks [50]. To efficiently comprehend data, the DL models incorporate several interpretation levels. Unlike machine learning, these deep learning models combine feature extraction and classification into a single module. These models achieve optimal results by incorporating complex methods based on massive data [51]. This study proposed a novel deep learning-based CNN architecture named CDLM for detecting driver drowsiness. This architecture divides the data into three categories: yawning, not yawning, and talking while driving. Visual information is used to extract features from input video frames which effectively detect drowsiness in unseen video frames. Further details related to CNN and the proposed architecture are described in the subsequent sections.

7) CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is still an authoritative model confined to several network layers in the DL domain. Each layer acquires unique information from input and passes it to the following layers [52]. The first layer of CNN architecture is an input layer that takes an image from an external source. The successive

TABLE 2. Proposed CNN architecture with 30 layers named CDLM.

Chunk No.	Layer No.	Layer Name	Maps and Neurons	Padding	Kernel Size
	0	Image Input	64 × 64 × 3	Valid	-
1	1	Convolutional	32	Same	3x3
	2	Activation	ReLU	-	-
	3	Batch Normalization	-	-	-
2	4	Convolutional	32	Same	3x3
	5	Activation	ReLU	-	-
	6	Batch Normalization	-	-	-
	7	Max Pooling	2×2	Valid	-
	8	Dropout	0.25	-	-
3	9	Convolutional	64	Same	3x3
	10	Activation	ReLU	-	-
	11	Batch Normalization	-	-	-
	12	Dropout	0.25	-	-
4	13	Convolutional	128	Same	3x3
	14	Activation	ReLU	-	-
	15	Batch Normalization	-	-	-
	16	Max Pooling	2x2	Valid	-
	17	Dropout	0.25	-	-
5	18	Flatten	-	Valid	-
	19	Dense	512	-	-
	20	Activation	ReLU	-	-
	21	Batch Normalization	-	-	-
	22	Dropout	0.5	-	-
	23	Dense	128	-	-
	24	Activation	ReLU	-	-
	25	Batch Normalization	-	-	-
	26	Dropout	0.5	-	-
	27	Dense	Unit=1	-	-
	28	Activation	Sigmoid	-	-
	29	Output	Binary_crossentropy	Same	-

layers include convolutional layers, pooling layers, activation layers, flattened and dense layers, etc. The main advantage of DL over ML models is self-organization and self-leaning. A detailed description of the CNN model is given below:

Image input layer: CNN architecture is incomplete without the image input layer. This layer receives a two-dimensional frame from an external source as input. The image's input size, on the other hand, must be set in the layer's parameters. The input layer's equation can be written as:

$$x = Input\ (V, D) \tag{6}$$

where x denotes the sequential model which takes input with dimension $V = m \times n$ size and D represents a depth of input image. Here, *Input* is a function that accepts the shape of input, passing to the first network layer.

Convolutional Layer: The convolutional layer performed convolutions on the input image. A general mathematical description of convolutional layers is given below:

$$output = conv(V, k, S_d, R) \tag{7}$$

where V is an input with dimension$V = m \times n$, k is kernel window size, S_d is the stride size, and R represents Rectified Linear Unit (ReLU). The ReLU is an activation function used to break up the model linearity. The following equation shows the convolutional process on the input. In this regard, a specified size kernel k is convolved with the input image to extract the feature map using the following formula.

$$G[m, n] = (V \times k)\,[m, n] = \sum_i \sum_j k\,[i, j]\, V(m - i, n - j) \tag{8}$$

In the above equation, V and k are image size and kernel, respectively. Moreover, m and n have represented the rows and columns of the input image, respectively.

Batch Normalization Layer: By increasing the learning rate, this layer is likely to improve the model's performance. It also normalizes prior layer activations for individual batches. It also kept the mean activation close to 0 and the standard deviation close to 1. It takes the activation from the feature map calculated with equation (8) and normalizes it

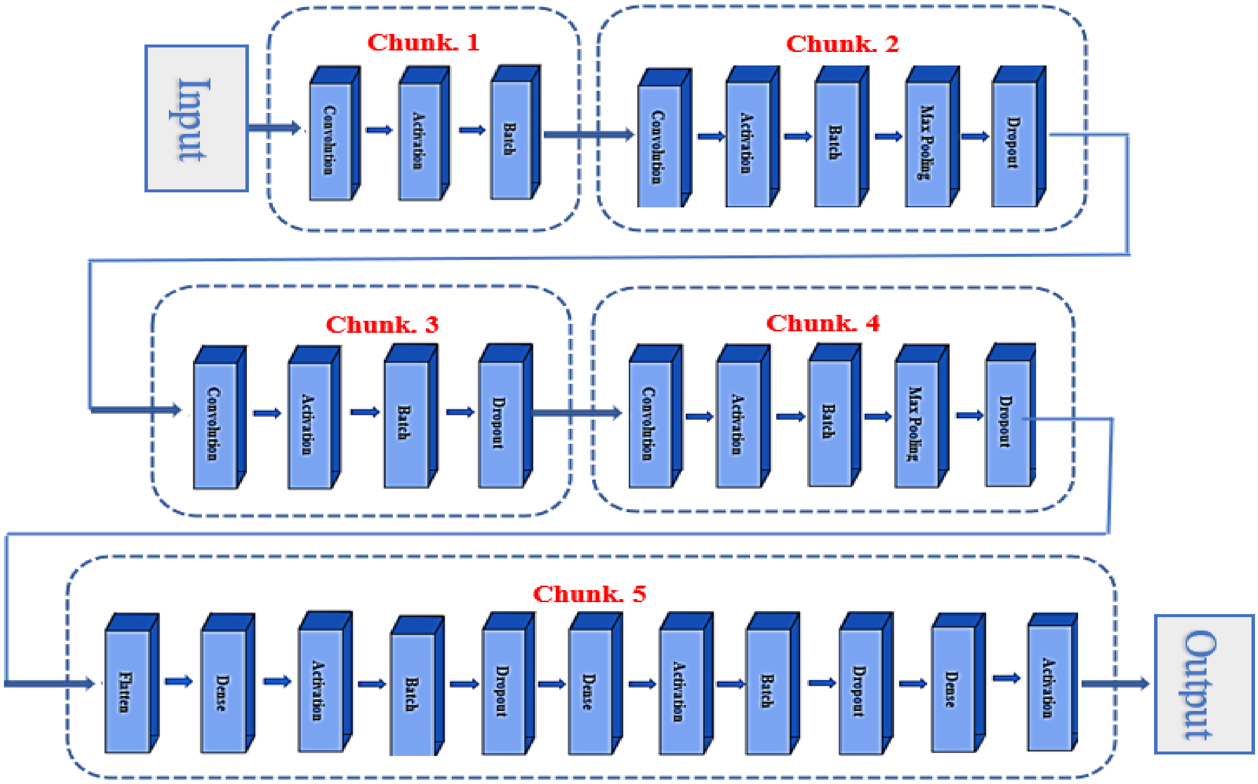


FIGURE 5. Proposed 30-layer CNN architecture named DLM/CDLM for D3.

with the following equation:

$$F_{C+1} = N\left(F_c, ax, \bar{M}\right) \tag{9}$$

where $N()$ denotes the normalization function and ax and $\bar{M}$ represent the axis (dimension) and momentum (parameter), respectively. The summary of the proposed 30-layer CNN architecture is shown in Table 2.

Max Pooling Layer: The normalized feature map computed using equation (9) is passed to the max-pooling layer for pooling operation. This layer reduces the size of the input image compared to the initial image size. The equation for the pooling operation is given below:

$$F_{C+1} = maxPool\left(F_{C+1}, P_s\right) \tag{10}$$

where Ps denotes the pool size.

Fully Connected (FC) Layer: The FC layer contains information about all layers and connects the previous layer's neuron with successive layers. These layers also acknowledged class labels and class predictions. The equation for FC is given below where X_i is the output of the FC layer.

$$\begin{aligned} X\_i = F\_\text{connect}(&conv(V, k, S\_d, R), \\ &N(F\_c, ax, M^{-}), maxPool(F\_(C+1), P\_s) \end{aligned} \tag{11}$$

8) PROPOSED CNN ARCHITECTURE

The traditional CNN model comprises two primary parts. The first part consists of multiple operations such as convolutional, pooling, normalization, ReLU, etc. FC takes input from all previous layers, and then a weights matrix is used to apply a linear transformation to the input vector and predicts the output class in the next part.

Keeping in view the CNN parts, a 30-layer CNN architecture CDLM is proposed in this work, where all network layers are represented with $L_0, L_2, L_3, \ldots.., L_{29}$ for D3. Overall, the proposed architecture is divided into five chunks. The first three chunks contain 12 layers $\{L_1, L_2, L_3, \ldots.., L_{12}\}$ excluding first input layer. The fourth chunk contains 5 layers $\{L_{13}, L_{14}, \ldots.., L_{17}\}$ and the fifth chunk contains 11 layers $\{L_{18}, L_{23}, L_{24}, \ldots.., L_{28}\}$ excluding last output layer. Table 2 and Figure 5 show the tabular form and graphical representation of the proposed DLM, respectively. In the first three chunks, the repeating layers are about the same as L_1. After the input layer is convolutional layer, which performs convolution operations on the input image, L_2 is a ReLU activation function that breaks the network linearity, L_3 is batch normalization layer, L_4 is again convolutional layer, L_5 is a ReLU activation function, L_6 is batch normalization layer, $L_7$is max pooling layer and L_8 the dropout layer is used to drop some instances tentatively from the architecture. Moreover, dropout also prevents the model from overfitting. An additional flattened and dense layer is used in the fourth and fifth chunks. The flattened layer converts two-dimensional features into a one-dimensional array for class prediction. Hence, the dense layer receives input from all neurons in the entire model. According to Figure 5, chunk 1 of the proposed DLM finds basic patterns including textures, edges, and corners along with other low-level features associated with the input image. Chunk 2 extends the initial features observed in Chunk 1 by inserting a new convolutional layer that learns more complicated patterns and edge combinations. This chunk also incorporates max pooling, which minimizes the spatial dimensions of the feature maps, allowing us to focus on the most essential characteristics while lowering computational strain. Dropout prevents overfitting by randomly discarding units during training, while batch normalization ensures consistent and efficient learning. The next chunk, 3, requires learning of intermediate features that are more critical than those offered by the preceding chunk(s). After this, chunk 4, initiates the learning of high-level features and capture of abstract representations. This convolutional layer further modifies the patterns while further reducing spatial dimensions, again rely on the most complex feature patterns. Dropout is carried out again to further boost the generalization. Moreover, batch normalization is applied again to handle the stability of training. This chunk further enhances the depth and level of abstraction for feature extraction. The last chunk, 5, includes a transition from convolutional feature extraction to fully connected layers that make the final output. The flattened layer converts 2D feature maps into 1D vectors, making them suitable for dense layers. Multiple thick layers, interleaved with activation functions, batch normalization, and dropout, refine and combine high-level features to form complex decision boundaries. This chunk ends with the output layer, which returns the final categorization results. Batch normalization and dropout remain crucial for ensuring consistent learning and minimizing overfitting in fully linked networks. The training settings include learning rate 0.001, bath size 32, optimizer Adam, and loss function. The proposed CDLM is trained using binary_crossentropy for binary classification and categorical_crossentropy for ternary classification.

C. HYBRID MODULE (HM)

This section discusses the hybrid module which consists of two parts: MLEM (PCA-controlled selected features set) and proposed deep CNN architecture. In the machine learning part, HOG and PCA approaches are used as feature extractors/descriptors. However, in the deep CNN part, the proposed 30 layers of CNN architecture are used for ternary and binary classification in a hybrid module. The details related to DLM are given in subsection III named proposed CNN architecture. Figure 6 shows the workflow of the hybrid module for D3. From the video dataset, frames are extracted with five frames per second (FPS). PCA-controlled HOG features (see subsection of III named MLEM-M) and 30-layer CNN architecture are jointly investigated for binary (normal and yawning) and ternary (normal, yawning, and talking) classification.

IV. EXPERIMENTAL SETUP AND RESULTS

Detailed experiments are conducted on a dataset to validate the performance of proposed machine learning, deep learning, and hybrid approaches. Below is the discussion about the dataset and performance evaluation measures.

A. DATASET DESCRIPTION

The YawDDvideo dataset was used for training and testing in this study. The YawDD dataset contains two video sets of the driver with multiple facial features [53]. These features are used to test yawning detection and face detection algorithms and systems. These videos were shot in a variety of lighting conditions in natural settings. The camera is mounted on the car's front mirror in the first set of videos. Every person in this video has three to four different mouth actions like usual, talking, yawning, etc. Statistics related to this dataset and sample per class are discussed in Table 3. This dataset contains 322 videos consisting of males and females using glasses and not using glasses from different angles. In the second set of videos, the camera is mounted on the car's dashboard. Everyone is videoed once with the same actions of normal/not yawning, talking, and yawning. There is a total of 29 male and female videos with or not using glasses from different backgrounds. The video format is 640 × 480.24-bit RGB having 30 frames per second, and its type is AVI without audio. This dataset has a total size of 5.1 GB with the help of the ACM multimedia system conference dataset archive, which can be accessed. Figure 7 depicts an example image from the applied dataset.

B. PERFORMANCE EVALUATION MEASURES AND IMPLEMENTATION SETTINGS

In this work, four commonly used evaluation metrics such as accuracy, precision, recall, and F1-measure or F1-score are used to evaluate the performance of the proposed method. A detailed description related to performance measures and their equations is given below. For the assessment of the proposed method, widely used training/testing combinations are taken such as 70% for training and 30% for testing. All the approaches such as MLEM-M, DLM, and HM are implemented using Python 3.7 and performed on a Desktop system core i5-4200 M with 16 GB RAM and GeForce GTX 1080 CPU.

Accuracy: Accuracy is a common evaluation standard for machine learning and deep learning algorithms. As shown in the following equation, accuracy is defined as the number of correctly identified outputs.

$$\text{Accuracy} = \frac{Correctlypredictedclass}{Totaltestingclass} \tag{12}$$

The following can also be calculated using the following equation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative rates.

Precision:Precision is a criterion for correctly separating real negative participants. To compute the precision, the

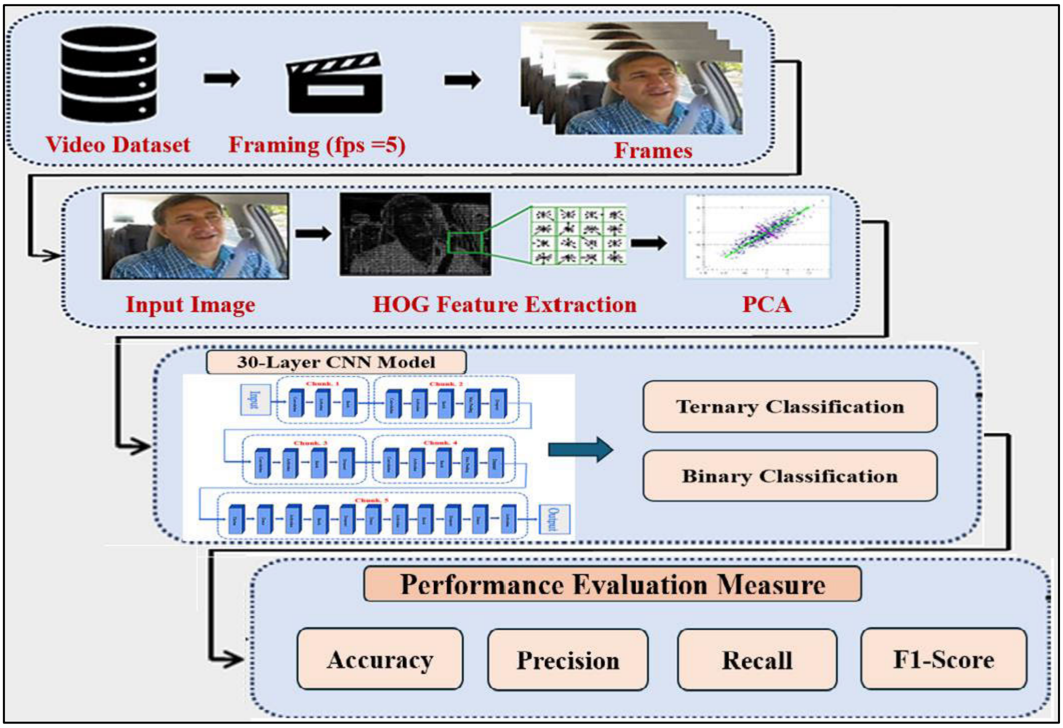


FIGURE 6. Flow diagram of proposed hybrid module.

FIGURE 7. Sample images selected from YawDD video dataset.

following equation is used.

$$\text{Precision} = \frac{TN}{TN+FP} \tag{14}$$

Recall: The recall is a touchstone of the true positive participants classified correctly and calculated using the equation.

$$\text{Recall} = TP/(TP + FN) \tag{15}$$

F1-measure: Equations 14 and 15 are used to calculate the harmonic mean of precision and recall, respectively. The following equation is used to calculate the F1 measure.

$$\text{F1\_measure} = \frac{2{*}Precision{*}Recall}{Precision + Recall} \tag{16}$$

C. RESULTS AND ANALYSIS

This section presents the experimental results obtained after a detailed experiment and empirical analysis of the proposed methodology for D3. The evaluation of each approach is discussed separately in subsequent sections.

1) PERFORMANCE OF MLEM MODULE FOR D3

In this section, different machine learning and ensemble methods are assessed for drowsiness detection. The selected features subset is supplied to these methods and outcomes are noted on YawDD video dataset, as shown in Table 3. The ensemble-based model XG Boost with HOG and PCA has attained 97.1% accuracy and outperformed other machine learning and ensemble-based approaches on YawDD video dataset. However, SVM, LRG, and DT attained 85.1%, 85.0%, and 94.3% accuracy, respectively. Among the ensemble-based approaches, RF acquires 94.8% accuracy and outperforms in terms of precision, recall, and F1-measure as well. The XG Boost model also achieves higher results as compared to applied machine learning

TABLE 3. Description of YAWDD video dataset.

Description	Statistics
Number of videos in 1st Video Set	322 (Camera installed on the front mirror)
Number of videos in the 2nd Video Set	29 (Camera installed on the dashboard)
Total number of classes	3 (Normal, talking, yawning)
Average duration of videos	30 seconds
Maximum duration of videos	120 seconds
Video Extension type	AVI
Number of frames per second (fps) general	30 fps
Size of video datasets	5.1 GB
Frame per second (Used)	5
No samples (Extracted and used)	Normal 21400, yawning 14272, talking 19925

and ensemble-based approaches. In this regard, the best results in terms of accuracy, precision, recall, and F1-measure for ternary and binary classification using ML module (HOG+PCA+classifier) are shown in bold in Table 3. According to Table 3, it is observed that HOG+PCA+XG_B produced better results in terms of overall accuracy as well as best class-wise true positives as 6310 with 98.4% accuracy, 5888 with 98.3% accuracy, and 4051 with 98.0% accuracy for normal, talking, and yawning classes, respectively. Similarly, the best results are attained for binary classification using the same settings (HOG+PCA+XG_B). For instance, better results in terms of overall accuracy as well as best class-wise true positives are obtained as 6360 with 98.4% accuracy, and 4180 with 98.4% accuracy for normal, and yawning classes, respectively. Detailed results in terms of accuracy, precision, recall, and F1-measure for both ternary and binary classification using ML module are presented in Table 3. These results confirm that PCA-controlled HOG features show improvement with different ML and ensemble methods, specifically better results with HOG+PCA+XG_B. Both HOG and PCA with applied classification methods support the better prediction of drowsiness from driver face images because the application of PCA-controlled features helps in reducing noise and computational efficiency. Moreover, HOG features capture essential patterns and robustness to variations such as pose and illuminations. The combination of PCA, HOG features and XGBoost takes advantage of the capabilities of each method: PCA reduces noise and dimensionality, HOG extracts significant patterns, and XGBoost robustly classifies the data. The combination of these components improves performance in D3, making it an effective method in machine learning pipelines.

2) PERFORMANCE OF DL MODULE FOR D3

This section discusses the evaluation results using proposed 30-layer CNN architecture to YawDD video dataset for binary and ternary classification. For comparison, Table 4 presents the results of ternary and binary classification where the proposed architecture is evaluated under a different number of epochs. This architecture achieves 99.7% accuracy for binary classification whereas 99.4% accuracy for ternary classification on YawDD dataset. With four different epochs, results in terms of accuracy, precision, recall, and F1-measure are presented in Table 4. According to Table 4, the proposed architecture has achieved 99.4% overall accuracy with 60 epochs for ternary classification. Furthermore, class-wise true positives are noted as 6395 with 99.6% accuracy, 5948 with 99.5% accuracy, and 4241 with 99.0% accuracy for normal, talking, and yawning classes, respectively. Similarly, the best results for binary classification are achieved using CDLM under 60 epochs. In this regard, better results are presented in terms of overall accuracy as well as best class-wise true positives as 6408 with 99.8% accuracy, and 4265 with 99.6% accuracy for normal, and yawning classes, respectively. Detailed results in terms of accuracy, precision, recall, and F1-measure for both ternary and binary classification using DL modules with different epochs are described in Table 4. Through results, it is also noted that poor precision is the reason for a low F1 score because poor precision and recall are indicators of a low F1 score. The machine learning model produces incorrect annotations and misses out on finding appropriate annotations.

3) PERFORMANCE OF HYBRID MODULE FOR D3

In a hybrid approach, a HOG feature extractor is applied with PCA and proposed 30 layers of CNN architecture for classification. Table 5 presents the results of the hybrid approach for ternary classification and binary classification. According to the results, the hybrid approach achieves 99.0% accuracy with only 10 epochs. However, with 60 epochs, the model attains 99.6% accuracy and outperforms the 30-layer CNN model without HOG feature extraction and PCA. Moreover, the hybrid model attains better results in terms of precision, recall, and F1-measure, as shown in Table 5.

Similarly, with four different epochs, results in terms of accuracy, precision, recall, and F1-measure for ternary classification using a hybrid module are presented in Table 5. The ternary and binary classification results related to all modules on test sample frames of YawDD videos are shown in Figure 8. In ternary classification using a hybrid approach,

TABLE 4. Results of ML and ensemble-based module for ternary and binary classification.

MLEM Module	Ternary Classification				Binary Classification			
	Accuracy	Precision	Recall	F1-measure	Accuracy	Precision	Recall	F1-measure
HOG+PCA+SVM	85.1	86.7	86.3	86.5	88.6	88.2	88.3	88.2
HOG+PCA+LGR	85.0	84.6	88.9	86.7	86.8	86.7	86.2	86.4
HOG+PCA+DT	94.3	92.2	87.7	89.9	96.1	95.4	95.8	95.5
HOG+PCA+RF	94.8	92.8	89.8	91.3	93.7	93.1	92.9	92.9
HOG+PCA+ XG_B	97.4	97.6	97.3	97.4	98.4	98.5	98.5	98.5

TABLE 5. Results of proposed 30-layer CNN architecture for ternary and binary classification.

Epochs	Ternary Classification				Binary Classification			
	Accuracy	Precision	Recall	F1-measure	Accuracy	Precision	Recall	F1-measure
10	89.4	61.9	96.4	75.4	94.2	98.1	80.4	88.4
20	99.2	62.0	96.0	75.3	93.1	98.1	77.1	86.3
40	99.1	62.4	97.1	76.0	99.6	98.1	98.1	98.1
60	99.4	61.7	98.0	75.5	99.7	99.8	99.9	99.2

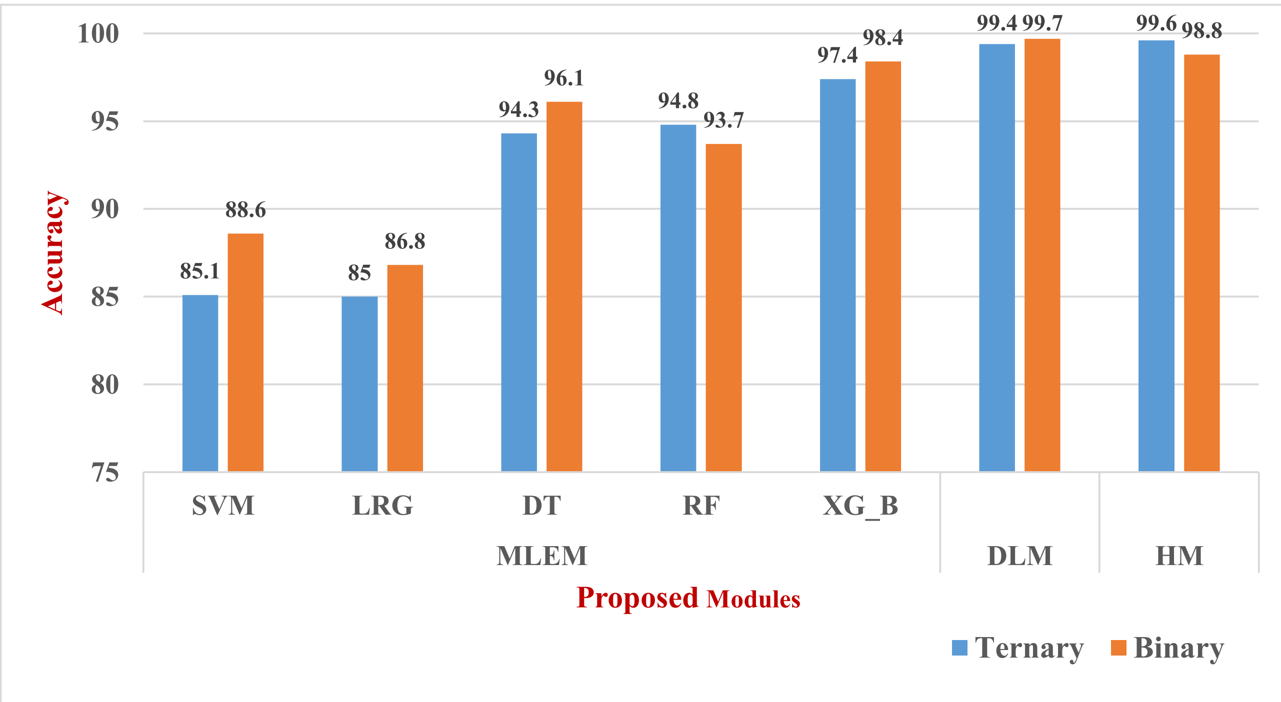


FIGURE 8. The comparison between the proposed MLEM module, DLM, and HM in terms of accuracy, best results for both ternary and binary classification.

class-wise true positives are observed as 6405 with 99.7% accuracy, 5955 with 99.6% accuracy, and 4251 with 99.2% accuracy for normal, talking, and yawning classes, respectively. Furthermore, experiments are conducted for binary classification where the proposed hybrid module achieved superior results in less than 60 epochs. Class-wise true positives are attained as 6408 with 98.8% accuracy, and 4265 with 97.7% accuracy for normal, and yawning classes, respectively. Detailed results in terms of accuracy, precision, recall, and F1-measure for both ternary and binary classification using DL modules with different epochs are given in Table 5. Figure 8 presents the comparison chart between machine learning and ensemble-based, deep learning-based, and hybrid approaches for binary and ternary classification. According to Figure 8, XG Boost obtained better accuracy because it is highly effective in handling complex and high-dimensional data than other machine learning and ensemble methods. XG Boost can also capture intricate

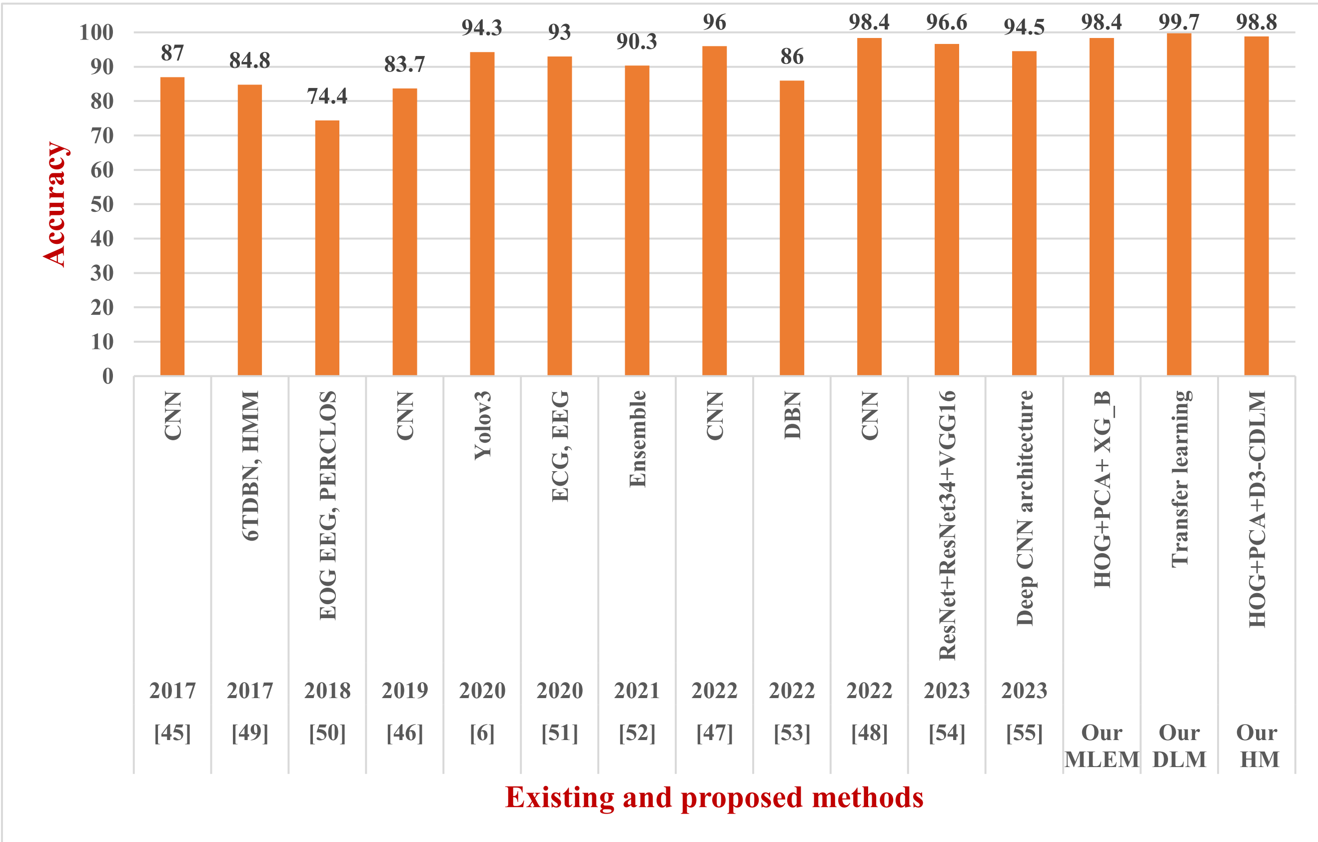


FIGURE 9. The comparison between the proposed MLEM module, DLM, and HM in terms of accuracy with existing methods.

patterns in the data as well. However, this combination produced lesser results as compared to the other proposed two modules DLM and HM.

D. COMPARISON WITH STATE-OF-THE-ART AND DISCUSSIONS

This section covers the discussions on results and comparisons with existing methodologies in which YawDD dataset is used for D3. To evaluate the performance of all proposed schemes (machine-learning based, deep-learning based, and hybrid), a comparison is conducted on relevant and recent methods such as CNN [28], [54], [55], [56], 6TDBN-HMM [29], EOG-EEG-PERCLOS [57], Yolov3 [6], ECG-EEG [58], Ensemble [59], DBN [60], ResNet+ResNet34+VGG16 [61], and Deep CNN architecture [30] presented by researchers for D3 on YawDD dataset. Despite the excellent performance of the proposed D3-CDLM, attained by 30-layer CNN architecture, it is not without limitations as we have well-organized and massive data for training purposes. Furthermore, the number of epochs also triggered performance measures. According to Tables 4 and 5, the hybrid approach, which combines the proposed 30 layers of CNN architecture with HOG and PCA, obtained 98.8% accuracy. However, without HOG and PCA, deep learning architecture attains 99.7% accuracy, whereas 98.4% accuracy is achieved by using machine learning and ensemble methods module with the combination of HOG+PCA+XG_B. Overall, the deep learning and hybrid modules obtain better accuracy, precision, recall, and F1-measure values than the machine learning module. A comparison with previous work is shown in Figure 9. Overall, the proposed HOG+D3-CDLM exceeded all state-of-the-art methods, for example, CNN, 6TDBN-HMM, EOG-EEG-PERCLOS, CNN, Yolov3, ECG-EEG, Ensemble, CNN, DBN, CNN, ResNet+ResNet34+VGG16, and Deep CNN architecture in terms of accuracy with an improvement of 11.8%, 14.0%, 24.4%, 15.1%, 4.5%, 5.8%, 8.5%, 2.5%, 12.8%, 0.6%, 2.8%, and 4.3%, respectively as depicted in Figure 9. Similarly, the proposed CDLM surpassed all state-of-the-art methods, for example, CNN, 6TDBN-HMM, EOG-EEG-PERCLOS, CNN, Yolov3, ECG-EEG, Ensemble, CNN, DBN, CNN, ResNet+ResNet34+VGG16, and Deep CNN architecture in terms of accuracy with an improvement of 12.7%, 14.9%, 25.3%, 16.0%, 5.4%, 6.7%, 9.4%, 3.4%, 13.7%, 1.5%, 3.7%, and 5.2%, respectively. The superior results of the proposed method are obtained because of PCA-controlled low-level features that are rotational-invariant as well as less sensitive to illumination changes. In addition, the proposed fine-tuned 30-layer architecture (CDLM) is more capable of learning both low and high-level features which

TABLE 6. Results of proposed 30-Layers CNN architecture on HOG and PCA processed images for ternary and binary classification.

Epochs	Ternary Classification				Binary Classification			
	Accuracy	Precision	Recall	F1-measure	Accuracy	Precision	Recall	F1-measure
10	99.0	60.5	96.2	74.2	97.2	70.4	84.1	76.6
20	98.9	60.1	96.2	74.0	96.7	71.2	85.5	77.6
40	99.2	59.8	96.3	73.7	97.4	71.0	84.7	77.2
60	99.6	59.7	96.8	73.8	98.2	72.6	86.1	78.7

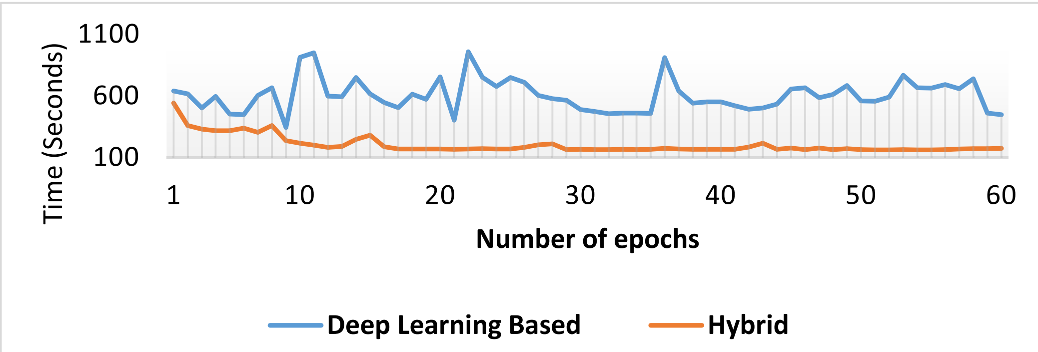


FIGURE 10. Comparison of proposed DLM and HM for ternary classification w.r.t training time.

are more robust to variations in input RGB video frames such as variations in lighting, scale, and rotation. The proposed architecture also has a strong tendency to generalize well to new and previously unknown data.

Moreover, Figure 10 shows the comparison of deep learning and hybrid approaches for ternary classification w.r.t training time. This graph clearly shows that the hybrid approach is taking less time as compared to deep-learning-based drowsiness detection. According to the findings, deep learning and hybrid approaches spent an average of 603 and 206 seconds on a single epoch. Both approaches significantly differ in training duration, as shown in the graph. The training time for a hybrid module, which involves 30 layers of CNN architecture with HOG and PCA, is less than the deep learning module without HOG and PCA. Moreover, according to the results in terms of accuracy, the hybrid approach achieves better results as compared to the deep learning approach. The overall results show that the proposed layers CNN architecture outperforms state-of-the-art machine learning and deep learning approaches. With highly deep networks, an issue known as the degradation problem might arise, in which adding additional layers results in increased training error. This occurs because the network gets more difficult to train properly, and additional layers do not always learn relevant features. More layers equal more parameters, which increases computational complexity and memory requirements. This results in more extended training times and higher computational costs, requiring more powerful equipment (GPUs). Therefore, we limit our network to 30 layers. The goal of this balance is to preserve training efficiency and stability while utilizing the depth for feature learning.

Conceptually, the proposed D3-CDLM can be more conceptual when dealing with the practical implementation of D3 in mobile phones which might make a better opportunity and cost-efficient solution due to their worldwide commonality and easy accessibility. The idea is straightforward but powerful: if a cell phone has an approach of CDLM-D3, it becomes an autonomous security assistant for all drivers in the vehicle. This approach minimizes the need for extra hardware costs by leveraging existing processing and sensors that are built into mobile phones. If a driver slaps it on a cell phone (and sets it up atop the dashboard), D3-CDLM may be more effective in detecting drowsy driving states. Additionally, this method can offer various bells and whistles such as true MRC (e.g., alarm) accessibility; scaled capabilities (read: cheaper to produce); nationwide availability; user-friendly. The proposed approach can also be embedded in an electronics-based system/device. The system has three components: a camera, an embedded system, and an alarm. The camera will help to take continuous monitoring of the driver; the embedded model can detect drowsiness effectively and efficiently as presented in this paper. Finally, in case of drowsiness detection, alarm generation will help to avoid accidents successfully.

V. CONCLUSION AND FUTURE DIRECTIONS

This study presents deep learning, machine learning, and a hybrid module for D3. Procedures are commonly used to detect driver drowsiness: behavioral, physiological, and vehicular-based measures. Focus is on the driver's drowsiness state rather than the vehicle in behavioral measures, and the camera is used to monitor the driver's physical conditions such as eye, face, head, and yawning. Driver's eye detection, eye blink rate, yawning analysis, and facial expression analysis provide behavioral measures. This system is trained and tested by using a publicly available dataset YawDD, which is explicitly used for yawning detection using facial videos of the driver. The proposed architecture also attained 99.4% accuracy by implementing the proposed technique to these images for training and validation purposes. The hybrid approach, which combines the proposed 30 layers of CNN architecture with HOG and PCA, obtained 99.6% accuracy. The accuracy achieved by using the proposed CNN model is 99.4%, and by using machine learning, the accuracy achieved is less than the other two approaches. In the future, the validity, application, and implementation of the proposed hybrid module can be expanded to include more diverse scenarios in abnormal behaviors such as intoxication while driving. It would need to establish a relationship between yawning behavior (as investigated in this study) and the individual's level of deficiency due to alcohol or other drugs. Meanwhile, some characteristics such as driver's attentiveness, reduced concentration, abrupt acceleration, aggressive driving style, and unsafe lanes in detecting drunkenness or yawning behavior can also be investigated while driving. Moreover, physiological and vehicle-based measures can be covered individually and with behavioral measures for D3.

ACKNOWLEDGMENT

We thank Princess Nourah bint Abdulrahman university Researchers Supporting Project number (PNURSP2024R513), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Authors are also thankful to Prince Sultan University, Riyadh Saudi Arabia for support.

REFERENCES

[1] M. Ahmed, M. Ramzan, H. U. Khan, S. Iqbal, M. A. Khan, J.-I. Choi, Y. Nam, and S. Kadry, "Real-time violent action recognition using key frames extraction and deep learning," *Comput., Mater. Continua*, vol. 69, no. 2, pp. 2217–2230, 2021.

[2] M. Ramzan, A. Abid, H. U. Khan, S. M. Awan, A. Ismail, M. Ahmed, M. Ilyas, and A. Mahmood, "A review on state-of-the-art violence detection techniques," *IEEE Access*, vol. 7, pp. 107560–107575, 2019.

[3] M. T. Khan, H. Anwar, F. Ullah, A. Ur Rehman, R. Ullah, A. Iqbal, B.-H. Lee, and K. S. Kwak, "Smart real-time video surveillance platform for drowsiness detection based on eyelid closure," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–9, Mar. 2019.

[4] A. Moujahid, F. Dornaika, I. Arganda-Carreras, and J. Reta, "Efficient and compact face descriptor for driver drowsiness detection," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114334.

[5] S. Chaabene, B. Bouaziz, A. Boudaya, A. Hökelmann, A. Ammar, and L. Chaari, "Convolutional neural network for drowsiness detection using EEG signals," *Sensors*, vol. 21, no. 5, p. 1734, Mar. 2021.

[6] F. You, Y. Gong, H. Tu, J. Liang, and H. Wang, "A fatigue driving detection algorithm based on facial motion information entropy," *J. Adv. Transp.*, vol. 2020, pp. 1–17, Jun. 2020.

[7] 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.

[8] T. Kundinger, A. Riener, and R. Bhat, "Performance and acceptance evaluation of a driver drowsiness detection system based on smart wearables," in *Proc. 13th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Stanford, CA, USA, 2021, pp. 49–58.

[9] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.

[10] H. Meda, J. M. P. Ganesh, and A. Sahani, "Machine learning models for drowsiness detection," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Glasgow, U.K., May 2021, pp. 1–5.

[11] S. H. Park, H. S. Yoon, and K. R. Park, "Faster R-CNN and geometric transformation-based detection of driver's eyes using multiple near-infrared camera sensors," *Sensors*, vol. 19, no. 1, p. 197, Jan. 2019.

[12] M. Ngxande, J.-R. Tapamo, and M. Burke, "Detecting inter-sectional accuracy differences in driver drowsiness detection algorithms," in *Proc. Int. SAUPEC/RobMech/PRASA Conf.*, Cape Town, South Africa, Jan. 2020, pp. 1–6.

[13] V. Vijayan and E. Sherly, "Real time detection system of driver drowsiness based on representation learning using deep neural networks," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 1977–1985, Mar. 2019.

[14] W. Liu, J. Qian, Z. Yao, X. Jiao, and J. Pan, "Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection," *Future Internet*, vol. 11, no. 5, p. 115, May 2019.

[15] B. Reddy et al., "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 121–128.

[16] M. Ngxande, J.-R. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," in *Proc. Pattern Recognit. Assoc. South Afr. Robot. Mechatronics (PRASA-RobMech)*, Nov. 2017, pp. 156–161.

[17] X. Zhu, W.-L. Zheng, B.-L. Lu, X. Chen, S. Chen, and C. Wang, "EOG-based drowsiness detection using convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Beijing, China, Jul. 2014, pp. 128–134.

[18] G. Li and W.-Y. Chung, "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier," *Sensors*, vol. 13, no. 12, pp. 16494–16511, Dec. 2013.

[19] O. K. Firke and D. M. Jain, "Analysis of driver drowsiness detection using EEG and EOG," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 46–51, Apr. 2018.

[20] B. Warwick, N. Symons, X. Chen, and K. Xiong, "Detecting driver drowsiness using wireless wearables," in *Proc. IEEE 12th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Dallas, TX, USA, Oct. 2015, pp. 585–588.

[21] K. A. Brookhuis and D. de Waard, "Monitoring drivers' mental workload in driving simulators using physiological measures," *Accident Anal. Prevention*, vol. 42, no. 3, pp. 898–903, May 2010.

[22] L.-L. Chen, Y. Zhao, J. Zhang, and J.-Z. Zou, "Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7344–7355, Nov. 2015.

[23] M. Hajinoroozi, Z. Mao, and Y. Huang, "Prediction of driver's drowsy and alert states from EEG signals with deep learning," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2015, pp. 493–496.

[24] M. K. Hussein, T. M. Salman, A. H. Miry, and M. A. Subhi, "Driver drowsiness detection techniques: A survey," in *Proc. 1st Babylon Int. Conf. Inf. Technol. Sci. (BICITS)*, Babil, Iraq, Apr. 2021, pp. 45–51.

[25] S. Arefnezhad, S. Samiee, A. Eichberger, M. Frühwirth, C. Kaufmann, and E. Klotz, "Applying deep neural networks for multi-level classification of driver drowsiness using vehicle-based measures," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113778.

[26] S. Yang, J. Xi, and W. Wang, "Driver drowsiness detection through a vehicle's active probe action," in *Proc. IEEE 2nd Connected Automated Vehicles Symp. (CAVS)*, Sep. 2019, pp. 1–7.

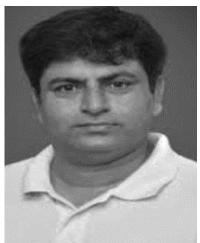
[27] K. Vasudevan, A. P. Das, B. Sandhya, and P. Subith, "Driver drowsiness monitoring by learning vehicle telemetry data," in *Proc. 10th Int. Conf. Human Syst. Interact. (HSI)*, Jul. 2017, pp. 270–276.

[28] H. Li, Y. Wang, and Y. Nan, "Motion fatigue state detection based on neural networks," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, Mar. 2022.

[29] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Proc. Amer. Conf. Comput. Vis. (ACCV)*. Taipei, Taiwan: Springer, Nov. 2016, pp. 117–133.

[30] E. Civik and U. Yuzgec, "Real-time driver fatigue detection system with deep learning on a low-cost embedded system," *Microprocessors Microsystems*, vol. 99, Jun. 2023, Art. no. 104851.

[31] M. S. Razeeth, R. Kariapper, and S. S. Nawaz, "Accident mitigation system with drowsiness detection: A machine learning and IoT with hybrid approach," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Jordan, Jul. 15, 2021, pp. 462–465.
[32] M. Fayyaz, M. Yasmin, M. Sharif, J. H. Shah, M. Raza, and T. Iqbal, "Person re-identification with features-based clustering and deep features," *Neural Comput. Appl.*, vol. 32, no. 14, pp. 10519–10540, Jul. 2020.
[33] M. Ahad and M. Fayyaz, "Pedestrian gender recognition with handcrafted feature ensembles," *Azerbaijan J. High Perform. Comput.*, vol. 4, no. 1, pp. 60–90, Jun. 2021.
[34] H. Zhou and G. Yu, "Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features," *Future Gener. Comput. Syst.*, vol. 125, pp. 604–615, Dec. 2021.
[35] D. Lakshmi and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders," *Microprocessors Microsystems*, vol. 82, Apr. 2021, Art. no. 103834.
[36] X. Wang, C. Jin, W. Liu, M. Hu, L. Xu, and F. Ren, "Feature fusion of HOG and WLD for facial expression recognition," in *Proc. IEEE/SICE Int. Symp. Syst. Integr.*, Dec. 2013, pp. 227–232.
[37] J. Fan, J. Lee, and Y. Lee, "A transfer learning architecture based on a support vector machine for histopathology image classification," *Appl. Sci.*, vol. 11, no. 14, p. 6380, Jul. 2021.
[38] P. Rasna and M. Smithamol, "SVM-based drivers drowsiness detection using machine learning and image processing techniques," in *Proc. Progress Adv. Comput. Intell. Eng. (ICACIE)*, vol. 2 Springer, 2021, pp. 100–112.
[39] M. Fayyaz, M. Yasmin, M. Sharif, T. Iqbal, M. Raza, and M. I. Babar, "Pedestrian gender classification on imbalanced and small sample datasets using deep and traditional features," *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11937–11968, Jun. 2023.
[40] A. Suresh, R. Udendhran, and M. Balamurgan, "Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers," *Soft Comput.*, vol. 24, no. 11, pp. 7947–7953, Jun. 2020.
[41] N. Nojiri, X. Kong, L. Meng, and H. Shimakawa, "Discussion on machine learning and deep learning based makeup considered eye status recognition for driver drowsiness," *Proc. Comput. Sci.*, vol. 147, pp. 264–270, Jan. 2019.
[42] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[43] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[44] S. Gündoğdu, "Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique," *Multimedia Tools Appl.*, vol. 82, no. 22, pp. 34163–34181, Sep. 2023.
[45] N. Q. K. Le, D. T. Do, T.-T.-D. Nguyen, and Q. A. Le, "A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features," *Gene*, vol. 787, Jun. 2021, Art. no. 145643.
[46] S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, "ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost," *Nucl. Eng. Technol.*, vol. 53, no. 2, pp. 522–531, Feb. 2021.
[47] M. Murtaza, M. Sharif, M. Yasmin, M. Fayyaz, S. Kadry, and M. Y. Lee, "Clothes retrieval using M-AlexNet with mish function and feature selection using joint Shannon's entropy Pearson's correlation coefficient," *IEEE Access*, vol. 10, pp. 115469–115490, 2022.
[48] F. Abbas, M. Yasmin, M. Fayyaz, M. A. Elaziz, S. Lu, and A. A. A. El-Latif, "Gender classification using proposed CNN-based model and ant colony optimization," *Mathematics*, vol. 9, no. 19, p. 2499, Oct. 2021.
[49] M. Fayyaz, M. Yasmin, M. Sharif, and M. Raza, "J-LDFR: Joint low-level and deep neural network feature representations for pedestrian gender classification," *Neural Comput. Appl.*, vol. 33, no. 1, pp. 361–391, Jan. 2021.
[50] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. Fayyaz, "Exploring deep learning approaches for Urdu text classification in product manufacturing," *Enterprise Inf. Syst.*, vol. 16, no. 2, pp. 223–248, Feb. 2022.
[51] J. Chen, S. Wang, E. He, H. Wang, and L. Wang, "Recognizing drowsiness in young men during real driving based on electroencephalography using an end-to-end deep learning approach," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102792.
[52] M. Hashemi, A. Mirrashid, and A. B. Shirazi, "Driver safety development real time driver drowsiness detection system based on convolutional neural network," 2020, *arXiv:2001.05137*.
[53] S. Abtahi et al., "YawDD: A yawning detection dataset," in *Proc. 5th ACM Multimedia Syst. Conf.*, Singapore, 2014, pp. 24–28.
[54] W. Zhang and J. Su, "Driver yawning detection based on long short term memory networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–5.
[55] N. A. Bajwa et al., "Detecting driver drowsiness in real time through deep learning based object detection," in *Proc. 15th Int. Work-Conf. Artif. Neural Netw. (IWANN)*. Gran Canaria, Spain: Springer, Jun. 2019, pp. 283–296.
[56] A. Rajkar, N. Kulkarni, and A. Raut, "Driver drowsiness detection using deep learning," in *Proc. Applied Inf. Process. Syst. (ICCET)*. Springer, 2021, pp. 73–82.
[57] S. Junaedi and H. Akbar, "Driver drowsiness detection based on face feature and PERCLOS," *J. Phys., Conf. Ser.*, vol. 1090, Sep. 2018, Art. no. 012037.
[58] A. A. Miah, M. Ahmad, and K. Z. Mim, "Drowsiness detection using eye-blink pattern and mean eye landmarks' distance," in *Proc. Int. Joint Conf. Comput. Intell. (IJCCI)*. Springer, 2020, pp. 111–121.
[59] R. M. Salman, M. Rashid, R. Roy, M. M. Ahsan, and Z. Siddique, "Driver drowsiness detection using ensemble convolutional neural networks on YawDD," 2021, *arXiv:2112.10298*.
[60] B. K. Savaş and Y. Becerikli, "Behavior-based driver fatigue detection system with deep belief network," *Neural Comput. Appl.*, vol. 34, no. 16, pp. 14053–14065, Aug. 2022.
[61] S. Afroze, M. R. Hossain, M. M. Hoque, and M. A. A. Dewan, "An empirical framework for detecting speaking modes using ensemble classifier," *Multimedia Tools Appl.*, vol. 83, no. 1, pp. 2349–2382, Jan. 2024.



MUHAMMAD RAMZAN received the Ph.D. degree in computer science from the University of Management and Technology, Lahore, Pakistan. Currently, he is a Lecturer with the CS & IT Department, University of Sargodha. Previously, he was a Lecturer with the Virtual University of Pakistan. He was with the Higher Education Department as a Database/Network Administrator. Before this, he was a Lecturer with Minhaj University Lahore (Sharia College). He has authored more than 40 research articles published in reputed peer-reviewed journals. His research areas of interests include medical imaging, deep learning, machine learning, video surveillance, activity recognition, and computer vision.

ADNAN ABID (Senior Member, IEEE) received the Ph.D. degree from the Politecnico di Milano, Italy, in 2012. He is a Professor of computer science with the Faculty of Computer Science and Information Technology, Virtual University of Pakistan, Pakistan. His research interests include adaptive and self-organizing, systems computational, linguistics computer education, data mining and machine learning, data science, databases, and software engineering. He is a member of ACM. He has been associated with editorial boards of well-reputed journals in computer science.

MUHAMMAD FAYYAZ received the Ph.D. degree in computer science from COMSAT University Islamabad (CUI), Wah Campus, Pakistan, in September 2021. He is currently an Assistant Professor and the Head with the Department of Computer Science, FAST–National University of Computer and Emerging Sciences (NUCES), Chiniot-Faisalabad Campus, Pakistan. Before this, he was a Lecturer with the Department of Computer Science, CUI, from September 2014 to January 2022. He remained a very active research member in fields related to pedestrian analysis, computer vision, pattern recognition, and machine/deep learning. He served as a TPC member, the organizing co-chair, and a reviewer for many leading international conferences and journals.

TAHANI JASER ALAHMADI received the B.S. degree in computer science and the M.S. degree in information technology (data management), and the Ph.D. degree from the Faculty of Information Technology, Griffith University, Australia, in 2019. She is currently an Assistant Professor with the IS Department, Faculty of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Saudi Arabia. Her research interests include innovative research methods in data analysis and mining, machine learning, pattern recognition, image processing, accessibility, usability, and sentiment analysis. She is a member of the Golden Key Society and Media Access Australia. She received multiple awards, such as the Google Doctoral Consortium Award and the Institute for Integrated and Intelligent Systems (IIIS) Award for Quality and Impact Research.

HAITHAM NOBANEE (Fellow, IEEE) received the Ph.D. degree from the University of Manchester. He is a Professor of finance and financial technology (Fintech) with Abu Dhabi University, United Arab Emirates. He is also an Honorary Professor with the University of Liverpool, U.K.; and a Visiting Research Professor with the University of Oxford, U.K. His work has been published in the *International Review of Economics and Finance*, *North American Journal of Economics and Finance*, *Renewable and Sustainable Energy Reviews*, *Big Data*, and *Corporate Governance*. He has also been published in professional journals, such as *Harvard Business Review*. He is a fellow of the Higher Education Academy (FHEA) and the Learning and Performance Institute (LPI). He is serving on the academic community as an Editor for *Arab Gulf Journal of Scientific Research* (Emerald), an Associate Editor for *International Review of Financial Analysis* (Elsevier and Heliyon-Elsevier), a Guest Editor for *International Review of Economic and Finance* (Elsevier), a member of the editorial board for *Financial Innovation* (Springer), and a reviewer for various international refereed journals.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degree, specializing in information security using image processing techniques from the Faculty of Computing, University Teknologi Malaysia (UTM), Malaysia, in 2010. He is currently an Associate Professor with CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several projects and completed projects funded by MoHE Malaysia and Saudi Arabia. His research interests include bioinformatics, the IoT, information security, and pattern recognition. He received the Rector Award for the 2010 Best Student from UTM Malaysia.