IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 22, 2021, accepted July 5, 2021, date of publication July 9, 2021, date of current version July 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3096032

ADMT: Advanced Driver's Movement Tracking System Using Spatio-Temporal Interest Points and Maneuver Anticipation Using Deep Neural Networks

SHILPA GITE[1,2], BISWAJEET PRADHAN[3,4], (Senior Member, IEEE), ABDULLAH ALAMRI[5], AND KETAN KOTECHA[1,2]

[1]Department of Computer Science and Information Technology, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India
[2]Symbiosis Centre of Applied A.I. (SCAAI), Symbiosis International (Deemed University), Pune 412115, India
[3]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and I.T., University of Technology Sydney, Ultimo, NSW 2007, Australia
[4]Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia
[5]Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

Corresponding author: Ketan Kotecha (head@scaai.siu.edu.in)

This work was supported in part by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and I.T., University of Technology Sydney (UTS), in part by the Researchers Supporting Project under Grant RSP-2021/14, King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Assistive driving is a complex engineering problem and is influenced by several factors such as the sporadic nature of the quality of the environment, the response of the driver, and the standard of the roads on which the vehicle is being driven. The authors track the driver's anticipation based on his head movements using Spatio-Temporal Interest Point (STIP) extraction and enhance the anticipation of action accuracy well before using the RNN-LSTM framework. This research tackles a fundamental problem of lane change assistance by developing a novel model called Advanced Driver's Movement Tracking (ADMT). ADMT uses customized convolution-based deep learning networks by using Recurrent Convolutional Neural Network (RCNN). STIP with eye gaze extraction and RCNN performed in ADMT on brain4cars dataset for driver movement tracking. Its performance is compared with the traditional machine learning and deep learning models, namely Support Vector Machines (SVM), Hidden Markov Model (HMM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and provided an increment of almost 12% in the prediction accuracy and 44% in the anticipation time. Furthermore, ADMT systems outperformed all of the models in terms of both the accuracy of the system and the previously mentioned time of anticipation that is discussed at length in the paper. Thus it assists the driver with additional anticipation time to access the typical reaction time for better preparedness to respond to undesired future behavior. The driver is then assured of a safe and assisted driving experience with the proposed system.

INDEX TERMS RCNN, advanced driver movement tracking system, Spatio-temporal interest points, eye gaze tracking, deep neural networks.

I. INTRODUCTION

Recent years have seen a rise in the research and the efforts to develop self-driving or autonomous vehicles, starting from prototypes to targeting a full-scale development [1]. The pioneer companies in this domain include Tesla and Google [2].

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

While there is no unanimous acceptance of such systems, the focal point of the research continues to be on improving their efficiency and effectiveness. Over 80% of fatalities and accidents reported on roads can be attributed to human errors introduced due to the driver's fault [3]. He often makes mistakes while changing the lanes on the highways. If he gets lane change assistance while driving, the possible accidents can be reduced drastically [4].

This paper proposes a solution to real-time lane change tracking drivers using deep learning frameworks. The demanding task of driver assistance systems is to provide a timely and appropriate response to the drivers. Thus, these systems must be aware of both the context and the situation [3], [4]. Accurate prediction of the possible events and corresponding driver maneuvers is possible only when the system considers the time sequence of the context [5]. The precise prediction helps in the prompt delivery of action suggestions to the driver. An alert and fast context analysis system can be developed using modern deep learning techniques, which work well in real-time situations like driving [6]. Because of increased computational power and new technological advancements, image processing, and computer vision have been widely evolving and deployed in assistive driving [7]. However, owing to multiple action sequences, automated moderation of driver activity anticipation is challenging for these techniques [8]. The authors present approaches that build upon the previously proposed situation-aware mechanisms [9]–[11]. Thus, the problem objective can be framed as a system that can anticipate driver action and alert drivers with good accuracy of action anticipation.

One of the challenges while designing the anticipation models for the driver is sufficient anticipation time with high accuracy [12]. Anticipation is referred to as reaction time that is 2-3 sec for a human being in an actual situation for any mishap. The anticipation of a driver's future action should be more using computer vision techniques to help the driver make real-time decisions [13]. The authors tried to improve the anticipation time with correct maneuver prediction by the ADMT system. More anticipation time means adding a few extra seconds to the driver's reaction time for individual actions while driving, thereby minimizing the possibility of an accident [14]. This research work is inspired by the Brain4cars research group [15]. Their sensory fusion architecture using deep learning gives a 3.5sec anticipation time with 86% action accuracy. The authors further improved these performance parameters by extracting Spatio-temporal features from video sequences, driver's eye gaze tracking, and deep learning for future action prediction. This paper has attempted to build a driver activity prediction system by combining convolutional and recurrence-based deep neural architectures. The proposed method is capable enough to provide accurate responses well within the required anticipation time so that drivers get a more reliable and trustworthy assistance system. The feature extraction from the Spatio-temporal video data is performed using the Spatio-Temporal Interest Points (STIP) method [16], [17]. The benefit is demonstrated using results obtained on the standard brain4cars dataset [15]. The obtained results comfortably outperformed the existing popular machine learning and deep learning approaches. Though machine learning /deep learning models are used in human activity tracking on a broad scale, they lack accuracy or head movement tracking or eye-gaze tracking, or future activity prediction [18]–[21]. Most systems use the proactive approach of driver action anticipation by incorporating either CNN or RNN [15], [18], [21]. The driver's activities like head movement and eye gaze identification can be tracked from the inside captured video of the dataset. The ADMT system also tried to cut down the computational resources by using only the internal context of the vehicle, as opposed to previous approaches, which used both inside and outside contexts for processing.

The key contributions in this paper can be highlighted as follows:

i. First, automatic detection of driver actions from the video sequences using the Spatio-temporal interest points tracking helps interpret the driver action's nature.

ii. Implementation of eye gaze tracking for driver's intention prediction to improve the action anticipation time.

iii. Design of ADMT (Advanced Driver's Movement Tracking) system to track the driver's actions and retrieve the driver's movement by processing the frames. The proposed ADMT system makes use of inside context for driver's action prediction.

iv. Implementation of ADMT with Recurrent Convolutional Neural Networks (RCNN) as a deep learning technique to enhance the action classification accuracy.

v. Detailed performance analysis and discussion of the proposed ADMT using RCNN with other ML and DL methods.

II. PREVIOUS WORK

A. DRIVER'S MOVEMENT TRACKING

Driver's movement tracking in a video has always been an interesting research problem to provide effective ADAS solutions [22]–[24]. Driver actions have been controlled and predicted using semi-autonomous methods [1]. Koppula *et al.* stated that Spatio-temporal tracking could be used for the activity anticipation, though it may not produce sufficient lead time [12]. Alert generation systems sometimes may lead to delayed warning messages that could not prevent any mishap [13]. Jain *et al.* combined recurrent and long short-term memory networks (RNN-LSTM) to develop a sensory fusion deep learning approach for pre-emptive anticipation of the driver's actions [15]. The authors are motivated by the recent development of assistive driving by [15]. However, it is computationally expensive to collect and merge both types of features (inside and outside) and process them for every instance of the videos and other data provided from both sources. The information given by the inside camera would be sufficient to track the driver's facial actions, so the authors get rid of the dependency on external features and save up processing time. Thus, only an internal context that portrays the driver's movements is considered, and the system has been developed accordingly.

B. VIDEO SEQUENCE ANALYSIS (VSA)

Video sequence means images are getting displayed continuously at a fixed rate. Video Sequence Analysis (VSA) has become a capable and potential area of research in the

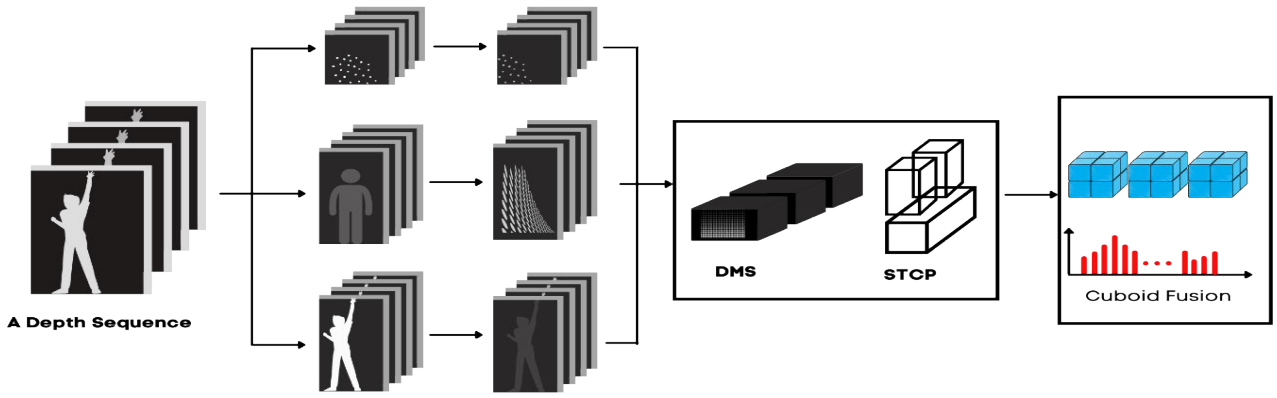

FIGURE 1. Example of Spatio-Temporal Cuboids Points extraction (STCP) [33].

computer vision and image processing domain because of its usefulness in numerous applications [25], [26]. It needs to know core digital image processing and related processes like image enhancement, image segmentation, morphological operations, feature extraction, image representation, image classification, and many others [27]. It works in both spatial-temporal manner and extracts information concerning spatial and temporal changes [16]. To perform the analysis of the video sequences, the Spatio-temporal approach is mainly used [17]. It primarily includes segmentation (spotting targeted objects or regions in the video scene) and feature extraction processes. The segmentation and feature extraction of the spatial and temporal dimensions is needed for the powerful representation of the object/region in the video sequence. However, there are challenges in the analysis of video sequences, such as quality of the videos, background noise, the recording settings, illumination variations, camera motion, viewpoint variations, foreground and background similarity of the object, high dimensionality, and also redundancy of the data, etc. [28]. Designing and developing a system that can deal with these issues is still an exciting computer vision problem, and STIP can effectively solve these issues. It can handle the detection of interest points from a video to process the Spatio-temporal domain information effectively [29], [30]. Hence, it has become adaptable for the research problem.

C. SPATIO-TEMPORAL ACTION RECOGNITION

The STIP method is used by Yanshan Li *et al.* [16] for the action detection problem but can also be used for driver's action recognition in our case. Spatio-temporal action recognition or action localization enables the action classification task performed in a sequence of frames (or video) and localizes each action both in space and time. Convolutional Neural Network (CNN) architectures have been primarily used in localization tasks and can be identified using bounding boxes or masks [20]. Figure 1 depicts the example of human action recognition using the STIP extraction approach. Then the fusion of all the cuboids of the spatiotemporal method occurs subsequently, and the final cuboid is produced as Depth Motion Sequence (DMS) by different layers. DMS generates Spatio-Temporal Cuboids Pyramids (STCP) that takes care of all the changes happening in the video process. Every depth map from a video sequence generates DMS by calculating the difference between two successive frames in the anticipated sequences of videos. It helps to keep the temporal details of depth maps and grip the problematic situation like self-occlusion. STCP subdivides DMS into multiple temporal segments and spatial sub-cuboids to capture the temporal order along with object and body shape, respectively. In the end, the cuboid feature fusion method effectively implemented the correlation between projected image planes. Therefore, a Spatio-temporal approach would effectively track the actions from a video and hence been used in the problem statement.

However, Spatio-temporal action recognition has problems like tracking the action in a video, object, or action localization [32]. Moreover, action localization becomes more challenging with the temporal dimension [33]. STIP extraction would be an effective solution for these types of problems, with the capability of moving object detection while reducing the need for background modeling with foreground segmentation. In addition, STIP extracts more features from the images and better video recognition accuracy to improve the video/action/image classification [34].

D. MACHINE LEARNING AND DEEP LEARNING FOR ACTION RECOGNITION

Probabilistic approaches were used in the recent past for action recognition problems, whereas recent developments are executed mainly by machine learning and deep learning techniques [20], [37], [38]. The researchers exploited popular data mining algorithms, like SVM, Bayesian, clustering, and decision tree for activity recognition [25]. Some clustering algorithms are shown reasonably good accuracy, and even SVM also offers good anticipation accuracy. However, to get

the desired anticipation time by using SVM or machine learning classifiers is a challenging task because it is not possible to process time-series data like a video. Therefore, ongoing research focuses on using artificial neural network (ANN) algorithms for such a non-linear real-time problem [39], [40]. Nevertheless, anticipation is of utmost importance in predicting driver's behavior while maintaining accuracy and precision in action anticipation; hence, adopting recent techniques generates the desired outcomes.

It has been observed that the majority of the machine learning and deep learning approaches in the past were implemented by combining the driver features with the external road data, thus leading to more trainable parameters from the video data [12]–[15]. Deep learning methods help optimizes the performance of anticipation/recognition systems significantly by intelligent algorithms such as CNN or RNN, or LSTM [37]. The most common fusion of deep learning methods for human activity recognition provides a pool of features automatically used in different application areas [41]. CNN is beneficial in image processing, capturing real-time videos, and other computer vision-related tasks [37], [41]. Convolutional Neural Network is a deep, interconnected layered structure to perform a convolutional operation on the input data using multiple hidden layers that facilitates robust feature extraction [41]. These hidden layers are merged to formulate deeply layered architectures for the feature extraction of correlated image data. RNNs work sequentially to predict the following action. LSTM has the memory gates and decides which part of the information to be excluded from the data pool [42]. An essential difference between these two algorithms is that while CNN is a core component of feed-forward propagation of visual data in the architecture, RNNs are more powerful to get the following sequence in the data because it is a sequential model [37]. A fusion of LSTM and RNN as a collective deep learning approach is used to classify the driver's actions, represented in the DMT algorithm [43], to model spatial-temporal dependencies in the continuous video data. It works well in a different contextual environment applicable for human activity recognition applications [40], [41], [44].

Some methods used the inside visual features and outside features as input to the sensory fusion deep learning model, but it takes more processing time. Hence to optimize this can be a research gap in this problem. Thus, the research gaps were identified, such as more accurate, more response time generating future action prediction systems for a driver. There is also substantial research potential in the Spatio-temporal domain by extracting and utilizing interest points for driver's movement tracking, head pose estimation, and eye gaze that would give important clues about his intent [45], [46]. When the combination of CNN with RNN takes place in ADMT, the representations obtained from the RNN are used to enhance CNN progressively [47]. STIP helps extract driver movements in the dynamic Spatio-temporal domain, whereas the eye tracker extracts eye gaze, and the system could work in an 'advanced' way for assistive driving.

III. METHODOLOGY

This paper presents two main contributions: STIP with eye gaze implementation and RCNN in ADMT for proposed driver anticipation architecture. As discussed earlier, the authors designed STIP-based techniques to improve the robustness in the continuous images and reduce feature extraction time. For visual feature extraction, rather than using the typical face detection, landmark points extraction, head-pose motion detection, the authors proposed the STIP detection-based techniques to develop robustness and decrease the time for feature extraction only the driver's inside the video [16], [32]. The proposed STIP for tracking the driver's movements is designed in which the filtering method is used to extract the STIPs from the input videos for noise removal. If the driver intends to take any lane change action, he first moves his head to that side to get the idea of other vehicles on the road. Here the authors assumed that the driver is moving his head in the desired direction only when he has to take any turn and ignoring the scenario where he is talking with his fellow members by moving his head towards the right as the driver's seating position on the left of the car in the dataset [15]. Therefore, the authors restrained processing the outside videos and only focused on inside videos with the driver's face. Then, the anticipation of the driver maneuver using a deep learning classification approach took place.

A. DATASET

The performance of proposed approaches is evaluated on the publicly available Brain4Cars video dataset [15] of around 700 video clippings ranging from 2-5sec with 25FPS (frames per second). It has a combination of videos from both the internal setup of the car and the external environment gathered under a standard design. For the ADMT system, only inside cabin videos are taken as input. This dataset was divided into the ratio of 70:30 as training and testing sets, respectively. Moreover, cross-validation was not used in this paper, so 70:30 has been adopted. The number of video samples was sufficiently large enough (700) to generate the desired model output that further justified the need for the split of 70:30. As such, there is no global standard rule for the selection of the sampling ratio for training-test data.

Moreover, [52] have used 70:30 split applied for multiple video datasets video containing faces. Additionally, in the earlier implementation papers of the authors, the same dataset split generated remarkable results [21], [43]. In the machine learning literature, it can be observed that different sample ratios have been used depending on the size of the dataset. The rationale behind this split was the nature, volume, and complexity of the training videos. Considering all the aspects related to the dataset, the optimal dataset split for the video dataset was finalized.

B. HYPERPARAMETERS AND SYSTEM DESIGN

In this section, the hyperparameters used in the proposed RCNN model are given below in Table 1.

FIGURE 2. Eye gaze tracking.

TABLE 1. Hyper parameters used in RCNN model.

Hyperparameter name	Hyperparameter Value
Hidden layers	5
Layer delay	1:2
Epoch	7
Performance delay	0.494
Gradient	0.655
Batch size	150
Initial Learning rate	0.01
Activation function	Softmax

When the complexity of the RCNN Model is compared with other ML/DL models used in this work, 5 hidden layers generated the optimal performance for the driver activity anticipation problem. The authors have tested the variation in the hidden layers (from 1 hidden layer to 25 hidden layers) to check the model's performance. However, a model with 5 hidden layers generated the best set of results in both maneuver accuracy and anticipation time.

The proposed framework's system design contains three steps: a driver's inside context is taken as an input, visual motion-based feature extractions of head movement with eye gaze tracking, and the future action prediction. Since the focus is only on the inside details, features from the driver's head movements are tracked, extracted, and fused with eye gaze. Then our RCNN framework processes the features to find out the probability for five classes. In the dataset, five classes are Left Lane (L.L.), Right Lane (R.L.), Left Turn (R.T.), Right Turn (R.T.), and Straight Drive (S.D.).

C. RNN_CNN MODEL IN ADMT

RNN has been used in maneuver anticipation because it works well in sequential data like ours [37]. However, CNN is the best suit for video and image data, and the authors used CNN also in ADMT. This approach is called ''Advanced'' as it takes care of both sequential and visual aspects of the data by fusing CNN with RNN. It is to classify driver action from the video input data. Driver's Movement Tracking (DMT) is proposed in [43] to track driver's movement by context fusion of inside and outside video data and the RNN-LSTM framework. RNNs and CNNs can be combined for more efficiency in classification. In this paper, the authors have worked upon recurrence-based convolutional networks [48].

The STIP extraction method is implemented to determine the variations in the subsequent images, forming the multiple cuboids. Finally, the cuboids are fused to generate the final array of visual features. The advantage of STIP-based movement is that it can be directly detected from video to describe moving objects [35]. It is the local invariant feature for video, and it can resist the changes such as rotation, scale variations, viewpoint change, etc. In the ADMT model, face tracking is done by Kanade-Lucas-Tomasi (KLT) tracker [49], and then Spatio-temporal points from the face can be tracked [50]. The framework is elaborated below subsections.

Eye gaze estimation is played a significant role in getting the intentions of the person [45], [46]. So the eye region features are also extracted from the videos once the face detection is performed with the Viola-Jones detector [51], as shown in figure 2. DMT proposed fast and accurate eye tracking with effective localization of the iris center [43]. Subsequently, iris boundary and eye gaze are tracked. Kalman filter is used for iris tracking, which also works for eye closure identification and eye corner recognition, leading to accurate eye-gaze recognition [52]. Thus, implementing STIPs and eye gaze features could effectively track the driver's movements and predict the driver's action. It also helps to generate sufficient reaction time for future action prediction.

For the set task of action detection, the authors proposed a convolutional network consisting of Recurrent CNN or RCNN [48]. As depicted in figure 3, the RCL or recurrent convolutional layers are the core component of our proposed architecture and are considered the fundamental idea of our methodology. Though the input connection is static in this framework, the network can develop with these connections,

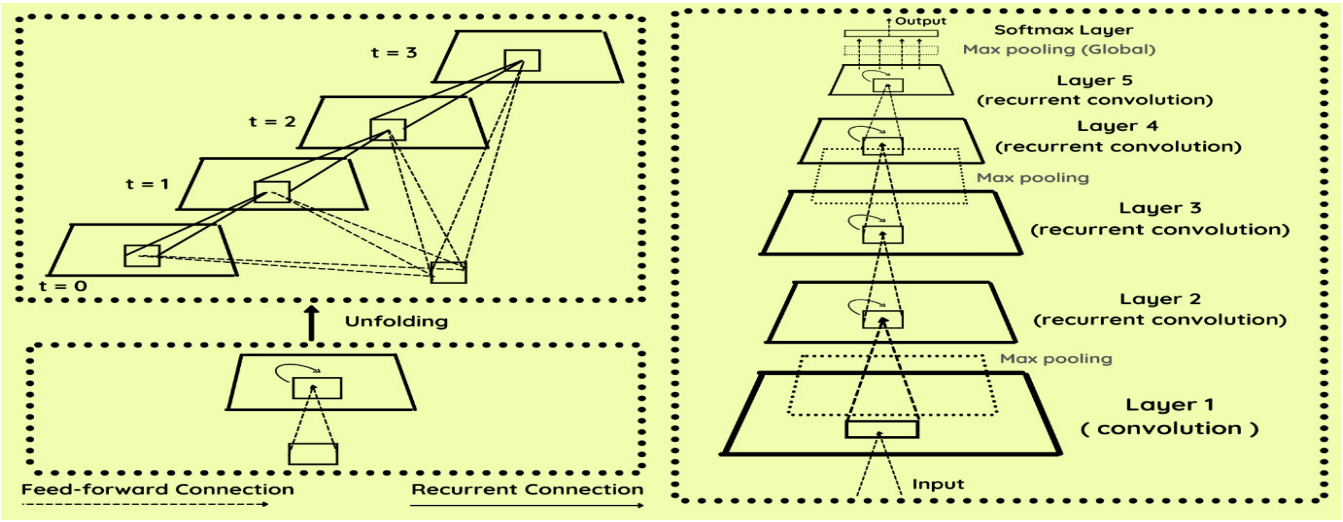

FIGURE 3. Propagation of data in the proposed RCNN [47].

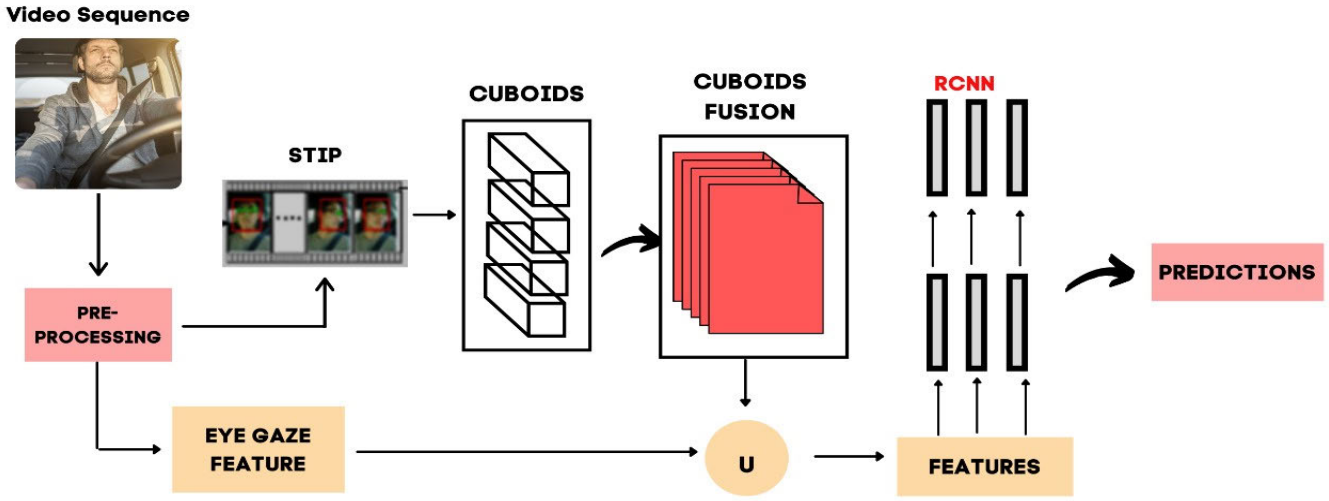

FIGURE 4. Proposed ADMT system using the RCNN.

and its adjacent units highly influence each network unit. First, the model structure of the proposed RCNN for driver activity prediction is visualized. Here, considering a triad of time steps of the recursive connections, the authors decode the working of the RCNN architecture. The architecture consists of a combination of convolution and pooling layer set such that the practical complexity of training is minor. Then use of max-pooling in the pooling layers to extract maximum features took place.

Further, four RCL layers are placed together, with a max-pooling layer placed after the first two. The terminal nodes of the RCL layers are connected to the final softmax layer via a global max-pooling layer. The stride values for pooling layers are set to 2 and 3. The feature map extracted from the global max-pooling layer helps to obtain the intermediate feature vector representation. A similar description for RCL is provided over 3-time steps, connected to an FFNN ranging in depth from one to four. Thus, the entire RCNN configuration consists of one convolution layer, three max-pooling layers, four recurrent convolutional layers, and a final softmax layer.

The block diagram of RCNN architecture for driver activity prediction is shown in Figure 4. The pre-processing steps and strategies used for the intermediate stages of feature extraction from STIPs and eye gazes are executed. During the prediction phase, the maneuver can be predicted, which is the driver's possible future action. However, the significant difference is that the class-wise probability for each input video frame is calculated using the RCNN model.

IV. ADVANCED DRIVER MOVEMENT TRACKING (ADMT)

Gite *et al.* [43] presented a Driver Movement Tracking (DMT) approach consisting of three steps: pre-processing techniques, STIP, tracking of the eye gaze, and classification

ADMT Algorithm 1: Maneuver Prediction algorithm using RCNN

Input:
Video Vint // Internal video frames
Features *Fint* // Determined features from internal context
Output:
Predicted Probabilities P (P1, P2,…Pn)

```
START
Consider a particular Vint
while (Vint)
        Tinst = extraction_frame (Vint);
    end while
    Ftr1 = STIP_extract (Tinst); //Spatio-temporal
    Ftr2 = eye_gaze (Tinst);
    Concat_frint = [Ftr1, Ftr2];
    Predictions P = model_predict (Concat_frint)
Action Anticipation using RCNN
return P (the probabilities for each driving action);
    END
```

FIGURE 5. Pseudocode of ADMT.

of actions using deep learning algorithms. The movement tracking process, which is done by STIPs, is then followed by the eye gaze estimation of the driver. It further enhances the accuracy of action prediction. In DMT, RNN with LSTM classifies the driver's maneuvers and is processed inside and outside contexts. The authors tried to explore CNN to solve the driver maneuver anticipation problem as CNN works best for image or video data.

A. MANOEUVRE ANTICIPATION USING ADMT

In the ADMT system, prediction of the driver activity is made by passing the visual input data to the deep learning model after extracting the STIPs and efficient eye-gaze features. Then LSTM is executed for action prediction, which reduces the overall inference time. Finally, the architecture predicts combining the elements from context frames and then propagating them further to the deep learning modules of the system. The system's working for driver activity prediction using a mix of algorithms is shown in Algorithm 1in figure 5, similar to presented in the papers [11], [43]. Initially, the Spatio-temporal and eye-gaze-based features are extracted, and then combined to form the intermediate vector representation. This representation is then fed to the RCNN model that returns the probability of the input video frame belonging to one of the possible class labels. This final prediction suggests the next feasible driver activity move.

In the proposed approaches, the authors only focus on the frames provided from the internal context, upon which all the processes are performed. These extracted internal features are used to derive the final prediction for the next activity of the driver. The output of the deep learning models is a probability for each of the class labels. Further, the obtained probability is normalized to a range between zero to five, where 0 indicates an entirely contrary predicted maneuver, while 5 shows a complete match to the expected safe maneuver. Finally, a limit threshold is applied to the derived probabilities to obtain the final prediction value post computation.

The performance of the proposed system is assessed via multiple metrics. Firstly, the classification accuracy is calculated with each possible maneuver as one of the class labels. Later, the authors also derive the F1 score, which is the harmonic mean of the system's precision and recall. The performance measures like precision, recall, accuracy, F1 score, and anticipation time have been considered for the system model. Two measures, namely accuracy in driver maneuver and anticipation time for predicting driver maneuver are the crucial parameters for our research problem. Jain *et al.* [15] have formulated the equations for precision and recall of the system for this task. Consider the following symbols and their definitions:

F.P. = False positive, i.e., activity is anticipated but the wrong one.

T.P. = True positive, i.e., maneuver predicted correctly.

T.N. = True negative, i.e., maneuver predicted, but no movement from the driver

F.N. = False negative, i.e., No activity anticipated, but the driver does perform a maneuver.

The primary reason for false-positive predictions could be samples of diverted driving. Driver's interactions with other passengers in the car or their distractions towards any neighboring visuals may get processed wrongly by the system. Consistent interpretation of the distractions caused by the driver is a challenging research task out of the scope of this research work.

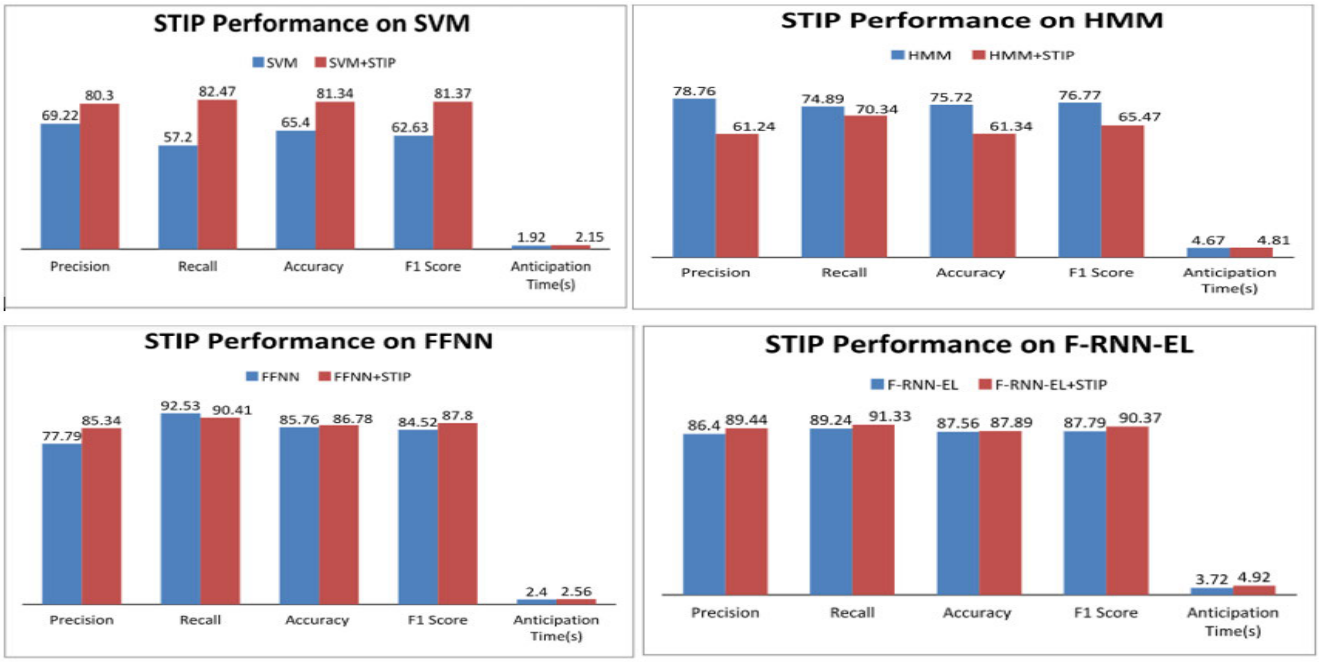

FIGURE 6. STIP performances on various algorithms.

In addition, the authors calculated the inference time that an algorithm takes to predict the activity and previously mentioned metrics. The authors term this time as the anticipation time and indicate it by T. Along with task-based performance. It is also essential to assess how early the system can predict a particular activity. The proposed deep learning-based model is expected to provide an improvement in not just the performance metrics but also make these predictions at a faster time. Here the authors focus on increasing it, as an early anticipation of the right maneuver prevents the driver from performing a risky activity and proactively assists him.

The anticipation time can be calculated as follows:

$$\mathrm{T} = \mathrm{TN} - \mathrm{T\ PN} \tag{1}$$

T.N. in the frame when the activity occurs, and TPN is when the maneuver is anticipated to occur at T.N. A higher value of prediction time indicates a better performing mode, the simple principle of deciding the better classification technique.

Having elucidated the terminologies and performance metrics, the authors are considering, we now present the proposed and standard approaches across these metrics for the determined dataset.

V. RESULTS AND ANALYSIS

The authors assessed the performance of the proposed RCNN in this section concerning the current state-of-the-art methods. They then conducted a comparative analysis of the earlier driver movement tracking algorithm [43].

A. CURRENT STANDARD METHODS

As baseline models for comparison and further evaluation, the authors considered a set of approaches. These standard sets of classifiers include Fusion of Recurrent Neural Network with LSTM (F-RNN-EL), SVM, HMM, along with the traditional feed-forward network [15]. Each of these methods considered for the study is:

- *Support Vector Machine (SVM):* It is one of the traditional approaches that involve a discriminative method for classification. The context videos from inside data for a 5 seconds frame are used for training the SVM. Then, the fusion of extracted features is done to get the probability of driver maneuver.
- *Hidden Markov Model (HMM):* HMM is a probabilistic classifier with one hidden layer but considers the only current context for prediction. It makes use of a Bayesian setup for the prediction of the driver activity. The fusion of extracted features is done to get the probability of driver maneuver.
- *Feed-Forward Neural Network (FFNN):* FFNN is a machine learning classifier with the feature of the discriminative classifier. The training of 5 seconds of driving context is given on FFNN, and the internal context features are combined to train the FFNN. The fusion of extracted features is done to get the probability of driver maneuver.
- *Fusion-RNN-Exp-Loss (F-RNN-E):* For more information on F-RNN-EL, interested readers are referred to [15]. It considers the RNN + LSTM for maneuver anticipation. LSTM is one of the proper models of sequential data for deep learning, making the driver activity prediction faster and more accurate.

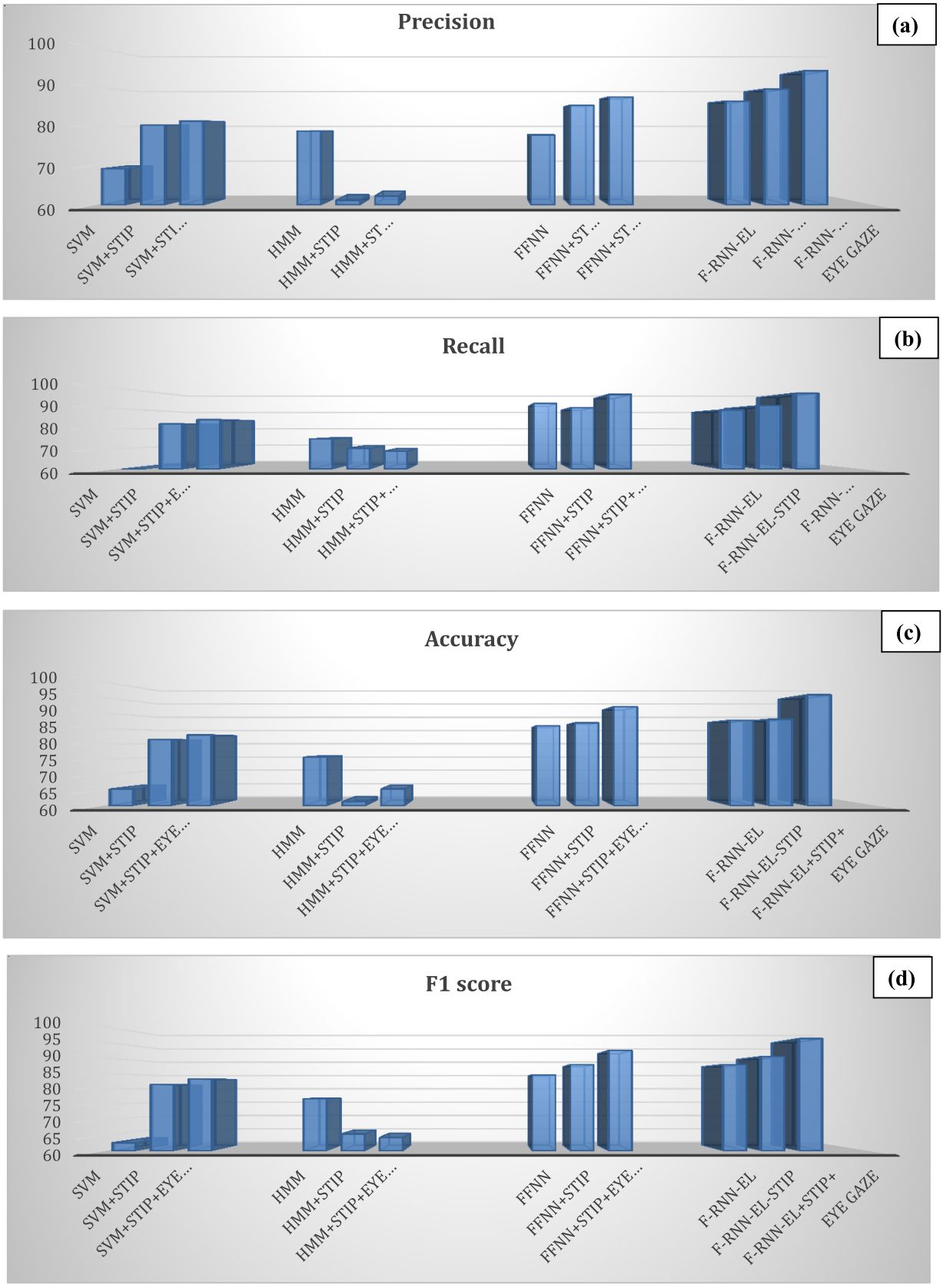

FIGURE 7. Performance evaluation metrics: (a) precision, (b) recall, (c) accuracy, (d) F1 score, and (e) anticipation time.

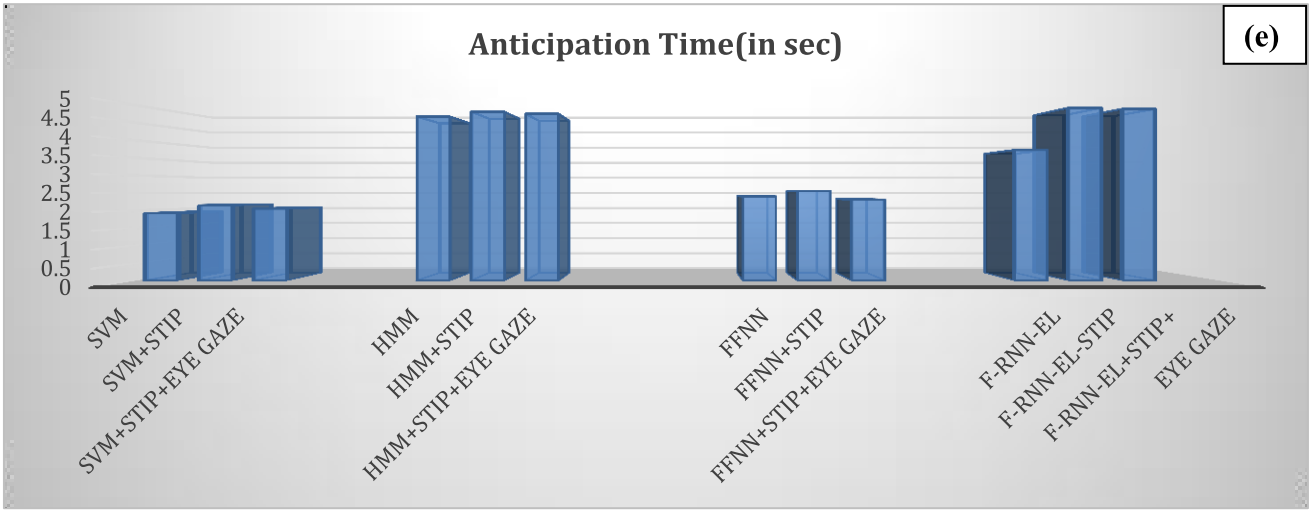

FIGURE 7. *(Continued.)* **Performance evaluation metrics: (a) precision, (b) recall, (c) accuracy, (d) F1 score, and (e) anticipation time.**

TABLE 2. Performance evaluation of F-RNN-EL-STIP and eye gaze tracking.

Techniques	Precision	Recall	Accuracy	F1 score	Anticipation time (in a sec)
SVM	69.22	57.2	65.4	62.63	1.92
SVM+STIP	80.3	82.47	81.34	81.37	2.15
SVM+STIP+Eye Gaze	81.3	84.5	82.86	83.22	2.05
% difference	17.45	47.72	26.66	32.87	6.77
HMM	78.76	74.89	75.72	76.77	4.67
HMM+STIP	61.24	70.34	61.34	65.47	4.81
HMM+STIP+Eye Gaze	62.3	68.9	65.43	64.3	4.75
% difference	**-20.89**	**-7.99**	**-13.32**	**-16.24**	**1.71**
FINN	77.79	92.53	85.76	84.52	2.4
FFNN+STIP	85.34	90.41	86.78	87.80	2.56
FFNN+STIP+EyeGaze	87.45	96.96	91.95	92.39	2.31
% difference	**12.41**	**4.78**	**7.21**	**9.31**	**3.75**
F-RNN-EL	86.4	89.24	87.56	87.79	3.72
F-RNN-EL-STIP	89.44	91.33	87.89	90.37	4.92
F-RNN-EL+STIP+ Eye Gaze	94.11	97.56	95.80	96.21	4.89
% difference	**8.92**	**9.32**	**9.41**	**9.59**	**31.45**

For minimizing the loss function, the exponential loss layer is applied. Then, the fusion of extracted features is done to get the probability of driver maneuver.

- *ADMT:* The authors proposed and implemented CNN for the driver maneuver anticipation. They used pre-processing techniques, Spatio-temporal interest point techniques for feature extraction, and CNN for the action classification.

B. EVALUATION OF THE F-RNN-EL-STIP MODEL

The performance of the F-RNN-EL-STIP model is compared against some popular approaches like SVM, HMM, and FFNN techniques. As observed in figure 6, the F-RNN-EL-STIP showed an essential improvement in evaluation parameters compared to existing methods due to an accurate motion tracking algorithm preferred over different visual features extraction techniques. Furthermore, the correct key point's extraction boosts the computation and maximizes the probability of the maneuver anticipation. Hence the improved system performance is observed as depicted in the below graphs.

The performance of STIP on SVM is improved in all evaluation parameters. Anticipation time is enhanced by almost 11% after using STIP, and accuracy is improved by 24%. Here STIP did work well in SVM for the desired performance. We noticed that STIP performance on HMM is not improved in almost all evaluation parameters except anticipation time, which is enhanced by nearly 3% after using STIP. The performance of STIP on FFNN is enhanced in all evaluation parameters. Anticipation time is improved by almost 1% after using STIP, and accuracy is improved by 6.66%. The performance of STIP on F-RNN-EL is enhanced in all evaluation parameters. Anticipation time is enhanced by almost 32% after using STIP. However, in HMM, STIP does not work well; instead, the performance is decreased. It is because of

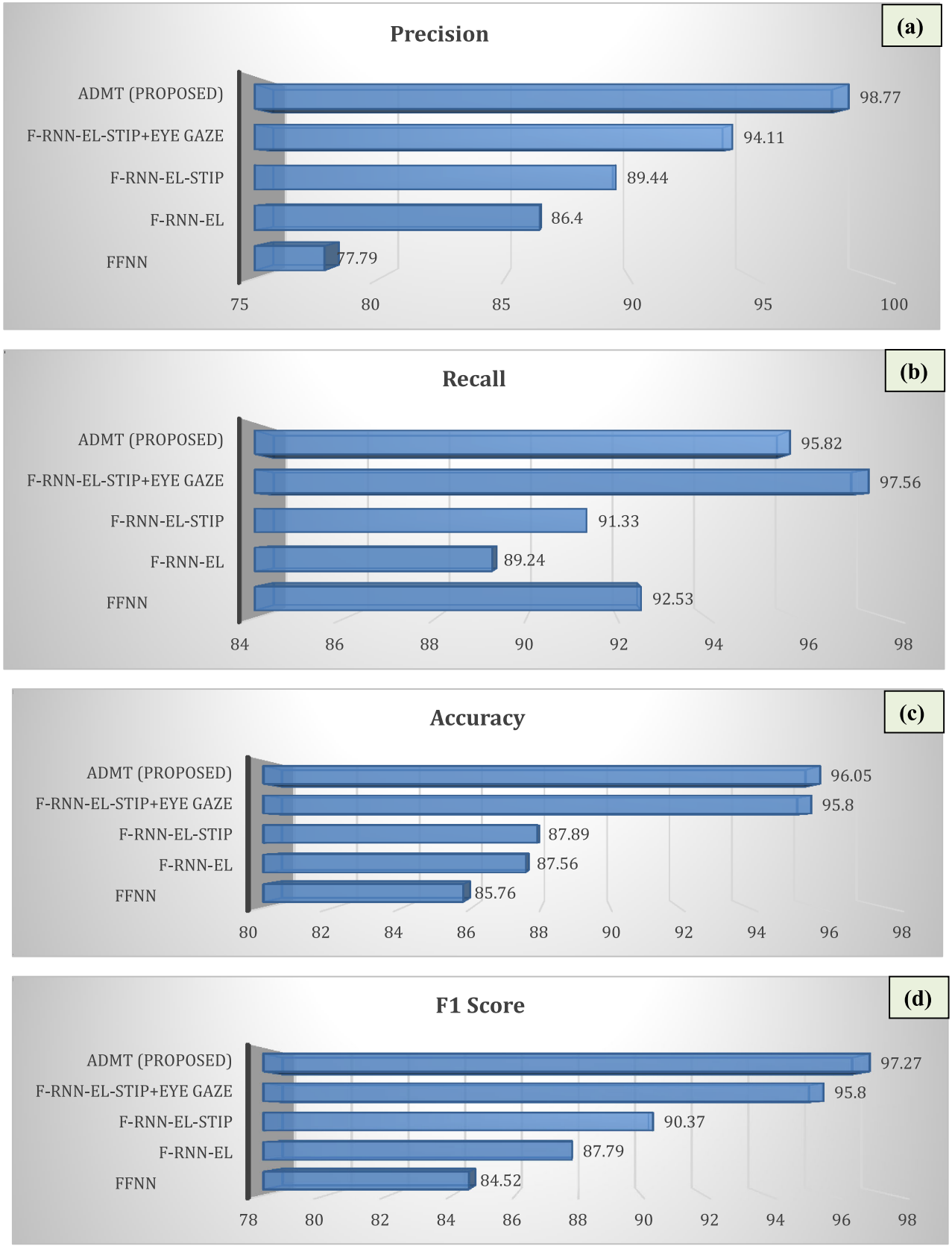

FIGURE 8. ADMT performance evaluation metrics: (a) precision, (b) recall, (c) accuracy, (d) F1 score, and (e) anticipation time.

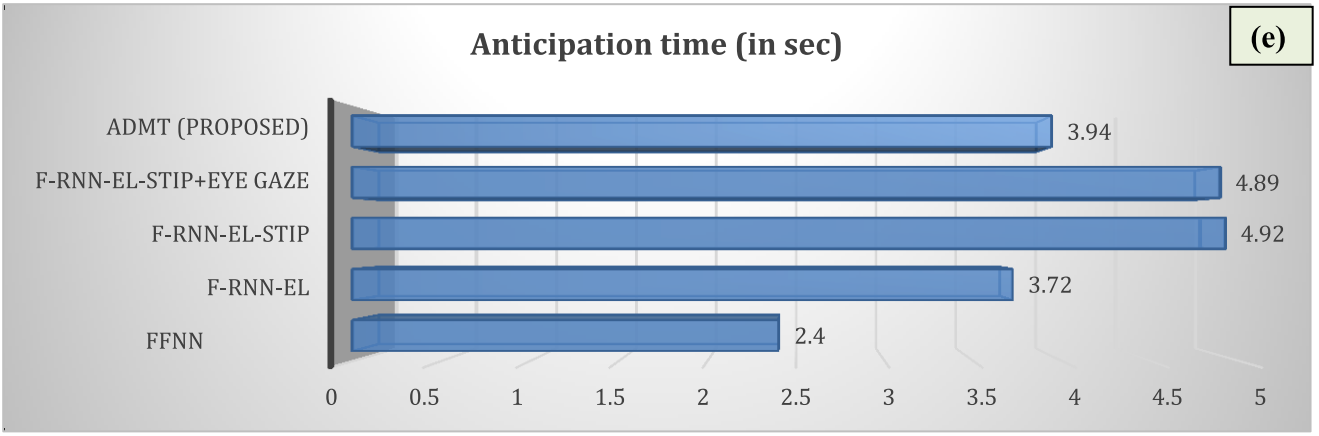

FIGURE 8. *(Continued.)* **ADMT performance evaluation metrics: (a) precision, (b) recall, (c) accuracy, (d) F1 score, and (e) anticipation time.**

the sequential video data that is given as the input to the system model. HMM handles only current data, and hence action recognition or prediction is not managed effectively and not a good solution for our kind of problem.

The thorough evaluation of the STIP and eye gaze methods are comprehended in table 2. The computed time average is 4.92 seconds before the actual maneuver may help the driver take timely action and using eye gaze 4.89 seconds. This is because of the reduced processing time required for only inside features, whereas other state-of-the-methods work on both inside and outside features computation. Further, the results of different classifiers using STIPs + Eye gaze features are presented, as demonstrated in table 2. For each classifier, eye gaze features with STIPs improve the precision, recall, and accuracy rates, with minimum impact on prediction time compared to just using the STIPs features. Eye gaze feature extraction has also seen improvements in the performance, as shown in table 2 and presented in diagrams 7 (a-e).

C. DISCUSSION ON STIP AND EYE GAZE IMPLEMENTATION

The main goal of using the STIP and eye gaze is to enhance the early anticipation time performance. SVM and FFNN are improved using STIP-based feature extraction techniques based on inside video compared to the performances discussed in table 2. STIPs with RNN-LSTM show the best performance compared to other classifiers.

The improvement in the anticipation accuracy is the focus on the inside context, which reduces processing efforts. The classifiers such as SVM and FFNN give insufficient anticipation time, whereas the HMM-based method gives excellent performance for anticipation time; however, the accuracy of maneuver anticipation is reduced for the HMM-based technique is not acceptable for the system. STIP with the deep learning methodology outperforms other standard algorithms in anticipation time and maneuver accuracy.

After applying the STIP and eye gaze to the state-of-the-art techniques, the results have shown an improvement. For example, suppose the percentage improvement of F-RNN-EL-STIP-Eye gaze is compared with F-RNN-EL increases around 9% inaccuracy and 34% in anticipation time. That concludes STIP with eye gaze is an effective technique when applied to F-RNN-EL.

D. EVALUATION OF ADMT

Table 3 presents the results of ADMT system implementation using all the performance metrics. The emphasis is to improve the anticipation time by processing only inside video data to track the driver's movements. The last row indicates the percentage of improvement in the various performance evaluation parameters of ADMT with the baseline FFNN. There is a clear indication of performance enhancement in terms of all the parameters, like precision, accuracy, recall, and anticipation time. The accuracy of the proposed ADMT is superior compared to the previous approaches, including the state-of-the-art ones. However, the anticipation time for STIP was the best so far compared to advanced techniques such as eye gaze extraction and RCNN. The probable reason behind this could be that the extraction of more features led to more processing and anticipation generation time. In RCNN, feature extraction through recurrent layers and then thorough convolutional layers might take more time, so ADMT took 3.94 seconds, which is still a 64% improvement compared with FFNN.

As shown in the above diagrams of ADMT performance in various evaluation parameters, ADMT is a superior solution for maneuver accuracy, one of the system's desired criteria. The ADMT approach gave a better performance as compared to F-RNN-EL and F-RNN-EL-STIP. However, anticipation time is not improved by ADMT; instead declined by almost 20% compared with F-RNN-STIP. A possible reason for the decrease in the anticipation time is the time required to run the RCNN networks. RCNN is computationally expensive compared to the other two approaches, where simple ANN

TABLE 3. Comparison of algorithms performance evaluation of ADMT.

Techniques	Precision	Recall	Accuracy	F1 Score	Anticipation time (in a sec)
FFNN	77.79	92.53	85.76	84.52	2.4
F-RNN-EL	86.4	89.24	87.56	87.79	3.72
F-RNN-EL-STIP	89.44	91.33	87.89	90.37	4.92
F-RNN-EL-STIP +Eye gaze	94.11	97.56	95.80	95.80	4.89
ADMT(Proposed)	98.77	95.82	96.05	97.27	3.94
% Difference	**26.97**	**3.55**	**11.99**	**6.27**	**64.16**

works on the system performance. Accuracy is improved by 8%. The generated anticipation time is 3.94sec by ADMT with RCNN and F-RNN-EL-STIP gives 4.92sec, which is sufficient for the driver to consider taking any further action without any haphazardness.

VI. CONCLUSION

In this paper, the authors carried out the task of driver activity anticipation using a deep learning method that combined convolution with recurrence to build an Advanced Driver Movement Tracking (ADMT) system. The comparison of STIP and eye gaze performance on SVM, HMM, FFNN and F-RNN-EL are presented, and it shows STIP improves the system performance in SVM, FFNN, and F-RNN-EL. It does not work well for HMM as it is not a good fit for non-linear data like the driver's video data. ADMT is both faster and accurate than the previous approaches. It has the novelty of requiring working only on the interior features, including the combination of the context video frame, the eye gaze movement, and the STIP values. STIP also performed well as the model could extract more Spatio-temporal features of the driver's activity, improving the system's performance. The RCNN obtains the best values for precision and recall. Therefore, it can be concluded to be the best performing method for this task. Our proposed deep learning-based approach has improved more than 12% accuracy (96%) and 62% in anticipation time (3.94 sec) compared to the basic DL model.

Some critical observations for this research work could be stated as the Brain4cars dataset contains only daytime videos, so this model's performance may be varied for night-time videos and can be considered an extension of this research work. Another limitation is the RCNN model's testing for robustness on some other datasets to formalize a generalized DL solution. Comparative study of the proposed model with similar DL ADAS-based models would lead to a new research foundation. Future research directions are to integrate bidirectional LSTM in the model architecture and implement transfer learning for activity anticipation problems. Similarly, the action recognition applications such as patient monitoring, mob monitoring, surveillance, etc., would be adopted using similar solutions. Being uncertain about the nature of the driver's behavior is a challenging research problem, and it would still attract many of the research frontiers in computer vision and deep learning.

AUTHOR CONTRIBUTIONS

Conceptualization: Shilpa Gite and Biswajeet Pradhan; methodology: Shilpa Gite and Biswajeet Pradhan; formal analysis and data curation: Shilpa Gite; writing – original draft preparation: Shilpa Gite; writing – review and editing: Shilpa Gite, Biswajeet Pradhan, Abdullah Alamri, and Ketan Kotecha; supervision: Shilpa Gite and Biswajeet Pradhan; funding: Biswajeet Pradhan and Abdullah Alamri; All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] V. Shia, Y. Gao, R. Vasudevan, K. D. Campbell, T. Lin, F. Borrelli, and R. Bajcsy, "Semi-autonomous vehicular control using driver modeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2696–2709, Dec. 2014.
[2] Accessed: Feb. 14, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7908960/, doi: 10.1007%2Fs43681-021-00041-8.
[3] M. Rezaei and R. Klette, "Look at the driver, look at the road: No distraction! No accident," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 129–136.
[4] B. Frohlich, M. Enzweiler, and U. Franke, "Will this car change the lane?—Turn signal recognition in the frequency domain," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 37–42.
[5] A. Doshi, B. T. Morris, and M. M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *IEEE Pervas. Comput.*, vol. 10, no. 3, pp. 22–34, May 2011.
[6] S. Al-Sultan, A. H. Al-Bayatti, and H. Zedan, "Context-aware driver behavior detection system in intelligent transportation systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4264–4275, Nov. 2013.
[7] R. Vasudevan, V. Shia, Y. Gao, R. Cervera-Navarro, R. Bajcsy, and F. Borrelli, "Safe semi-autonomous control with enhanced driver modeling," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 2896–2903.
[8] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1110–1118.
[9] F. Tango and M. Botta, "Real-time detection system of driver distraction using machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 894–905, Jun. 2013.
[10] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.
[11] S. Gite and H. Agrawal, "On context awareness for multisensor data fusion in IoT," in *Proc. 2nd Int. Conf. Comput. Commun. Technol.*, 2016, pp. 85–93.
[12] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 14–29, Jan. 2016.
[13] H. Koppula and A. Saxena, "Learning spatiotemporal structure from RGB-D videos for human activity detection and anticipation," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 792–800.

[14] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3182–3190.
[15] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, "Brain4Cars: Car that knows before you do via sensory-fusion deep learning architecture," 2016, *arXiv:1601.00740*. [Online]. Available: https://arxiv.org/abs/1601.00740
[16] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, "Survey of spatio-temporal interest point detection algorithms in video," *IEEE Access*, vol. 5, pp. 10323–10331, 2017.
[17] O. Mur, M. Frigola, and A. Casals, "Modelling daily actions through hand-based spatio-temporal features," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, Jul. 2015, pp. 478–483.
[18] S. Virmani and S. Gite, "Developing a novel algorithm for identifying driver's behavior in ADAS using deep learning," *IJCTA*, vol. 10, no. 8, pp. 573–579, 2017.
[19] D. Bhatt and S. Gite, "Novel driver behavior model analysis using hidden Markov model to increase road safety in smart cities," in *Proc. 2nd Int. Conf. Inf. Commun. Technol. Competitive Strategies (ICTCS)*, 2016, pp. 1–6.
[20] W. Dong, J. Li, R. Yao, C. Li, T. Yuan, and L. Wang, "Characterizing driving styles with deep learning," 2016, *arXiv:1607.03611*. [Online]. Available: http://arxiv.org/abs/1607.03611
[21] S. Gite and H. Agrawal, "Early prediction of driver's action using deep neural networks," *Int. J. Inf. Retr. Res.*, vol. 9, no. 2, pp. 11–27, Apr. 2019, doi: 10.4018/IJIRR.2019040102.
[22] S. Martin, S. Vora, K. Yuen, and M. M. Trivedi, "Dynamics of driver's gaze: Explorations in behavior modeling and maneuver prediction," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 2, pp. 141–150, Jun. 2018.
[23] D. Wu, N. Sharma, and M. Blumenstein, "Recent advances in video-based human action recognition using deep learning: A review," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2865–2872.
[24] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no, 13, pp. 666–676, Mar. 2016.
[25] A. Bux, "Vision-based human action recognition using machine learning techniques," Ph.D. dissertation, School Comput. Commun., Lancaster Univ., Lancaster, U.K., 2017.
[26] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, A. Patsekin, J. Kindelsberger, L. Ding, S. Seaman, A. Mehler, A. Sipperley, A. Pettinato, B. Seppelt, L. Angell, B. Mehler, and B. Reimer, "MIT advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation," 2017, *arXiv:1711.06976*. [Online]. Available: http://arxiv.org/abs/1711.06976
[27] I. Pitas, *Digital Image Processing Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 200.
[28] D. Liu, Y. Yan, M.-L. Shyu, G. Zhao, and M. Chen, "Spatio-temporal analysis for human action detection and recognition in uncontrolled environments," *Int. J. Multimedia Data Eng. Manage.*, vol. 6, pp. 1–18, Jan. 2015.
[29] J. Li, X. Liu, W. Zhang, M. Zhang, J. Song, and N. Sebe, "Spatio-temporal attention networks for action recognition and detection," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2990–3001, Nov. 2020.
[30] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognit.*, vol. 61, pp. 295–308, Jan. 2017.
[31] H. Wei and N. Kehtarnavaz, "Simultaneous utilization of inertial and video sensing for action detection and recognition in continuous action streams," *IEEE Sensors J.*, vol. 20, no. 11, pp. 6055–6063, Jun. 2020.
[32] A. Sasithradevi and S. M. M. Roomi, "Video classification and retrieval through spatio-temporal radon features," *Pattern Recognit.*, vol. 99, Mar. 2020, Art. no. 107099.
[33] L. Xia and J. K. Aggarwal, "Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2834–2841.
[34] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. IEEE Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, Oct. 2005, pp. 65–72.
[35] M. S. Ryoo and J. K. Aggarwal, "Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1593–1600.
[36] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognitionn," in *Human Behavior Understanding* (Lecture Notes in Computer Science), vol. 7065, A. A. Salah and B. Lepri, Eds. Berlin, Germany: Springer, Nov. 2011, pp. 29–39.
[37] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," 2016, *arXiv:1604.08880*. [Online]. Available: http://arxiv.org/abs/1604.08880
[38] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018.
[39] V. Veeriah, N. Zhuang, and G.-J. Qi, "Differential recurrent neural networks for action recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4041–4049.
[40] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, "Temporal segment networks: Towards good practices for deep action recognition," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 20–36.
[41] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *Proc. Int. Workshop Hum. Behav. Understand.*, Berlin, Germany: Springer, Nov. 2011, pp. 29–39.
[42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[43] S. Gite, H. Agrawal, and K. Kotecha, "Early anticipation of driver's maneuver in semiautonomous vehicles using deep learning," *Prog. Artif. Intell.*, vol. 8, no. 3, pp. 293–305, Sep. 2019.
[44] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[45] N. H. Cuong and H. T. Hoang, "Eye-gaze detection with a single WebCAM based on geometry features extraction," in *Proc. 11th Int. Conf. Control Autom. Robot. Vis.*, Dec. 2010, pp. 7–10.
[46] A. George and A. Routray, "Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images," *IET Comput. Vis.*, vol. 10, no. 7, pp. 660–669, Oct. 2016.
[47] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3367–3375.
[48] P. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2014, pp. 82–90.
[49] R. Boda, M. J. P. Priyadarsini, and J. Pemeena, "Face detection and tracking using KLT and Viola Jones," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 23, pp. 13472–13476, 2016.
[50] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2010-66, 2010.
[51] M. Jones and P. Viola, "Fast multi-view face detection," Mitsubishi Electr. Res. Lab, Cambridge, MA, USA, Tech. Rep. TR-20003-96, 2003, vol. 3, no. 14.
[52] A. Handa, R. Agarwal, and N. Kohli, "A comprehensive video dataset for multi-modal recognition systems," *Data Sci. J.*, vol. 18, Nov. 2019.

SHILPA GITE received the Ph.D. degree from Symbiosis International (Deemed University), Pune, India, in 2019, with a focus on deep learning for assistive driving in semi-autonomous vehicles. She is currently working as an Associate Professor with the Department of Computer Science, Symbiosis Institute of Technology, Pune. She is also working as an Associate Faculty with the Symbiosis Centre of Applied A.I. (SCAAI). She has around 13 years of teaching experience. She is currently guiding Ph.D. students in biomedical imaging, self-driving cars, and natural language processing areas. She has published more than 50 research articles in international journals and 20 Scopus indexed international conferences. Her research interests include deep learning, machine learning medical imaging, and computer vision. She was a recipient of the Best Paper Award at 11th IEMERA Conference held virtually at Imperial College London, London, in October 2020.

BISWAJEET PRADHAN (Senior Member, IEEE) received the Habilitation degree in remote sensing from the Dresden University of Technology, Germany, in 2011. He is currently the Director of the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT. He is also a Distinguished Professor with the University of Technology Sydney. He is also an internationally established Scientist in geospatial information systems (GIS), remote sensing and image processing, complex modeling/geo-computing, machine learning, soft-computing applications, natural hazards, and environmental modeling. From 2015 to 2021, he served as the Ambassador Scientist for the Alexander Humboldt Foundation, Germany. He has published more than 650 articles and more than 550 articles in Science Citation Index (SCI/SCIE) technical journals. In addition, he has authored eight books and 13 book chapters. He was a recipient of the Alexander von Humboldt Fellowship from Germany. He has been received 55 awards in recognition of his excellence in teaching, service, and research, since 2006. He was also a recipient of the Alexander von Humboldt Research Fellowship from Germany. From 2016 to 2020, he was listed as the World's Most Highly Cited Researcher by Clarivate Analytics Report as one of the world's most influential mind. From 2018 to 2020, he was awarded as the World Class Professor by the Ministry of Research, Technology and Higher Education, Indonesia. He is an associate editor and an editorial member of more than eight ISI journals. He has widely traveled abroad and visiting more than 52 countries to present his research findings.

ABDULLAH ALAMRI received the B.S. degree in geology from King Saud University (KSU), in 1981, the M.Sc. degree in applied geophysics from the University of South Florida, Tampa, in 1985, and the Ph.D. degree in earthquake seismology from the University of Minnesota, USA, in 1990. He is currently a Professor of earthquake seismology and the Director of the Seismic Studies Center at KSU. His research interests are in the area of crustal structures and seismic micro zoning of the Arabian Peninsula. His recent projects involve also applications of E.M. and M.T. in deep groundwater exploration of empty quarter and geothermal prospecting of volcanic Harrats in the Arabian shield. He has published more than 150 research articles, achieved more than 45 research projects, and authored several books and technical reports. He is a Principal and the Co-Investigator in several national and international projects, such as KSU, KACST, NPST, IRIS, CTBTO, U.S. Air force, NSF, UCSD, LLNL, OSU, PSU, and Max Planck. He is also a member of Seismological Society of America, American Geophysical Union, European Associate for Environmental and Engineering Geophysics, Earthquakes Mitigation in the Eastern Mediterranean Region, National Communication for Assessment and Mitigation of Earthquake Hazards, Saudi Arabia, and Mitigation of Natural Hazards Com at Civil Defense. He has also chaired and co-chaired several SSG, GSF, RELEMR workshops, and forums in the Middle East. He obtained several worldwide prizes and awards for his scientific excellence and innovation. He is the President of the Saudi Society of Geosciences and the Editor-in-Chief of the *Arabian Journal of Geosciences (AJGS)*.

KETAN KOTECHA is an administrator and a teacher of deep learning. He has expertise and experience of cutting-edge research and projects in A.I. and deep learning for last 25 years. He has published more than 100 articles widely in several excellent peer-reviewed journals on various topics ranging from cutting edge A.I., education policies, teaching learning practices, and A.I. for all. He has published three patents and delivered key note speeches at various national and international forums, including at Machine Intelligence Laboratory, USA, IIT Bombay under the World Bank Project, and International Indian Science Festival organized by the Department of Science and Technology, Government of India. His research interests include artificial intelligence, computer algorithms, machine learning, and deep learning. He was a recipient of the two SPARC projects worth INR 166 lacs from MHRD Government of India in A.I., in collaboration with Arizona State University, USA, and the University of Queensland, Australia. He was also a recipient of numerous prestigious awards like Erasmus+ faculty mobility grant to Poland, the DUO-India professors fellowship for research in responsible A.I., in collaboration with Brunel University, U.K., LEAP Grant at Cambridge University, U.K., UKIERI Grant with Aston University, U.K., and a Grant from the Royal Academy of Engineering, U.K., under Newton Bhabha Fund. He is currently working as an Associate Editor of IEEE Access journal.