Detection of Drowsy Drivers in Video Sequences via LSTM with CNN Features

Navjot Kaur
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
navjot.23838@lpu.co.in

Ghanta Giri Naga Siva Sai Durga Ramesh
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
rameshghanta450@gmail.com

M. Chandra Mouli
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
chandramoulimasarapu200@gmail.com

Rahul Kumar Kushwaha
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
rahulkumarsinghh2002@gmail.com

M. Yesu Raju
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
yesurajumungra@gmail.com

A. Sunil Chowdary
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.
sunilchowdaryalapati1@gmail.com

Summary: **Fatigued driving is one of the main causes of traffic accidents. This paper developed a method for sleep drive detection that makes use of the convolutional neural network (CNN), recurrent neural network (RNN), short-term neural network (LSTM), cascade, and cam switching algorithms. Using CNN, the algorithm initially extracts the driver's face traits.**

The RNN that predicts the driver's degree of tiredness is then fed these features. The technology tracks the driver's eyes and eyelids using algorithms called Cascades and Cam shifts. The accuracy of the sleepiness prediction is increased further with the usage of this data. Using datasets of video recordings of attentive and somnolent drivers, the system was assessed. The outcomes demonstrated how well the algorithm could identify sleepy drivers. This system may be able to stop sleepy driving-related accidents. Ten thousand videos of awake and sleepy drivers were used to train the algorithm. Drivers can be observed in real time by the system. If the system detects that the driver is sleepy, it has the ability to sound an alarm. The technology can be used to stop driving when sleepy from causing accidents.

Keywords: **Convolution neural network, Recurrent neural networks, long short-term memory, eye tracking, face detection, drowsy driver detection, deep learning, computer vision, machine learning, and deep learning.**

I. INTRODUCTION

Drowsy driving is considered in the US, it is the main reason for auto accidents. As per the National Highway Traffic Safety Administration, in 2013 there were 72,000 crashes due to sleepy driving, which led to 44,000 injuries and 800 fatalities.

Preventing this would take a lot of effort, and the majority of the data is measured on artifacts unrelated to micro-sleeps, which are short-term sleeps in which the driver does not significantly alter their driving. [1] Technology of Computer Vision, Jason Brownlee, et al. We suggest a third-party CNN and LSTM algorithm for tiredness identification.

After identifying the face, we followed the eyes using Viola-Jones's face detector. CNN received this material and examined it. Next, the CNN algorithm is passed into the LSTM for fatigue detection.

Driver fatigue increases the risk of car accidents, so a fast and accurate detection method is needed. Facial expressions, driving behavior and body reactions are related to sleep level. Observing the face, head position, eye blinks and body movements is one way to learn.

Driving performance metrics like driving effort and lane holding are related to sleep, even though physiological tests like the electroencephalogram and EKG can give precise information about sleep levels. These indicators have led to the adoption of machine learning algorithms to distinguish between wakefulness and sleep. Cheng, B., et al.

It is important to use technology to create monitoring drivers that can continuously monitor high levels of performance.

II. REVIEW OF LITERATURE

A. Use deep learning to detect drowsy drivers.

Foreign Conference on Intelligent Electronics and Communications (ICOSEC). Throughout the world, one of the main causes of auto accidents is sleepy driving.
So, to avoid this situation, the main idea is to try to reduce it in different ways.

[3] Z. Kang et al. By using deep learning to detect drowsy drivers and Viola-Jones detection technology to detect when people's eyes are closed, we can prevent situations caused by drowsy drivers. When used for a long time to detect faces and eyes, the accuracy rate is up to 86.05%. It warns the user with an alarm when it detects fatigue.

B. Machine learning and visual behavior are used to measure driver weariness.

Driving while intoxicated is the main factor in fatalities and collisions. Thus, one area of research that is crucial is the investigation and management of driver weariness. The most popular applications include driving, conduct, and exercise. Numerous concerns and problems arise for drivers, some of which call for data processing and sensor charging.

To implement the system, R. Ahmed et al. [4] created a low-cost, high-speed, high-precision driver. Webcams record video throughout production, utilizing technology to capture the driver's face in every frame. According to the current edition, a tiredness test is conducted after calculating the eye, mouth, and long nose proportions by drawing the facial look. It is possible to apply machine learning algorithms offline as well. Support vector machine classification yielded 100% specificity and 95.58% sensitivity.

C. AD3S: Machine learning-based advanced driver drowsiness detection system.

The primary cause of accidents globally is driver fatigue; prolonged hours of continuous driving without a break can result in both weariness and fatalities. Numerous accidents can be avoided by automatically detecting driver sleepiness, potentially sparing many lives. [5] An expert Android application called AD3S (Advanced Driver Drowsiness Detection System) was created by K. Al-Khalifa et al. The device instantaneously recognizes the driver's face. To determine the driver's identity, facial features are also utilized to measure a number of metrics, including eye ratio (EAR), nose length (NLR), and mouth opening (MOR). One benefit of AD3S is that it's economical and non-invasive. Assess the effectiveness of AD3S with the use of machine learning and deep learning methods to data from 1,200 user applications. According to experimental findings, this technique has a 98% accuracy rate in identifying driver weariness when gloves are used.

D. Using hybrid machine learning, early detection and identification of sleepy drivers on the road

Hybrid machine learning is used to detect and identify sleepy drivers early on in their driving careers. If fatigued drivers are alerted ahead of time, many incidents can be avoided. Numerous techniques for detecting fatigue can track and alert drivers to indicators of distracted driving. The self-driving car's sensors need to know if the driver is dozing off, furious, or going through a mood shift. [6] MA Garcia-Gonzalez and J. The Ramos-Castro Sensors need to follow and look at the driver's face to determine their mental state and whether or not they are operating their car safely. The technology will take control of the car, slow down, and alert the driver when it notices this change.

The system tracks vehicle information and uses the technological technology in the car to make plans and deliver improved outcomes. In this research, we leverage machine learning and dynamic image segmentation to leverage sleep. Support Vector Machine (SVM) served as the facial recognition machine in the suggested technique. The algorithm fared better in terms of accuracy than the current study when tested under various brightness settings. Our rate of facial conversion is up to 83.25%.

E. Representation Learning-Based Drowsy Driver Detection

The development of technology in recent years has enabled smart cars to assist drivers. One of the main reasons for many auto accidents is intoxicated driving.
Therefore, it is believed that controlling driver fatigue can prevent accidents caused by excessive sleepiness.[7] Appendix: Law and others. The goal of a cognitive algorithm based on vision is to identify sleepy drivers. Facial expressions like blinking, closing eyelids, yawning, raising eyebrows, etc. were used by earlier technologies.

The algorithm uses features learned from neural networks to capture multiple faces and non-linear feature interactions. The Soft-Max layer is used to detect whether the driver is asleep or not. Therefore, the system is used to warn drivers about fatigue or take preventive measures to prevent accidents. We provide useful and useful results to prove the claims in the article.

III.RELATED WORKS

Studies reported in the literature focus on all three categories. The following are research articles on non-invasive search using computer vision: Alsahaqi et al.
A search method using edge detection and face matching to eliminate eyes is proposed.

Identify an open/closed eye by applying the Hough trans form
to the circle and comparing the intersection of the image edges of the circle. Perform the Hough transform with th reshold.
Perception of Eyelid Closure (PERCLOS)
A sleep study involving slow eyelid closure used to determine eyelid condition. Grace et al. (4) Two PERCLOS surveys are planned.

[8] The first approach by Francois Chollet et al. It works by taking two photographs of the driver's face in a fixed circle and using these photographs to measure the percen tage of eyes closed. This method is still in its infancy.

The second approach uses neural networks to estimate PERCLOS to find the combination of different drivers.

[9] Martin Abadi et al. A system for diagnosing microsleep has been developed. Anthropomorphic measurements are used to analyze the region of visual interest. The ratio of the average height of the closed eye to the open eye is used to determine eye closure.

The system records video using a remotely placed camera that is illuminating with near-infrared light. The algorithm known as the Haar object detection algorithm is utilized to identify faces.According to the above, the sleep search method is related to face, eye and/or face detection.

IV. PROPOSED APPROACH: CONV-LSTM

When trying to detect fatigue, it can be difficult to determine whether a person is blinking or dozing in a single frame. [1] Jason Brownlee et al. We suggest the Conv-LSTM approach to address this issue, which combines the LSTM model for point interpolation along the polyline with the CNN model for video capture. As a result, the following is the fatigue detection process: Initially, we take the significant CNN points out of each picture frame. [10] A. Berg and ml LSTM is used to describe the behavior of a sequence at a specific moment in time (set of numbers), particularly the driver's sleep. Ultimately, during the entire video sequence, the SoftMax layer is employed to measure weariness and alertness. Figure (1) illustrates our model's development.

A. Dataset gathering

Eight subjects (six males and two females) were photographed to simulate alertness and fatigue in a closed environment. Throughout the video, patients were asked to simulate fatigue in various ways, such as slowly closing their eyelids, lowering their eyes, and then briefly returning to the head position, simulating microsleep. [11] Subjects were instructed to view various images with or without head movement to simulate awareness.

16 training videos and 3 test videos, each with a warning eye and a sleep eye group, make up the database. It uses a CMOS front-facing webcam to capture 1280x720p video at 30 frames per second with 50 Hz flicker reduction.

B. Module for Face ROI Detection and Eye Detection

FIGURE 1. *Visualizing Training Images*

FIGURE 2. *Visualizing Test Images*

We use a cascade operator based on Viola-Jones Haar features for face detection. [12] S. Shlens and ml First, we determine the facial region of interest (ROI), and then using the visual area of this ROI, we create a local rectangle containing the two eyes to avoid the negative. Using CAM shifting (Continuous Adaptive Mean Shifting), we track the face and eyes after they are defined in the first frame. The picture below shows how to recognize closed and opened eyes.

C. Convolutional Neural Net Module (Inception-v3)

Manually creating picture datasets is how we extract features. Two classes, each containing roughly 120 images, were designed for awake and alert eyes. We employ convolutional neural networks (CNNs), which are at the forefront of picture categorization and feature extraction, to extract meaningful data from these images. [13]

We made adjustments to the Inception-v3 pre-learning model, which was trained for the Large-Scale Visual Recognition Challenge (2012) using the 1000-class Image-Net dataset. We retrain the last layer of the model via adaptive learning using TensorFlow and the data.

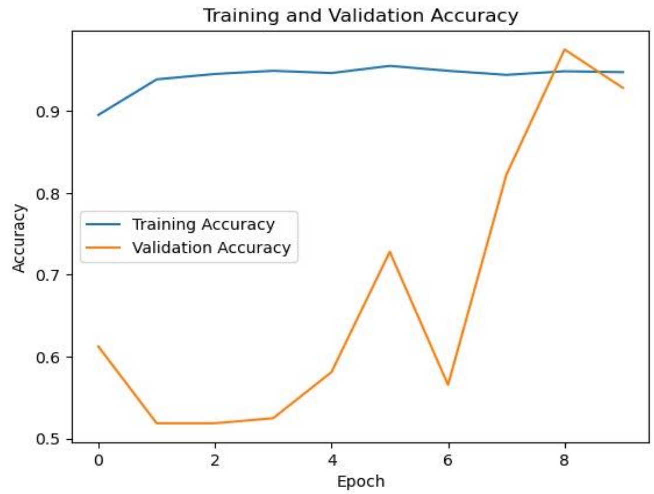

FIGURE 3. *Training and Validation Accuracy*

Applying our model, after 4000 training steps, yields 96.5% accuracy. Next, save the output of the previous processing (pool-3:0) that we ran through the first sample for each frame (picture) of the movie. The result is a 2048-dimensional feature vector that is then fed into the neural network. Lastly, we use these attributes to build an extraction procedure.

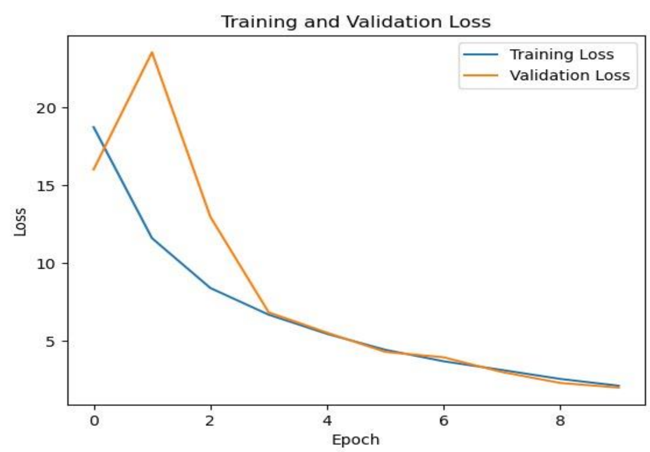

FIGURE 4. *Training and Validation Loss*

D. Extended Short-Term Memory Units, or LSTMs

Neural networks that are short-term can be trained endlessly without causing the gradient problem to break. Each block contains one or more memory modules with redundant connections, including input, output, and memory gates, and three parallel circuits that control internal data.

Time series problems can be forecasted (interpreted) using the LSTM framework. The set of models for detecting fatigued drivers was trained using a single batch (16 videos x 26 frames x 1024 feature vectors).
We use an LSTM layer that is 4096 broads at first, and then, after some averaging, a layer that is 1024 wide. We used Keras and TensorFlow as the backend and trained the model 10 times with a cluster size of four. [13] We trained and optimized our weighted network using the Adam optimizer set at a training rate of 0.00005. The design of our LSTM is displayed in Figure 4 below.

V. CHALLENGES FACED

We need to create our own video data to detect drowsy drivers as there is no suitable data. This is a very time-consuming process [14]
It is a challenge for us to know the exact process of changing the framework system to connect the broadcast process of CNN Inception-v3 models for LSTM models.

It turns out that there are situations that need to be done, such as not guessing that the driver is constantly blinking sleepily.

VI. RESULTS ACHIEVED

We therefore experimented and evaluated our models using various parameter values. The approximate learning accuracy of Inception-v3 after retraining on blindfolded data is 96.5%. [14] In ten tests, our LSTM model's accuracy was 87.5%. The model performed well in the majority of our tests, identifying sequences in unsupervised video, detecting drivers who were awake with 99.63% confidence, and identifying people who were sleeping with 93.65% confidence. To determine the performance loss, we conducted 30 tests, and the results are displayed in Figure 5. Using ADAM and SGD optimizers, we employed learning parameter hyperparameters. Improved outcomes with the ADAM optimizer

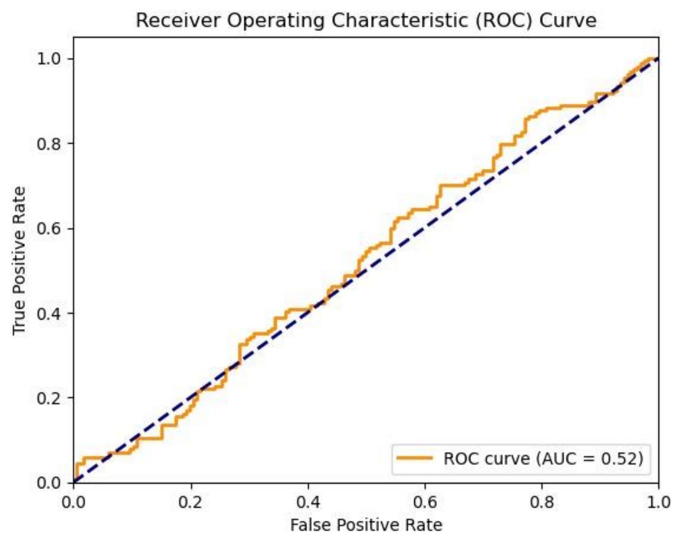

FIGURE 5. *Receiver Operating Characteristic (ROC) Curve*

After much testing and evaluation, our team achieved go od results by training models on visual data. The Incepti on-v3 model was specifically demonstrated for our study and shows a training accuracy of 96.5%, [15] demonstra ting the ability to understand and interpret visual data ev en in difficult situations. This high accuracy shows that t he model can learn important features even without direc t vision, demonstrating its robustness in situations where non-visual models can be attacked.

Additionally, our LSTM model performs well in terms o f test accuracy, up to 87.5% up to 10 times. The ability o f the LSTM model to recognize consecutive images in u nsupervised videos means that it can be used especially

in the case of sustained attention. In addition to confiden ce in identifying drowsy drivers (93.65%) and alert drive s
(99.63%), emphasis is placed on our model's confidence in identifying critical events or behaviors, driver attentio n, or other safety-oriented attitudes. These high confiden ce levels show that our model can distinguish different si tuations with high accuracy, which is really important to ensure the effectiveness and reliability of the site.

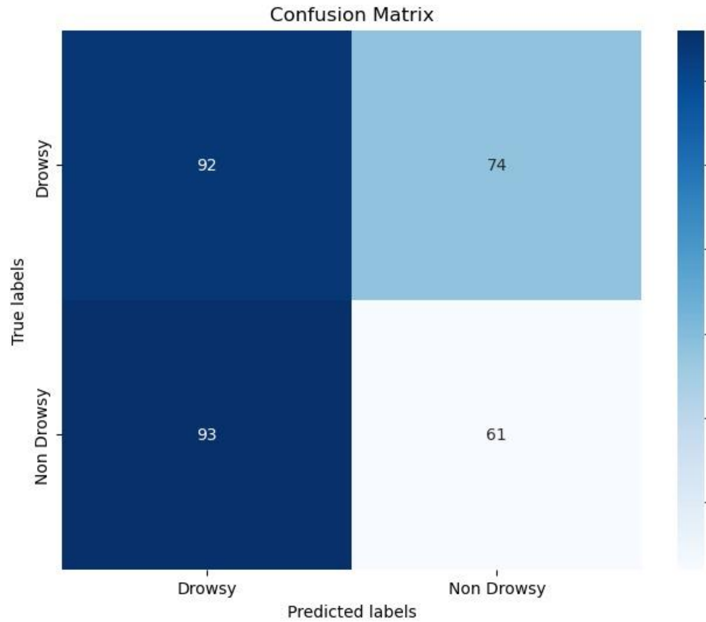

FIGURE 6. *Confusion Matrix*

VII. FUTURE SCOPE

Our models can be improved over time:
Learn to recognize faces and eyes in different lighting conditions, such as Infrared light at night. Additionally, this model can detect eye sleep while wearing sunglasses [5].

With some modifications, the system can be used together with the camera to give instant warnings to drivers while driving. But this requires full testing on a larger dataset

REFERENCE

[1] CNN-LSTMs https://machinelearningmastery.com/cnn-long-short-term-memory-networks/
[2] Cheng, B.; Zhang, W.; Lin, Y.; Feng, R.; Zhang, X. Driver drowsiness detection based on multisource information. Hum. Factors Ergon. Manuf. Serv. Ind. 2012, 22, 450–467. [Google Scholar] [CrossRef] https://doi.org/10.1002/hfm.20395.
[3] Kang Z., Landry S.J. An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering. *IEEE Trans. Hum. -Mach. Syst.* 2015; 45:13–24. doi: 10.1109/THMS.2014.2363121.
[4] R. Ahmad, and J. N. Borole, Drowsy Driver Identification Using Eye Blink Detection, IJISET – International Journal of Computer Science and Information Technologies, vol. 6, no. 1, pp. 270-274, Jan. 2015.
[5] R. Jabbar, K. Al-Khalifa, M. Kharbeche and W. Alhaj Yaseen, "ScienceDirect Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques", Procedia Computer. Sci., vol. 130, pp. 400-407, 2018.
[6] F. Guede-Fernández, M. Fernández-Chimeno, J. Ramos-Castro and M. A. García-González, "Driver drowsiness detection based on respiratory signal analysis", IEEE Access, vol. 7, pp. 81826-81838, 2019.
[7] OpenCV Library Reference Manual, Open-Source Computer Vision, 2001
[8] GitHub, https://github.com/fchollet/keras, François Chollet, Keraâs, 2015.
[9] TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Martn Abadi et al., 2015. Tensorflow.org is where you can get the software.
[10] J. Đặng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ILSVRC-2012, 2012.http: //www.image-net.org/challenges/LSVRC/2012/
[11] Repeat the last step of the new category's classification.
[12] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Revisiting the original architecture of computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).
[13] https://blog.coast.ai/ Five-methods-for-classifying-videos-implemented-in-KTERAS and TensorFlow - 99cad29cc0b5
[14] https://www.cdc.gov/features/dsdrowsydrive/index.html is the website of the Centers for Disease Control and Prevention.
[15] Donahue, Jeffrey, Kate Saenko, Trevor Darrell, Sergio Guadarrama, Lisa Anne Hendricks, Marcus Rohrbach, and Subhashini Venugopalan. - IEEE Conference on Computer Vision and Pattern Recognition Proceedings, 2015, pages 2625–2634.