Real Time Drowsy Driver Monitoring and Detection System Using Deep Learning Based Behavioural Approach

P.E. Kekong*, I.A. Ajah2, U. Chidiebere3

1,2Department of Computer Science, Ebonyi State University, Abakaliki, Nigeria
3Desnet Research Group, New layout, Enugu, Nigeria

*Corresponding Author: piuskekong2019@gmail.com

ABSTRACT- This work presents real time drowsy driver detection and monitoring system using deep learning based behavioral approach. The aim is to design and implement software which captures live driver’s behavior during driving and train using convolutional neural network (CNN) to predict the behavior’s of the driver. This was achieved developing a drowsy driver dataset; intelligent video based device and the CNN architecture and configurations. The designs were implemented using deep learning tool and MATHLAB. The system was tested and the result showed a detection accuracy of 99.8%. MATHLAB was used to develop a prototype model of the system.

KEYWORDS- Drowsy behavior, Convolutional Neural Network, Training, Deep learning, MATHLAB

I. INTRODUCTION

According to [1] an estimated 90% of passengers depend on road network for transportation. This implies that the passengers reach out to their respective commercial service based, habitat or domain via vehicles. Some of the commercial driver’s piloting these vehicles spend most of their times on the steering wheel driving, since their pay checks depend on the number of passengers they convey to various destinations daily. This alone accumulates enough mental stress and fatigue on the driver’s brain and induces drowsiness while still on high way after some hours. This is the reason why some drivers take precautionary measures like eating bitter kola, drinking coffee, drinking alcohol, taking few minute refreshment breaks, eating chewing gums and even smoking sometimes with hope of countering this effect of drowsiness. However one cannot manipulate nature as these traditional techniques are not good enough or reliable. The use of alcohol for instance induces more fatigue to the brain cells and increases the rate of drowsiness contrary to the drivers aim and as a result can have a devastating effect on the driver concentration, thus endangering the lives of the passengers on board and most time leading to series of road accidents.

For instance in the final quarter of 2015 according to a report released by the federal road safety corp (FRSC) of Nigeria, over 12,077 road crashes were recorded with 75400 fatalities within the year [2]. A more recent annual report by the same agency recorded 3,947 road crashes involving 6448 vehicles which results to 1758 deaths and 11250 injuries between the first quarter of the year 2020, with all investigations pointing to drowsy drowsiness and reckless driving. Therefore there is need for intelligent based drowsy driver monitoring and detection systems which is cost effective so as to affordable and monitor driver’s action when driving with précised result.

To solve these problems, various research works have been developed on drowsy driving detection system over the past decades, employing different methods such as physiological, behavioral and vehicle-based methods respectively. In physiological methods [3], electrocardiography, electroencephalography and physiological sensors are used to access the driver’s conditions, but despite the success, this method is not widely employed due to the implementation complexities and non user friendly condition.

In vehicle-based methods [4], drowsiness is analyzed based on the nonlinear dynamics of the vehicle state during translation and then control based on automatic cruise control techniques. This method despite the success is limited by huge implementation cost, false alarm and sometimes delayed response time. [5] Reported that accident takes just some seconds of drowsiness to occur and hence there is need for a real time monitoring and detection system.

The behavioral method [6] on the other hand employs sensors like camera to capture the driver’s condition and process to detect driver’s action while on the steer wheel. This method presents better performance because it is actually designed to understand and process the driver’s action directly compared to the vehicle based and physiological methods respectively. Due to this reason, most recent papers published to solve this problem of driver drowsiness have been based on this method (behavioral) employing various approaches with image processing [7] and the use of artificial intelligence [8]
among research favorites. These approaches employed real time images from drowsy driver dataset to process and predict driver behavior, and if drowsiness is sensed, the system alerts the driver. In the image processing approach, various techniques like segmentation, edge detection, computer vision, and similarity differential models to distinguish between driver’s state and drowsy detection results. This approach despite the success gives high rate of false alarm, especially when the driver blinks continuously within intervals. The use of machine learning (ML) techniques like support vector machine, fuzzy logic, hidden markov model, k-nearest neighbor, and clustering approach on the other presents the better detection accuracies when compared to the image processing counterpart [8], [9] and [10], but lacks common sense lack the intelligence of the exact type of data to collect and process and as a result has hindered the system success and global adoption of the technique for now. This research therefore presents a user friendly, reliable, cheap and real time drowsy driver behavior monitoring and control system using deep learning technique.

Deep learning (DL) is a kind of machine learning function which mimics the human brain power for processing of data and decision making. This DL has the ability to learn on its own without the supervision of human for both unlabelled and unstructured data. This technique is also the best to classify large amount of data like the case study where 9000 frames of images is captures within five minute drive [11].

Research in [12] and [13] among other have adopted this DL technique to solve the problem of drowsy driving, but despite the success, still lacks common sense, and poor quality of training data resources.

Common sense implies that the system should have the intelligence of the problem at hand and also the type of data to collect by the camera, just like human eye and brain functions, but this coordination is lacking in the existing DL techniques and as a result has hindered their complete success. Hence there is need for an intelligent image acquisition device which has the intelligence of the exact type of data to search for drowsiness and collect for training. Secondly most deep learning technique are designed to train data based on eye blinking especially, even though it is not the only drowsy sign that can be dangerous. Therefore, there is need for training dataset which considers all drowsy behaviors such as micro sleep, yawning, eye closed with glasses, continuous eye, heads down, blinking with and without glasses, yawning with glasses face down as attributes. This will highly improve the accuracy of the DL detection and system reliability.

In DL techniques, there are the convolutional neural network (CNN), long short term memory, recurrent neural network, deep belief network among others, however the main advantage of CNN compared to other DL techniques is its ability to automatically detect an extract important feature vectors via a hierarchical model which works by building various network layers connected as a fully connected layer to produce a processed output with very high accuracy. This technique will therefore be employed in this research to develop a real time drowsy driver monitoring and detection system with the following contribution to knowledge.

- A drowsy driver system with common sense will be developed
- A reliable, easy to use, intelligent and real time system which won’t interfere with the driving process when installed will be presented.
- A new drowsy driver dataset will be developed
- An intelligent video based drowsy monitoring device will be developed
- A new deep learning model will be developed and configured

II. LITERATURE REVIEW

[14] presented a vehicle based drowsy detection system using steering and lane data based features. However this technique suffers false alarm as other factors such as pot holes for instance can cause lane diversion or nonlinearity in the vehicle dynamics which does not necessary indicate drowsiness. [15] presented a driving stress detection system using physiological sensors, however this approach despite the success, interferes with the driving condition and as a result can be a distraction to drivers concentration. There is need for a drowsy driving detection system which has no direct contact with the driver, but still performs the monitoring and detection system effectively. [16], presented a paper on a comprehensive study on behavioral parameters based drowsy detection techniques. The paper identified the behavioral approach as the best for drowsy detection when compared to other approaches like the behavioral and physiological. From the paper it was observed that the best result achieved so far using deep learning technique is 96.7%. However the result is limited to only sleep and yawning as drowsy features based on eye and mouth changes. [17] Used a combination of CNN and SVM monitor eye behavior during driving condition. The accuracy achieved is 94.80%. [18] Detected drowsiness based on facial features such as head, mouth and eyes and achieved an accuracy of 95.8%, while in [19] viola jones algorithm was used for the detection of eye features like sleep and blinking as drowsy attributes and achieved an accuracy of 93.79%. In [20] fuzzy logic based approach was used and furnished with ANN, and feature descriptor.. [21] Developed an effective eye blink detection system for the disabled using open CV, eye aspect ratio and adaboost classifier and achieved an eye blink detection accuracy of 85.7%. The system was tested and result achieved is 94%. The literature reviewed revealed generally that despite the causes of the existing system, they are all limited by training dataset or training performance.

III. THE PROPOSED SYSTEM

The proposed system was developed using a training dataset of video clips made up of various driving attributes
characterized as critical, minor and perfect driving behaviors. These data were processed to have the same specifications such as length, weight and then used to train the deep learning technique to develop the proposed system.

The video acquisition device was made intelligent and used for the data collection; this intelligence was possible using computer vision to detect drivers face (since the face is the main part of the body where drowsiness can be sensed) and automatically search for facial points with drowsy features. The data collected is set to the same size as those in the training dataset. The data collected were extracted feed forward for training using deep learning technique. After the training the result was classified to detect driver behavior in three outputs as follows; if critical drowsy symptoms like sleep for instance are detected, then a warning sign indicating danger is displayed, else if classified as perfect driving condition, then it is displayed, else if any minor drowsy behavior is classified such as yawning for instance, then sensing drowsy behavior is displayed.

The activity diagram of the proposed system is presented as shown below which further explained the process flow of the system and the interactions within each step as shown below.

![Activity Diagram](image)

**Figure 1: The proposed system activity diagram**

## IV. METHODS AND SYSTEM MODELLING

This section presented the step by step processed used to design and implement the proposed drowsy driver detection system as shown below;

**Video Acquisition**: this is a process of collecting real time driving behavior from the driver in a video format. The process was facilitated using a hardware device (video camera) whose operation was made intelligent via computer vision. The computer vision was developed
using viola jones algorithm in [22] and then incorporated into the camera to have object detection intelligence with main focus on human face for data collection. The camera collects 30 frames of pictures per seconds in the size of 200 x 200 resolutions. This implies that the quality of each frame is 200 pixels in height and 200 pixels in weight; hence the total number of pixels for each image quality is 40000.

**Face detection and tracking:** this is a process of detecting a human face from other objects. This process was made possible using computer vision algorithm as also disclosed. The algorithm equipped the camera with facial detection and tracking capabilities and then collects information in series of frames for processing. Data flow diagram (DFD) was used to model the video acquisition and intelligent data collection process as shown below;

![Diagram](image)

**Computer vision**
This is an intelligent technology which was used to improve the camera object detection ability via viola Jones Algorithm. The algorithm was developed in [19] and [22] for detecting, tracking, and collecting facial features. This process is called face detection.

**Training dataset**
A sample size of 10 self volunteered drivers were used as the primary data source with each displaying drowsy symptoms such as yawning, frequent eye blinking, face down while driving, eye close (micro sleep). These data were collected using video camera and store three classes which are critical drowsy class, containing the most dangerous drowsy symptom which is micro sleep. Minor drowsy class which contains attributes like yawning, eye blinking and face down; then the perfect driving class with attributes of eye open and focused. The data were captured in a resolution of 200 x 200 AVI format and then stored as video in the table below;

<table>
<thead>
<tr>
<th>Drivers</th>
<th>Minor drowsy attributes</th>
<th>Critical drowsy attributes</th>
<th>Perfect driving attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/N</td>
<td>Yawning</td>
<td>Frequent eye blinking</td>
<td>Blinking eye with glasses</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Total video</td>
<td>200</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Total image frames</td>
<td>18000</td>
<td>27000</td>
<td>27000</td>
</tr>
</tbody>
</table>
In the table 1, each video has a time lapse of three seconds and contains ninety frames of images. The total video collected from each driver is 130 clips, comprising of 11700 frames and then stored in three classes to make up the training dataset. The class of minor drowsy attributes characterized with drivers yawing, frequent eye blinking, and eye blinking with glasses has 80 video clips of 7200 frames. The class of critical drowsy attribute characterized with eye closed, micro sleep (sleep for 3 seconds), and face down with and without glasses has 30 video clips of 2700 frames and then the class of perfect driving attribute with eyes open with and without glasses has 20 video clips of 1800 frames respectively. The total video in the training dataset is 1300 video clips of three second each with 117000 frames and the dataset is designed using the class diagram below;

**Feature Extraction**

This is a process of extracting images frame from the video clips collected by camera and then feed forward to the deep learning technique for training. These images were set to be compatible with the training images in terms of size as (200 x 200) before feeding forward for training.

**Deep learning algorithm used:** Deep learning is a branch of machine learning techniques used for image classifications. It is employed for training the dataset in figure 3 and also new features extracted from the testing dataset. The deep learning algorithm used is called convolutional neural network (CNN). CNN is a deep neural network class structured with neurons of learnable weights and biases which intelligently process large sets of data and make correct predictions.

**The Convolutional Neural Network Design Architecture**

The CNN was designed with input layer, convolutional layers, pooling layers, fully connected layers and the output layers to form a complex architecture for the deep learning process as shown in the architecture below in figure 4.
The model in figure 4 presents how the CNN collects and train data to classify driving behavior. The various layer and specifications of the CNN is discussed below;

**Input image layer**
The input layer is the first layer of the CNN architecture which collects the image frame extracted from the video data. This layer of the CNN identifies the image using the number of pixels and channels (RGB color size) which in this case is (200 x 200 x 3) where 3 presents the channel size and then feed forward to the convolutional layers via the neurons. The neuron weight is computed using the channels size and image resolutions as 12000. After the scaling process, the data were feed forward to the convolutional layer.

**Convolutional layer**
This part of the CNN was designed using multiple convolutions of neurons which in this case are five different convolutional layers aligned in series to each other to learn the features collected from the input layers by scanning and incrementally updating itself with the scanned features in each convolutional layer until the final layer which is the fifth convolutional layer. This scanning process was achieved using filters, which are function of the input data dimensioned, bias function (which is 1) and filter dimension which is (5 x 5).

The filter dimension is the receptive filed which the filter can scan in the query main image in one convolution and the output is the feature map, extracted in the 5 x 5 part of the drowsy data. During this scanning process, in the case where the filter does not fit in properly in the map image, padding was used to drop the part of the image not valid in the filter size. Rectified linear unit was used to introduce nonlinearity to the CNN to ensure that the output of each convolution process is non zero values.

This first feature map represents the result of one scanning process in the 5 x 5 part of the image. The scanning process continues with an increment of 1 unit until the filter scans the whole other parts of the images and then output result as an array of feature maps. The feature map is the output collected from each convolutional layer to another and pooled to the next layer (see discussion of pooling layer and the type used below). This scanning process continuous until the fifth convolutional layers by which the best part o the image data have been extracted.

**Pooling layer**
This is a down sampling operation which minimizes the spatial size of the feature maps to eliminate redundant spatial information, thereby increasing the filters in the deeper convolutional layers with computational delay. The pooling approach used is the maximum pooling technique. This selected only the part of the feature maps with the biggest pixel values from each array of filtered pixels during each convolution and feed forward as strides. The strides were feed forward to a fully connected neural network structure as discussed in the fully connected layer below;

**Fully connected layer**
The fully connected layer is responsible for flatten the matrix from the last convolutional layer and the feed to a neural network structure for classification in the output layer using softmax activation function. The neural network is presented in the diagram below.

The figure 5 presents the neural network structure of the fully connected layer, showing how the neurons of all the activation units in the proceeding layer are used to train the network. During this training process, the deep learning training tool automatically splits the feature vectors into test and training sets in the ratio of 30:70. The 70% are learned as already discussed in the convolutional layer and then the 30% test features are used to self examine the training accuracy using various epoch values until the best result is achieved and presented as result the
result is measured based on loss function which measures how well the training process occurred using the model below;

\[ L = \sum_{i=1}^{K} (P_i - D_i)^2 \]  

equation 1

Where \( L \) is the loss function, \( k \) is the number of observations, \( P \) is prediction and \( D \) is the training target. The model presented the loss between \( P \) and \( D \), where \( K \) is the number of classes; \( l \) is the output loss which is scalar.

**Output or prediction layer**

This is the final layer of the network which produces the desired output of the training process. This layer is designed using a softmax activation function in [22] which transforms the input feature vectors into probability distributions consisting of various probabilities proportional to the various exponential of the input video data captured from the drivers behavior.

V. IMPLEMENTATION

The system was implemented using deep learning network designer tool and MATHLAB. The deep learning toolbox was used for the training process while the MATHLAB was used to implement the algorithms as application software, using necessary toolbox like the image acquisition toolbox, computer vision toolbox, deep learning toolbox, neural network toolbox to implement the models designed.

VI. RESULTS

The results is presented in two sections which are result of the training processes measured using the loss function in equation 3.1 and the result of the prototype system when deployed for real time drowsy monitoring and detection.

6.1 Training Results

This section presents the result of the deep learning training process. The deep learning tool uses the loss function in equation 1 to measure the training performance of the system using the training parameters in table 2; and the results presented in figure 3;

**Table 2: Deep learning training parameters**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of input channels</td>
<td>3</td>
</tr>
<tr>
<td>Filter dimension</td>
<td>5 x 5</td>
</tr>
<tr>
<td>Image size</td>
<td>200 x 200</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Momentum factor</td>
<td>0.75</td>
</tr>
<tr>
<td>Gradient</td>
<td>4.79</td>
</tr>
<tr>
<td>Weight of neurons</td>
<td>1200</td>
</tr>
</tbody>
</table>

The result in figure 6, presents the training performance of the deep learning toolbox. The result was obtained when the training dataset was feed forward to the CNN architecture (see figure 4) and trained to achieve a convolutional drowsy intelligence in the fully connected layer which was used for the classification of future input video data from the camera when applied on real time driving conditions. From the result it was observed that the
training process iterates progressively and achieved the best training performance at epoch value of 30 at a learning rate of 0.001 with an accuracy of 99.8%. Having successfully trained and learned the necessary problem of drowsy driver under study, the training algorithm was used to build a prototype system and installed in a vehicle as shown below;

![System Installation](image)

**Figure 7: System installation**

In the figure 7, the camera connected to the laptop system which serves as the monitoring screen for the sake of this research demonstration was mounted on the dash board and focused to collect the driver behavior in real time. The performance of the software and camera was then tested when the driver drives and displays various behaviors as discussed in the next section.

6.2 Result on real time driving scenario

This section of the result present a real time application of the prototype system develop, using various drowsy features attributed to the data model designed in figure 3 (see section 4) as test features. The section also presents the result of the computer vision algorithm used for face tracking and detection. The results when deployed and tested are presented as shown below;

![Result for Eye Closed While Driving](image)

**Figure 8: Result for eye closed while driving**
In the result of figure 8, the system was tested when the driver is sleeping (eye closed for some seconds) and it was observed that the software was able to capture the driver’s behavior using computer vision controlled camera and then train to classify the drowsy status using the CNN architecture (see figure 4). In the CNN the softmax layer was used for the classification and the output layer was used to predict the behavioral status as a normal probability distribution function labeled as “Warning”. This warning label was used in this research as the control response for demonstration sake, however the researcher suggest in future implementation to use alarm, or cruise control applications which are more effective to alert the driver as a control method.

The next result presented in figure 9, shows the performance of the software when tested with the same driver while yawing. Yawing was choose here as a drowsy behavior because it was one of the attributes in the minor drowsy dataset as stated in the dataset (see table 1). During this act, the driver was monitored in real time and the data collected and processed is presented below;

From the result presented in figure 9, the software displays the data collected from the camera which was simultaneously trained and classified as shown in the display “Sensing Drowsy Behavior”. The implication of this result shows that the software was able to capture the driver’s facial behavior using compute vision and then feed forward for training which classified the data collected with the reference convolutional model and then predict as shown in the result displayed.

The next result was used to test the software with the driver displaying a perfect driving attribute. Recall from the dataset designed using the data model in figure 3, that a perfect driving condition is characterized with a driving state when the driver’s eye is open. The behavior was monitored as shown below;
From the result presented in figure 10, the driver behavior was tracked using the computer vision tool to read the facial features and then acquire the live data using the video acquisition camera. This data was feed forward to the CNN and trained for classification with the softmax function. The classified result was predicted in the output layer as a normal probability distribution and identified using the label “Driving Condition Perfect”.

VI. DISCUSSIONS

This work have presented the result for the training and testing the new system. The training result was performed using the data collected from the training dataset to generate a drowsy model which was used by the software for the classification of future drowsy behavior. The result shows that the system when tested with drowsy driver features and perfect driving features was able to train and predict the correct behavioral status in real time with an accuracy of 99.8%. The result is compared with the existing system studied in the literature review and the analyzed as shown below;

From the result in figure 11, it was observed that the new system performs better than the existing deep learning technique with an improved accuracy of 18.8%.

VIII. CONCLUSIONS

This paper has successfully developed an intelligent drowsy driver detection system based on behavioral approach. The research reviewed various literatures drowsy driving detection systems, identifying the technical challenges, contributions and setbacks. From the literature it was observed that the drowsy driver detection approaches are classified into behavioral, physiological and vehicle based approaches. But among all these approaches, the behavioral is the current trend, due to its capacity to comprehend the problem of drowsiness directly from the source (driver) without interruption of the driving process.

In the behavioral approaches various techniques like machine learning, image processing were all used to solve drowsy problem, however despite their success, suffers false alarm, poor dataset design, unreliability, among others. Nevertheless, the application of ML algorithms such as DL in the literatures revealed that this technique if properly designed will detect drowsy driving in real time. There are many types of DL algorithms but CNN was choose due to its operating success in processing large image or video base data, when compared to other DL algorithms. The CNN was designed using the appropriate number of layers and design specifications for each training parameters and then trained with a training dataset of 117000 frames characterized with various drowsy features using deep learning tool. The system was implemented using MATHLAB and tested on a real time driving scenario. The result showed a detection accuracy of 99.8%. The result was also compared with the existing system and an improved accuracy of 18.8% was recorded for the new system.

REFERENCE

[5]. H. Parker “Distracted Driver” 3 seconds is all it takes for accident to happen; 2019.


[10]. B.K. Savas, & Y. Becerikli, Real Time Driver Fatigue Detection Based on SVM Algorithm. 6th International Conference on Control Engineering & Information Technology (CEIT), 2018. doi:10.1109/ceit.8751886.


[13]. Manishi and N. Kumari ”a comprehensive study on behavioral parameters based drowsy detection techniques; international journal of computer science and engineering; Vol 8, issue 4, ISSN, 2347 – 2693. 2020.


AUTHORS PROFILES

Kekong is currently a postgraduate student at the Department of Computer Science, Ebonyi State University, Abakaliki, Nigeria.

Dr. Ajah is a Head of Department (HOD); Computer Science, Ebonyi State University, Abakaliki, Nigeria.

Chidiebere is a member, Desnet Research Group, Enugu, Nigeria.