

An Intelligent Deep-Learning based Drunk Driving Detection System

AJITH GEE VARGHESE ¹, BIBIN BINOY ¹, DEEPANSHU TARAGI ¹, Dr. C SAHAYA KINGSLEY ²

¹Department of Computer Science, Mangalam College of Engineering, Ettumanoor, Kerala

²Associate Professor, Department of Computer Science, Mangalam College of Engineering, Ettumanoor, Kerala

Abstract — Drunk Driving accidents have been increasing rapidly, every year many people suffer due to this problem. Even though police use a breath-alcohol tester for the detection of drunk driving state of a person, but due to the residue left on the mouthpiece it may not be accurate and isn't an hygienic method. The study proposes a two-stage neural network for recognition of drunk driving. The first stage uses the simplified VGG network to determine the age range of the subject. The second stage uses the simplified Dense-Net to identify the facial features of drunk driving.

Keywords— *Machine Learning; Drunk Driving Detection; VGG-16; DenseNet, Embedded System*

I. INTRODUCTION

Driving under the influence of alcohol is a significant problem worldwide, resulting in countless accidents, injuries, and fatalities each year. Despite efforts to increase public awareness and strengthen laws and regulations surrounding drunk driving, this dangerous behavior continues to be a serious public safety concern. In order to address this issue, there has been growing interest in developing more reliable and accurate methods for detecting signs of intoxication in drivers.

Breath alcohol detectors, commonly used by police officers to detect signs of intoxication in drivers, have been found to pose potential hygiene risks. The devices are typically designed to be used multiple times by different individuals, and the mouthpiece can become contaminated with saliva, bacteria, and other substances over time. This can lead to the spread of infectious diseases, and may also cause discomfort or reluctance among individuals who are required to use the device.

One approach to detecting drunk driving that has shown promise in recent years is the use of deep learning models. Deep learning is a subfield of machine learning that utilizes neural networks to automatically learn representations of data, and has been applied successfully to a wide range of tasks in computer vision, natural language processing, and other domains. In the context of drunk driving detection, deep learning models can be trained to recognize subtle changes in facial expression, speech patterns, and other features that are indicative of intoxication.

In this research paper, we propose a novel approach to drunk driving detection using deep learning models, specifically VGG for age detection and DenseNet for detecting signs of intoxication. VGG is a popular deep convolutional neural network architecture that has been

widely used for image classification tasks, while DenseNet is a more recent architecture that is designed to improve the flow of information between layers in the network.

Despite the potential benefits of deep learning models for detecting drunk driving, there are several challenges that must be addressed in order to develop a reliable and effective system. These challenges include variations in the ways that different people show signs of intoxication, the need to process large amounts of data quickly, and the ethical considerations surrounding the use of facial recognition technology for public safety purposes.

Nevertheless, the potential benefits of a more reliable and accurate system for detecting drunk driving are substantial, both in terms of reducing the number of accidents, injuries, and fatalities caused by this dangerous behavior, and in promoting responsible behavior and accountability among drivers. This research paper represents a step forward in the development of such a system, and contributes to the growing body of work aimed at leveraging the power of deep learning to improve public safety and the well-being of communities around the world.

II. RELATED WORK

A. Deep expectation of real and apparent age from a single image without facial landmarks (R. Rothe et. al, 2018) [2]

Age estimation from a single face image is an important task in human and computer vision which has many applications such as in forensics or social media. It is closely related to the prediction of other biometrics and facial attributes tasks such as gender, ethnicity, hair color and expressions. A large amount of research has been devoted to age estimation from a face image under its most known form - the real, biological, age estimation. This paper proposes a deep learning solution to age estimation from a single face image without the use of facial landmarks and introduce the IMDB-WIKI dataset, the largest public dataset of face images with age and gender labels. If the real age estimation research spans over decades, the study of apparent age estimation or the age as perceived by other humans from a face image is a recent endeavor. We tackle

both tasks with our convolutional neural networks (CNNs) of VGG-16 architecture which are pretrained on ImageNet for image classification. We pose the age estimation problem as a deep classification problem followed by a softmax expected value refinement.

The key factors of our solution are:

deep learned models from large data, robust face alignment, and expected value formulation for age regression. We validate our methods on standard benchmarks and achieve state of the art results for both real and apparent age estimation. We mainly use a convolutional neural network (CNN) to predict the age of a person starting from a single input face image. This takes an aligned face with context as input and returns a prediction for the age. The CNN is trained on face images with known age.

B. *Imagenet Classification with Deep Convolutional Neural Networks (A. Krizhevsky et. al, 2012) [3]*

Current approaches to object recognition make essential use of machine learning methods. To improve their performance, we can collect larger datasets, learn more powerful models, and use better techniques for preventing overfitting. This paper describes about a deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry. ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories. The images were collected from the web and labeled by human labelers using Amazon's Mechanical Turk crowdsourcing tool. Starting in 2010, as part of the Pascal Visual Object Challenge, an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. ImageNet consists of variable-resolution images, while our system requires a constant input dimensionality. Therefore, we down-sampled the images to a fixed resolution of 256×256 . Given a rectangular image, we first rescaled the image such that the shorter side was of length 256, and then cropped out the central 256×256 patch from the resulting image. We did not pre-process the images in any other way, except for subtracting the mean activity over

the training set from each pixel. So we trained our network on the (centered) raw RGB values of the pixels.

C. *Drunk Person Identification Using Thermal Infrared Images (G. Koukiou et. al, 2015) [4]*

Biometrics is a hot research area with numerous publications the last few years. Applications are met in medicine, financial transactions, person identification and mainly in security issues. Research has been carried out (Jain et al., 1999) in several biometric problems, such as face and fingerprint recognition, facial expression classification and iris identification with high rate of success. Drunk person identification is carried out using thermal infrared images. Two different approaches are proposed for distinguishing a drunk person by means of radiometric values on its face. The features used in the first approach are simply the pixel values of specific points on the face of the person. It is proved that the cluster of a specific person moves in the feature space as the person consumes alcohol. Fisher linear discriminant approach is used for space dimensionality reduction. The feature space is found to be of very low dimensionality. The majority of the clusters moves towards the same direction and the feature space can easily be separated into the 'sober' and 'drunk' regions. Thus the 'drunk' feature space is introduced. In the second approach thermal differences between various locations on the face are evaluated and their value is monitored. It was found that specific areas in the face of a drunk person present an increased thermal illumination. These areas are the best candidates for identifying a drunk person. The concept behind this second proposed approach relies on a physiology-based face identification procedure. In this paper, we present two different approaches to the problem of identifying drunk people using thermal images. The first approach was based on the fact that the representation of a specific person into the feature space (cluster of feature vectors) moves as the person consumes alcohol. The feature vector is formed by considering the values of 20 specific pixels on the face (Koukiou et al., 2009). The feature space is easily reduced to two dimensions and it is proved that the clusters of the persons move towards the same direction with alcohol consumption. The feature space is separated into two regions, i.e., the one containing the clusters of the sober and that with the clusters of the drunk persons. Accordingly, this feature space is called 'the drunken space'.

The second approach was based on the fact that the thermal differences between specific locations on the face increase as the person consumes alcohol.

D. *Experiments on detection of drinking from face images using neural networks (K. Takahashi et. al, 2015) [5]*

Drunk driving is a severe problem in most countries around the world, and the governments of all countries have established punishments and methods to prevent drunk driving activities, including increasing penalties for driving under the influence, joint penalties for passengers, revocation of driving license, and imprisonment, among others.

According to national statistics, although there is a downward trend in incidents of drunk driving accidents year on year, the results are still limited, and the suggested measured have not completely discouraged drunk driving. Breath-alcohol meters are still mainly used as the means to detect drunk driving status; however, there are several problems with such an exhalants detection method, including the high price of the instrument, the mouthpiece is a consumable, which is not environmentally friendly, hygiene concerns, and inconvenient usage. Further, misjudgements are easily possible with the results owing to residual moisture in the instrument under continuous use conditions. This paper proposes a method to detect drinking state of a driver by using a camera, in order to prevent drunken driving. The method detects whether a driver has drunken alcohol or not by using face images and their parts captured by a camera. In experiments, a 3-layered neural network is employed to detect drinking and determine the parts of a face useful for detection. Through experiments the accuracy of the method is shown and is compared with that by the decision tree C4.5. It is also shown that parts, cheek, neck, and hand are useful to detect drinking.

E. A Precise Drunk Driving Detection Using Weighted Kernel Based On Electrocardiogram (C. Wu et. al, 2016) [6]

Globally, 1.2 million people die and 50 million people are injured annually due to traffic accidents. These traffic accidents cost \$500 billion dollars. Drunk drivers are found in 40% of the traffic crashes. Existing drunk driving detection (DDD) systems do not provide accurate detection and pre-warning concurrently. Electrocardiogram (ECG) is a proven biosignal that accurately and simultaneously reflects human’s biological status. In this letter, a classifier for DDD based on ECG is investigated in an attempt to reduce traffic accidents caused by drunk drivers. At this point, it appears that there is no known research or literature found on ECG classifier for DDD. To identify drunk syndromes, the ECG signals from drunk drivers are studied and analyzed. As such, a precise ECG-based DDD (ECG-DDD) using a weighted kernel is developed.

From the measurements, 10 key features of ECG signals were identified. To incorporate the important features, the feature vectors are weighted in the customization of kernel functions. Four commonly adopted kernel functions are studied. Results reveal that weighted feature vectors improve the accuracy by 11% compared to the computation using the prime kernel. Evaluation shows that ECG-DDD improved the accuracy by 8% to 18% compared to prevailing methods. Drunk driving detection (DDD) is an efficient way to reduce the tragedies and the expenditures due to traffic accidents. In each year, more than 1.2 million people die and more than 50 million people are injured due to traffic accidents. The expenditures on these traffic accidents account for 1% to 3% of the world’s GDP, and are more than \$500 billion dollars. The World Health Organization (WHO) cautioned that traffic accidents will become the fifth leading cause of death.

Among all traffic crashes, about 40% of the crashes are due to drunk driving which also counts about 22% of the

traffic injury costs. WHO concludes that 70% of the world’s population can be protected if strong restrictions are imposed on drunk drivers. The development of DDD is vital because it will help to reduce traffic accidents and thus reduce expenditure accordingly. DDD provides an early protection to both drivers and pedestrians by alerting the drunk drivers when the drunk status is detected. DDD can be categorized into three types: direct detection (which is also known as blood alcohol test), drivers’ behavior-based detection and biosignal-based detections. The first two types cannot meet the requirements of pre-warning, fully automated and high accuracy simultaneously while the third type suffices. The plethysmogram signal, which measures changes in an organ’s volume and thus resulting in a fluctuation in blood or air content, was used. However, this detection is time consuming and the accuracy has not been specified. In addition to plethysmogram, other biosignals such as Electrocardiogram (ECG) and Electroencephalography (EEG) may also accurately reflect the human’s status.

III. METHODOLOGY

This study suggests a two-stage deep neural network for identification. The first step uses a basic VGG deep neural network to identify the age group, and the second stage uses a simplified Dense-Net deep neural network to further refine that identification. It also uses a embedded system which will send a SMS to specific number stored in the device. The SMS will contain the current location and the state of the person.

Stage 1: A portion of the photos used to train the network are taken from the IMDB-WIKI dataset [1], and the remainder data were uniquely gathered for this purpose. The necessary training images are acquired after the age image preprocessing stage in Fig. 1 and additional data is produced by data augmentation to form the age dataset. Finally, the age discrimination component of the network is trained and classified using the condensed VGG. As given in fig. 1.

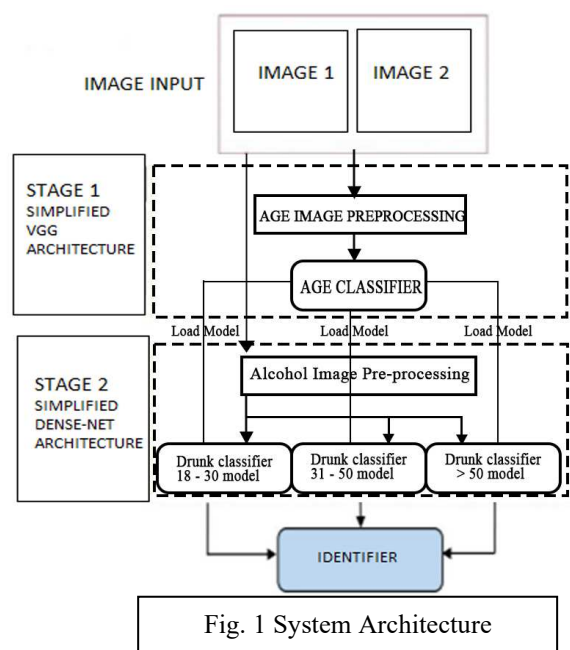


Fig. 1 System Architecture

Stage 2: Here we extract the facial image ROI through some image preprocessing methods, than using data augmentation a diverse data set is generated. The test data generated are classified for age-group 18-30, 31-50, >51.[1]

A. IMAGE PREPROCESSING :-

The first stage involves the steps before age discrimination, where the training data includes images from the IMDB-WIKI dataset as well as self-collected images. To ensure that the properties of the two data are similar, five pre-processing steps are required: face detection, face calibration, histogram equalization, median filtering, and grayscale. The second stage is the preprocessing of the alcohol test results identification. The pre-processing consists of two steps, namely video frame extraction and face detection.

B. DATA AUGMENTATION :-

To increase the diversity and amount of original and overall data to avoid over fitting during training, it is necessary to perform data augmentation. This work uses the data augmentation method for the following actions: random rotation 5–10°, random enlargement of 5–10%, random left and right flips, and random cropping of 5–10%.

C. DATA SET :-

The age dataset of the first stage of this work uses partial images from the IMDB-WIKI dataset and self-collected image data, with the age dataset being divided into three categories, namely 18–30 years, 31–50 years, and 51 years. The original training set has data from a total of 189 people aged 18–30 and includes 322 pictures, 281 people aged 31–50 with 313 pictures, and 262 people 51 years with 541 pictures. Using data augmentation the total number of images used for training was 19,072, and a total of 1,664 images were used for validation.

TABLE 1. Dataset

	AGE	DATA AUGMENTATION	TOTAL FRAME	
TRAINING SET	Below 21	9122	9345	40,000
	21 – 29	11,145	11,254	
	30 – 45	9455	9574	
	45 Above	9632	9827	
VALIDATION SET	Below 21	2342	2445	10,000
	21 – 29	3411	3542	
	30 – 45	2312	2441	
	45 Above	1485	1572	

The second stage of the drunk driving recognition uses self- recorded data, including images from 124 subjects: 77 people aged 18–30, 25 people aged 31–50, and 22 people aged 51 years, of which 110 subjects are used for training

and validation, and 14 subjects for testing. Each recorded movie is on average 120 seconds in length, including nondrinking (BrAC = 0), drinking within standard (0<BrAC<0.15), drinking exceeding standard (0.15≤BrAC<0.25), and drinking severely exceeding standard (BrAC ≥ 0.25).

D. EMBEDDED SYSTEM

Alcoholic Input Sensor :- It is the output of the alcohol detection, which is given as the input in Arduino nano.

Arduino Nano :- It is used for communicating with other microprocessor. It lacks a DC supply. It has a Mini-B USB cable. It has almost same function as Arduino Duemilanove but uses different packages.

GSM :- The Global System for Mobile communication or GSM will store the number of the user who will get the notification regarding the position of the drunk driver.

GPS :- It will send the location of the drunk driver to the person whose details are stored in GSM.

RELAY :- It is used to provide independent low power signals, or to control several devices with one signal.

Buzzer LED indicator :- The Buzzer LED indicates turns red, yellow or green based on the input from Alcohol Input Sensor.

E. ALGORITHM :-

- 1) The user’s video will be recorded using the camera.
- 2) The frames will be extracted from the video.
- 3) Using MTCNN we detect the face in the image frame.
- 4) Extract the coordinates of the bounding box around each detected face.
- 5) Calculate the center point of each bounding box using the coordinates of the top-left and bottom-right corners.
- 6) Determine the area of interest by applying the distance formula to find the distance between the center point of each bounding box and the center of the image.
 - a) $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
- 7) Crop the image to the area of interest.
- 8) To increase the diversity and amount of original and overall data to avoid over fitting during training, it is necessary to perform data augmentation.
 - a) Random rotation ±5–10°
 - b) Random enlargement of 5–10%
 - c) Random left and right flips
 - d) Random cropping of 5–10%
- 9) Resize the image to the appropriate input size required by the VGG 16 model (224 x 224 pixels).
- 10) Load the pre-trained VGG 16 model and use it to extract features from the preprocessed image.
- 11) Load the pre-trained DenseNet model that has been trained to classify drunk or sober states using images of faces.
- 12) Feed the preprocessed image into the model for prediction.

- 13) Obtain the predictions for the image from the output layer of the model.
- 14) Calculate the sensor value, If the prediction for drunk state is greater than the threshold value classify the person as drunk. Otherwise, classify them as sober
 - a) $sensorValue = analogRead(SENSOR_PIN);$
 - b) $voltage = sensorValue * (5.0 / 1023.0);$
 - c) $Alcohol = (voltage * 100) - 35; .$
- 15) Display the classification result as output.

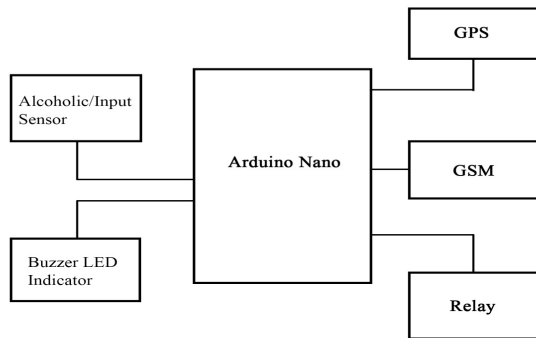


Fig. 2 Embedded System

IV. RESULT

Two deep neural networks were combined in this study for a two-stage identification system. The system determines the age in the first stage and loads the corresponding drunk driving model to the second stage to perform the final alcohol test evaluation. The Result analysis is given in fig 3.

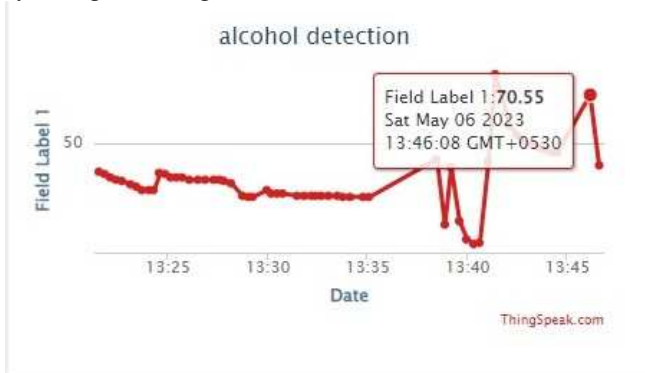


Fig. 3 Result Analysis

Comparing the results of this study with those from literature, it is observed that it uses self-collected images, with a total of 16 training and 24 testing frames, including regional images of cheeks, chin, neck, ears, and hands, and applying multilayer preceptor.

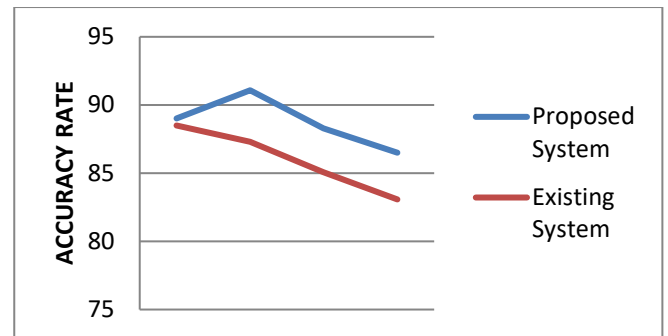


Fig. 4 Analysis Comparison

This proves that the proposed system has high accuracy and practicality.

V. FUTURE SCOPE

Although the results of this paper are better, the accuracy needs to be further improved in the future. A challenging road is still ahead for a real-world application due to the challenges of geographic, gender, and age generalization. Furthermore different deep-learning algorithms can be used in future to increase the accuracy of the overall system.

VI. CONCLUSION

In this paper, artificial intelligence is used to detect intoxication without physical touch via facial picture analysis. The suggested solution uses a two-stage deep neural network, with the first stage using a simplified VGG to assess the subject's age range and the second stage using a simplified Dense-Net to recognize the facial traits of drunk driving for alcohol test identification. The IMDB-WIKI dataset and self-collected information are used to conduct age discrimination. The experimental results (fig. 4) show that the overall alcohol test accuracy rates are above 90%. This proves that the proposed system has high accuracy and practicality.

VII. ACKNOWLEDGMENT

The authors wish to thank Principal Dr. Vinodh P Vijayan, Neethu Maria John, H.O.D. Computer Science Department, for the proper guidance, valuable support, and helpful comments during the proofreading.

REFERENCES

- [1] R. C. -H. Chang, C. -Y. Wang, H. -H. Li and C. -D. Chiu, "Drunk Driving Detection Using Two-Stage Deep Neural Network," in IEEE Access, vol. 9, pp. 116564-116571, 2021, doi: 10.1109/ACCESS.2021.3106170.
- [2] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," Int. J. Comput. Vis., vol. 126, nos. 2-4, pp. 144-157, Apr. 2018.
- [3] C. Wu, K. Tsang, H. Chi, and F. Hung, "A precise drunk driving detection using weighted kernel based on

electrocardiogram,” *Sensors*, vol. 16, no. 5, p. 659, May 2016.

[4] K. Takahashi, K. Hiramatsu, and M. Tetsuishi, “Experiments on detection of drinking from face images using neural networks,” in *Proc. 2nd Int. Conf. Soft Comput. Mach. Intell. (ISCFMI)*, Nov. 2015, pp. 97–101.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

[6] G. Koukiou and V. Anastassopoulos, “Drunk person identification using thermal infrared images,” *Int. J. Electron. Secur. Digit. Forensics*, vol. 4, no. 4, pp. 229–243, 2012.