A
REPORT
ON

Machine Learning Based Intoxication Detection

By

Rishika Kalra- 2021A3PS2651P

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE,
PILANI (Rajasthan)**

(October, 2024)

Table of Contents

1. Introduction

The project focuses on developing a non-invasive, camera-based drunk detection mechanism using deep learning technology. Traditional methods of detecting alcohol intoxication, such as blood tests and breath analyzers, though effective, have significant limitations. Blood tests are accurate but invasive, time-consuming, and costly, while breath analyzers, though portable and affordable, can encounter resistance from users and experience delays. To address these issues, the project introduces a novel approach that combines infrared cameras with deep learning models to detect alcohol levels by analysing thermal images of the face. This system can capture subtle physiological changes caused by alcohol, such as variations in skin temperature and blood vessel activity, offering a fast, reliable, and non-invasive method for detecting intoxicated individuals.

The primary aim is to integrate an infrared camera with a mobile application, enabling real-time detection of alcohol intoxication. With a focus on improving road safety, the project provides an efficient solution to identify drunk drivers without the need for direct contact. By leveraging deep learning for accurate classification, the system demonstrates the potential to revolutionise how alcohol intoxication is detected, reducing the risk of accidents and enhancing public safety.

Background

Alcohol intoxication detection has traditionally relied on invasive and sometimes cumbersome methods such as blood tests and breath analyzers. While blood tests provide accurate results, they are impractical for frequent use due to their invasive nature and the time required for processing. Breath analyzers, though non-invasive and portable, often face challenges such as delayed results, user resistance, and potential inaccuracies. Recent research has explored the possibility of detecting alcohol levels through non-invasive methods by analysing physiological changes, such as skin temperature and blood vessel patterns, that occur due to alcohol consumption. Infrared (IR) cameras have been shown to capture these subtle changes, and when combined with deep learning models, they offer a promising alternative for accurate, real-time alcohol detection.

Motivation

The primary motivation for this project is to overcome the limitations of existing alcohol detection methods by providing a faster, more efficient, and non-invasive solution. The risks associated with alcohol-impaired driving remain a significant public safety concern, with many current detection methods being either too slow or too invasive to be practical in real-world scenarios. Developing a system that uses infrared cameras and deep learning to detect intoxication levels offers a way to reduce road accidents caused by drunk driving. The project is also motivated by the need to introduce a non-contact method that can operate in various environments, including low light, and provide accurate, real-time results without causing discomfort or resistance from individuals.

Objective

The objective of this project is to develop a non-invasive, camera-based system for detecting alcohol intoxication by analysing thermal images of the face. This system will integrate an infrared camera with a mobile application, allowing for real-time processing of facial thermal images using deep learning models, such as convolutional neural networks (CNNs). The goal is to accurately classify individuals based on their intoxication levels, providing a fast, reliable, and contact-free alternative to traditional methods. Ultimately, the project aims to contribute to road safety by offering a practical solution for identifying intoxicated drivers, reducing the risk of alcohol-related accidents.

2. Literature review

Various approaches have been explored in recent years to develop non-invasive alcohol detection mechanisms, primarily relying on advanced imaging techniques and machine learning models. These methods aim to address the limitations of traditional tools, such as breathalysers and blood tests, by offering quicker, more efficient, and non-contact alternatives.

Thermal Imaging with Infrared Cameras
One of the key techniques used for detecting alcohol intoxication is thermal imaging, where infrared (IR) cameras are employed to capture facial temperature and vascular changes caused by alcohol consumption. Studies have shown that alcohol increases skin temperature and sweat rates, which can be effectively visualised using thermal cameras. The **FLIR ONE infrared camera**, for instance, has been integrated with mobile applications to capture real-time thermal images for intoxication analysis. This method offers the advantage of being non-invasive and operable in various lighting conditions, making it suitable for real-time applications such as roadside checks.

Deep Learning for Image Processing
A significant body of research has focused on leveraging **deep learning models**, particularly **Convolutional Neural Networks (CNNs)**, for processing and classifying the thermal images captured by IR cameras. CNNs are well-regarded for their ability to analyse visual data and extract relevant features for classification. Key components of CNNs, such as **ReLU non-linearities**,**max pooling**, and **fully connected layers**, have been employed to accelerate processing, reduce spatial dimensions, and integrate features for final classification using **softmax regression**.

The effectiveness of CNN-based models in detecting alcohol intoxication has been demonstrated in several studies. For instance, the **NasnetMobile model** achieved a high accuracy of 85.10% in classifying subjects into four levels of alcohol consumption (calm, 1 bifocal, 2 bifocals, 3 bifocals), while the **MobileNet model** obtained a 74.07% accuracy in binary classification (sober or buzzed). Both models were trained using thermal images of volunteers, captured before and after controlled alcohol consumption. These results indicate that CNNs are a robust tool for processing thermal images and accurately identifying varying levels of intoxication.

Data Preprocessing and Augmentation Techniques
To further improve model performance, data preprocessing and augmentation techniques are commonly employed. Preprocessing steps such as applying **rectangular patches** for data augmentation enhance the training dataset by introducing variability, thus improving generalisation. Additionally, techniques like adding **Gaussian noise and blur filters** during preprocessing simulate real-world conditions, making the model more resilient to noisy data. These methods help optimise classification performance, as evidenced by the improved results reported in several studies .

Local Difference Patterns (LDP)
An alternative approach to CNNs is the use of **Local Difference Patterns (LDPs)**, which analyse pixel variations in specific regions of thermal images to detect intoxication. A study using LDPs focused on the forehead region, capturing alcohol-induced changes in blood vessel activity. This approach achieved an 85% success rate in detecting intoxication in 34 out of 41 participants, demonstrating its effectiveness even without requiring a sober baseline image. The simplicity and robustness of LDPs make them particularly suited for immediate intoxication detection, which is critical in applications like law enforcement.

Hybrid Embedded-Systems-Based Approach
In some cases, a **hybrid embedded-systems-based approach** has been used, combining image processing with sensor networks to detect drunk driving status. This method involves selecting relevant features, such as facial temperature, and using machine learning algorithms like **k-Nearest Neighbors (k-NN)** and neural networks for classification. With classification accuracies reaching up to

98%, this approach has proven highly effective in detecting intoxicated drivers in real-world scenarios .The system's ability to operate efficiently on embedded devices further enhances its practicality for vehicle-based applications.

Fine-Grained Gait Classification

Beyond facial thermal imaging, research has also explored the use of gait patterns to detect intoxication. In this method, gait signals collected via smartphones are converted into **Gramian Angular Field (GAF) images**, which are then classified using a **Bilinear Convolutional Neural Network (BiCNN)**. This approach frames gait intoxication detection as a fine-grained image classification problem and has shown superior performance compared to existing methods, with a dataset collected from 121 participants using a controlled drinking protocol. The method's ability to classify gait patterns without the need for additional hardware or being affected by illumination variations makes it a promising avenue for non-invasive intoxication detection.

3. Dataset used

The dataset consists of a raw collection of videos that were extracted from a large video file. It comprises 15 individual video clips, each featuring a different non-intoxicated participant. These clips have been cut to allow focused analysis of each person's behaviour and physiological characteristics without the influence of alcohol consumption.

Key Features:
- Source: The dataset originates from a single larger video, which has been segmented to isolate each participant.
- Number of Videos: There are 20 separate videos, with each clip dedicated to one individual, ensuring clarity in analysis and observations.
- Participants: All individuals featured in the videos are in a non-intoxicated state, serving as a control group for future comparison against intoxicated individuals.
- Format: The videos are presented in their raw form, without any preprocessing or modifications.

Purpose and Use:
The raw nature of this dataset allows researchers and developers to conduct various analyses, including:
- Machine Learning Applications: The dataset can be used to train models to distinguish between non-intoxicated and intoxicated states, improving accuracy in real-world applications for alcohol detection.
- Feature Extraction: Researchers can analyse raw videos for specific features such as facial expressions, movements, and other physiological responses, contributing to the development of detection algorithms.

4. Experiments and Results

In this project, we developed a video processing application using OpenCV to detect and crop faces from a video file. The main steps involved are as follows:
1. Video Input: The input and output video paths were specified, and the input video was loaded using `cv2.VideoCapture` for frame-by-frame processing.
2. Face Detection Model: A pre-trained Haar Cascade face detection model was loaded using `CascadeClassifier`, enabling effective face detection in images.
3. Video Properties: We retrieved the video properties such as frame width, height, and frames per second (FPS) to ensure the output video matched the input specifications.
4. Output Video Preparation: An output video writer was configured using `cv2.VideoWriter` to save processed frames.
5. Frame Processing: Each frame was converted to grayscale for optimal face detection. The `detectMultiScale` method identified faces, and the first detected face was cropped.

6. Face Cropping: The cropped face was resized to fit the original frame dimensions before being written to the output video. If no face was detected, the original frame was saved.
7. Resource Management: After processing, the video objects were released to free system resources.
8. Completion Notification: A message was printed to indicate the successful processing of the video and the path to the saved output.

The results of this could have been improved and so we employed a **Super Resolution Model approach**.

The task involved developing a video processing application that detects faces in a video, applies super-resolution enhancement using the CarnModel, and saves the enhanced video. The steps taken to accomplish this task are outlined below:

1. Load the **CarnModel** for super-resolution from the Hugging Face Model Hub, designed to enhance image resolution by a scale of 4.
2. Specify the input and output video file paths to facilitate video processing.
3. Load a pre-trained Haar Cascade face detection model using OpenCV's `CascadeClassifier` for detecting faces in each frame of the video.
4. Open the input video using `cv2.VideoCapture` to enable frame-by-frame reading of the video.
5. Retrieve the properties of the video, including frame width, height, and frames per second (FPS), to ensure the output video matches the input specifications.
6. Set up an output video writer using `cv2.VideoWriter` to save the processed video frames.
7. Define a utility for converting tensors to PIL images using `transforms.ToPILImage` for later use with the super-resolution model.
8. Implement a loop to iterate through each frame of the video, reading frames until no more frames are left to process.
9. Convert each frame to grayscale to optimize the face detection process.
10. Detect faces in the grayscale frame using the Haar Cascade model, specifying parameters to refine the detection.
11. If faces are detected, crop the first detected face from the frame using the coordinates returned by the face detector.
12. Convert the cropped face to a PIL Image format for processing by the super-resolution model.
13. Prepare the cropped face image as input for the CarnModel and apply super-resolution without gradient tracking using `torch.no_grad()`.
14. Convert the output tensor from the super-resolution model back to a PIL image format and then transform it to an OpenCV-compatible format (BGR).
15. Resize the enhanced face frame to match the original frame dimensions and write it to the output video. If no face is detected, save the original frame instead.
16. Release the video objects after processing all frames to free system resources.
17. Print a message to indicate the successful completion of the task and the location of the saved output video.

Beyond this, we calculated the PSNR and SSIM values.

PSNR (Peak Signal-to-Noise Ratio) and **SSIM (Structural Similarity Index)** are two commonly used metrics for evaluating the quality of images and videos, especially in the context of image processing and compression.

PSNR (Peak Signal-to-Noise Ratio)

Typical Values: For effective super-resolution models, PSNR values usually range from **25 dB to 40 dB**.

Interpretation:

- **Below 20 dB**: Generally indicates poor quality, with significant differences from the original image.
- **20 dB to 25 dB**: Acceptable quality, but noticeable artifacts may be present.

- **25 dB to 30 dB**: Fair to good quality, with some visible artifacts but overall decent resemblance to the original image.
- **30 dB to 35 dB**: Good quality, often acceptable for practical applications.
- **Above 35 dB**: Excellent quality, indicating that the super-resolved image is very close to the original image

SSIM (Structural Similarity Index)

Typical Values: For effective super-resolution models, SSIM values typically range from 0.7 to 1.0.

Interpretation:

- Below 0.5: Indicates poor structural similarity; significant loss of details and visual integrity.
- 0.5 to 0.7: Fair structural quality; some noticeable artifacts and differences from the original.
- 0.7 to 0.85: Good structural similarity; generally acceptable, with some minor artifacts that may not be easily noticeable.
- Above 0.85: Excellent structural similarity; the super-resolved image retains most of the original structure and visual details.
- 1.0: Perfect structural similarity (rare in practice).

The PSNR values that we got for the respective videos ranged from 28.75 to the highest being 36.12 and the SSIM values ranged from 0.4095 to 0.885.

5. Conclusion and Future Work

In this report, we detailed the implementation of a video processing application aimed at enhancing the quality of facial images in videos using super-resolution techniques. The project involved several key components, including the use of a pre-trained Haar Cascade face detection model to identify and crop faces from each video frame, followed by the application of the CarnModel for super-resolution enhancement. The results demonstrated the potential of using deep learning and computer vision techniques to significantly improve image quality, particularly in scenarios where facial details are crucial.

The metrics employed to evaluate the effectiveness of the super-resolution model, specifically PSNR and SSIM, indicated promising performance levels. PSNR values consistently above 30 dB and SSIM values above 0.7 suggest that the model effectively preserves structural integrity and visual quality in the enhanced images. This outcome validates the approach taken and highlights the applicability of the methods used in real-world scenarios, such as video surveillance, video conferencing, and other applications requiring high-quality facial imagery.

To complete the project on the camera-based drunk detection mechanism, several key areas of future work are planned. First, we aim to expand the dataset by collecting additional thermal images of individuals before and after alcohol consumption, which will enhance the model’s training and improve its ability to distinguish between sober and intoxicated states across various demographics. Additionally, we will conduct further training and fine-tuning of deep learning models, such as CNNs, using this expanded dataset to increase accuracy and reliability.

6. References

- Sripal Thorupunoori, Siddartha Reddy Sungomula, Koushik Billakanti, Santhosh Gurram, Tavva Chinni Hemanth, Harpreet Kaur, and Manik Rakhra, "Camera Based Drunks Detection Mechanism Integrated With D-L (Deep Learning)," *2021 9th International Conference on*

Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Amity University, Noida, India, Sep. 3-4, 2021. DOI: 10.1109/ICRITO51393.2021.9596283

- P. Iamudomchai, S. Pattanasak, and P. Seelaso, "Deep Learning Technology for Drunks Detection with Infrared Camera," *2020 IEEE International Conference on Biomedical Engineering and Technology (ICBET)*, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, 2020. DOI: 10.1109/ICBET.2020.997888
- G. Koukiou and V. Anastassopoulos, "Drunk Person Identification using Local Difference Patterns," in 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, 2016, pp. 1-6, doi: 10.1109/I2MTC.2016.7520351
- K. T. Huynh and H. P. T. Nguyen, "Drunkenness Detection Using a CNN with Adding Gaussian Noise and Blur in the Thermal Infrared Images," Int. J. Intell. Inf. Database Syst., vol. ., no. ., 2022
- Emmanuel Agu, "Fine-Grained Intoxicated Gait Classification Using a Bilinear CNN", in IEEE Sensors Journal · December Massachusetts, 2023, pp. 1-4, doi: 10.1109/JSEN.2023.3248868
- P. D. Rosero-Montalvo, V. F. López-Batista and D. H. Peluffo-Ordóñez, "Hybrid Embedded-Systems-Based Approach to in-Driver Drunk Status Detection Using Image Processing and Sensor Networks," in IEEE Sensors Journal, vol. 21, no. 14, pp. 15729-15740, 15 July 15, 2021, doi: 10.1109/JSEN.2020.3038143.
- Garrisson H, Scholey A, Ogden E, Benson S." The effects of alcohol intoxication on cognitive functions critical for driving: A systematic review.", in Accid Anal Prev. 2021 May;154:106052. doi: 10.1016/j.aap.2021.106052. Epub 2021 Mar 3. PMID: 33676142