# Neural Networks for Intoxication Detection


**Julian Burgos**
Department of Computer Science
University of Maryland
College Park, MD
jburgos@terpmail.umd.edu

**Joseph Jubilee**
Department of Computer Science
University of Maryland
College Park, MD
jjubi133@terpmail.umd.edu

**Matthewos Gashaw**
Department of Computer Science
University of Maryland
College Park, MD
mattgashaw2@gmail.com

**Olade Kounou**
Department of Computer Science
University of Maryland
College Park, MD
okounou@terpmail.umd.edu

**Jacob Lane**
Department of Computer Science
University of Maryland
College Park, MD
jlane123@terpmail.umd.edu

**Trevor Lorin**
Department of Computer Science
University of Maryland
College Park, MD
tlorin@terpmail.umd.edu



## Abstract

Given the significant dangers posed by alcohol-impaired driving, which accounts for thousands of fatalities each year, this study explores the application of neural networks for detecting intoxication using facial recognition technology, addressing the critical issue of alcohol-impaired driving. Leveraging datasets containing sober and intoxicated facial images, we developed and evaluated three models: a custom convolutional neural network (CNN), a fine-tuned MobileNet, and a pre-trained FaceNet. The custom CNN achieved a validation accuracy of 90.71%, effectively identifying intoxication-related features, while the fine-tuned MobileNet demonstrated similar performance with a validation accuracy of 90.92%. The FaceNet model, tested for its ability to differentiate sober and intoxicated faces, delivered an accuracy above 99.5%, highlighting its robustness.

Key contributions include advanced preprocessing techniques to optimize dataset usability and a comparative analysis of model architectures, emphasizing the strengths of task-specific designs. These findings demonstrate the potential of neural networks in real-time intoxication detection and highlight their value in public safety. Future efforts will focus on diversifying datasets, integrating multi-modal inputs such as voice or thermal imaging, and improving model generalization through ensemble methods. This work lays the groundwork for machine learning solutions to reduce alcohol-related harm and support proactive intervention strategies.


## 1 Introduction

Facial recognition technology has rapidly evolved, becoming a pivotal tool in various sectors such as security, law enforcement, and personal device authentication. With the rise of artificial intelligence, neural networks have shown remarkable success in accurately identifying and analyzing facial features. Our project seeks to build upon these advancements by leveraging neural networks to confirm identity from facial images and extract additional observations such as identifying whether a


38th Conference on Neural Information Processing Systems (NeurIPS 2024).

person is intoxicated.

The need for enhanced detection of intoxication is particularly urgent given the ongoing issue of alcohol-impaired driving. In 2022, more than 13,500 people were killed in alcohol-impaired crashes in the United States alone, accounting for 32% of all traffic-related fatalities [6]. Furthermore, around 1 million drivers are arrested each year in the U.S. for driving under the influence (DUI) of alcohol or drugs [1]. These alarming statistics highlight the potential life-saving applications of improved facial recognition systems capable of identifying intoxicated individuals in real-time.

In practical terms, this technology could be deployed by law enforcement or security personnel at strategic locations such as the exits of bars or events, providing a proactive measure to assess whether individuals are fit to drive before they leave the premises. By integrating identity verification with intoxication detection, these models could offer a multi-dimensional approach to facial recognition, ensuring that potentially impaired drivers are identified before they take to the road. This would save lives and reduce the burden on law enforcement through early intervention. Ultimately, this project aims to contribute to public safety by providing law enforcement with innovative tools to make real-time, accurate assessments in high-risk situations.

## 2 Related works

"Drunk person identification using thermal infrared images" [4] is an excellent first step in alcohol intoxication detection. Using infrared images, the author uses their hypothesis that certain areas of the face heat up when intoxicated to detect whether the infrared image is of someone intoxicated. These specific areas were reduced to two dimensions, which showed that the clusters moved in the same direction when a sober person became intoxicated. Infrared images however are not easy to collect, making this process difficult to have useful real-life applications.

The report "Drunkenness Face Detection using Graph Neural Networks" [3] tests a neural network trained on a large database of facial images labeled as sober or intoxicated to identify if a facial image is of an intoxicated or sober person. The author used a GNN, stacking graph convolutional network layers with a gated linear unit in order to create a model to classify images. The model yielded an average accuracy of 88 percent in their final findings. This model has proven to not have a bias based on skin color, which was an issue in previous intoxication detection systems. Since this model uses facial images, it also provides much more practical use.

"Detection of Alcohol Intoxication Using Voice Features: A Controlled Laboratory Study" [7] collected voice spectrographic signatures in order to create a model that could determine intoxication from a small audio recording of spoken English. The study collected data by having sober and then intoxicated people read tongue twisters in which they split the data into one-second segments. These segments were labeled and used to train a model that yielded an accuracy of 98 percent. This study is unique as it is one of few that has not used the data from the Alcohol Language Corpus to create a model for drunk speech recognition. Using voice samples along with facial features could create an excellent ensemble for neural network intoxication identification.

## 3 Method

We began our investigation of intoxication detection by testing a pre-trained facial recognition model. We chose to use the the FaceNet model from the Pytorch library. This model uses a MTCNN (Multi-Task Cascaded Convolutional Network) to conduct facial recognition, which can then be used to determine if two distinct facial images are of the same person or not. We chose to test if this model's accuracy on recognizing the faces of individuals was at all affected when given images of drunk subjects. To do so, we used the *Three Drinks Later* image dataset which contains sets of images of a subject after consuming none, one, two, and three drinks of wine.

Using this image dataset, we were able to have an extensive set of images of the same subjects with varying levels of intoxication. This allowed us to create different random sets of data to use to test the model on its effectiveness when given sober facial images in comparison to when it was given either one or two images of an intoxicated subject.

Figure 1: 2x2 grid of photo in 3 Drinks Later dataset

Figure 2: A secondary image from the dataset

To get a larger view of the project and to challenge our learning, we decided to test with two different models: a CNN architecture that we'd create and a fine-tuned MobileNet model. Our motivation in choosing the CNN model was the effectiveness seen from convolutional networks for a variety of image tasks, leading us to believe that it could be effective for our drunk detection binary classification task. MobileNet is a type of neural network that is used in a variety of real-world applications, including facial recognition and object detection [2]. Our aim was to make use of the knowledge contained in this pre-trained model and transfer that knowledge to our intoxication detection task. The goal of both these models was to create a model that could accurately classify an image of a person's face as either the person is sober or the person is intoxicated. From there, we'd compare the results of the two models to determine which architecture performs better for the problem at hand.

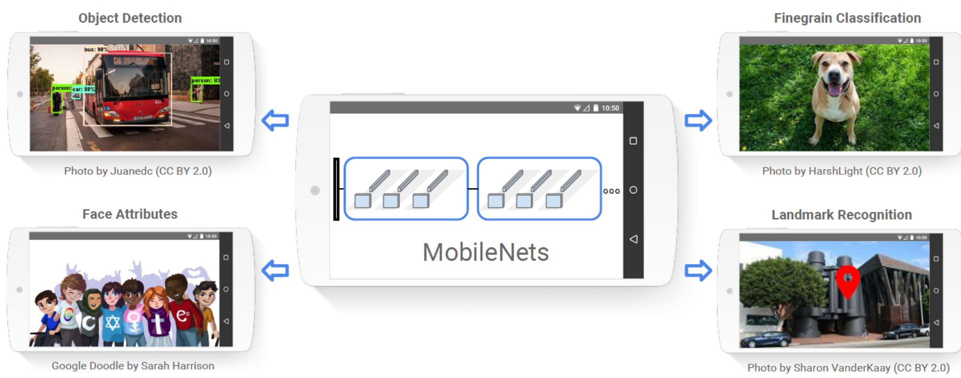


Figure 3: Diagram of MobileNet applications from MobileNet paper [2]

Both of these models would take in the same input of preprocessed datasets using OpenCV and MTCNN. We used MTCNN for face detection and OpenCV for consistent image formatting. Images would be centered and cropped around the subject's face to remove as much background noise as possible. We have a dataset of 2608 images of intoxicated faces from Universe Roboflow [5] and another dataset of 2128 images of sober faces from UTKFace [8]. We used these datasets as they have similar image quality to reduce as much bias as possible in the classification.

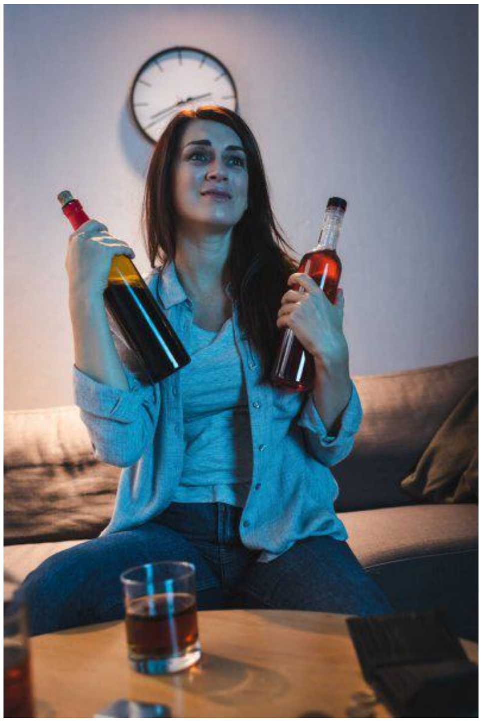

Figure 4: Sample image of an intoxicated person prior to preprocessing

Figure 5: Sample image of an intoxicated person after preprocessing

## 4 Experiments

To properly train our model, we needed to pre-process the images from the *Three Drinks Later* dataset, which shows individuals becoming progressively more intoxicated. The dataset came in a 2x2 grid format, where each grid contained four pictures of the same person at different stages. Since our model required individual images, we wrote code to split these grids into separate images. We also had to review the images to make sure they all had a clear facial image. In a few images, the subject completely covered their face, rendering the image invalid for our use. To circumvent this, we would use the image for two drinks in to replace the invalid three drink image.

The code handled separating the images while keeping the labeling intact so that each picture could still be tied to its respective individual. Out of the four images in the grid, we only needed three for training, so the code also selected the relevant ones and ignored the unnecessary fourth image. We also designed the labeling system to be simple and clean, making it easy to integrate into the training process. This step helped turn the *Three Drinks Later* dataset into something our model could work with, allowing it to learn from the progression we wanted to focus on.


Figure 6: Parsed version showing 3 images with labels


Figure 7: Another parsed version showing drink progression

With the images separated and labeled, we then wrote a script to use these images to evaluate the FaceNet model. Since the model takes approximately one second to compare two images, it would take an unsuitably long time to compare all possible combinations of images. This led to us testing the model using semi-random samples of images. In this process, we tested in two rounds: one testing sober images to other sober images, and another testing drunk images to other drunk images and sober images.

In the first round, we would test each labeled individual with ten randomly selected images from the sober labeled group. We also tested each individual against their one glass image that way we ensured the model would be tested for correctly identifying the same individual from two different images. The accuracy of the model in this round was 99.83%. In the second round, we would test each drunk individual against their own sober image and then ten randomly selected individuals from the sober labeled group and ten from the drunk labeled group. This tested not only comparing images when both subjects were intoxicated but also when only one individual was intoxicated as well. The accuracy of the model from this round of testing was 99.56%. The model had a lower accuracy when dealing with images of intoxicated subjects but not at a statistically significant level.

The FaceNet model operated at an extremely high accuracy when dealing with images of intoxicated subjects, highlighting neural networks' ability to create excellent models for detecting facial features. Since a neural network model could have such accuracy in identifying these features and linking them to an individual, we proposed that it would follow that an effective model could be created to classify if the features belonged to an intoxicated or sober individual. This led to the creation of our two intoxication detection models.

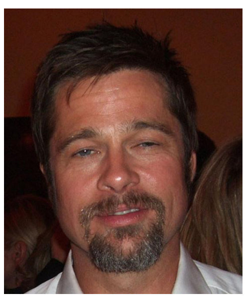

Figure 8: Example intoxicated face image

Figure 9: Example sober face image

We created a classic TensorFlow model using a CNN architecture for our first model. We experimented with different numbers of layers and kernel sizes and we decided to finalize an architecture that consists of four 2D convolutional layers to detect features of the images with max pooling between each of the layers in order to reduce the computational load while retaining only the most important features and discarding unnecessary details. The last layer uses a sigmoid activation function for the binary classification of sober or intoxicated faces.

To test the performance of this model, we split our dataset into training and validation datasets, with the validation dataset comprising of 20% of the images. After training the model with the training dataset over 3 epochs, we achieved a validation accuracy of 90.71%, as shown in Figure 8. The plots also illustrate a consistent decline in both training and validation loss on a per-epoch and batch-wise basis, indicating the model's effectiveness in differentiating features of intoxicated and sober faces.

Our second model used the pre-trained MobileNet model, with weights distilled from ImageNet. To fine-tune this model for our task, we removed the last five layers of the model and replaced them with a sigmoid layer to give the binary output for our classifier. We then froze all but the final 10 layers of the model to train at a reasonable speed while keeping the knowledge contained within the first layers of the MobileNet model.

The model quickly saw increasing training accuracy, so we trained for just 2 epochs, giving a validation accuracy of 90.92%.

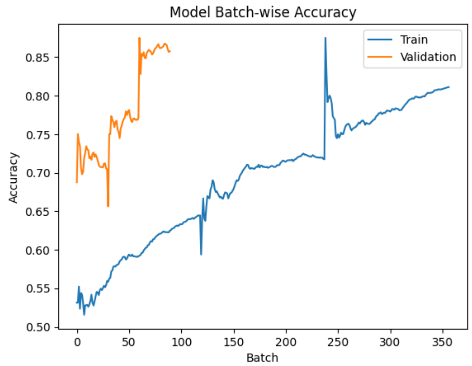


Figure 10: Accuracy per batch from CNN Model


Figure 11: Loss per batch from CNN Model

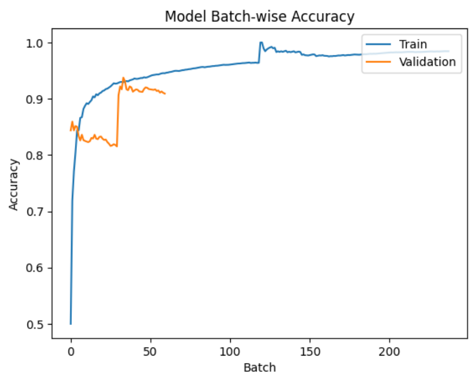


Figure 12: Accuracy per batch from MobileNet Model

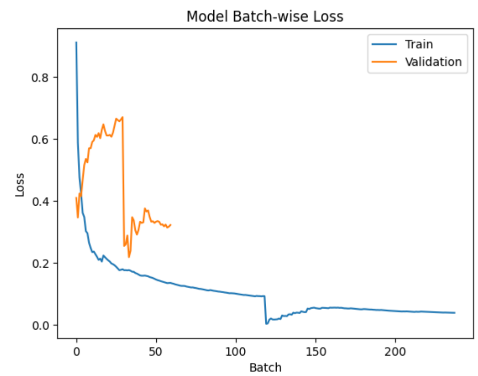


Figure 13: Loss per batch from MobileNet Model

## 5 Conclusion

This experiment explored the application of neural networks for detecting alcohol intoxication through facial images, employing three distinct models: a custom CNN, a fine-tuned MobileNet, and a pre-trained FaceNet for facial recognition.

The custom CNN model achieved a validation accuracy of **90.71%**% after three epochs, demonstrating its ability to effectively extract and classify features associated with intoxication. The fine-tuned MobileNet model, pre-trained on general image tasks and fine-tuned on our data, reached a validation accuracy of **90.92%**, demonstrating the power of transfer learning for this application.

For our facial recognition task, the FaceNet model, tested on its ability to employ facial recognition with both drunk and sober images, achieved exceptionally high accuracies in both rounds of testing (over **99.5%**), indicating there was really no effect on whether a person was intoxicated or not.

**Key findings include:**

- The FaceNet model showed high accuracy when comparing sober and intoxicated faces.
- Both the custom CNN and the fine-tuned MobileNet models were able to achieve a respectable level of accuracy, suggesting that both custom and pre-trained CNN-based models can be effective at classifying drunk facial images.
- Effective pre-processing of the dataset, including splitting grid images and validating labels, played a critical role in ensuring model performance and reducing noise in training data.

The results demonstrate the feasibility of employing neural networks for real-world intoxication detection systems, with potential applications in traffic safety, public event monitoring, and law enforcement. The robustness of the FaceNet model in distinguishing drunk and sober individuals provides a benchmark for future studies in this domain.

Future work should focus on expanding the dataset to include more diverse demographics and varying environmental conditions to improve model generalizability. Additionally, exploring multi-modal inputs, such as voice features or thermal imaging, could enhance detection accuracy and applicability. Comparative analyses of neural network models across multiple modalities would further solidify the understanding of their strengths and limitations.

We also recognize that working with real-world datasets, there may be unidentified confounding factors such as location or lighting that may not have been fully eliminated by our pre-processing and normalization. For future work, data experimentally collected under consistent conditions could provide a "cleaner" dataset that would ensure the generalization of our results.

In conclusion, this project represents a possible avenue toward the integration of AI in public safety, showcasing the potential of neural networks to reduce alcohol-related harm.

## References


[1] CDC. Impaired driving facts, n.d. Retrieved from `https://www.cdc.gov/impaired-driving/facts/index.html`.

[2] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[3] V. B. Kamath, S. S. Pai, S. S. Poojary, A. Rastogi, and K. S. Srinivas. Drunkenness face detection using graph neural networks. In *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pages 1–6, Bangalore, India, 2021.

[4] G. Koukiou, G. Panagopoulos, and V. Anastassopoulos. Drunk person identification using thermal infrared images. In *2009 16th International Conference on Digital Signal Processing*, pages 1–4, 2009.

[5] new-workspace 8swzs. Drunk dataset. `https://universe.roboflow.com/new-workspace-8swzs/drunk`, sep 2021. visited on 2024-12-09.

[6] NHTSA. Drunk driving, n.d. Retrieved from `https://www.nhtsa.gov/risky-driving/drunk-driving`.

[7] B. Suffoletto, A. Anwar, S. Glaister, and E. Sejdic. Detection of alcohol intoxication using voice features: A controlled laboratory study. *Journal of Studies on Alcohol and Drugs*, 84(6):808–813, 2023.

[8] Song Yang Zhang, Zhifei and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.