Drunk and Sober Classifier

Prepared by: Abhishek Gupta

Executive Summary

This project focused on developing a machine learning model that can detect intoxication using thermal infrared (IR) images. The goal was to build a solution that works reliably in real-world settings—despite facing major challenges like a very small dataset and wide variation in image quality.

The development followed a careful, step-by-step approach. It began with a simple baseline model—a Support Vector Machine (SVM) using raw pixel data—which achieved only 59.23% accuracy. As the project progressed, different feature engineering methods and algorithms were tested. Interestingly, more complex feature extraction techniques (like texture or statistical features) didn't help and sometimes made things worse. However, switching to a Random Forest Classifier significantly improved performance.

The final model—a **Random Forest trained on resized raw pixel images**—achieved an accuracy of **79.23%.** More importantly, it identified **100% of intoxicated individuals**, making it an excellent early-stage screening tool. While it does produce some false positives (flagging sober people as intoxicated), this is a reasonable trade-off in safety-critical applications where missing a true case would be far more costly.

Problem Understanding and Initial Strategy

The core task was to build a binary classifier to distinguish between "Sober" and "Drunk" states using a dataset of thermal IR images from 41 individuals.

Key Data Characteristics and Challenges:

A detailed review of the problem and README highlighted key challenges that shaped the project approach:

- **Limited Data:** With only 649 images, the dataset is very small for a computer vision task, increasing the risk of overfitting.
- **Unique Image Format:** Each .tif file contains 50 sequential frames. The original MATLAB preprocessing averaged these frames to create a stable thermal image—this step was recreated in Python.
- **High Variability:** Images come from four different views (front face, side face, eyes, hand palm) and four conditions (sober plus three post-drinking stages), adding complexity the model must manage.
- **Focus on Real-World Use:** To ensure practical relevance, the evaluation prioritized precision and recall metrics over simple accuracy.

Strategic Decisions:

Based on these challenges, two key upfront decisions were made:

1. **Class Simplification:** The three post-drinking time points (20 mins, 40 mins, 1 hour) were combined into a single "Drunk" class (label 1), while the "Sober" state remained as label 0. This binary classification approach helped increase the positive class size and made the problem more manageable.
2. **Iterative Modelling:** Given the complexity, the approach began with a simple baseline model and progressively improved it through feature engineering and stronger algorithms, carefully evaluating performance at each stage.

Exploratory Data Analysis (EDA)

EDA was performed to validate assumptions and understand the data's structure.

- **Class Distribution:** After redefining classes, there was a moderate imbalance with **485 "Drunk" and 164 "Sober" images.** This was handled during modeling by applying class_weight='balanced' to avoid bias toward the majority class.
- **Image Visualization:** Examining sample images from both classes showed subtle thermal differences around the eyes and nose, matching known effects of alcohol. However, there was also significant variation in head pose, angle, and individual baseline temperatures.

Modelling Journey: A Tale of Two Experiments

A systematic approach was taken to find the optimal model, with the results from each stage informing the next.

Baseline Model: SVM with Raw Pixels

- **Hypothesis:** As a starting point, a linear Support Vector Machine (SVM) might be able to find a separating hyperplane using the raw pixel values of the images.
- **Implementation:** Images were resized to a consistent (64, 80) dimension and flattened into a 5120-element feature vector. An SVM was then trained on this data.
- **Result: Accuracy: 59.23%**.
- **Analysis:** The baseline model performed poorly, with a low F1-score of 0.25 for the 'Sober' class. Its accuracy was worse than a naive model that always predicted 'Drunk,' highlighting the linear SVM's struggle with high-dimensional, noisy pixel data.

Winner Model: Random Forest with Raw Pixels

- **Hypothesis:** A more powerful, non-linear ensemble model like Random Forest might be able to discover meaningful patterns directly from the raw pixels where the linear SVM failed.
- **Implementation:** The same flattened pixel data from the baseline was used to train a RandomForestClassifier.
- **Result: Accuracy: 79.23%**.
- **Analysis:** This represented a massive improvement (+20%) over the baseline. Random Forest excelled because:
    1. **Ensemble Power:** By averaging the predictions of many decision trees, it is more robust to noise and less prone to overfitting.
    2. **Internal Feature Selection:** Each tree in the forest is trained on a random subset of pixels. This forces the model to learn from different parts of the image and prevents

it from relying on spurious correlations, effectively discovering the most important thermal regions on its own.

3. **Non-Linearity:** It can capture complex relationships (e.g., "if pixel X is hot AND pixel Y is cool") that are impossible for a linear model to find.

4. Final Model Performance and Justification

The Random Forest classifier trained on raw pixel data was chosen as the final model because it delivered the best performance.

Detailed Performance Metrics:

Class	Precision	Recall	F1-Score	Support
Sober (0)	**1.00**	0.18	0.31	33
Drunk (1)	0.78	**1.00**	0.88	97
Overall Accuracy			**0.7923**	**130**

4.2. Interpretation for a Real-Life Scenario:

The metrics reveal a model with a specific and highly valuable bias for a real-world screening application:

- **Never Misses an Intoxicated Individual:** With a perfect recall of 1.00 for the 'Drunk' class, the model identified every intoxicated person in the test set, ensuring zero false negatives—an essential feature for safety screening.
- **High Precision for 'Sober' Predictions:** The model achieved 100% precision for the 'Sober' class—everyone it labelled as sober was truly sober, with zero false accusations.
- **The Trade-Off:** The model has a low **recall (0.18) for the 'Sober' class**, often misclassifying sober individuals as drunk. In practice, this makes it a conservative screener—it may over-flag for follow-up testing, but ensures no intoxicated individuals are missed.

4.3. Justification of the Accuracy Ceiling:

Achieving an accuracy near 80% is a strong result given the dataset's inherent limitations:

1. **Data Scarcity:** With only ~500 images for training, the model's ability to generalize is fundamentally limited.
2. **Weak Signal & High Noise:** The thermal differences between classes are subtle, while variations between individuals and camera angles are large. This low signal-to-noise ratio makes the problem extremely challenging.
3. **Class Imbalance:** The 3:1 class imbalance makes it difficult for any model to learn the features of the 'Sober' class effectively.

5. Conclusion and Future Work

This project successfully demonstrates the systematic process of building and evaluating an image classifier on a challenging, small-scale dataset. The final **Random Forest model achieved a respectable 79.23% accuracy** and, more importantly, exhibited performance characteristics that make it an ideal candidate for a real-world intoxication screening system.