

THE BOUNCER

Development of a Face Recognition and Key Word Detection Security System

Luca Pivetti
914199



Martin Gerini
876695

Table of Contents

1. Goal
2. Solution
3. Face Recognition
4. Key Word Detection
5. Results
6. Problems encountered
7. IRL Tests
8. Further Developments

Goal

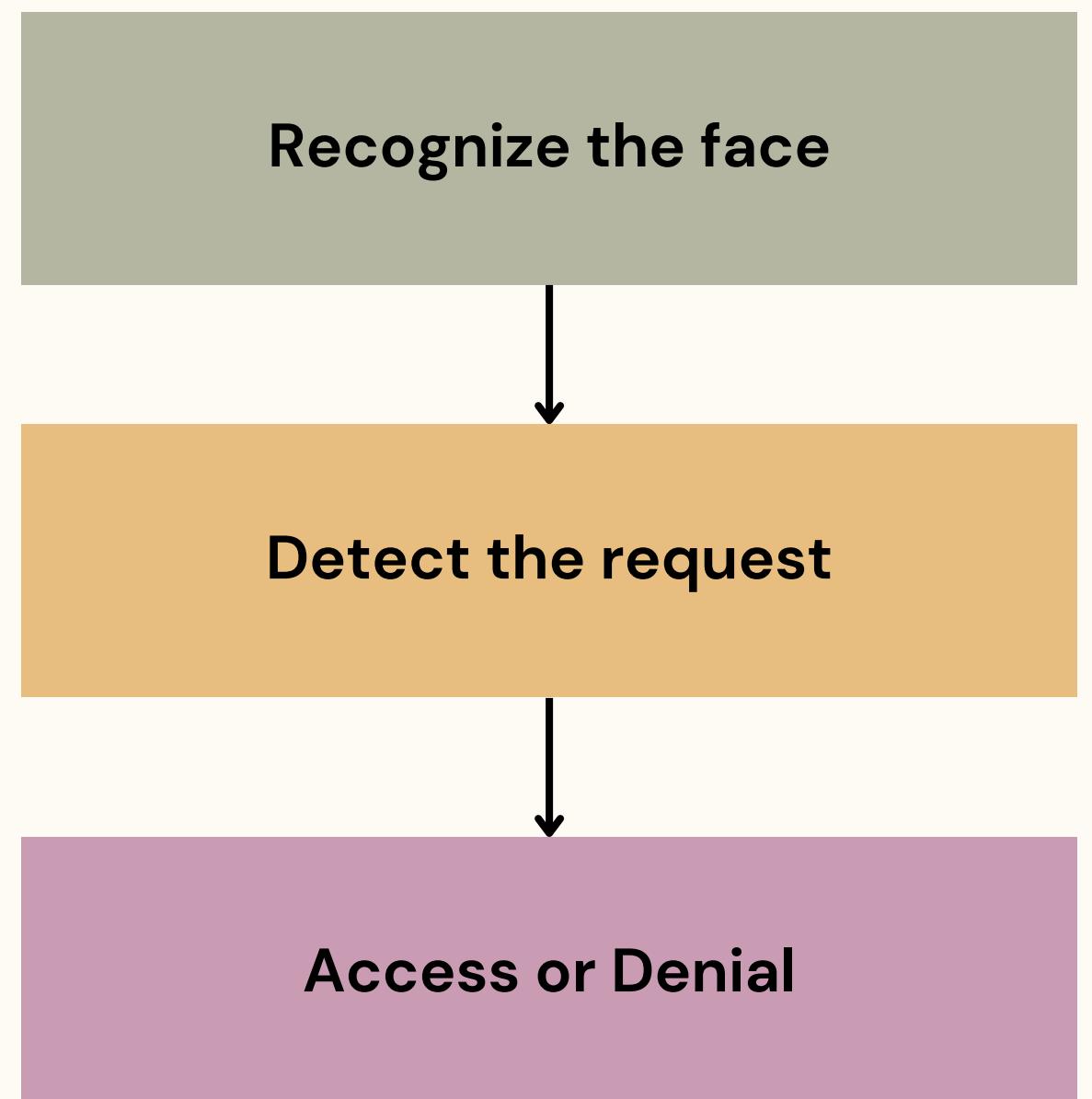
Security system based on linking
individuals and their possible **requests**

Applications

- **Elevator** where each employee has access to certain floors
- **Access doors** to different departments in the facility
- **Smart Home** applications

Solution

We rely on **Face Recognition** to identify the person and **Key Word Detection** to detect the required action



Face Recognition

Data Processing

Video Recording

Capturing every frame

Face Detection

Looking for a face and its confidence in each frame

Face Continuity

The face with the highest confidence score must remain for at least 5 frames

Face Processing

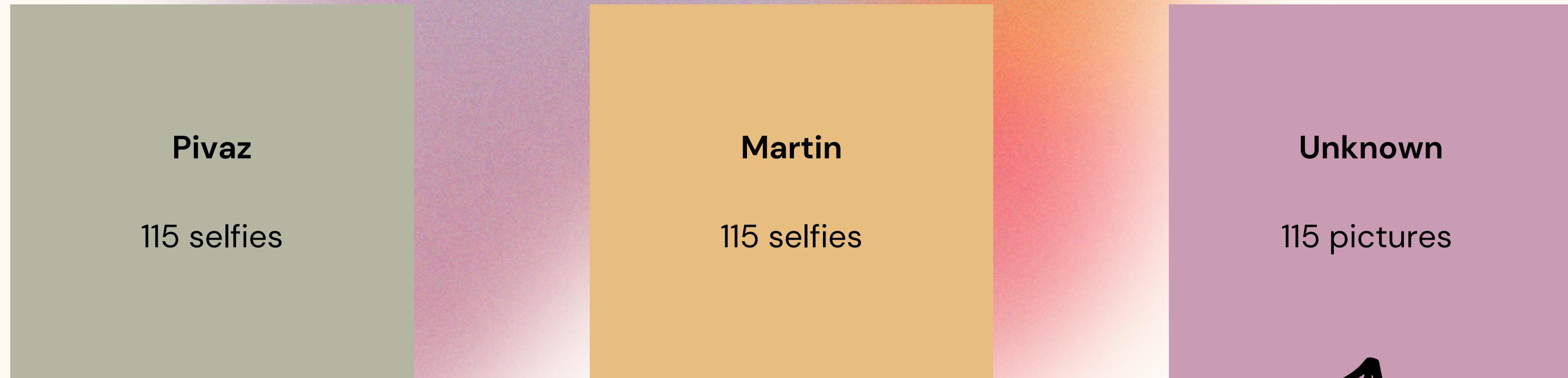
Alignment of detected face

Face Recognition

Aligned face is passed to the model for recognition (5x)

Face Recognition

Dataset



*Labelled Face in the Wild
(LFW)*

Face Recognition

Models

Face Detection

Model from dlib
composed by **HOG** and **Linear SVM**

Being a consumer system, efficiency is important:
therefore, a **fast and efficient model** was chosen
despite not invariant on rotations and viewing
angles, considered as negligible given the context

Face Recognition

sphere20 model
modified for **classification** purposes

The original model was born with the purpose of
creating embedding; to deal with classification,
**we modified the last two layers and did fine
tuning on them freezing the previous ones**

Key Word Detection

Data Processing

Audio Recording

Capturing audio for a time of 3 seconds

Audio Processing (Training)

Applying pad/trim, noise augmentation and normalization

Audio Processing (Application)

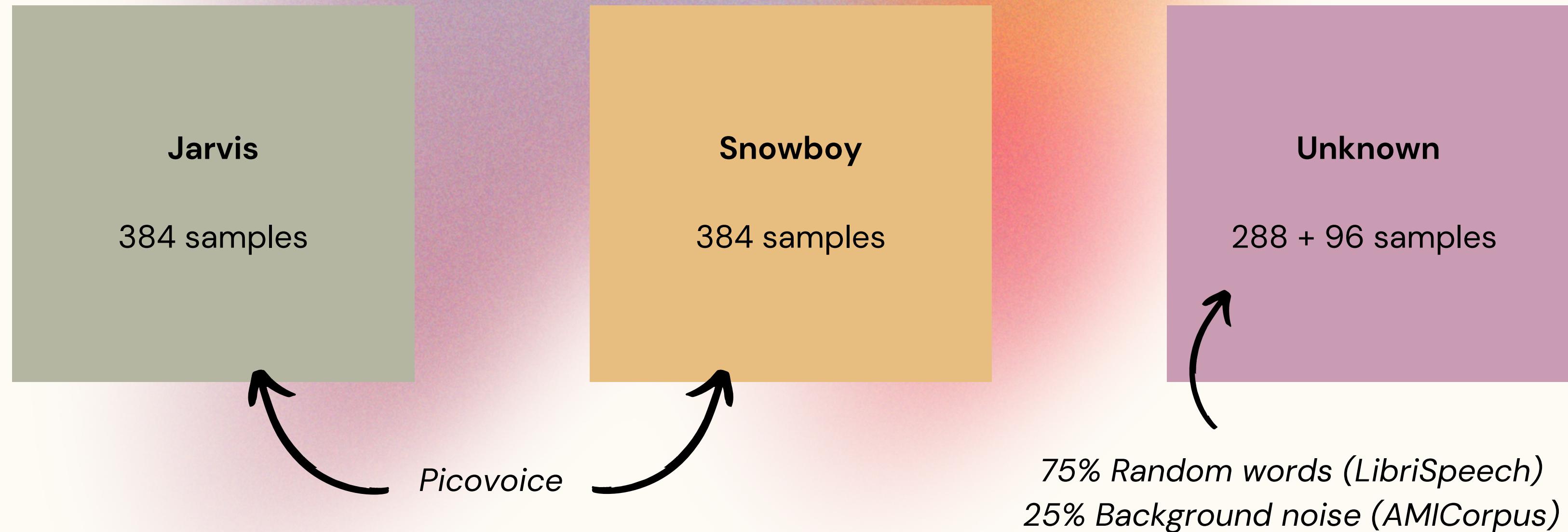
Noise gate filter, centering and normalization

Key Word Detection

Processed audio is passed to the model for detection

Key Word Detection

Dataset



Key Word Detection

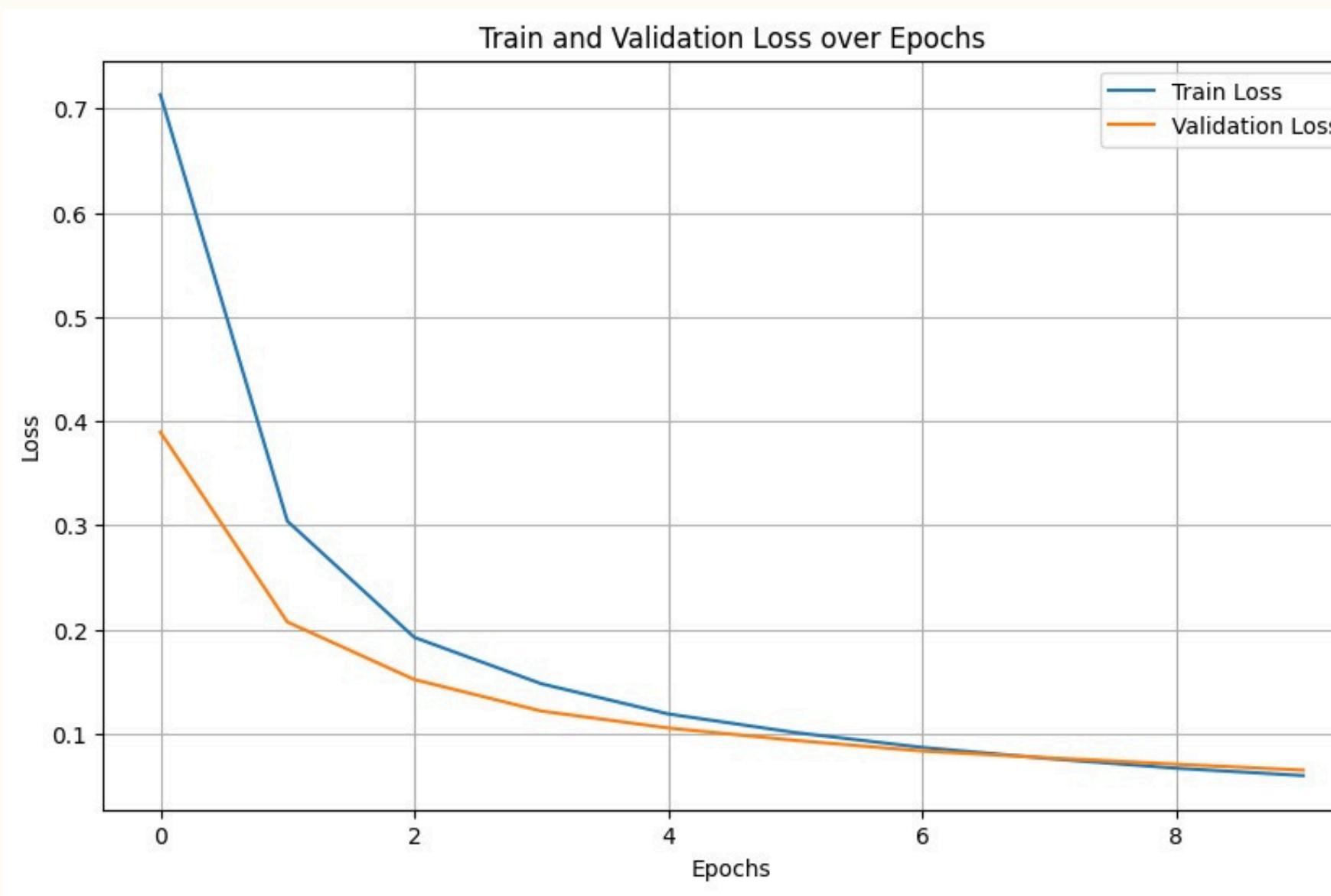
Model

The model for Key Word Detection is divided in 3 sections:

- **MEL Spectrogram:** audio is transformed into 2D, highlighting the characteristic features of human language
- **Feature Extraction:** audio passes through 3 convolutional layers to extract features from the previously obtained image
- **Classification:** the audio then passes through 2 linear layers for final classification

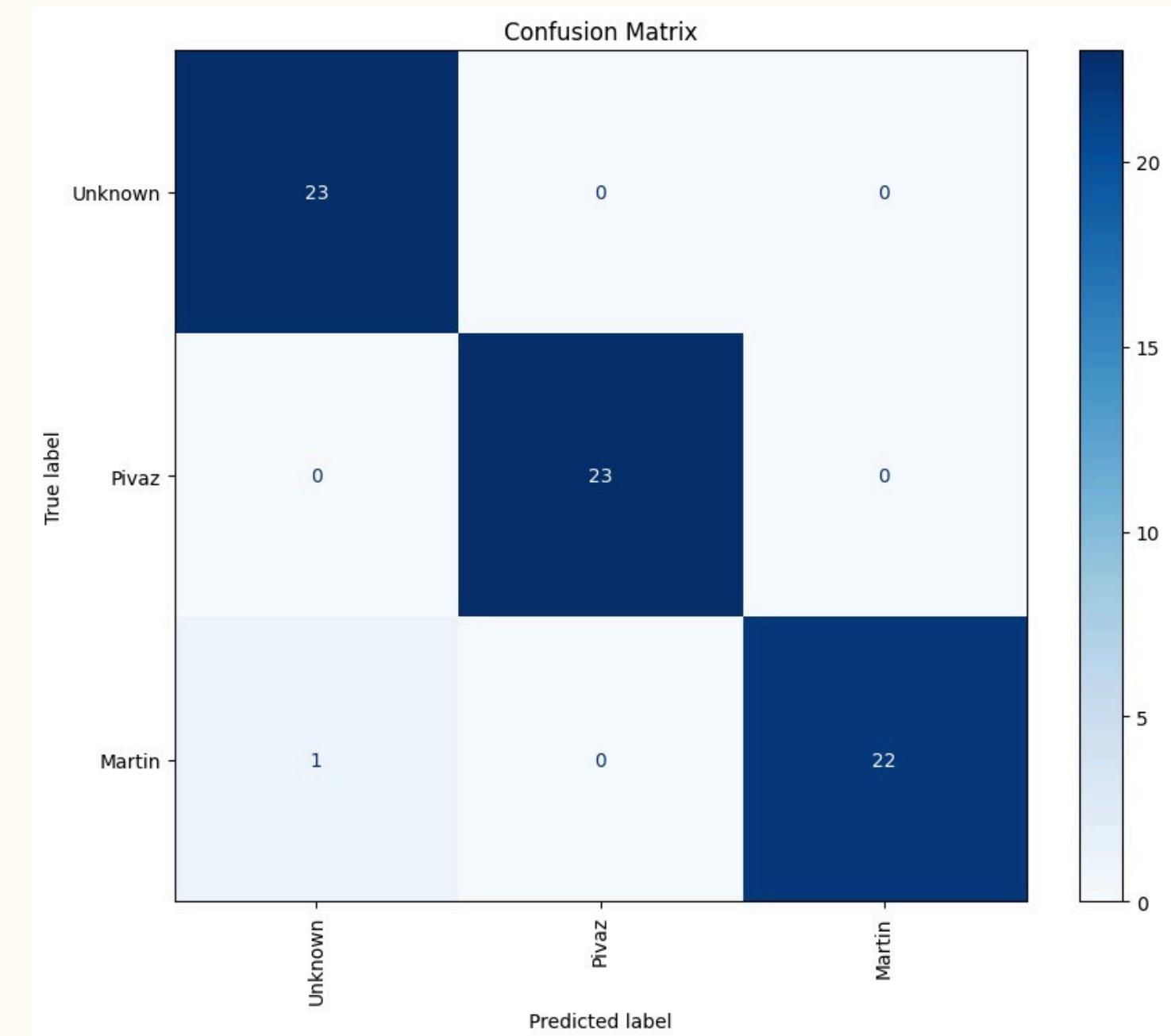
Results

Face Recognition



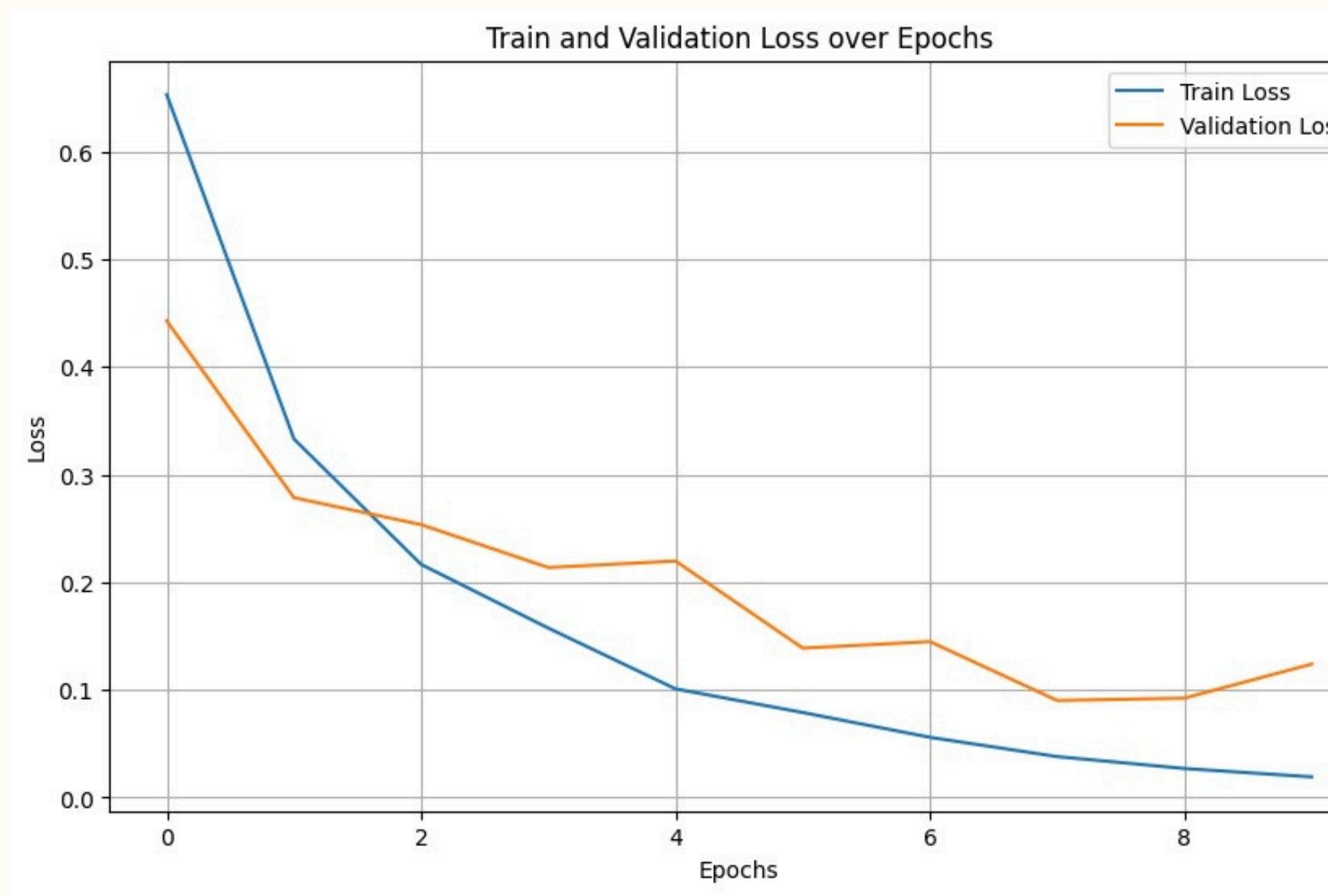
Acc: 0.9855
F1 Score: 0.9710

72% Train
8% Val
20% Test



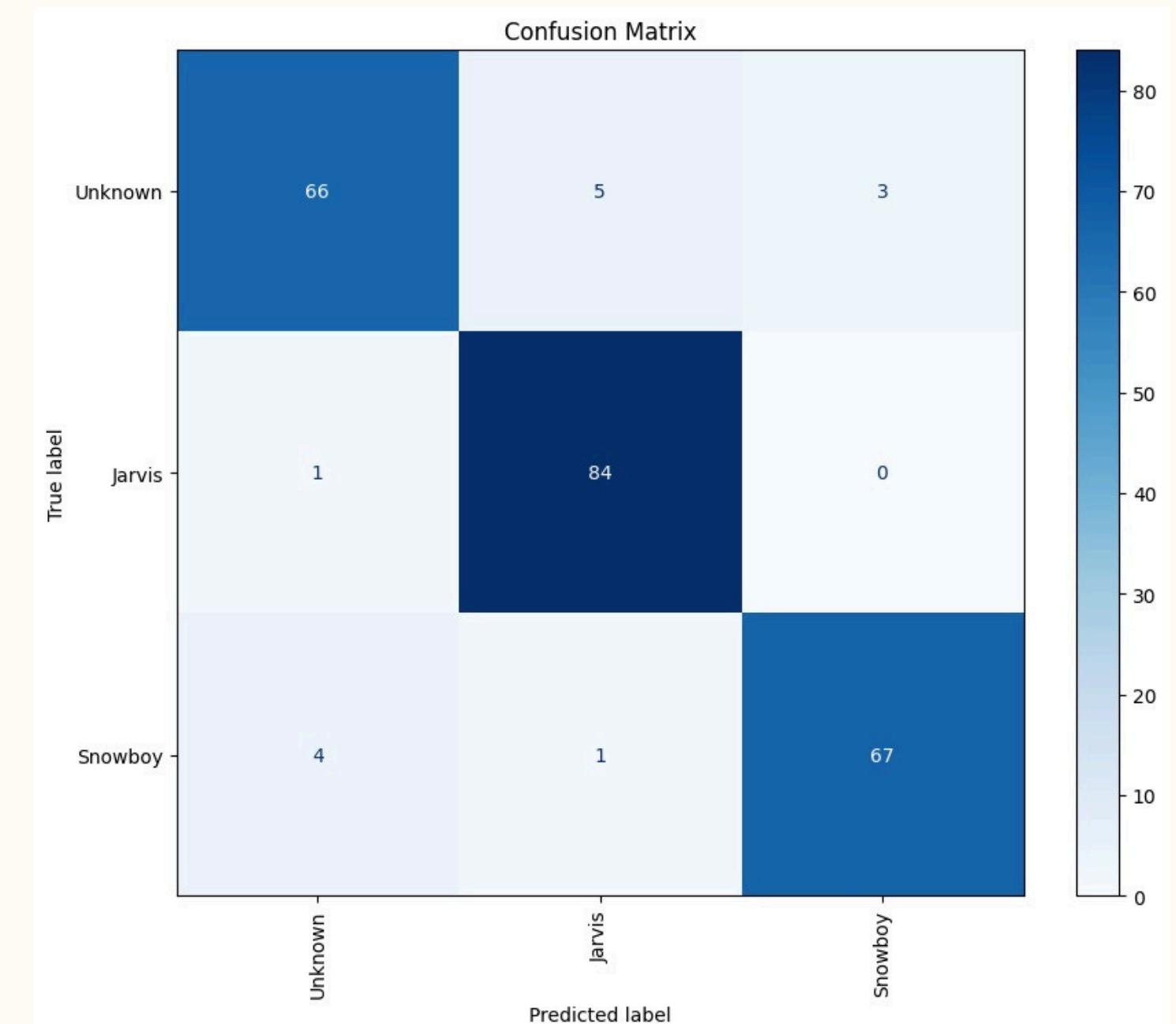
Results

Key Word Detection



Acc: 0.9394
F1 Score: 0.9295

72% Train
8% Val
20% Test



Problems encountered

Face Recognition

- **Data Scarcity:** self-acquisition of photos was needed
- **Data Acquisition:** difficulty in acquiring images in many different contexts and scenarios (brightness issue)
- **Devices Compatibility:** functions for adaptation were needed

Problems encountered

Key Word Detection

- **Data Scarcity:** dataset only contains around 400 samples for each label
- Samples of Dataset were **studio recorded:** addition of noise for training process was needed
- Samples of Dataset were **perfectly centered:** centering of the recorded audio was needed
- Training on **not recognized words:** background noise and random words were used as negatives
- **Accents:** difficulty to generalize across different people's accents
- **Devices Compatibility:** functions for adaptation were needed

IRL Tests

The Bouncer was tested on 10+1 different subjects through iOS, Android, MacOS and Windows OS.

	<i>Face Recognition</i>	<i>Key Word Detection</i>
Pivaz	✓	✓
Pivaz's mum	✓	✓
Pivaz's dad	✓	✓
Matteo	✓	✓
Elisa	✓	✓

	<i>Face Recognition</i>	<i>Key Word Detection</i>
Martin	✓	✓
Martin's mum	✓	✓
Martin's dad	✓	✓
Sabrina	✓	✓
Thomas	✓	✓



Further Developments

- **Models robustness**
- **Pruning/Knowledge Distillation** for more efficient systems
- **Enrollment** for new employees
- **Anti-spoofing** for Face Recognition
- **Rectification** of key words
- **Interaction** improvements
- **Privacy and security**

And now, let's test it!
