KOMISJA INFORMATYKI I AUTOMATYKI
ODDZIAŁU PAN W POZNANIU

POLSKA AKADEMIA NAUK

POZNAN UNIVERSITY OF TECHNOLOGY

A lightweight software application for increasing users well-being

prof. PP, dr hab. inż. Anna Kobusińska

Sztuczna inteligencja w zapobieganiu chorobom, wspomaganiu leczenia i dla poprawy jakości życia

Outline

- Motivation
- Goal of the work
- General idea of the proposed solution
- Speach emotion recognition and facial emotion recognition
- Experimental evaluation
- Conclusions

Motivation

- The **fields of well-being** such as psychiatry and psychology have significantly **benefited from technological developments**
- By **gathering the data regarding particular cases**, where such problems occurred, **models of particularly vulnerable individuals can be created** to find the repeating pattern.
- **Development of the system** that may **help to understand the human emotions better, find the factors influencing** their **occurrence** or frequency, and to **warn in case of accumulation of emotions**
- The choice of technology used to implement such a system is of a great importance.

Goal of the work

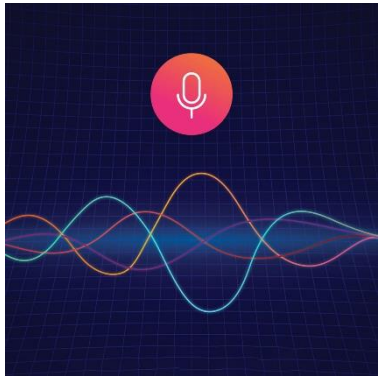
- Design and implementation of a system that collects and processes information about users' emotions that:
  - **provides reports** for **subsequent analysis** or **inspection of emotions**
  - is equipped with simple functions that **advise actions** based on the **current users' emotions and mood**.
- The solution leverages a range of modern technologies, including micro-services, cloud processing, deep learning, and human-device interaction

The idea of the proposed solution

- Primarily based on **Amazon Web Services**
- Author's extension of an Alexa voice service, utilizing **Speech Emotion Recognition model** to detect emotions from voice recordings sampled from users' commends or utterances
- Visual source of emotions, images of facial expressions are captured by the device's camera and analyzed by the **AWS Rekognition service**
- The **metadata regarding user's environment**, such as weather, web activity (e.g. Spotify), time of a day/week during user-system interaction are also included

Speach emotion recognition

- Source of signs allowing us to interpret them - speech characterized by speed, loudness, stability of voice
- **Speech Emotion Recognition (SER)** model determines based on the given data in the form of recording, speech-streaming or live speech at the time of an interaction between the user and the device, the emotions accompanying particular human utterance

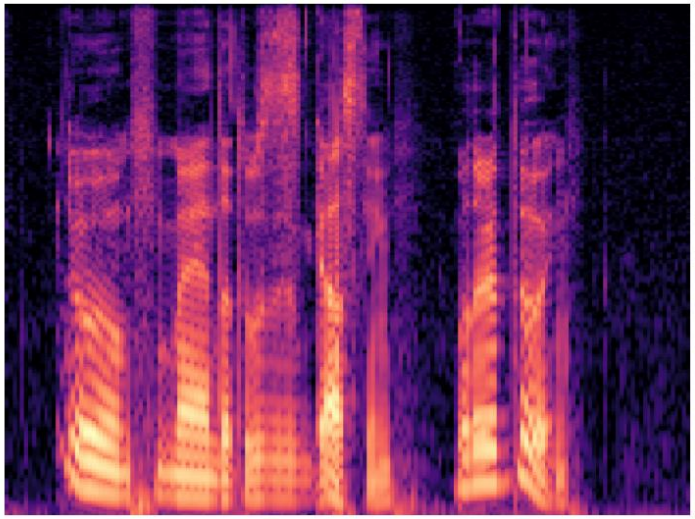


Features of speach emotion recognition

- **Prosodic features** — all the sonic properties of a language characterizing syllables or their subsequences in the course of utterance
  - data extracted from the prosodic data are Pitch, Energy, and Duration
- **Characteristics of the vocal tract system**
  - Cepstral coefficients, like Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC), and Discrete Fourier Transform (DFT)

Features of speach emotion recognition

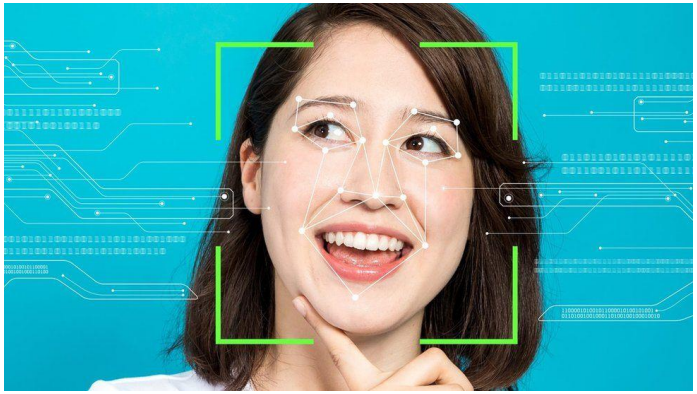
- **Visualization of features distributions** - creating, e.g. Mel Spectrograms, where colours indicate the energy of particular frequencies in the domain of time
- This step enables transferring the classic SER problem into the Computer Vision image classification task, where deep learning may **increase the accuracy of recognition**



Example Mel spectrogram generated from an (disgust) utterance recording

Facial Emotion Recognition

- **Face Emotion Recognition (FER)** is a branch of AI focused on recognizing specific features in photos, that allows drawing conclusions about the emotional state of the person presented in that photos
- It can be done based on the layout of important facial points, such as eyes, brows, nose and mouth.
- Solutions to that uneasy task currently can reach results surpassing 90% with the usage of Convolutional Neural Networks (CNN)

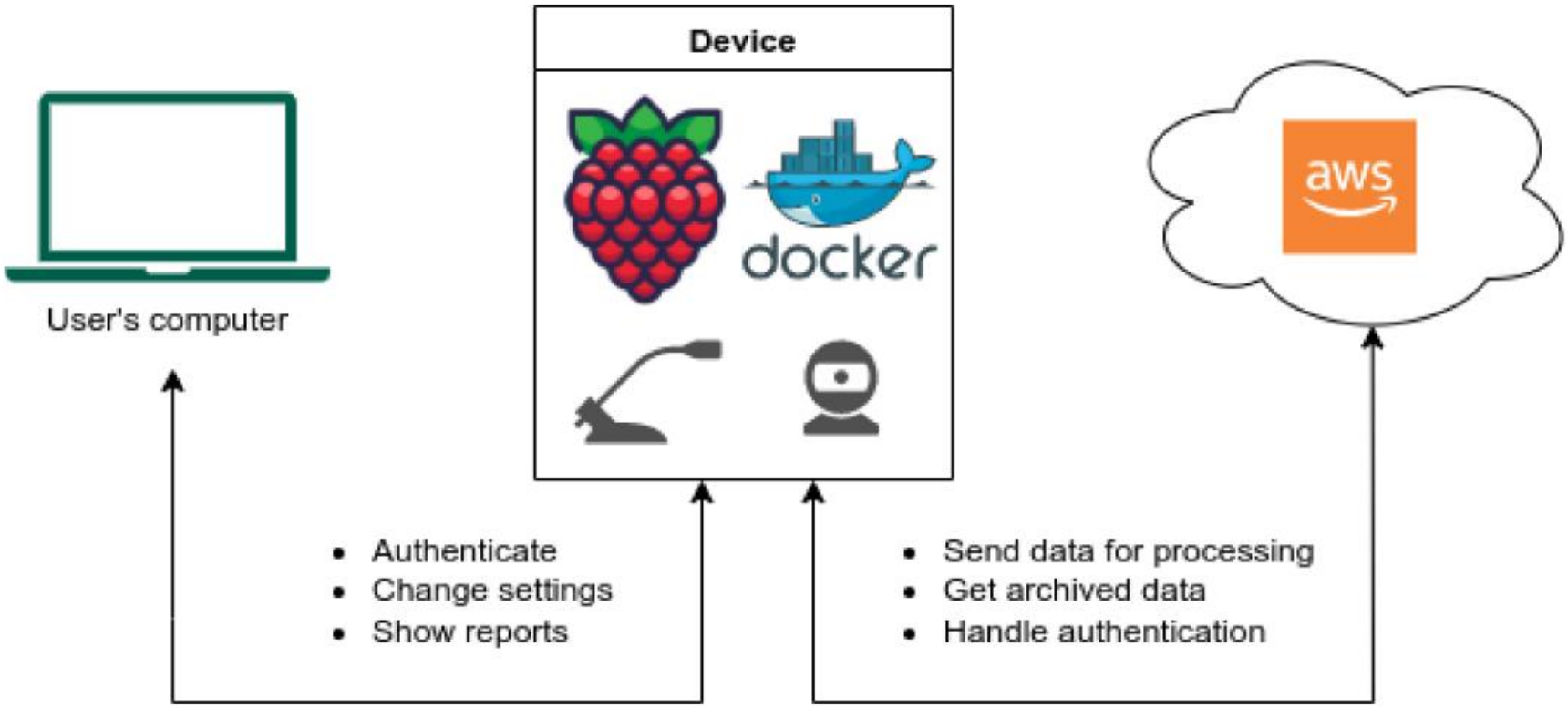


Existing applications keeping track of emotions

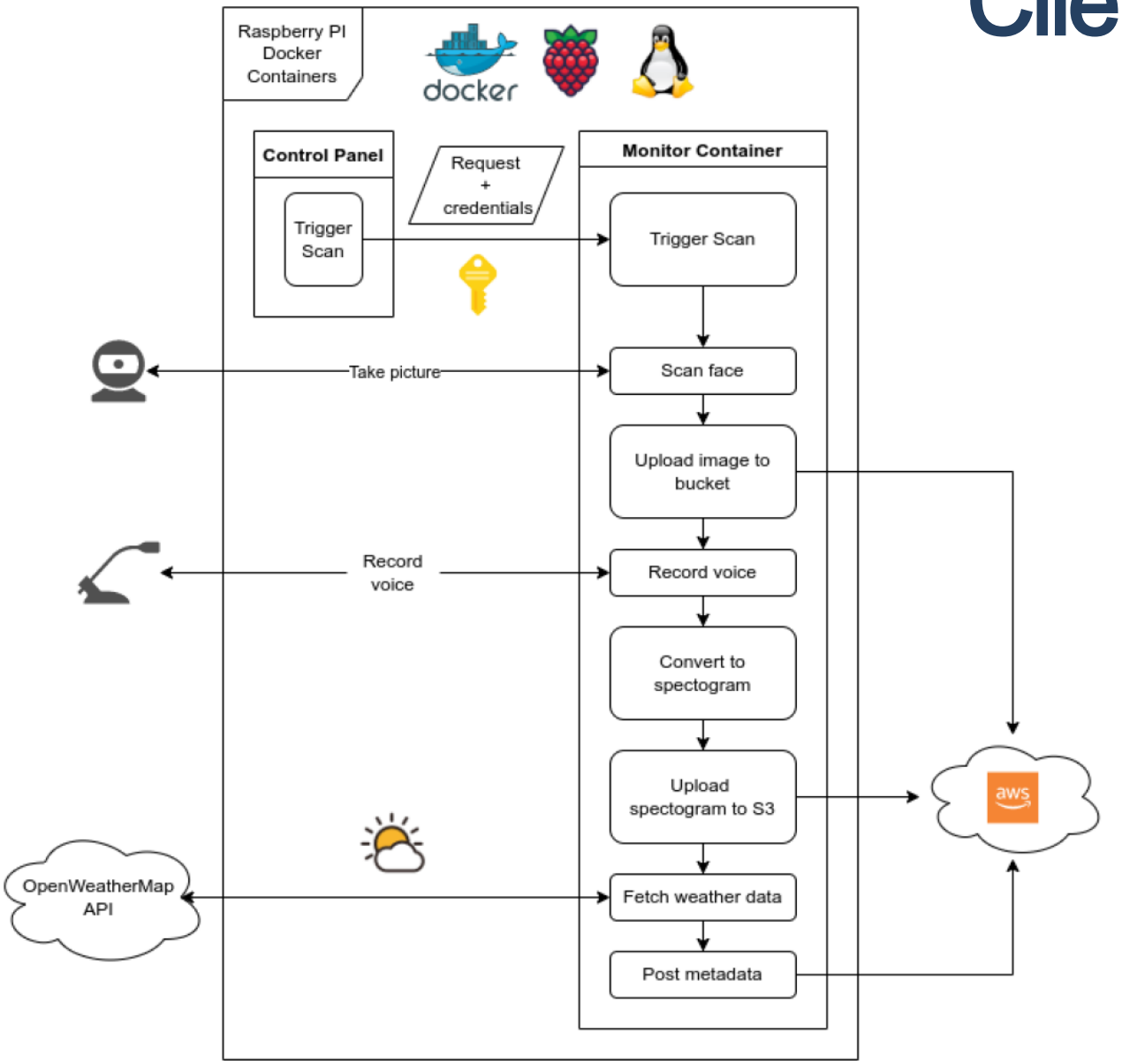
Existing Android applications allowing users to keep track of their emotions:

- **Moodfit** — an application allowing to track user's mood, meditation practices, notes, medicines, etc.
- **eMoods Bipolar Mood Tracker** — a mood tracking application that sends reports to the user's doctor
- **Bearable** — an application allowing to track moods and connect the moods with the user's activities

General concept of the architecture



Client application



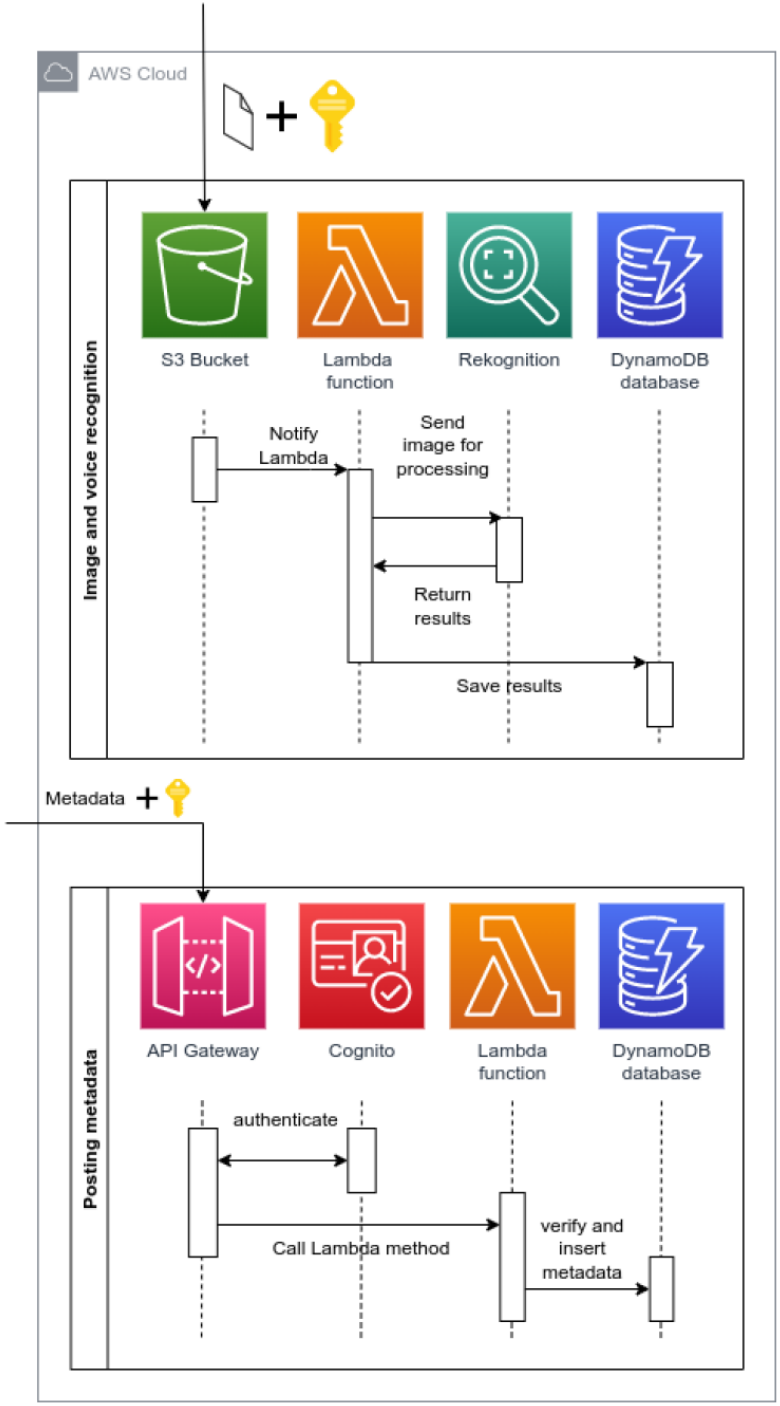
Report generation and presentation

- **Monthly report** – the data is presented in the form of a calendar with an emoticon representing the emotion for each day allows user to quickly see and analyse the whole month
- **Daily report** – data is presented in the form of table of hours, each with most prevalent emotion, again presented in the form of an emoticon with tooltip for details to allow more detailed look on emotions influencing the user
- The set of emotions consists of: HAPPINESS, SADNESS, ANGER, CONFUSION, DISGUST, SURPRISE, CALM, FEAR, UNKNOWN

Trends detection

- Correlations with **part of the day** — with parts of day defined as: Morning (07:00-13:00), Afternoon (13:00-18:00), Evening (18:00-23:00) and Night (23:00-07:00)
- Correlations with **day of week**
- Correlations with **location** — using location set by user
- Correlations with **weather** — offering further customization, choosing between temperature, pressure and humidity
- **Series of days** in the **same mood**

AWS backend application



Speach emotion recognition

Training data - datasets used for training are shortly characterized below:

- **RAVDESS** - Each of the actors was recorded in three modes: audio-only, audio-video, and video-only
- **EMO-DB** - The set of emotions consists of happiness, sadness, anger, fear, boredom, disgust, neutral.
- **CREMA-D** - six emotion categories: happiness, sadness, anger, fear, disgust, and neutral.
- **TESS** – Set of emotions: happiness, sadness, anger, fear, surprise, disgust, neutral.
- **SAVEE** – Set of emotions consists of: happiness, sadness, anger, fear, surprise, disgust, neutral.

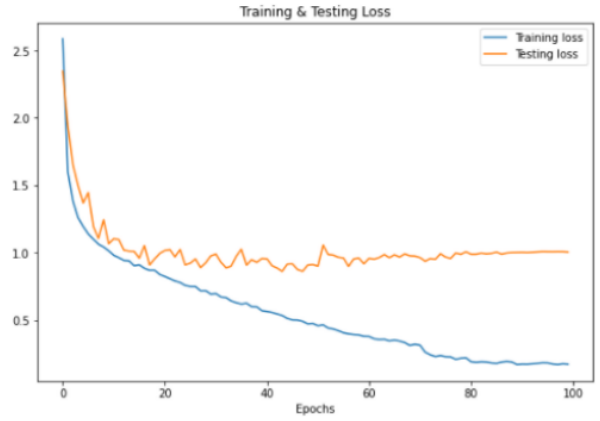
SER model implementation and training

- The deep learning model was implemented with the **Keras API** for **TensorFlow framework**,
- For the deep learning model structure, the convolutional neural network was chosen based on the concept proposed by **Bonaventure and Yeno**
- Data was split into train and validation sets in 8:2 proportion, fed into the network in batches of 64 images per batch.

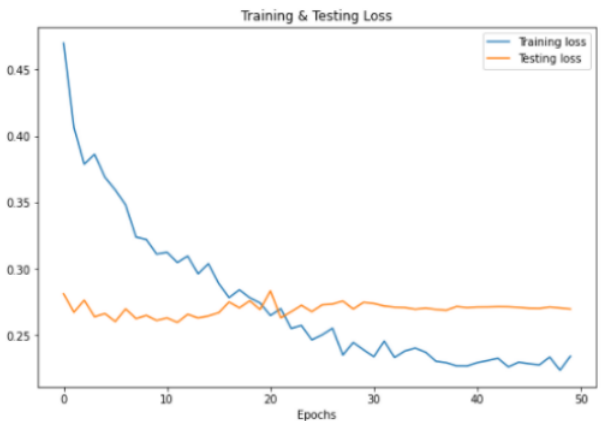
SER model implementation and training

The three variants reaching top results after 100 epochs:

- • clean RGB data (train. acc.: 0.94, test. acc.: 0.69),
- • clean extended dataset (0.95, 0.66)
- • clean higher resolution data (0.74, 0.64)



Accuracy and loss values registered during training on not augmented RGB images



Accuracy and loss values registered during fine tuning

Conclusions

- The final result is a **fully functional application** allowing users to keep a record of their emotions in time and observe the occurring trends.
- There are some interesting **directions of future works** that, even though are currently exceeding the scope of that project, may be added in the future:
  - capturing done automatically throughout the day.

[Thank you for your attention]