Multimodal Playlist Generation

Using emotion, physiology & activity data to recommend the perfect music

Gabby Cordelli, Anjali Goel, and John Hope

Why this project?

- Music boosts mood, motivation, and focus
- But playlists don’t always match what we feel
- Can AI personalize music using **multimodal signals**?

Data Sources

FER-2013

- Facial emotion images
- 7 emotion classes

WESAD

- Biometric time-series (HR, EDA)
- Label: stress, neutral

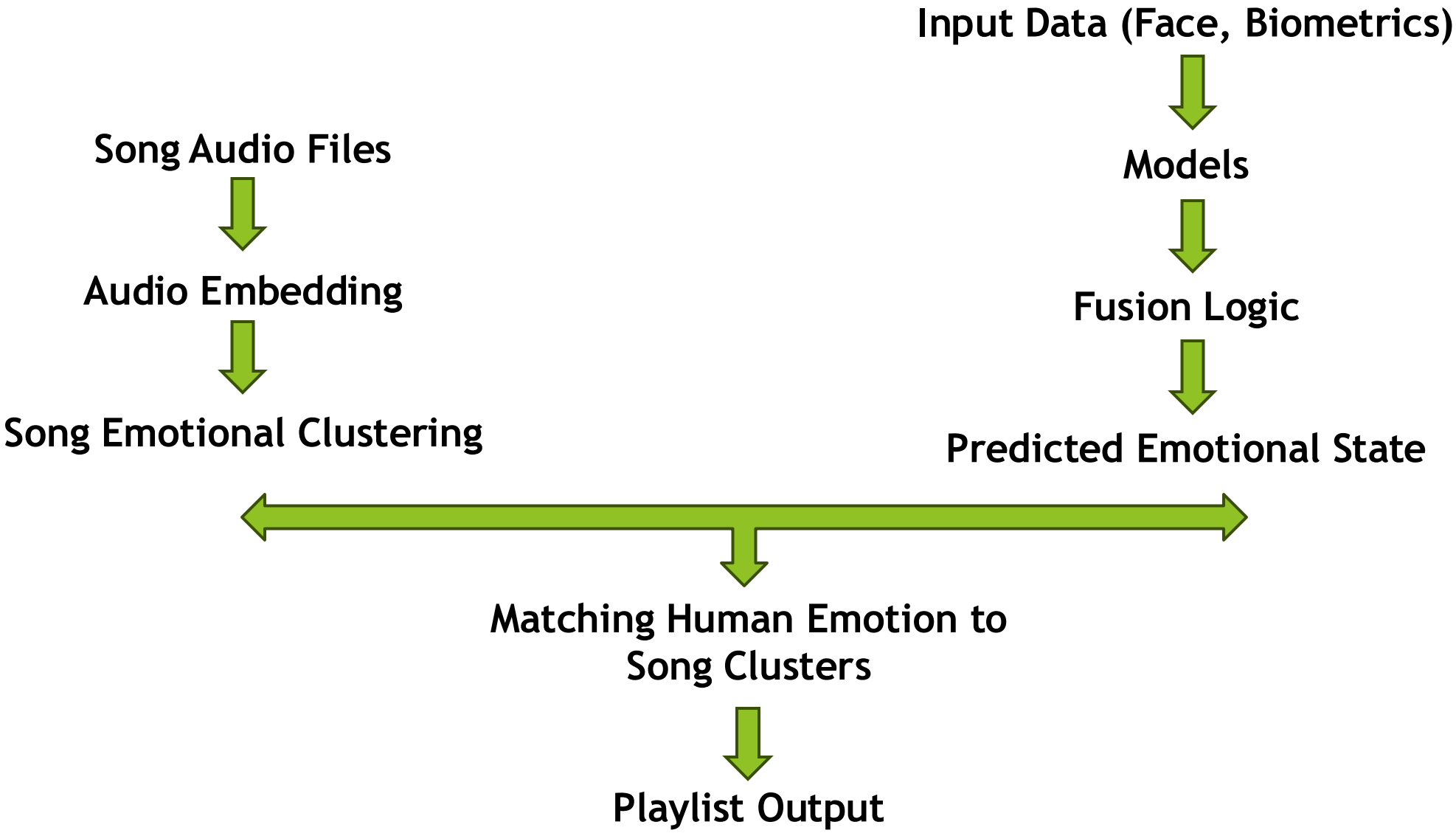
Capture-24 / HAR

- Accelerometer data
- Labels: walking, sitting, etc.

Spotify Song Data

- Pulled 1,000+ popular songs from the past 25 years using Spotify API, and associated information
- Used **yt-dlp** API to download song audios from YouTube
- Goal: classify each song by emotion + vibe

System Pipeline



Emotion Classifier (FER-2013)

- CNN-based image classifier
- Trained on grayscale facial images
- 7 emotion classes (happy, sad, angry, neutral, disgust, fear, surprise)
- Validation Acc: 66.3%

WESAD Stress & Emotion Classifier

- Time-series data from chest and wrist-worn sensors, detecting key biometrics
- Developed a custom CNN-based regression model
- Predicted each of these emotional states from 1-5:
  - ['Stressed', 'Angry', 'Happy', 'Sad', 'Inspired', 'Excited', 'Nervous']

WESAD Classifier - Evaluation

Emotion	Pearson Correlation
Angry	**r = 0.843**
Sad	r = 0.833
Nervous	r = 0.819
Inspired	r = 0.764
Stressed	r = 0.753
Excited	r = 0.718
Happy	**r = 0.659**

Mean Absolute Error (MAE): 0.0355

Classifying Songs by Vibe

- Ran audio files through OpenL3 model to retrieve detailed low-level audio features (BPM, valence, overall sound, etc.)
- Used K-Means clustering to map all songs to 7 distinct emotional categories (Stressed, Angry, Happy, Sad, Inspired, Excited, Nervous) and create emotion vectors on each song

Fusion + Spotify Integration

1. Predictions on input data (face, biometrics)

2. Map emotions of inputs to each other, and **combine into single input vector**

3. Run **cosine similarity** of combined input human emotion vector against individual song emotion vectors

Sample Outputs – Happy

*[0.000, 0.000, 0.000, **0.995**, 0.002, 0.000, 0.002]*

Predicted Happy

Biometric Input Vector

[-0.7123, -0.7659, ..., 0.3253, -0.4465]

*[-0.255, -0.298, **-0.251**, -0.288, -0.250, **-0.241**, -0.258]*

Predicted Excited/Happy

Song	Similarity
The Door - CYRIL Remix - Teddy Swims, CYRIL	0.800712
Kisses (feat. bbyclose) - BL3SS, CamrinWatsin, bbyclose	0.798134
BATTITO - Fedez	0.792892
Call Me When You Break Up (with Gracie Abrams) - Selena Gomez, benny blanco, Gracie Abrams	0.782491
505 - Arctic Monkeys	0.762684
MILLION DOLLAR BABY - Tommy Richman	0.758073
sobelove - Beéle	0.753966
GOOD CREDIT (with Kendrick Lamar) - Playboi Carti, Kendrick Lamar	0.751938
Scream & Shout - will.i.am, Britney Spears	0.751934
Diet Pepsi - Addison Rae	0.751828
Chiquita - Neton Vega, Tito Double P	0.749925
Gang Baby - NLE Choppa	0.748114
Forever Young - David Guetta, Alphaville, Ava Max	0.727988
Feather - Sabrina Carpenter	0.727280
I Wonder - Kanye West	0.726722
Nada Com Nada - Ao Vivo - Gustavo Mioto, Grupo Menos É Mais	0.723432
ANXIETY (feat. Doechii) - Sleepy Hallow, Doechii	0.722538
Anxiety - Doechii	0.716818
Sexy Bitch (feat. Akon) - David Guetta, Akon	0.706531
Something About You - Eyedress, Dent May	0.698444

Sample Outputs – Nervous / Stressed

*[0.186, 0.004, 0.255, 0.009, 0.186, **0.351**, 0.009]*

Predicted Fear

Biometric Input Vector

[-1.194, -0.783, ..., 0.189, 1.054]

*[**0.007**, -0.186, -0.233, -0.249, -0.234, -0.124, **0.006**]*

Predicted Nervous / Stressed

Song	Similarity
BATTITO - Fedez	0.847726
The Door - CYRIL Remix - Teddy Swims, CYRIL	0.835113
Kisses (feat. bbyclose) - BL3SS, CamrinWatsin, bbyclose	0.820183
Call Me When You Break Up (with Gracie Abrams) - Selena Gomez, benny blanco, Gracie Abrams	0.808726
Scream & Shout - will.i.am, Britney Spears	0.802241
505 - Arctic Monkeys	0.799782
sobelove - Beéle	0.799525
Chiquita - Neton Vega, Tito Double P	0.789706
ANXIETY (feat. Doechii) - Sleepy Hallow, Doechii	0.788605
I Wonder - Kanye West	0.787646
Diet Pepsi - Addison Rae	0.787504
Nada Com Nada - Ao Vivo - Gustavo Mioto, Grupo Menos É Mais	0.785544
Gang Baby - NLE Choppa	0.784830
Forever Young - David Guetta, Alphaville, Ava Max	0.776476
Anxiety - Doechii	0.774981
Something About You - Eyedress, Dent May	0.770619
GOOD CREDIT (with Kendrick Lamar) - Playboi Carti, Kendrick Lamar	0.769897
MILLION DOLLAR BABY - Tommy Richman	0.769731
Sexy Bitch (feat. Akon) - David Guetta, Akon	0.765638
Mãe Solteira - DG e Batidão Stronda, Mc Davi, J. Eskine, MC G15	0.753532

Manual Outputs – Excited / Surprise

*[0.007, 0.000, 0.008, 0.019, 0.028, 0.002, **0.937**]*

Predicted Surprise

Manual Excited Emotion Vector

*[0.100, 0.050, 0.100, 0.050, **0.950**, 0.800, 0.050]*

Song	Similarity
Too Much - Dove Cameron	0.847178
FXCK UP THE WORLD (feat. Future) - LISA, Future	0.842028
I Smoked Away My Brain - A$AP Rocky, Imogen Heap, Clams Casino	0.840972
Cópia Proibida - Léo Foguete	0.839458
Kiss Me Thru The Phone - Soulja Boy, Sammie	0.836395
Under Your Spell - Snow Strippers	0.830006
FOMDJ - Playboi Carti	0.827926
The Door - Teddy Swims	0.827648
blackout 🧊 - Emilia, TINI, Nicki Nicole	0.814923
CARNIVAL - ¥$, Kanye West, Ty Dolla $ign	0.806491
Shivers - Ed Sheeran	0.802609
i like the way you kiss me - Artemas	0.801658
Drag Me Down - One Direction	0.799887
Quevedo: Bzrp Music Sessions, Vol. 52 - Bizarrap, Quevedo	0.799581
Waiting For Love - Avicii	0.797821
BATTITO - Fedez	0.794299
Something About You - Eyedress, Dent May	0.793114
Nada Com Nada - Ao Vivo - Gustavo Mioto, Grupo Menos É Mais	0.765638
APT. - ROSÉ, Bruno Mars	0.753532

Where this could go

- Limitations to our project
  - Amount of data (not enough songs, accelerometer data is overwhelming)
  - Loose mapping rules (between inputs, and within song clusters)
  - Lack of contextual awareness (physical activity)
- Integration into wearables (e.g. Apple Watch, etc.)

References

[1] A. Ospina-Bohórquez, A. B. Gil-González, M. N. Moreno-García, and A. de Luis-Reboredo, "Context-Aware Music Recommender System Based on Automatic Detection of the User's Physical Activity," in *Distributed Computing and Artificial Intelligence, 17th International Conference (DCAI 2020)*, Y. Dong, E. Herrera-Viedma, K. Matsui, S. Omatsu, A. González Briones, and S. Rodríguez González, Eds., Cham: Springer, 2021, vol. 1237, pp. 142–151. doi: 10.1007/978-3-030-53036-5_15. SpringerLink

[2] J. S. Lee and J. C. Lee, "Music for My Mood: A Music Recommendation System Based on Context Reasoning," in *Smart Sensing and Context (EuroSSC 2006)*, P. Havinga, M. Lijding, N. Meratnia, and M. Wegdam, Eds., Berlin, Heidelberg: Springer, 2006, vol. 4272, pp. 190–203. doi: 10.1007/11907503_14. SpringerLink+1ACM Digital Library+1

[3] S. Wang, Y. Wang, and J. Wang, "A Context-Aware Music Recommendation System Using Fuzzy Logic," in *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, Beijing, China, Oct. 2020, pp. 123–128. doi: 10.1109/ICAICE51518.2020.00030.

[4] Spotify, "Spotify Web API," [Online]. Available: https://developer.spotify.com/documentation/web-api. [Accessed: Apr. 21, 2025].

[5] yt-dlp Contributors, "yt-dlp: A youtube-dl fork with additional features and fixes," GitHub, 2024. [Online]. Available: https://github.com/yt-dlp/yt-dlp. [Accessed: Apr. 21, 2025].

[6] J. Cramer, H. H. Wu, and J. Salamon, "Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2019. [Online]. Available: https://github.com/marl/openl3. [Accessed: Apr. 21, 2025].

[7] I. Goodfellow, D. Erhan, L. Carrier, A. Courville, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015. [FER2013 Dataset]. [Online]. Available: https://www.kaggle.com/datasets/msambare/fer2013. [Accessed: Apr. 21, 2025].

[8] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for Wearable Stress and Affect Detection," in *Proc. 20th ACM Int. Conf. on Multimodal Interaction*, Boulder, CO, 2018, pp. 400–408. [Online]. Available: https://ubicomp.eti.uni-siegen.de/home/datasets/. [Accessed: Apr. 21, 2025].

Thank you!

Questions?