Topic: Behavioral and Physiological Cues in Depression Detection: A Multimodal Machine Learning Perspective

Depression affects more than 300 million people around the world and is one of the leading causes of disability, with significant impacts on cognition, emotion, and daily functioning [1, 2]. The prevalence of depression continues to increase, while traditional diagnostic approaches are time-intensive, require significant clinician expertise, and are not always scalable due to limited availability of trained professionals [3]. These limitations underscore the urgent need for scalable and accessible screening solutions. Automated multimodal approaches could complement clinical practice by identifying at-risk individuals earlier and supporting referral to specialist care [4, 5, 6]. In this context, machine learning offers a promising alternative by enabling automated analysis of multimodal behavioral and physiological data, including speech, facial expressions, and textual content, thus supporting earlier detection, continuous monitoring, and reduced dependency on clinical judgment [7, 8, 9].

Depression detection using multimodal data, such as audio, facial expressions, and text, has been extensively explored in the literature [10, 11, 12]. Researchers often rely on feature extraction tools tailored to each modality. For audio signals, common approaches include extracting low-level descriptors like Mel-Frequency Cepstral Coefficients (MFCCs) or higher-level feature sets such as extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) using the Open-source Speech and Music Interpretation by Large-space Extraction (openSMILE) toolkit [13]. Facial behavior features, including action units, head pose, and eye gaze, are frequently derived using tools such as Open Dynamic Behavioral Measure (OpenDBM) [14]. In the linguistic domain, lexical and psycholinguistic features can be obtained through tools such as LIWC, while semantic and contextual features are captured using pretrained transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) [15]. These handcrafted features are typically modeled using classical machine learning algorithms such as Support Vector Machines (SVMs), Random Forests, or Gradient Boosting Decision Trees (GBDTs) [3, 11]. These models offer advantages in interpretability, robustness on small datasets, and low computational cost, making them well-suited for initial screening tasks or deployment in resource-limited environments. However, they often struggle to capture the intricate temporal patterns within each modality and the inter-dependencies across modalities [16].

Few studies have systematically compared early and late fusion strategies across multiple modalities using classical machine learning. Furthermore, many models have not been benchmarked on datasets that integrate biosignals with behavioral data [17, 18, 19]. There is a need for a scalable, interpretable, and clinically relevant approach to multimodal depression detection that provides for both emotional and physiological dimensions [20].

This thesis investigates the application of machine learning models for multimodal depression screening, with a focus on systematically comparing early and late fusion strategies. The aim is not to replace traditional diagnostic procedures, but to improve scalable screening tools that can support clinicians in early detection and referral. The analysis will be based on the EmpkinS-EKSpression dataset [21], collected as part of the CRC EmpkinS project, and will incorporate a diverse set of digital biomarkers across four modalities: video (facial behavior), audio (speech characteristics), text (interview transcripts), and physiological biosignals, including electrocardiography (ECG), electromyography (EMG), and Respiration Patterns (RSP).

To assess the performance and clinical relevance of the model, we used primary diagnostic labels derived from SCID-5-CV (structured clinical interview for DSM-5 disorders) [22]. In addition, we will evaluate alignment with clinically validated screening and severity scales, including the PHQ-9 (Patient Health Questionnaire-9) [23] and CES-D(Center for Epidemiologic Studies Depression Scale) [24], which are widely used for depression screening, as well as HRSD(Hamilton Rating Scale for Depression) [25], which is commonly applied for severity rating.

The proposed work consists of the following parts:

- **Feature Extraction:**
  - **Audio:** various features are extracted, including MFCCs and eGeMAPS, using the openSMILE toolkit, complemented by additional acoustic descriptors obtained through other feature extraction methods.
  - **Video:** OpenFace and OpenDBM are used to derive Action Units (AUs), head pose, eye gaze features, and facial landmarks.
  - **Text:** open-source large language models are utilized to extract semantic and contextual features from interview transcripts, with a focus on identifying depression-related linguistic patterns.
  - **Biosignals:** NeuroKit2 is employed to extract statistical, morphological, and frequency-domain features from ECG, EMG, and RSP signals, including heart rate variability metrics, EMG amplitude and frequency measures, and respiration rate characteristics.
- **Machine Learning Development:**
  - Train machine learning models (e.g., SVM, Random Forests, GBDTs).
  - Compare early fusion (concatenating features from all modalities before training) with late fusion (training separate models per modality and combining their predictions via ensembling).
- **Evaluation Strategy:**
  - For regression tasks (e.g., predicting PHQ-9 scores): evaluate performance using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).
  - For classification tasks (e.g., binary depression status): report Accuracy, F1-score, precision, recall, and Area Under the Curve (AUC).

Advisors: Misha Sadeghi M.Sc., Mahdis Habibpourfatideh M.Sc., Prof. Dr. Björn M. Eskofier
(Machine Learning and Data Analytics Lab (MaDLab), FAU)
Lydia Helene Rupp M.Sc., Prof. Dr. Matthias Berking
(Lehrstuhl für Klinische Psychologie und Psychotherapie (KliPs), FAU)
Student: Yasaman Moradi Fard
Start: 01.09.2025 – 01.03.2026

References

[1] Zahraa Al Sahili, Ioannis Patras und Matthew Purver. "Multimodal Machine Learning in Mental Health: A Survey of Data, Algorithms, and Challenges". In: *arXiv preprint arXiv:2407.16804* (2024).

[2] Ming Fang u. a. "A multimodal fusion model with multi-level attention mechanism for depression detection". In: *Biomedical Signal Processing and Control* 82 (2023), S. 104561.

[3] Lin Sze Khoo u. a. "Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches". In: *Sensors* 24.2 (2024), S. 348.

[4] Hamdi Dibeklioğlu, Zakia Hammal und Jeffrey F Cohn. "Dynamic multimodal measurement of depression severity using deep autoencoding". In: *IEEE journal of biomedical and health informatics* 22.2 (2017), S. 525–536.

[5] Sahar Harati u. a. "Depression severity classification from speech emotion". In: *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2018, S. 5763–5766.

[6] World Health Organization u. a. “Global health estimates: depression and other common mental disorders”. In: (2017).

[7] Sharifa Alghowinem u. a. “Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors”. In: *IEEE Transactions on Affective Computing* 9.4 (2016), S. 478–490.

[8] Sahar Harati u. a. “Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression”. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, S. 2254–2257.

[9] Le Yang u. a. “Hybrid depression classification and estimation from audio video and text information”. In: *Proceedings of the 7th annual workshop on audio/visual emotion challenge*. 2017, S. 45–51.

[10] Mariia Nykoniuk u. a. “Multimodal data fusion for depression detection approach”. In: *Computation* 13.1 (2025), S. 9.

[11] Anastasia Pampouchidou u. a. “Automatic assessment of depression based on visual cues: A systematic review”. In: *IEEE Transactions on Affective Computing* 10.4 (2017), S. 445–470.

[12] Zhenwei Zhang u. a. “Multimodal sensing for depression risk detection: Integrating audio, video, and text data”. In: *Sensors* 24.12 (2024), S. 3714.

[13] Florian Eyben, Martin Wöllmer und Björn Schuller. “Opensmile: the munich versatile and fast open-source audio feature extractor”. In: *Proceedings of the 18th ACM international conference on Multimedia*. 2010, S. 1459–1462.

[14] Tadas Baltrušaitis, Peter Robinson und Louis-Philippe Morency. “Openface: an open source facial behavior analysis toolkit”. In: *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2016, S. 1–10.

[15] Jacob Devlin u. a. “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019, S. 4171–4186.

[16] Muhammad Muzammel, Hanan Salam und Alice Othmani. “End-to-end multimodal clinical depression recognition using deep neural networks: A comparative analysis”. In: *Computer Methods and Programs in Biomedicine* 211 (2021), S. 106433.

[17] Kaining Mao u. a. “Prediction of depression severity based on the prosodic and semantic features with bidirectional LSTM and time distributed CNN”. In: *IEEE transactions on affective computing* 14.3 (2022), S. 2251–2265.

[18] Michel Valstar u. a. “Avec 2016: Depression, mood, and emotion recognition workshop and challenge”. In: *Proceedings of the 6th international workshop on audio/visual emotion challenge*. 2016, S. 3–10.

[19] Arnab Kumar Das und Ruchira Naskar. “A deep learning model for depression detection based on MFCC and CNN generated spectrogram features”. In: *Biomedical Signal Processing and Control* 90 (2024), S. 105898.

[20] Meiling Li u. a. “Enhancing multimodal depression detection with intra-and inter-sample contrastive learning”. In: *Information Sciences* 684 (2024), S. 121282.

[21] Marie Keinert u. a. “Facing depression: evaluating the efficacy of the EmpkinS-EKSpression reappraisal training augmented with facial expressions–protocol of a randomized controlled trial”. In: *BMC psychiatry* 24.1 (2024), S. 896.

[22] Michael B First u. a. “Structured clinical interview for DSM-5 disorders”. In: *Clinician Version (SCID-5-CV)* (2015).

[23] Kurt Kroenke, Robert L Spitzer und Janet BW Williams. "The PHQ-9: validity of a brief depression severity measure". In: *Journal of general internal medicine* 16.9 (2001), S. 606–613.

[24] Peter M Lewinsohn u. a. "Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults." In: *Psychology and aging* 12.2 (1997), S. 277.

[25] Max Hamilton. "The Hamilton rating scale for depression". In: *Assessment of depression*. Springer, 1986, S. 143–152.