

Nom et prénom	Etablissement	Grade	Qualité
Pr. BOSSOUFI Badre	Faculté des Sciences Dhar El Mahraz, Fès	PES	Président
Pr. EL KHOUKHI Fatima	Faculté des Lettres et des Sciences Humaines, Meknès	PES	Rapporteur
Pr. EL OAHBI Rachid	Faculté des Lettres et des Sciences Humaines, Meknès	PES	Rapporteur
Pr. RIFFI Jamal	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Rapporteur
Pr. OUSSAMA Mohamed Reda	Faculté des Sciences, Rabat	PES	Examineur
Pr. YAHYAOUY Ali	Faculté des Sciences Dhar El Mahraz, Fès	PES	Examineur
Pr. AARAB Abdellah	Faculté des Sciences Dhar El Mahraz, Fès	PES	Expert
Pr. SABRI My Abdelouahed	Faculté des Sciences Dhar El Mahraz, Fès	PES	Directeur de thèse

Mots-clés : Intelligence artificielle ; détection de la dépression ; apprentissage profond ; traitement automatique du langage ; analyse du comportement facial ; fusion multimodale ; modélisation temporelle.



Artificial Intelligence for Depression Detection: Computer Vision, Speech, Language, and Multimodal Behavioral Analysis

Abstract

Depression represents a major public health challenge, and its identification remains complex due to the diversity of symptoms and their variability across individuals. While clinical interviews and standardized questionnaires remain essential for assessment, advances in artificial intelligence offer new opportunities for analyzing linguistic, vocal, facial, and behavioral cues associated with depression.

This thesis investigates deep learning approaches for automated depression detection using heterogeneous data. It aims to develop informative representations and architectures capable of exploiting modality complementarity and temporal dependencies in observed behaviors. The research is organized into five complementary contributions.

The first contribution proposes a video-based facial emotion recognition approach using transfer learning and prediction aggregation. Evaluated on RAVDESS, it provides a methodological foundation for facial expression analysis, a task distinct from depression detection.

The second contribution investigates linguistic indicators of depression through different language representation and modeling approaches, ranging from lexical methods and recurrent neural networks to Transformers and large language models.

The third contribution develops a depression detection approach based on three-dimensional facial landmarks, transformed into pseudo-image sequences and processed using a CNN–LSTM architecture to capture spatial and temporal information.

The fourth contribution proposes a multimodal fusion framework integrating visual, acoustic, behavioral, and textual information. An adaptive attention mechanism weights the different modalities to exploit their complementarity.

The fifth contribution introduces MHTCI-Net, a hybrid architecture combining Transformers and BiGRU networks to model global and local temporal dependencies. It incorporates a temporal cross-interaction mechanism, adaptive representation refinement, and attention-based aggregation. Evaluations on DAIC-WOZ, E-DAIC, and LMVD, supported by comparative experiments and ablation studies, provide insights into the effectiveness of the investigated approaches.

Overall, this research contributes to the development of computational methods for analyzing depression-related manifestations. It opens new perspectives for tools supporting depression screening and assessment, subject to further external and clinical validation.

Keywords: Artificial intelligence; depression detection; deep learning; natural language processing; facial behavior analysis; multimodal fusion; temporal modeling.