Topic: Depression Detection Using ML/DL on 3D Full-Body Skeleton Data

Globally, approximately 350 million individuals suffer from depression, with the condition often persisting for several years. When considering both disability and mortality, depression ranks as the ninth leading cause of overall disease burden [1]. Beyond its significant impact on mortality, depression is associated with a range of comorbid conditions, including anxiety, sleep disturbances, and other mental health disorders. The impact of depression can vary in severity, affecting not only the individuals experiencing it but also their families and social circles [2].

Multiple studies suggest that early detection and timely intervention are among the most effective strategies for managing depression, as they facilitate prompt initiation of treatment and improve patient outcomes [4–6]. Despite being a major global health burden, the accuracy of clinical diagnosis for depression remains relatively low. One possible explanation for this challenge is the heterogeneous nature of depression, which has not yet been fully elucidated [7]. Recent studies have highlighted the strong potential of machine learning (ML) and deep learning (DL) techniques in accurately identifying complex and heterogeneous conditions, including those with characteristics similar to depression [8].

Mao et al. [9] conducted a systematic review on automated clinical depression diagnosis, revealing that the majority of studies in this field have primarily relied on speech, text, and facial expression analysis to identify depressive symptoms. However, the correlation between depression and body movements has been relatively underexplored, and the accuracy of existing studies in this area remains suboptimal.

A study by Joshi et al. [10] highlights that upper-body expressions, gestures, and head movements can serve as significant features in automated depression detection, offering an alternative to traditional facial dynamics. Liu et al. [16] showed that analyzing body movements in a multimodal setup improves depression detection accuracy. Extracting human poses from video data becomes essential for capturing and analyzing movement patterns. Horigome et al. [14] confirmed that RGB videos are effective for deriving both 2D and 3D body poses in depression detection, while Lincke et al. [15] showed that 3D skeleton analysis offers notable improvements over 2D methods in similar contexts.

In this thesis, we train, test, and optimize a range of deep learning models using 3D human poses extracted from the Empkins D02 dataset. Our video dataset is significantly larger than those used in comparable studies. To support effective training, we introduce a systematic method, similar to Relative Parts Movement (RPM) [13] and Action Entropy [16], for quantifying and representing body movements in a form suitable for ML/DL input. Furthermore, we define an output metric similar to Behavioral Depression Degree (BDD) [16] to represent depression severity in alignment with clinical standards such as the Hamilton Depression Rating Scale (HAM-D).

Early research in depression detection primarily focused on traditional machine learning techniques such as Random Forest (RF), Support Vector Machines (SVM) [14], and K-Nearest Neighbors (KNN) [11]. More recent studies, particularly those analyzing head movements, have explored the potential of deep learning methods, including Convolutional Neural Networks (CNN) [16], Temporal Convolutional Networks (TCN) [12], and Transformer-based models. There is also a growing trend toward hybrid architectures [3] and multimodal approaches [16]. However, to date, no study has exclusively applied deep learning techniques to analyze full-body 3D human poses for assessing depression severity. This thesis therefore introduces a novel approach to address this gap.

The Empkins D02 dataset comprises nearly 300 RGB videos of free style interviews conducted in real clinical settings, along with corresponding depression severity scores based on the

HAM-D scale. This expanded dataset is expected to enhance model accuracy and robustness in depression detection, addressing the limitations of prior studies that often relied on relatively small video datasets for training.

The proposed work consists of the following parts:

- Complete a literature review on 2D pose estimation, 3D lifting, and movement based depression detection by Feb 29.
- Extract 2D poses from video data using ViTPose, lift to 3D with MotionBERT, and clean the data by Mar 25.
- Define a movement metric for quantifying skeletal movements in a 3D plane and a depression severity score aligned with HAMD by Apr 15.
- Train 2 ML and 3 DL models to predict depression severity from 3D poses, and evaluate using MAE, RMSE, R^2 with 5-fold cross-validation by May 15.
- Optimize models and improve performance by June 15.

The thesis must contain extended research on literature, existing patents and related work in the corresponding areas that has to be performed. The thesis will include detailed descriptions of all developed and employed algorithms, along with a thorough evaluation and discussion of the results. The implemented code has to be documented and provided.

Advisors: Amirreza Asemanrafat M.Sc., Misha Sadeghi M.Sc., Prof. Dr. Bjoern Eskofier
Student: Shanaka Anuradha Samarakoon Samarakoon Mudiyanselage Adikaramlage Walawwe
Start—End: 15.02.2025 - 15.08.2025

References

[1] K. Smith, "Mental health: A world of depression," Nature, vol. 515, pp. 180–181, 2014. DOI: 10.1038/515180a.

[2] S. M. Stahl, "Does depression hurt?," J. Clin. Psychiatry, vol. 63, no. 4, pp. 273-304, Apr. 2002.

[3] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Measurement: Sensors*, vol. 25, p. 100587, 2023, doi: 10.1016/j.measen.2022.100587.

[4] A. Halfin, "Depression: the benefits of early and appropriate treatment," Am. J. Manag. Care, vol. 13, no. 4 Suppl, pp. S92–S97, Nov. 2007.

[5] A. Picardi, I. Lega, L. Tarsitani, M. Caredda, G. Matteucci, M. P. Zerella, and the SET-DEP Group, "A randomized controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care," J. Affect. Disord., vol. 198, pp. 96–101, Dec. 1, 2016.

[6] K. Rost, J. L. Smith, and M. Dickinson, "The effect of improving primary care depression management on employee absenteeism and productivity: A randomized trial," Med. Care, vol. 42, no. 12, pp. 1202–1210, Dec. 2004.

[7] X. Liu and K. Jiang, "Why is diagnosing MDD challenging?," Shanghai Arch. Psychiatry, vol. 28, no. 6, pp. 343–345, Dec. 25, 2016. DOI: 10.11919/j.issn.1002-0829.216073. PMID: 28638210; PMCID: PMC5434292.

[8] T. Naito, K. Inoue, S. Namba, et al., "Machine learning reveals heterogeneous associations between environmental factors and cardiometabolic diseases across polygenic risk scores," Commun. Med., vol. 4, p. 181, 2024. DOI: 10.1038/s43856-024-00596-7.

[9] K. Mao, Y. Wu, and J. Chen, "A systematic review on automated clinical depression diagnosis," npj Ment. Health Res., vol. 2, p. 20, 2023. DOI: 10.1038/s44184-023-00040-z.

[10] J. Joshi, R. Goecke, G. Parker, and M. Breakspear, "Can body expressions contribute to automatic depression analysis?," in Proc. 10th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), Shanghai, China, 2013, pp. 1–7. DOI: 10.1109/FG.2013.6553796.

[11] Z. Guo, N. Ding, M. Zhai, Z. Zhang, and Z. Li, "Leveraging domain knowledge to improve depression detection on Chinese social media," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1528–1536, 2023.

[12] Y. Yu, W. Li, Y. Zhao, J. Ye, Y. Zheng, X. Liu, and Q. Wang, "Depression and severity detection based on body kinematic features: Using Kinect recorded skeleton data of simple action," Frontiers in Neurology, vol. 13, p. 905917, 2022, doi: 10.3389/fneur.2022.905917. arXiv:1906.08172, 2019. [Online]. Available: https://arxiv.org/abs/1906.08172.

[13] J. Joshi, A. Dhall, R. Goecke, and J. F. Cohn, "Relative body parts movement for automatic depression analysis," in Proc. 2013 Humaine Assoc. Conf. Affective Comput. Intell. Interact. (ACII), Geneva, Switzerland, 2013, pp. 492–497. DOI: 10.1109/ACII.2013.87.

[14] T. Horigome, B. Sumali, M. Kitazawa, M. Yoshimura, K.-C. Liang, Y. Tazawa, T. Fujita, M. Mimura, and T. Kishimoto, "Evaluating the severity of depressive symptoms using upper body motion captured by RGB-depth sensors and machine learning in a clinical interview setting: A preliminary study," Compr. Psychiatry, vol. 98, p. 152169, 2020. DOI: 10.1016/j.comppsych.2020.152169.

[15] A. Lincke, C. Fagerström, M. Ekstedt, W. Löwe, and S. Backåberg, "A comparative study of the 2D- and 3D-based skeleton avatar technology for assessing physical activity and functioning among healthy older adults," Health Informatics J., vol. 29, no. 4, pp. 14604582231214589, Oct.–Dec. 2023. DOI: 10.1177/14604582231214589. PMID: 37924210.

[16] D. Liu, B. Liu, T. Lin, G. Liu, G. Yang, D. Qi, Y. Qiu, Y. Lu, Q. Yuan, S. C. Shuai, X. Li, O. Liu, X. Tang, J. Shuai, Y. Cao, and H. Lin, "Measuring depression severity based on facial expression and body movement using deep convolutional neural network," Front. Psychiatry, vol. 13, 2022. DOI: 10.3389/fpsyt.2022.1017064.