[1] A. Vita, S. Barlati, L. D. Peri, G. Deste, and E. Sacchetti, "Schizophrenia," *The Lancet,* vol. 388, 2016.

[2] G. Arbanas, "Diagnostic and Statistical Manual of Mental Disorders (DSM-5)," *Alcoholism and psychiatry research,* vol. 51, pp. 61-64, 2015.

[3] T. Gonzalez and C. Chiodo, "ICD 10," *Foot & Ankle International,* vol. 36, pp. 1110-1116, 2015.

[4] C. F. Benitez-Quiroz, R. Srinivasan, and A. Martínez, "EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5562-5570, 2016.

[5] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," *IEEE Transactions on Image Processing,* vol. 28, pp. 2439-2450, 2019.

[6] S. Li, W. Deng, and J. Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," in 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 2584-2593, 2017.

[7] X. Feng, "Facial expression recognition based on local binary patterns and coarse-to-fine classification," in *The Fourth International Conference on Computer and Information Technology (CIT),* pp. 178-183, 2004.

[8] L. Zhong, Q. Liu, P. Yang, J. Huang, and D. N. Metaxas, "Learning Multiscale Active Facial Patches for Expression Analysis," *IEEE Transactions on Cybernetics,* vol. 45, no. 8, pp. 1499-1510, 2015.

[9] S. Eleftheriadis, O. Rudovic, and M. Pantic, "Discriminative Shared Gaussian Processes for Multiview and View-Invariant Facial Expression Recognition," *IEEE Transactions on Image Processing,* vol. 24, no. 1, pp. 189-204, 2015.

[10] M. Singh, B. B. Naib, Aand A. K. Goel, "Facial Emotion Detection using Action Units," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1037-1041, 2020.

[11] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, pp. 91-110, 2004.

[12] A. Kläser, M. Marszalek, and C. Schmid, "A Spatio-Temporal Descriptor Based on 3D-Gradients," in *British Machine Vision Conference (BMVC)*, 2008.

[13] P. Ekman, "An argument for basic emotions," *Cognition & Emotion,* vol. 6, pp. 169-200, 1992.

[14] E. K. Gray and D. Watson, "Assessing positive and negative affect via self-report," *Handbook of Emotion Elicitation and Assessment*, 2007.

[15] K. Schindler, L. Gool, and B. Gelder, "Recognizing emotions expressed by body pose: A biologically inspired neural model," *Neural networks : the official journal of the International Neural Network Society,* vol. 21 9, pp. 1238-46, 2008.

[16] C. Chen, Z. Wu, and Y.-G. Jiang, "Emotion in Context: Deep Semantic Feature Fusion for Video Emotion Recognition," in *Proceedings of the 24th ACM international conference on Multimedia,* 2016.

[17] R. Kosti, J. M. Alvarez, A. Recasens, and À. Lapedriza, "Emotion Recognition in Context," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1960-1968, 2017.

[18] J. Lee, S. Kim, S. Kim, J.-I. Park, and K. Sohn, "Context-Aware Emotion Recognition Networks," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10142-10151, 2019.

[19] T. Mittal, P. Guhan, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "EmotiCon: Context-Aware Multimodal Emotion Recognition Using Freg's Principle," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 14222-14231, 2020.

[20] M. Valstar, B. Schuller, K. Smith, F. Evben, B. Jiang, S. Bilakhia, S. Schieder, R. Cowie, and M. Pantic, "AVEC 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, Barcelona, Spain, 2013.

[21] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Evben, B. Jiang, J. Krajewski, R. Cowie, and M. Pantic, "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[22] A. Beck, R. Steer, R. Ball, and W. Ranieri, "Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients," *Journal of Personality Assessment,* vol. 67 3, pp. 588-597, 1996.

[23] N. Cummins, J. Joshi, A. Dhall, V. Sethu, R. Goecke, and J. Epps, "Diagnosis of depression by behavioural signals: a multimodal approach," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge,* 2013.

[24] L. Yang, D. Jiang, W. Han, and H. Sahli, "DCNN and DNN based multi-modal depression recognition," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII),* pp. 484-489, 2017.

[25] Y. Zhu, Y. Shang, Z. Shao, and G. Guo, "Automated Depression Diagnosis Based on Deep Networks to Encode Facial Appearance and Dynamics," *IEEE Transactions on Affective Computing,* vol. 9, pp. 578-584, 2018.

[26] M. Ding, Y. Huo, J. Hu, and Z. Lu, "DeepInsight: Multi-Task Multi-Scale Deep Learning for Mental Disorder Diagnosis," in *British Machine Vision Conference (BMVC)*, 2018.

[27] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS),* 2012.

[28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F.-F. Li "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision,* vol. 115, no. 3, pp. 211-252, 2015.

[29] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *European Conference on Computer Vision (ECCV)*, pp. 818-833, 2014.

[30] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 580-587, 2014.

[31] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *ArXiv*, vol. abs/2004.10934, 2020.

[32] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[33] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional Two-Stream Network Fusion for Video Action Recognition," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1933-1941, 2016.

[34] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, C. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[37] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks,* vol. 2, no. 5, pp. 359-366, 1989.

[38] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 6, pp. 107-116, 1998.

[39] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[40] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[41] G. Li, M. Muller, A. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs Go As Deep As CNNs?," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV),* pp. 9266-9275, 2019.

[42] Q. Xu, X. Sun, C. Y. Wu, P. Wang, and U. Neumann, "Grid-GCN for Fast and Scalable Point Cloud Learning," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5660-5669, 2020.

[43] J. Chen, B. Lei, Q. Song, H. Ying, D. Z. Chen, and J. Wu, "A Hierarchical Graph Network for 3D Object Detection on Point Clouds," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 389-398, 2020.

[44] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[45] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *European Conference on Computer Vision (ECCV)*, pp. 499-515, 2016.

[46] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6738-6746, 2017.

[47] H. Wang *et al.*, "CosFace: Large Margin Cosine Loss for Deep Face Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5265-5274, 2018.

[48] J. Deng and S. Zafeririou, "ArcFace for Disguised Face Recognition," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 485-493, 2019.

[49] K. Zhao, J. Xu, and M. Cheng, "RegularFace: Deep Face Recognition via Exclusive Regularization," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1136-1144, 2019.

[50] C. Wu, R. Manmatha, A. J. Smola, and P. Krähenbühl, "Sampling Matters in Deep Embedding Learning," in *2017 IEEE International Conference on Computer Vision (ICCV)* , pp. 2859-2867, 2017.

[51] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815-823, 2015.

[52] A. Hermans, L. Beyer, and B. J. A. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," *ArXiv*, vol. abs/1703.07737, 2017.

[53] K. Sohn, "Improved Deep Metric Learning with Multi-class N-pair Loss Objective," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1857-1865, 2016.

[54] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep Metric Learning via Lifted Structured Feature Embedding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4004-4012, 2016.

[55] X. Wang, Y. Hua, E. Kodirov, G. Hu, R. Garnier, and N. Robertson, "Ranked List Loss for Deep Metric Learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5202-5211, 2019.

[56] X. Wang, X. Han, W. Huang, D. Dong, and M. Scott, "Multi-Similarity Loss With General Pair Weighting for Deep Metric Learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 5017-5025, 2019.

[57] D. Yi, Z. Lei, and S. Li, "Deep Metric Learning for Practical Person Re-Identification," *ArXiv,* vol. abs/1407.4979, 2014.

[58] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zhang, Z. Wang. and Y. Wei, "Circle Loss: A Unified Perspective of Pair Similarity Optimization,"in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 6397-6406, 2020.

[59] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[60] Y.-J. Li, F.-E. Yang, Y.-C. Liu, Y.-Y. Yeh, X. Du, and Y.-C. Wang, "Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 172-178, 2018.

[61] M. Geng, Y. Wang, T. Xiang, and Y. Tian, "Deep transfer learning for person re-identification," *ArXiv, vol. abs/1611.05244,* 2016.

[62] P. Peng, T. Xiang, Y. Wang, and M. Pontil, "Unsupervised cross-dataset transfer learning for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1306-1315, 2016.

[63] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. V. Gool, "Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 777-784, 2011.

[64] J. Redmon, S. Divvala, R. B. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 779-788, 2016.

[65] E. Svetieva and M. G. Frank, "Empathy, emotion dysregulation, and enhanced microexpression recognition ability," *Motivation and Emotion,* vol. 40, pp. 309-320, 2016.

[66] C. Hurley, A. E. Anker, M. G. Frank, D. Matsumoto, and H. C. Hwang, "Background factors predicting accuracy and improvement in micro expression recognition," *Motivation and Emotion,* vol. 38, pp. 700-714, 2014.

[67] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature," *Geoscientific Model Development,* vol. 7, pp. 1247-1250, 2014.

[68] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4489-4497, 2015.

[69] M. Sidorov and W. Minker, "Emotion Recognition and Depression Diagnosis by Acoustic and Visual Features: A Multimodal Approach," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014.

[70] H. Espinosa, H. Escalante, L. Pineda, M. Montes-y-Gómez, D. Pinto, and V. Reyes-Meza, "Fusing Affective Dimensions and Audio-Visual Features from Segmented Video for Depression Recognition: INAOE-BUAP's Participation at AVEC'14 Challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014.

[71] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang, and S. Turabzadeh, "Automatic Depression Scale Prediction using Facial Expression Dynamics and Regression," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014.

[72] H. Kaya, F. Çilli, and A. A. Salah, "Ensemble CCA for Continuous Emotion Prediction," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014.

[73] M. A. Jazaery and G. Guo, "Video-Based Depression Level Analysis by Encoding Deep Spatiotemporal Features," *IEEE Transactions on Affective Computing,* pp. 1-1, 2018.