

Identifikasi Depresi Berdasarkan Ekspresi Wajah Menggunakan Arsitektur VGGNet

Ipram Zanuar Prakhastama¹, Fajar Astuti Hermawati²

Program Studi Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya

Jl. Semolowaru No.45, Menur Pumpungan, Surabaya 60118, Jawa timur, Indonesia

Email: ¹ipramzanuar12@gmail.com, ²emaildua@blabla.edu, ³emailtiga@blabla.edu

Abstract. *Depression is a common mental health disorder that affects various aspects of an individual's life worldwide. Many cases of depression go undiagnosed in a timely manner due to traditional methods such as clinical interviews and questionnaires, which are time-consuming and prone to subjective bias. With the advancement of technology, computer vision-based approaches have made it increasingly possible to detect depression automatically, efficiently, and objectively. This research explores the VGG19 model to detect depression through facial expression images. The implementation uses the FER2013 dataset, which is re-categorized into two classes: depression (anger, disgust, fear, sadness) and non-depression (neutral, happiness, surprise). The implementation of the best model achieved a training accuracy of 78% and a validation accuracy of 77% on test data. The model will be implemented in the form of a web-based application to assist in the visual and interactive classification of depression.*

Keywords: *Detect Depression, Facial Expression, CNN, VGG19, Computer Vision.*

Abstrak. *Depresi merupakan gangguan kesehatan yang paling umum menyerang aspek kehidupan dari individu di dunia. Banyak kasus pasien depresi tidak terdiagnosis tepat waktu karena metode tradisional seperti wawancara klinis dan kuesioner yang memakan waktu serta rentan bias subjektif. Seiring perkembangan teknologi, pendekatan berbasis computer vision semakin memungkinkan deteksi depresi dilakukan secara otomatis, efisien, dan objektif. Penelitian ini mengeksplorasi model VGG19 untuk mendeteksi depresi melalui citra ekspresi wajah. Implementasi menggunakan dataset FER2013, yang dikategorikan ulang menjadi dua kelas: depresi (marah, jijik, takut, sedih) dan tidak depresi (netral, bahagia, terkejut). Hasil implementasi dari model terbaik menghasilkan nilai akurasi training sebesar 78% dan nilai validasi terhadap data uji sebesar 77%. Penerapan model ini akan diimplementasikan dalam bentuk aplikasi berbasis website untuk membantu proses klasifikasi depresi secara visual dan interaktif.*

Kata Kunci: *Deteksi Depresi, Ekspresi Wajah, CNN, VGG19, Computer Vision.*

1. Pendahuluan

Kesehatan mental merupakan nilai intrinsik dan instrumental yang sangat penting bagi kehidupan semua orang. Pada tahun 2019, 970 juta orang di seluruh dunia hidup dengan gangguan kesehatan mental, dengan gangguan kecemasan dan depresi yang paling umum terjadi [1]. Banyak penderita depresi tidak memperoleh pengobatan yang tepat waktu dan efektif akibat kegagalan diagnosis, karena metode tradisional seperti wawancara klinis dan kuesioner cenderung subjektif, serta memerlukan waktu dan sumber daya yang besar [2].

Gejala depresi kerap tercermin melalui ekspresi wajah seperti kesedihan, kelelahan, dan gangguan tidur, sehingga ekspresi wajah menjadi indikator penting dalam mengenali kondisi tersebut [3]. Kemajuan teknologi telah membuka peluang baru untuk mendeteksi depresi dengan cara yang lebih efisien dan objektif dengan teknologi *computer vision*. Deteksi depresi berbasis computer vision mampu menganalisis ekspresi wajah melalui data gambar atau video, dan berbagai penelitian sebelumnya telah menunjukkan efektivitas pendekatan ini dalam mengidentifikasi gangguan depresi. [4].

Penerapan machine learning telah dilakukan untuk mendeteksi subjek "kondisi mental yang berisiko mengalami psikosis" (ARMS) dari data video rekaman wajah [5]. Analisis terkomputerisasi juga diimplementasikan terhadap data ekspresi wajah dari rekaman video untuk mengungkap keterkaitan antara ekspresi wajah dan penilaian klinis mengenai *blunted facial affect*

pada individu dengan gangguan mental serius [6]. Model VGGNet dan *Fuzzy Logic* yang dikombinasikan dengan CNN mendapat nilai akurasi yang cukup baik dalam mendeteksi depresi pada citra wajah [4] [3]. *Part-and-Relation Attention Network (PRA-Net)* untuk mendeteksi depresi secara otomatis berhasil mencapai MAE dan RMSE yang lebih rendah dan mencapai tingkat akurasi yang lebih tinggi [2], serta model *two-stream feature fusion model (TSSFM)* juga mencapai nilai akurasi yang baik [7].

Penelitian ini bertujuan mengaplikasikan arsitektur model VGGNet untuk mendeteksi depresi melalui citra ekspresi wajah, sebagai upaya mengatasi keterbatasan metode diagnosis tradisional maupun klinis. Model ini memanfaatkan citra wajah sebagai input utama yang dianalisis untuk mengidentifikasi pola-pola ekspresi yang khas terkait gejala depresi, seperti menurunnya intensitas ekspresi wajah. Hasil dari pemrosesan ini berupa klasifikasi yang menunjukkan apakah seseorang termasuk dalam kategori depresi atau tidak, sehingga diharapkan dapat menjadi alat bantu bagi tenaga klinis dalam proses diagnosis.

2. Tinjauan Pustaka

Penelitian ini menggabungkan Fusion Fuzzy Logic (FFL) dengan Convolutional Neural Network (CNN) untuk mendeteksi depresi melalui ekspresi wajah. Model memanfaatkan data gambar, teks, video, dan audio, serta memproses beberapa frame video secara berurutan untuk meningkatkan akurasi deteksi. Hasil pengujian menunjukkan bahwa pendekatan ini mampu mengidentifikasi depresi dengan akurasi tinggi, dibuktikan melalui evaluasi menggunakan metrik seperti RMSE, MSE, MAE, persentase error, dan skor F1 [2].

Studi ini menerapkan algoritma *deep learning*, khususnya model CNN berbasis Keras, untuk mendeteksi, mengenali wajah, dan mengklasifikasikan emosi seperti bahagia, netral, marah, sedih, jijik, dan terkejut. Proses dilakukan dalam tiga tahap yaitu deteksi wajah menggunakan *Haar Cascade*, analisis fitur wajah berdasarkan *database*, dan klasifikasi emosi. Sistem ini dibangun dengan bantuan *OpenCV*, dataset gambar, dan *Python*, serta diuji secara *real-time* pada beberapa mahasiswa [8].

Deteksi emosi dengan satu model sering kali kurang akurat, sementara penggunaan banyak model dapat menyebabkan keterlambatan dan kesalahan jika data tidak difilter dengan baik. Studi ini mereview 15 metode ekspresi wajah dan 17 metode suara dalam pengenalan emosi, membandingkan kelebihan, kekurangan, serta akurasinya, termasuk penggunaan metode seperti CNN, KNN, dan SVM dalam model tunggal maupun gabungan. Tujuannya adalah membantu peneliti baru memilih metode yang paling tepat untuk deteksi emosi berbasis wajah dan/atau suara [9].

Penelitian ini mengusulkan metode deteksi depresi otomatis berbasis audio menggunakan model *hybrid* yang menggabungkan CNN untuk ekstraksi fitur dan SVM untuk klasifikasi. Menggunakan dataset DAIC-WOZ dari AVEC 2016, model ini mencapai akurasi 68%, lebih tinggi dibanding CNN murni (58,57%), serta menunjukkan peningkatan signifikan dalam akurasi, presisi, dan *recall* dibanding studi audio-based sebelumnya [10].

Studi ini mengusulkan metode klasifikasi kondisi mental berdasarkan pengenalan emosi wajah menggunakan VGGNet. Prosesnya mencakup pra-pemrosesan, ekstraksi fitur dengan PCA dan VGGNet, serta klasifikasi menggunakan SVM dan MLP. Dengan menguji beberapa model pada dataset FER2013, model terbaik kombinasi VGGNet dan SVM mencapai akurasi 66%, menunjukkan efektivitas pendekatan ini dalam mendeteksi kondisi kesehatan mental melalui ekspresi wajah [11].

Studi ini mengembangkan model hybrid CNN-RNN untuk mendeteksi emosi dari ekspresi wajah dalam video, menggunakan *Emotional Wearable Dataset 2020* yang mencakup emosi baru seperti *amusement*, *enthusiasm*, *awe*, dan *liking*. Tiga model diuji yaitu MobileNetV2-RNN, InceptionV3-RNN, dan CNN-RNN kustom, dengan akurasi masing-masing 59%, 66%, dan 63% [12].

Studi ini membangun dataset besar dari rekaman audio-video dan ekspresi wajah 3D menggunakan *Kinect*, diperoleh melalui lima tugas pemicu emosi di lingkungan rumah sakit jiwa. Metode deteksi risiko depresi yang dikembangkan menggunakan dua model *Deep Belief Network*

(DBN), satu untuk fitur wajah 2D dan satu untuk fitur dinamis 3D. Hasil evaluasi menunjukkan bahwa kombinasi model 2D dan 3D lebih akurat dibanding model tunggal, terutama dalam respons terhadap stimulus emosional positif dan negatif, dengan tingkat akurasi lebih tinggi pada perempuan [13].

Studi ini mengembangkan metode klasifikasi biner berbasis citra wajah 2D untuk membedakan penderita depresi dari individu sehat. Lima model CNN-FCN, VGG11, VGG19, ResNet50, dan InceptionV3 dilatih menggunakan 1020 gambar pasien dan 1100 gambar kontrol. Model FCN menunjukkan akurasi tertinggi sebesar 98,23%, diikuti VGG19 (97,35%) dan InceptionV3 (97,10%). Hasilnya membuktikan efektivitas CNN dalam deteksi depresi secara cepat dan akurat [14].

Studi ini mengusulkan DLGA-CNN, kerangka kerja berbasis deep learning yang menggabungkan CNN dengan mekanisme perhatian (*attention*) dan *Weighted Spatial Pyramid Pooling (WSPP)* untuk mendeteksi depresi secara lebih efisien. Model ini memiliki dua cabang: LA-CNN untuk fokus lokal dan GA-CNN untuk pola global. Gabungan keduanya, LGA-CNN, menangkap informasi komplementer. Hasil uji pada dataset AVEC2013 dan AVEC2014 menunjukkan bahwa pendekatan ini unggul dibanding metode deteksi depresi berbasis video sebelumnya [15].

Studi ini mengusulkan sistem deteksi depresi berbasis analisis ekspresi wajah, dengan menggunakan fitur mikro-ekspresi yang diekstraksi melalui *Facial Action Coding System (FACS)*. Deteksi dilakukan menggunakan model SVM dan CNN. Data dikumpulkan dari pasien di Bahawal Victoria Hospital berdasarkan skala PHQ-8. Hasil eksperimen menunjukkan akurasi validasi 99,9% pada model CNN dan 100% pada SVM, melampaui metode deteksi depresi lainnya [16].

2.1 VGG Network

VGG Network merupakan salah satu arsitektur *Convolutional Neural Network (CNN)* yang dikembangkan oleh Karen Simonyan dan Andrew Zisserman dari *Visual Geometry Group, Department of Engineering Science, University of Oxford*. Arsitektur ini menggunakan input berupa citra RGB dengan resolusi 224×224 piksel. Dalam jaringan ini, terdapat dua jenis *convolutional layer*, yaitu *layer* dengan filter berukuran 3×3 (conv3) dan filter berukuran 1×1 (conv1). Ukuran output dari layer-layer konvolusi bervariasi, mencakup dimensi 64×64 , 128×128 , 256×256 , hingga 512×512 .

3. Metodologi Penelitian

Metodologi yang diusulkan terdiri dari beberapa langkah untuk mengklasifikasikan depresi dari citra ekspresi wajah.

3.1. Dataset

Penelitian ini menggunakan dataset FER2013 (*Facial Expression Recognition 2013*), yang berisi gambar-gambar wajah dengan resolusi 48×48 piksel dalam skala abu-abu. Dataset ini terdiri dari tujuh kelas ekspresi, yaitu marah, jijik, takut, bahagia, sedih, terkejut, dan netral. Training image terdiri dari 28.709 gambar dan test image terdiri dari 3.589 gambar. Spesifikasi dataset FER2013 dapat dilihat pada tabel 1.

Dalam konteks penelitian ini, tujuh ekspresi tersebut dikelompokkan ulang menjadi dua kategori berdasarkan relevansinya terhadap gejala depresi. Tabel 2 menjelaskan ekspresi marah, jijik, takut, dan sedih dikategorikan sebagai indikasi depresi, sedangkan ekspresi netral, bahagia, dan terkejut dikategorikan sebagai tidak depresi. Pengelompokan ini didasarkan pada studi literatur terkait karakteristik emosional yang sering muncul pada individu dengan gangguan depresi.

Tabel 1. Spesifikasi dataset FER2013

Jumlah Data	Ukuran Gambar	Format
32.298	48×48 pixel	JPG

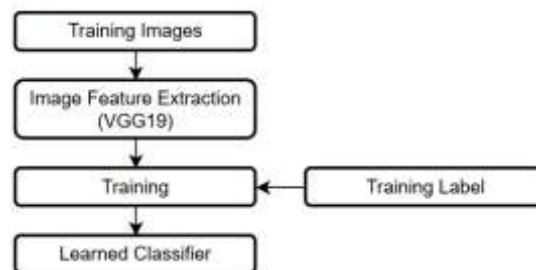
Tabel 2. Dataset Depresi

Depresi	Tidak Depresi
Marah	Netral
Jijik	Bahagia
Takut	Terkejut
Sedih	

3.2. Tahapan Penelitian

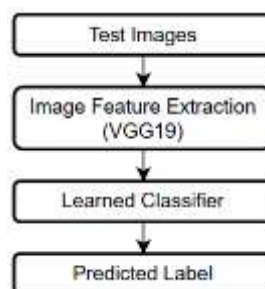
Penelitian ini menerapkan tahapan pelatihan dan pengujian model secara terstruktur untuk mendeteksi kondisi depresi melalui analisis ekspresi wajah dengan menggunakan arsitektur VGGNet. Proses dimulai dari pengolahan data pelatihan guna membentuk model klasifikasi, yang selanjutnya dievaluasi dengan menggunakan data uji yang belum pernah dikenali oleh model sebelumnya. Seluruh rangkaian proses mencakup pemberian input berupa citra wajah, ekstraksi fitur, hingga menghasilkan keluaran berupa klasifikasi kondisi depresi.

3.2.1. Tahapan Pelatihan

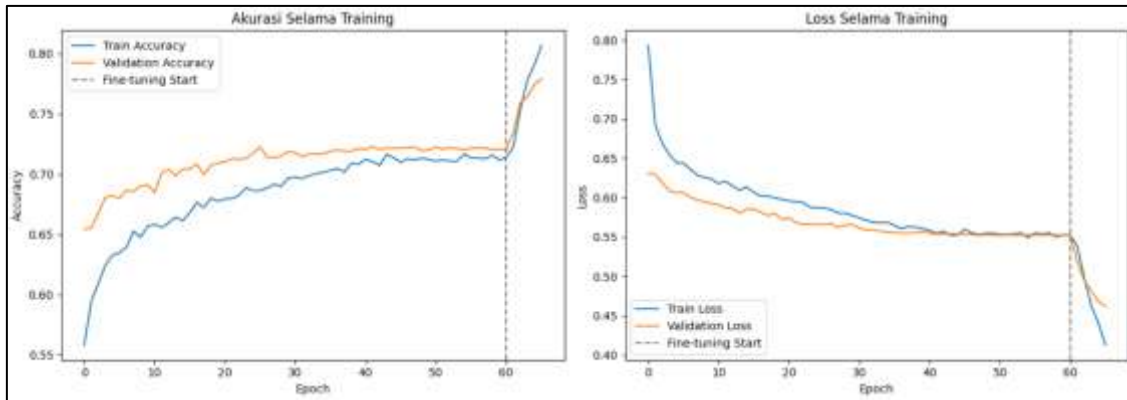
**Gambar 1. Diagram Tahapan Pelatihan**

Gambar 1 memperlihatkan beberapa tahapan pelatihan yang akan dilakukan dalam penelitian ini, diantaranya yaitu: (1) *Training Images*: Kumpulan citra wajah yang digunakan sebagai data pelatihan dan telah diklasifikasikan ke dalam dua kategori, yaitu depresi dan tidak depresi, (2) *Image Features Extraction*: Ekstraksi fitur visual penting dari citra, seperti bentuk dan tekstur wajah, dilakukan secara otomatis menggunakan arsitektur VGG19, (3) *Training Label*: Label diberikan pada setiap citra untuk menunjukkan kategori kondisi yang sesuai, (4) *Training*: Model dilatih menggunakan fitur dan label tersebut agar dapat mengenali pola ekspresi yang menunjukkan gejala depresi dan tidak depresi, (5) *Learned Classifier*: Model klasifikasi akhir yang telah terlatih dan siap digunakan untuk memprediksi kondisi depresi pada citra baru.

3.2.2. Tahapan Pengujian

**Gambar 2. Diagram Tahapan Pengujian**

Gambar 2 memperlihatkan mengenai beberapa tahapan pengujian yang akan dilakukan dalam penelitian ini, diantaranya yaitu: (1) *Test Images*: Citra uji yang belum pernah dianalisis oleh model digunakan untuk menguji kemampuan prediksi model, (2) *Image Features Extraction*: Fitur visual dari citra uji diekstraksi menggunakan arsitektur VGG19, seperti pada tahap pelatihan, (3) *Learned Classifier*: Model terbaik yang telah dilatih digunakan untuk memproses fitur dari citra uji, (4) *Predicted Label*: Hasil akhir berupa label prediksi yang menunjukkan apakah gambar tersebut termasuk kategori depresi atau tidak.



Gambar 3. Grafik Accuracy dan Loss Selama Training

Setelah model selesai dilatih dan dilakukan *fine-tuning*, langkah selanjutnya adalah pengujian terhadap data uji yang belum pernah dilihat sebelumnya oleh model terbaik VGG19. Berdasarkan hasil pengujian pada tabel, model memperoleh akurasi sebesar 77%, dengan nilai *precision*, *recall*, dan *F1-score* yang juga konsisten di angka 77%. Konsistensi nilai metrik ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dalam membedakan ekspresi wajah yang mengindikasikan kondisi depresi dan tidak depresi. Dengan kata lain, model tidak hanya bekerja optimal pada data pelatihan, tetapi juga mampu mempertahankan kinerja yang stabil saat dihadapkan pada data baru.

Tabel 4. Hasil Testing

Accuracy	Precision	Recall	F1-score
0.77	0.77	0.77	0.77

5. Kesimpulan dan Saran

Pada penelitian ini, telah dikembangkan sebuah model deteksi depresi berbasis ekspresi wajah menggunakan model VGG19. Model ini dapat mengklasifikasikan individu ke dalam kategori depresi atau tidak depresi berdasarkan ekspresi wajah mereka, yang diambil dari dataset FER2013 yang dikelompokkan menjadi dua kategori yaitu ekspresi wajah yang terkait dengan depresi dan ekspresi wajah yang tidak terkait dengan depresi. Model VGG19 yang diterapkan pada penelitian ini menunjukkan hasil yang memadai dengan akurasi mencapai 77% pada data uji, serta dapat membedakan dengan cukup baik antara ekspresi depresi dan tidak depresi.

Sebagai saran untuk pengembangan penelitian selanjutnya, penerapan teknik lain seperti *ensemble learning* atau *transfer learning* dari model-model arsitektur yang lebih kompleks, misalnya ResNet atau Inception, dapat dipertimbangkan untuk memperoleh hasil klasifikasi yang lebih optimal. Selain itu, integrasi data multimodal, seperti audio dan teks, dapat memperkaya fitur yang digunakan dan berpotensi meningkatkan akurasi sistem dalam mendeteksi kondisi depresi secara lebih menyeluruh.

6. Ucapan Terima Kasih

Penulis menyampaikan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan dan arahan selama proses penelitian ini berlangsung. Ucapan terima kasih juga disampaikan kepada Fakultas Teknik Universitas 17 Agustus 1945 Surabaya atas dukungan fasilitas dan finansial serta lingkungan akademik yang mendukung kelancaran penelitian ini. Dukungan dari berbagai pihak sangat berarti dalam mewujudkan penelitian ini hingga tuntas.

Referensi

- [1] World Health Organization, "World mental health report: Transforming mental health for all," Jun. 2022.

- [2] Z. Liu, X. Yuan, Y. Li, Z. Shangguan, L. Zhou, and B. Hu, "PRA-Net: Part-and-Relation Attention Network for depression recognition from facial expression," *Comput Biol Med*, vol. 157, May 2023, doi: 10.1016/j.combiomed.2023.106589.
- [3] A. S. Rajawat, P. Bedi, S. B. Goyal, P. Bhaladhare, A. Aggarwal, and R. S. Singhal, "Fusion Fuzzy Logic and Deep Learning for Depression Detection Using Facial Expressions," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 2795–2805. doi: 10.1016/j.procs.2023.01.251.
- [4] H. Tufail, S. M. Cheema, M. Ali, I. M. Pires, and N. M. Garcia, "Depression Detection with Convolutional Neural Networks: A Step Towards Improved Mental Health Care," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 544–549. doi: 10.1016/j.procs.2023.09.079.
- [5] A. A. Loch *et al.*, "Detecting at-risk mental states for psychosis (ARMS) using machine learning ensembles and facial features," *Schizophr Res*, vol. 258, pp. 45–52, Aug. 2023, doi: 10.1016/j.schres.2023.07.011.
- [6] T. Cowan, M. D. Masucci, T. Gupta, C. M. Haase, G. P. Strauss, and A. S. Cohen, "Computerized analysis of facial expressions in serious mental illness," *Schizophr Res*, vol. 241, pp. 44–51, Mar. 2022, doi: 10.1016/j.schres.2021.12.026.
- [7] X. Li *et al.*, "TSFFM: Depression detection based on latent association of facial and body expressions," *Comput Biol Med*, vol. 168, Jan. 2024, doi: 10.1016/j.combiomed.2023.107805.
- [8] S. A. Hussain and A. Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jan. 2020. doi: 10.1088/1742-6596/1432/1/012087.
- [9] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," *Journal of Informatics and Web Engineering*, vol. 1, no. 2, pp. 45–54, Sep. 2022, doi: 10.33093/jiwe.2022.1.2.4.
- [10] A. Saidi, S. Ben Othman, and S. Ben Saoud, "Hybrid CNN-SVM classifier for efficient depression detection system," in *Proceedings of the International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 229–234. doi: 10.1109/IC_ASET49463.2020.9318302.
- [11] A. Aref, A. Al-Zanam, O. Hussein, A. Elsayed, H. Alhomery, and C. Peng Tan, "Mental Health State Classification Using Facial Emotion Recognition and Detection," vol. 13, no. 6, 2023.
- [12] H. V. Manalu and A. P. Rifai, "Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm," *Intelligent Systems with Applications*, vol. 21, Mar. 2024, doi: 10.1016/j.iswa.2024.200339.
- [13] W. Guo, H. Yang, Z. Liu, Y. Xu, and B. Hu, "Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks," *Front Neurosci*, vol. 15, Apr. 2021, doi: 10.3389/fnins.2021.609760.
- [14] X. Kong, Y. Yao, C. Wang, Y. Wang, J. Teng, and X. Qi, "Automatic Identification of Depression Using Facial Images with Deep Convolutional Neural Network," *Medical Science Monitor*, vol. 28, Jul. 2022, doi: 10.12659/MSM.936409.
- [15] L. He, J. C. W. Chan, and Z. Wang, "Automatic depression recognition using CNN with attention mechanism from videos," *Neurocomputing*, vol. 422, pp. 165–175, Jan. 2021, doi: 10.1016/j.neucom.2020.10.015.
- [16] G. Gilanie *et al.*, "An Automated and Real-time Approach of Depression Detection from Facial Micro-expressions," *Computers, Materials and Continua*, vol. 73, no. 2, pp. 2513–2528, 2022, doi: 10.32604/cmc.2022.028229.