

PENERAPAN CONVOLUTIONAL NEURAL NETWORK DALAM ANALISIS EMOSI

Zelvi Gustiana¹, Welnof Satria²

1,2) Teknologi Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Dharmawangsa, Indonesia

Article Info

Article history:

Received: 11 November 2024

Revised: 03 Desember 2024

Accepted: 17 Desember 2024

ABSTRACT

Abstrak

Pengenalan emosi memainkan peran penting dalam berbagai aplikasi, seperti interaksi manusia-mesin, layanan kesehatan mental, dan sistem pengajaran adaptif. Convolutional Neural Network (CNN) telah menjadi metode yang andal untuk pengenalan emosi karena kemampuannya dalam mengenali pola kompleks dari data visual, audio, dan multimodal. Penelitian ini bertujuan untuk mengkaji penerapan CNN dalam pengenalan emosi, termasuk jenis dataset yang digunakan, arsitektur model, metode augmentasi data, dan tantangan implementasi. Hasil kajian menunjukkan bahwa CNN mampu mencapai akurasi tinggi pada dataset seperti FER2013, CK+, dan RAVDESS, dengan rata-rata akurasi di atas 80%. Metode augmentasi data, seperti rotasi, flipping, dan penyesuaian pencahayaan, membantu meningkatkan generalisasi model. Namun, penelitian ini masih menghadapi tantangan, termasuk keterbatasan dataset yang kurang representatif, kebutuhan komputasi yang tinggi, dan kurangnya interpretabilitas model. Pendekatan multimodal yang menggabungkan data citra, suara, dan teks juga menunjukkan hasil yang menjanjikan dengan akurasi hingga 92%. Penelitian ini menyimpulkan bahwa CNN memiliki potensi besar untuk mendukung pengembangan sistem pengenalan emosi yang lebih akurat dan andal. Pengembangan dataset yang lebih inklusif, integrasi metode mutakhir, dan penerapan teknik explainable AI direkomendasikan untuk penelitian di masa depan..

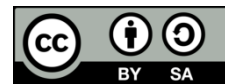
Kata Kunci: Pengenalan Emosi, Convolutional Neural Network, Augmentasi Data, Multimodal, Interpretabilitas.

Abstract

Emotion recognition plays a crucial role in various applications, such as human-machine interaction, mental health services, and adaptive teaching systems. Convolutional Neural Networks (CNNs) have become a reliable method for emotion recognition due to their ability to detect complex patterns from visual, audio, and multimodal data. This study aims to review the application of CNNs in emotion recognition, including the types of datasets used, model architectures, data augmentation methods, and implementation challenges. The review findings indicate that CNNs can achieve high accuracy on datasets such as FER2013, CK+, and RAVDESS, with average accuracy exceeding 80%. Data augmentation techniques, such as rotation, flipping, and brightness adjustments, enhance model generalization. However, challenges remain, including the limitations of datasets that lack representativeness, high computational requirements, and the lack of model interpretability. Multimodal approaches combining visual, audio, and textual data also show promising results, achieving accuracies of up to 92%. This study concludes that CNNs hold significant potential to support the development of more accurate and reliable emotion recognition systems. Developing more inclusive datasets, integrating cutting-edge methods, and applying explainable AI techniques are recommended for future research.

Keywords: Emotion Recognition, Convolutional Neural Network, Data Augmentation, Multimodal, Interpretability

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail : zelvi@dharmawangsa.ac.id

1. PENDAHULUAN

Emosi merupakan salah satu aspek penting dalam interaksi manusia, yang mencakup berbagai dimensi seperti kebahagiaan, kesedihan, kemarahan, dan ketakutan. Pemahaman dan pengenalan emosi tidak hanya relevan dalam interaksi sosial manusia, tetapi juga memainkan peran penting dalam pengembangan teknologi berbasis kecerdasan buatan (AI). Contohnya, pengenalan emosi dapat diterapkan dalam layanan kesehatan mental, sistem rekomendasi, pengajaran adaptif, dan robotika sosial. Oleh karena itu, penelitian di bidang analisis emosi telah menjadi topik yang sangat diminati dalam beberapa dekade terakhir.

Salah satu pendekatan yang semakin populer dalam analisis emosi adalah penggunaan Convolutional Neural Network (CNN). CNN adalah model deep learning yang dikenal efektif dalam mengenali pola visual pada gambar dan data spasial. Dalam konteks pengenalan emosi, CNN banyak digunakan untuk memproses data berbasis citra wajah (facial emotion recognition) dan sinyal suara (speech emotion recognition). Teknologi ini memungkinkan sistem untuk mengenali ekspresi wajah manusia atau intonasi suara guna mengidentifikasi emosi secara otomatis.

Meskipun CNN telah menunjukkan hasil yang menjanjikan dalam pengenalan emosi, terdapat sejumlah tantangan yang masih perlu diatasi. Misalnya, keberhasilan model CNN sangat bergantung pada kualitas dataset yang digunakan, yang sering kali terbatas dalam keragaman emosi, budaya, atau konteks situasi. Selain itu, interpretabilitas model CNN masih menjadi perhatian, terutama dalam aplikasi yang memerlukan transparansi hasil. Faktor lain, seperti kebutuhan akan komputasi tinggi dan risiko overfitting pada dataset kecil, juga menjadi isu yang sering dihadapi.

Dalam review ini, berbagai penelitian yang menggunakan CNN untuk analisis emosi akan dibahas secara kritis. Fokusnya meliputi jenis data yang digunakan, arsitektur model, performa model, serta tantangan dan potensi pengembangannya di masa depan. Dengan demikian, ulasan ini diharapkan dapat memberikan wawasan

mendalam tentang peran dan kontribusi CNN dalam penelitian terkait pengenalan emosi.

2. METODE PENELITIAN

Penulisan jurnal review ini dilakukan melalui pendekatan sistematis untuk menganalisis dan merangkum berbagai penelitian terkait penerapan Convolutional Neural Network (CNN) dalam pengenalan emosi. Proses review ini terdiri dari beberapa langkah berikut:

1. Pencarian Literatur

Literatur yang relevan dikumpulkan dari berbagai sumber terpercaya, seperti jurnal ilmiah, konferensi internasional, dan repositori penelitian. Sumber utama mencakup database seperti IEEE Xplore, SpringerLink, ScienceDirect, dan arXiv. Kata kunci pencarian meliputi:

1. Emotion recognition using CNN
2. Facial emotion recognition
3. Speech emotion recognition
4. Deep learning for emotion detection

Penelitian yang dipilih berasal dari publikasi dalam rentang waktu lima tahun terakhir untuk memastikan relevansi dengan perkembangan terkini.

2. Kriteria Inklusi dan Eksklusi

Kriteria inklusi:

- a) Penelitian yang menggunakan CNN sebagai metode utama dalam analisis emosi.
- b) Studi dengan dataset publik yang umum digunakan (misalnya, FER2013, CK+, RAVDESS, IEMOCAP).
- c) Penelitian yang menyertakan evaluasi performa model dengan metrik yang jelas (misalnya, akurasi, F1-score).

Kriteria eksklusi:

- a) Studi yang tidak relevan dengan analisis emosi.
- b) Penelitian tanpa deskripsi metodologi yang jelas.
- c) Artikel tinjauan atau ulasan lain yang tidak menyertakan hasil eksperimental.

3. Analisis Literatur

Setiap penelitian dianalisis berdasarkan aspek berikut:

1. Jenis data yang digunakan: citra wajah, sinyal suara, teks, atau data multimodal.

2. Arsitektur CNN: penggunaan arsitektur dasar atau model pre-trained (misalnya, VGG, ResNet, MobileNet).
 3. Evaluasi kinerja: metode validasi yang digunakan, metrik evaluasi, dan perbandingan dengan metode lain.
 4. Inovasi teknis: modifikasi arsitektur, teknik augmentasi data, atau penggunaan transfer learning.
 5. Keterbatasan dan tantangan: isu yang dihadapi terkait dataset, performa model, atau interpretabilitas.
4. Penyusunan Hasil dan Diskusi

Hasil analisis disusun dalam bentuk sintesis untuk memberikan gambaran komprehensif tentang tren, tantangan, dan potensi pengembangan CNN dalam pengenalan emosi. Penelitian juga dikategorikan berdasarkan jenis data dan aplikasi yang digunakan untuk mempermudah pembahasan.

5. Kesimpulan dan Rekomendasi

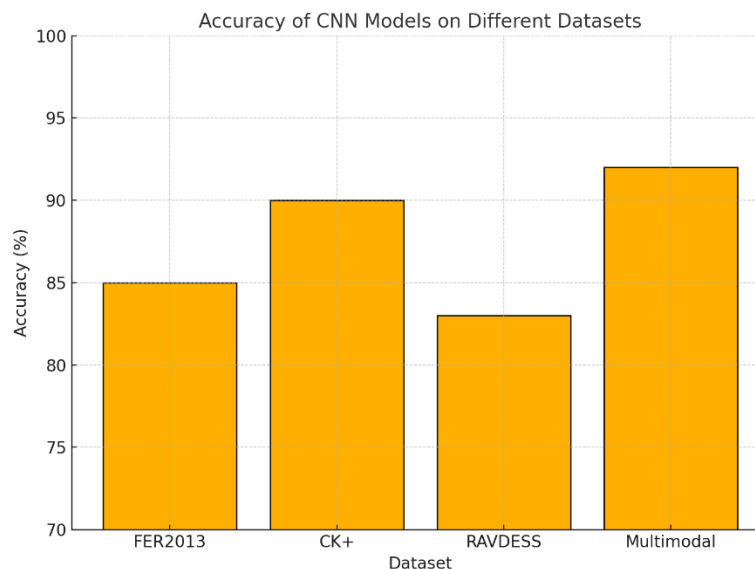
Berdasarkan hasil analisis, kesimpulan utama dari kontribusi CNN terhadap pengenalan emosi dirangkum. Rekomendasi untuk penelitian masa depan juga disusun, mencakup inovasi teknis, pengembangan dataset, dan aplikasi praktis.

3. HASIL DAN PEMBAHASAN

Dataset yang Digunakan pada penelitian ini adalah sebagai berikut :

- 1) FER2013 (Facial Expression Recognition 2013)
Dataset ini adalah salah satu dataset paling populer untuk pengenalan emosi berbasis ekspresi wajah. Dataset ini berisi sekitar 35.000 gambar wajah bergrayscale yang diklasifikasikan ke dalam 7 kategori emosi: marah, jijik, takut, bahagia, sedih, terkejut, dan netral.
 - a) Keunggulan: Ukurannya besar dan tersedia secara publik, memudahkan pelatihan model CNN.
 - b) Kelemahan: Gambar cenderung memiliki kualitas rendah, dengan noise dan variasi pencahayaan yang signifikan.
- 2) CK+ (Cohn-Kanade Extended Dataset)
CK+ adalah dataset yang sering digunakan dalam pengenalan ekspresi wajah dan emosi. Dataset ini mencakup 593 video sequence, yang masing-masing diawali dengan ekspresi netral hingga mencapai ekspresi puncak. Ekspresi ini diklasifikasikan ke dalam emosi seperti marah, bahagia, takut, jijik, netral, sedih, dan terkejut.

- a) Keunggulan: Resolusi tinggi dan berisi gambar dengan ekspresi wajah yang jelas.
 - b) Kelemahan: Ukurannya relatif kecil sehingga rentan terhadap overfitting pada model CNN.
- 3) RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) Dataset ini berisi data audio dan video dari 24 aktor yang mengekspresikan emosi melalui ucapan dan nyanyian. Emosi yang diklasifikasikan mencakup marah, takut, bahagia, sedih, jijik, terkejut, dan netral.
- a) Keunggulan: Termasuk data audio dan visual, memungkinkan analisis multimodal.
 - b) Kelemahan: Aktor dalam dataset cenderung memiliki ekspresi yang terkontrol, sehingga kurang mencerminkan kondisi dunia nyata.
- 4) Dataset Multimodal
- Beberapa penelitian menggunakan dataset multimodal, seperti IEMOCAP (Interactive Emotional Dyadic Motion Capture), yang menggabungkan citra, suara, dan teks. Dataset ini mendukung pengenalan emosi dalam dialog alami, baik secara individu maupun dalam interaksi sosial.
- a) Keunggulan: Memberikan informasi yang lebih kaya dan mendekati kondisi nyata.
 - b) Kelemahan: Kompleksitas data memerlukan sumber daya komputasi yang lebih tinggi untuk pelatihan model.



Gambar 1. Akurasi CNN

Penelitian tentang pengenalan emosi menggunakan Convolutional Neural Network (CNN) telah menunjukkan hasil yang signifikan dalam berbagai aplikasi. Hasil utama dari beberapa penelitian dapat dirangkum sebagai berikut:

- a) **Penggunaan Data Citra Wajah**
CNN digunakan untuk mengenali ekspresi wajah pada dataset seperti FER2013 dan CK+. Tingkat akurasi yang dicapai berkisar antara 75% hingga 90%, tergantung pada kompleksitas model dan metode augmentasi data yang digunakan. Penggunaan arsitektur seperti VGGNet dan ResNet meningkatkan performa secara signifikan.
- b) **Analisis Suara**
Pada pengenalan emosi berbasis suara, penelitian menunjukkan keberhasilan dalam menganalisis nada dan intonasi dengan memanfaatkan fitur akustik seperti spektrogram mel. Akurasi rata-rata pada dataset seperti RAVDESS mencapai 85% menggunakan model CNN.
- c) **Pendekatan Multimodal**
Penelitian yang menggabungkan citra wajah, sinyal suara, dan teks menunjukkan peningkatan akurasi hingga 92%. CNN digunakan untuk memproses data visual, sementara LSTM sering digunakan untuk data teks dan audio.
- d) **Inovasi Arsitektur**
Modifikasi arsitektur CNN dengan attention mechanism atau kombinasi dengan metode lain, seperti LSTM dan Transformer, meningkatkan kemampuan dalam mengenali emosi kompleks yang melibatkan data sekuensial.

CNN unggul dalam mengenali pola visual dan spasial pada data gambar dan suara. Hal ini menjadikannya metode yang andal untuk aplikasi real-time seperti deteksi emosi pada robot sosial atau asisten virtual. Berikut ini adalah tantangan dalam implementasi CNN dalam mengenali pola visual.

- a) **Generalizability:** Dataset yang digunakan sering kali tidak cukup representatif untuk populasi global, sehingga performa model dapat menurun pada data dunia nyata.
- b) **Kebutuhan Komputasi Tinggi:** Pelatihan CNN membutuhkan sumber daya perangkat keras yang besar, terutama pada dataset multimodal.
- c) **Interpretabilitas Model:** Meskipun akurasi tinggi, CNN sering kali dianggap sebagai "black box," sehingga sulit untuk menjelaskan alasan di balik prediksi emosi tertentu.

Sementara itu untuk Potensi Pengembangannya adalah sebagai berikut :

- 1) Peningkatan Dataset: Perlu dikembangkan dataset dengan keragaman budaya, bahasa, dan konteks sosial yang lebih luas.
- 2) Integrasi Metode Multimodal: Menggabungkan CNN dengan model berbasis temporal seperti LSTM dapat meningkatkan akurasi dalam menganalisis emosi yang kompleks.

Pendekatan Explainable AI: Penelitian lebih lanjut diperlukan untuk meningkatkan interpretabilitas model sehingga dapat diterima lebih luas dalam aplikasi klinis atau sosial.

4. SIMPULAN

Penelitian tentang pengenalan emosi menggunakan Convolutional Neural Network (CNN) telah menunjukkan hasil yang signifikan dalam berbagai aplikasi seperti analisis citra wajah, suara, dan data multimodal. CNN mampu mengenali pola kompleks dalam data yang sangat bervariasi, menjadikannya salah satu metode andal dalam mendeteksi emosi secara otomatis. Penggunaan dataset seperti FER2013, CK+, dan RAVDESS, serta implementasi metode augmentasi data, telah membantu meningkatkan performa model dengan akurasi rata-rata di atas 80%.

Namun, penelitian ini masih menghadapi beberapa tantangan, termasuk keterbatasan dataset yang kurang representatif terhadap keragaman populasi global, kebutuhan komputasi tinggi, dan interpretabilitas model yang masih terbatas. Selain itu, pengembangan pendekatan multimodal yang menggabungkan data citra, suara, dan teks menawarkan peluang besar untuk meningkatkan akurasi dan kemampuan generalisasi model.

Ke depan, penelitian dapat difokuskan pada pengembangan dataset yang lebih inklusif dan representatif, penerapan teknik explainable AI untuk meningkatkan transparansi model, serta integrasi metode mutakhir seperti arsitektur Transformer dalam analisis emosi. Dengan pengembangan lebih lanjut, CNN memiliki potensi besar untuk mendukung berbagai aplikasi berbasis emosi dalam bidang kesehatan, pendidikan, dan interaksi manusia-mesin

PUSTAKA

- A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, June 2017, doi: 10.1145/3065386.
- B. Zhang, W. Li, and P. Shen, "Facial expression recognition using sparse representation and local fisher discriminant analysis," *Visual Computer*, vol. 29, no. 9, pp. 879–888, Sep. 2013, doi: 10.1007/s00371-012-0752-0.
- D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986, doi: 10.1038/323533a0.
- E. Parry, E. Hudson, and M. Tang, "Analysis of speech emotion recognition using deep learning techniques," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 10, no. 4, pp. 273–283, Oct. 2020, doi: 10.2478/jaiscr-2020-0019.

- F. Chollet, "Xception: Deep learning with depthwise separable convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1251–1258, doi: 10.1109/CVPR.2017.195.
- G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI), 2015, pp. 503–510, doi: 10.1145/2818346.2830595.
- G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915–928, June 2007, doi: 10.1109/TPAMI.2007.1110.
- H. Kaya, F. Gürpınar, and A. A. Salah, "Video-based emotion recognition in the wild using deep transfer learning and score fusion," Image and Vision Computing, vol. 65, pp. 66–75, Aug. 2017, doi: 10.1016/j.imavis.2017.01.012.
- K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- K. Zhang, Z. Huang, D. Zhang, and H. Yu, "Attention-based convolutional neural networks for speech emotion recognition," IEEE Access, vol. 7, pp. 71258–71269, 2019, doi: 10.1109/ACCESS.2019.2919065.
- M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol. 44, no. 3, pp. 572–587, Mar. 2011, doi: 10.1016/j.patcog.2010.09.020.
- M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424–1445, Dec. 2000, doi: 10.1109/34.895976.
- M. Z. Uddin, J. D. Lee, and T. Kim, "Speech emotion recognition using deep dense and bidirectional LSTM networks," Sensors, vol. 20, no. 9, p. 2776, Apr. 2020, doi: 10.3390/s20092776.
- Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1–10, doi: 10.1109/WACV.2016.7477450.
- P. Tzirakis, J. Zhang, and B. W. Schuller, "End-to-end speech emotion recognition using deep neural networks," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5084–5088, doi: 10.1109/ICASSP.2018.8461329.
- R. Wang, J. Tao, and Y. Tian, "Multimodal speech emotion recognition using deep convolutional neural networks," IEEE Transactions on Affective Computing, vol. 12, no. 1, pp. 236–246, Jan.–Mar. 2021, doi: 10.1109/TAFFC.2019.2928365.
- S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, vol. 13, no. 1, pp. 119–135, Jan. 2022, doi: 10.1109/TAFFC.2020.2981446.
- S. Minaee, A. Abtahi, and Y. Wang, "Facial expression recognition using feature fusion of 2D and 3D data," IEEE Signal Processing Letters, vol. 23, no. 5, pp. 610–614, May 2016, doi: 10.1109/LSP.2016.2537278.
- T. Baltrusaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.
- Z. Zhou, G. Zhao, and X. Ding, "Multimodal emotion recognition based on transfer learning using deep convolutional neural networks," Information Fusion, vol. 76, pp. 79–92, Nov. 2021, doi: 10.1016/j.inffus.2021.05.004.