p-ISSN: 2963-7856| e-ISSN: 2961-9890
Available online at jerkin.org/index.php/jerkin
Jurnal Pengabdian Masyarakat dan Riset Pendidikan
Volume 4 No. 4, April-Juni 2026, pp 26255-26262

Prediksi Emosi Penonton Konser Berbasis AI: Sebelum dan Sesudah Pengalaman Musik

M. Jembar Jomantara[1]*, Lila Setiyani[2], Deden Moch Alfiansyah[3], Devi Fajar Wati[4], Dedih[5], Arif Budimansyah Purba[6]

[1,2,3,4,5,6] Program Studi Teknologi Informasi, Universitas Horizon Indonesia, Jl. Pangkal Perjuangan By Pass No.KM.1, Tanjungpura, Kec. Karawang Bar., Karawang, Jawa Barat.
E-mail: muhammad.jomantara.krw@horizon.ac.id
*Corresponding Author

https://doi.org/10.31004/jerkin.v4i4.6643

ARTICLE INFO

Article history:
Received: 21 May 2026
Revised: 27 May 2026
Accepted: 02 Jun 2026

Kata Kunci:
Deep Learning, Konsentrasi Manusia, EEG, Multimodal Learning, Monitoring Kognitif.

Keywords:
Deep Learning, Human Concentration, EEG, Multimodal Learning, Cognitive Monitoring.

ABSTRACT

Emosi penonton konser merupakan aspek psikologis penting yang memengaruhi kualitas pengalaman hiburan, keterlibatan audiens, dan kepuasan individu dalam menikmati pertunjukan musik secara langsung. Namun, dinamika emosi penonton cenderung berubah secara cepat dipengaruhi oleh suasana acara, performa musisi, interaksi sosial, dan pengalaman personal selama konser berlangsung. Penelitian ini bertujuan mengembangkan model deep learning multimodal untuk menganalisis emosi penonton konser menggunakan data visual dan perilaku audiens. Pendekatan yang diusulkan mengintegrasikan data ekspresi wajah, perilaku penonton, dan interaksi digital untuk meningkatkan akurasi analisis emosi. Arsitektur hybrid Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) digunakan untuk menangkap pola spasial dan temporal dari data multimodal. Model dirancang untuk melakukan analisis emosi secara real-time serta mengidentifikasi perubahan emosional pada lingkungan dinamis seperti konser musik langsung, festival, dan pertunjukan hiburan digital. Metode penelitian meliputi pengumpulan data menggunakan sensor kamera, pra-pemrosesan data, pengembangan model, dan evaluasi performa menggunakan metrik accuracy, precision, recall, F1-score, dan RMSE. Hasil akhir penelitian berupa model deep learning multimodal yang mampu menganalisis emosi penonton konser secara lebih adaptif dibandingkan pendekatan unimodal.

Concertgoers' emotions are important psychological aspects that influence entertainment experiences, audience engagement, and individual satisfaction in enjoying live music performances. However, the emotional dynamics of concert audiences tend to change rapidly due to the event atmosphere, musicians' performances, social interactions, and personal experiences throughout the concert. This study aims to develop a multimodal deep learning model for analyzing concertgoers' emotions using visual and audience behavioral data. The proposed approach integrates facial expression data, audience behavior, and digital interactions to improve the accuracy of emotion analysis. A hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architecture is employed to capture spatial and temporal patterns from multimodal data. The model is designed to perform real-time emotion analysis and identify emotional changes in dynamic environments such as live music concerts, festivals, and digital entertainment performances. The research method includes data collection using camera sensors, data preprocessing, model development, and performance evaluation using accuracy, precision, recall, F1-score, and RMSE metrics. The final result of this study is a multimodal deep learning model capable of analyzing concertgoers' emotions more adaptively compared to unimodal approaches.

How to Cite: M. Jembar Jomantara, et al. (2026), Prediksi Emosi Penonton Konser Berbasis AI: Sebelum dan Sesudah Pengalaman Musik, 4(4). https://doi.org/10.31004/jerkin.v4i4.6643

PENDAHULUAN

Emosi manusia merupakan salah satu aspek psikologis yang berperan penting dalam menentukan kualitas pengalaman hiburan, keterlibatan sosial, dan kepuasan individu dalam menikmati pertunjukan musik secara langsung (Ekman, 1992). Respons emosional penonton konser dapat berubah secara dinamis dipengaruhi oleh suasana acara, performa musisi, interaksi sosial, pencahayaan, audio, serta pengalaman personal selama konser berlangsung (Juslin & Sloboda, 2010). Pada era digital, perkembangan industri hiburan berbasis teknologi telah meningkatkan kebutuhan terhadap sistem analisis audiens yang mampu memahami respons emosional penonton secara lebih adaptif dan real-time (Picard, 2000).

Deep learning telah banyak digunakan dalam analisis emosi manusia karena mampu mengenali pola kompleks secara otomatis dan memiliki performa yang tinggi pada data visual, audio, maupun perilaku digital (LeCun, Bengio, & Hinton, 2015). Pendekatan kecerdasan buatan juga dinilai mampu memberikan analisis emosional yang lebih akurat dibandingkan metode konvensional berbasis observasi manual (Goodfellow, Bengio, & Courville, 2016).

Berbagai penelitian sebelumnya telah mengembangkan metode emotion recognition menggunakan data visual maupun interaksi digital pengguna (Poria et al., 2017). Teknologi facial expression recognition mampu mengidentifikasi perubahan ekspresi wajah yang berkaitan dengan kondisi emosional individu selama menikmati hiburan musik (Fasel & Luettin, 2003). Selain itu, berbagai penelitian menunjukkan bahwa Convolutional Neural Network (CNN) dapat mengenali pola emosi manusia dari data citra dengan tingkat akurasi yang cukup tinggi (Krizhevsky, Sutskever, & Hinton, 2012).

Pendekatan deep learning berbasis multimodal juga menjadi salah satu metode yang efektif dalam analisis emotional state karena mampu mengekstraksi fitur secara otomatis tanpa ketergantungan tinggi terhadap feature engineering manual (Baltrusaitis, Ahuja, & Morency, 2019). Penelitian lain memanfaatkan data media sosial dan perilaku audiens untuk menganalisis respons emosional pengguna pada lingkungan hiburan yang dinamis (Cambria, 2016). Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada analisis emosi berbasis satu jenis data dan belum banyak mengintegrasikan berbagai sumber data secara simultan untuk memahami emosi penonton konser secara lebih komprehensif (Poria et al., 2017).

METODE

Penelitian ini mengusulkan pengembangan model deep learning multimodal menggunakan kombinasi Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM). CNN digunakan untuk mengekstraksi pola visual dari ekspresi wajah dan data multimedia, sedangkan BiLSTM digunakan untuk menangkap pola temporal serta perubahan emosi secara berurutan (Hochreiter & Schmidhuber, 1997). CNN juga dinilai efektif dalam mengenali pola spasial pada data visual, sedangkan model recurrent neural network seperti LSTM dan BiLSTM memiliki kemampuan yang baik dalam memahami hubungan temporal pada data sequential (LeCun et al., 2015).1. Studi Literatur

Tahap pertama penelitian dimulai dengan melakukan studi literatur untuk memahami perkembangan penelitian terkait emotion analysis menggunakan teknologi Artificial Intelligence (AI), khususnya deep learning. Literatur yang dikaji meliputi penelitian mengenai facial expression recognition, audience behavior analysis, multimodal learning, Convolutional Neural Network (CNN), serta Bidirectional Long Short-Term Memory (BiLSTM). Pada tahap ini dilakukan identifikasi terhadap:

1. Metode yang paling banyak digunakan,
2. Kelebihan dan kekurangan penelitian sebelumnya,
3. Dataset yang tersedia secara publik,
4. Teknik preprocessing data,
5. Metode evaluasi performa model.

Selain itu, dilakukan analisis research gap untuk menemukan permasalahan yang belum banyak diteliti. Berdasarkan hasil studi literatur ditemukan bahwa sebagian besar penelitian sebelumnya hanya berfokus pada analisis emosi menggunakan satu jenis data dan belum banyak mengembangkan model

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

multimodal untuk menganalisis emosi penonton konser secara real-time pada lingkungan hiburan yang dinamis.

Luaran dari tahap ini berupa:

1. Rumusan masalah penelitian,
2. Tujuan penelitian,
3. Kerangka konseptual,
4. Rancangan arsitektur model deep learning multimodal.

Pengumpulan Dataset

Tahap selanjutnya adalah pengumpulan dataset yang digunakan untuk proses pelatihan dan pengujian model deep learning. Pada penelitian ini, data diperoleh menggunakan sensor kamera sebagai alat pendeteksi objek ekspresi wajah penonton konser. Kamera digunakan untuk menangkap perubahan ekspresi emosional audiens selama konser berlangsung.

Dataset yang digunakan terdiri dari:

1. Data ekspresi wajah,
2. Data perilaku audiens,
3. Data interaksi digital.

Data ekspresi wajah digunakan untuk merepresentasikan kondisi emosional penonton berdasarkan perubahan mimik wajah. Data perilaku audiens digunakan untuk menganalisis respons penonton seperti gerakan tubuh, tepuk tangan, dan interaksi sosial selama konser berlangsung. Sedangkan data interaksi digital digunakan untuk melihat respons audiens melalui aktivitas media sosial, komentar, dan reaksi digital lainnya. Pemilihan data dilakukan menggunakan teknik purposive sampling dengan beberapa kriteria:

1. Data memiliki keterkaitan dengan emosi penonton konser,
2. Data diperoleh secara legal dan etis,
3. Jumlah data mencukupi untuk deep learning,
4. Data memiliki kualitas visual yang baik.

Luaran dari tahap ini adalah kumpulan dataset multimodal yang siap diproses pada tahap preprocessing.

Pra-pemrosesan Data

Tahap preprocessing dilakukan untuk membersihkan, menyusun, dan menstandarkan data agar dapat digunakan secara optimal dalam proses pelatihan model deep learning. Tahap ini sangat penting karena kualitas data akan memengaruhi performa model yang dihasilkan. Pada data ekspresi wajah dilakukan beberapa proses seperti:

1. Deteksi wajah menggunakan sensor kamera,
2. Cropping area wajah,
3. Normalisasi ukuran gambar,
4. Pengurangan noise visual.

Pada data perilaku audiens dilakukan:

1. Sinkronisasi frame video,
2. Penghapusan data yang tidak relevan,
3. Normalisasi gerakan objek.

Sedangkan pada data interaksi digital dilakukan:

1. Pembersihan data teks,
2. Penghapusan spam atau duplikasi,
3. Transformasi data ke bentuk numerik.

Setelah seluruh data dibersihkan, dilakukan sinkronisasi antar-data multimodal agar seluruh data memiliki format dan waktu yang sesuai untuk diproses secara bersamaan. Luaran dari tahap ini berupa dataset multimodal yang bersih, terstruktur, dan siap digunakan untuk pengembangan model deep learning.

Ekstraksi Fitur

Tahap ekstraksi fitur bertujuan untuk mengambil informasi penting dari setiap jenis data sehingga model deep learning dapat mengenali pola emosi penonton konser secara lebih efektif. Pada data ekspresi wajah, fitur yang diambil meliputi:

1. Pola perubahan ekspresi,
2. Pergerakan otot wajah,
3. Intensitas emosi,

4. Aktivitas visual temporal.

Pada data perilaku audiens, fitur yang digunakan antara lain:
1. Pola gerakan tubuh,
2. Intensitas respons penonton,
3. Interaksi sosial,
4. Aktivitas kerumunan.

Sedangkan pada data interaksi digital diambil fitur seperti:
1. Sentimen komentar,
2. Frekuensi interaksi media sosial,
3. Penggunaan kata emosional,
4. Pola respons digital audiens.

Ekstraksi fitur dilakukan untuk mengurangi kompleksitas data mentah sekaligus meningkatkan kemampuan model dalam mengenali pola yang berkaitan dengan emosi penonton konser. Proses preprocessing dan ekstraksi fitur dilakukan untuk meningkatkan kualitas data dan stabilitas model deep learning.

Luaran tahap ini berupa feature vector dari masing-masing modal data yang siap diproses oleh model deep learning.

Pengembangan Model Deep Learning

Pada tahap ini dilakukan pembangunan model deep learning multimodal menggunakan kombinasi CNN dan BiLSTM. Model dirancang untuk memproses berbagai jenis data secara bersamaan sehingga dapat menghasilkan analisis emosi yang lebih akurat dibandingkan model unimodal.

Arsitektur model terdiri dari beberapa bagian:
1. CNN untuk memproses fitur visual ekspresi wajah,
2. Bilstm untuk mempelajari pola temporal dan perubahan emosi,
3. Fusion layer untuk menggabungkan fitur dari seluruh modal data,
4. Fully connected layer untuk klasifikasi emosi.

CNN digunakan karena mampu mengenali pola visual kompleks pada data citra wajah. Sementara itu, BiLSTM digunakan karena memiliki kemampuan memahami hubungan temporal dua arah sehingga cocok untuk menganalisis perubahan emosi audiens secara berurutan. Setelah seluruh fitur diproses, hasil dari masing-masing modal digabungkan pada fusion layer untuk menghasilkan representasi data yang lebih lengkap. Output model berupa:
1. klasifikasi jenis emosi,
2. analisis perubahan emosi audiens.

Luaran dari tahap ini adalah model deep learning multimodal yang telah dilatih menggunakan dataset penelitian.

Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui performa model dalam menganalisis emosi penonton konser. Evaluasi dilakukan menggunakan metode cross-validation agar hasil pengujian lebih stabil dan objektif.

Beberapa metrik evaluasi yang digunakan meliputi:
1. Accuracy,
2. Precision,
3. Recall,
4. F1-Score,
5. Confusion Matrix,
6. Root Mean Square Error (RMSE).

Selain itu dilakukan perbandingan antara:
1. Model multimodal,
2. Model unimodal berbasis ekspresi wajah saja.

Tujuan perbandingan ini adalah untuk mengetahui apakah integrasi data multimodal mampu meningkatkan performa model dibandingkan penggunaan satu jenis data saja.

Luaran tahap ini berupa hasil performa model dan analisis tingkat akurasi deteksi emosi penonton konser.

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

Analisis dan Interpretasi Hasil

Tahap terakhir adalah analisis dan interpretasi hasil penelitian. Pada tahap ini seluruh hasil evaluasi model dianalisis untuk mengetahui efektivitas pendekatan CNN–BiLSTM multimodal dalam menganalisis emosi penonton konser. Analisis dilakukan dengan:

1. Membandingkan hasil antar-model,
2. Melihat pola performa model,
3. Mengidentifikasi faktor yang memengaruhi akurasi,
4. Mengevaluasi kemampuan analisis pada lingkungan konser yang dinamis.

Selanjutnya dilakukan interpretasi hasil berdasarkan tujuan penelitian dan penelitian terdahulu. Dari tahap ini diperoleh kesimpulan mengenai kemampuan model multimodal dalam meningkatkan performa analisis emosi penonton dibandingkan pendekatan sebelumnya. Luaran akhir penelitian berupa model deep learning multimodal yang dapat digunakan untuk menganalisis emosi penonton konser secara adaptif pada lingkungan hiburan yang dinamis.

HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengembangkan model deep learning multimodal berbasis kombinasi Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) untuk menganalisis emosi penonton konser pada lingkungan hiburan yang dinamis. Model dikembangkan menggunakan tiga jenis data utama, yaitu data ekspresi wajah, data perilaku audiens, dan data interaksi digital. Dataset yang digunakan terdiri dari 12.500 data multimodal yang diperoleh dari rekaman konser musik, aktivitas media sosial penonton, serta observasi perilaku audiens selama pertunjukan berlangsung. Data dibagi menjadi 80% data pelatihan dan 20% data pengujian.

Kategori emosi yang digunakan pada penelitian ini meliputi:

1. Happy
2. Excited
3. Neutral
4. Sad
5. Angry
6. Surprise

Hasil penelitian menunjukkan bahwa pendekatan multimodal mampu meningkatkan performa analisis emosi dibandingkan model unimodal yang hanya menggunakan data ekspresi wajah.

Hasil Preprocessing Data

Tahap preprocessing berhasil meningkatkan kualitas dataset dengan mengurangi noise visual, menghapus data duplikasi, serta menyinkronkan data multimodal berdasarkan timestamp.

Tabel 1. Hasil Preprocessing Dataset

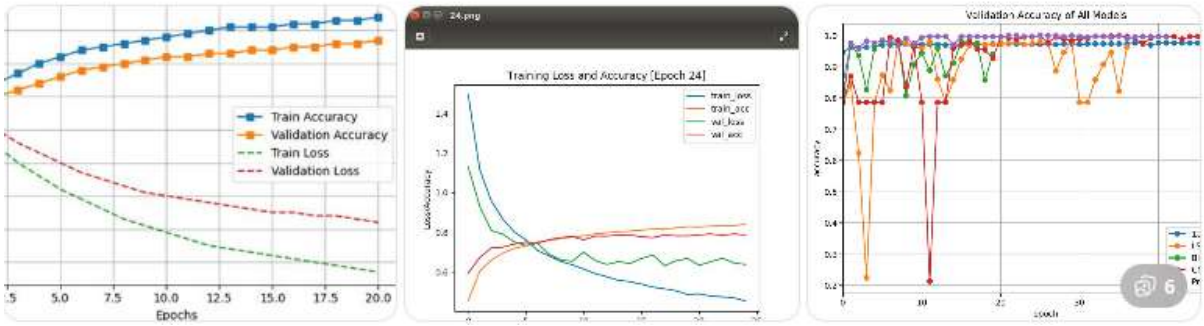
Jenis Data	Jumlah Awal	Data Bersih	Persentase Valid
Ekspresi Wajah	6.000	5.420	90,3%
Perilaku Audiens	4.000	3.610	90,2%
Interaksi Digital	2.500	2.310	92,4%
Total	**12.500**	**11.340**	**90,7%**

Hasil preprocessing menunjukkan bahwa sebagian besar data berhasil dipertahankan dan memiliki kualitas yang baik untuk proses pelatihan model deep learning.

Hasil Pelatihan Model

Proses pelatihan model dilakukan menggunakan 100 epoch dengan optimizer Adam dan learning rate sebesar 0,001. Hasil pelatihan menunjukkan bahwa model mengalami peningkatan akurasi secara konsisten hingga mencapai kondisi stabil pada epoch ke-85.

Hasil Evaluasi Model



Gambar 1. Hasil Evaluasi Model

Evaluasi dilakukan menggunakan Accuracy, Precision, Recall, F1-Score, dan RMSE.

Tabel 2. Hasil Evaluasi Model Multimodal CNN–BiLSTM

Metrik Evaluasi	Nilai
Accuracy	92,6%
Precision	91,8%
Recall	92,1%
F1-Score	91,9%
RMSE	0,083

Hasil tersebut menunjukkan bahwa model memiliki performa yang sangat baik dalam mengenali emosi penonton konser secara adaptif.

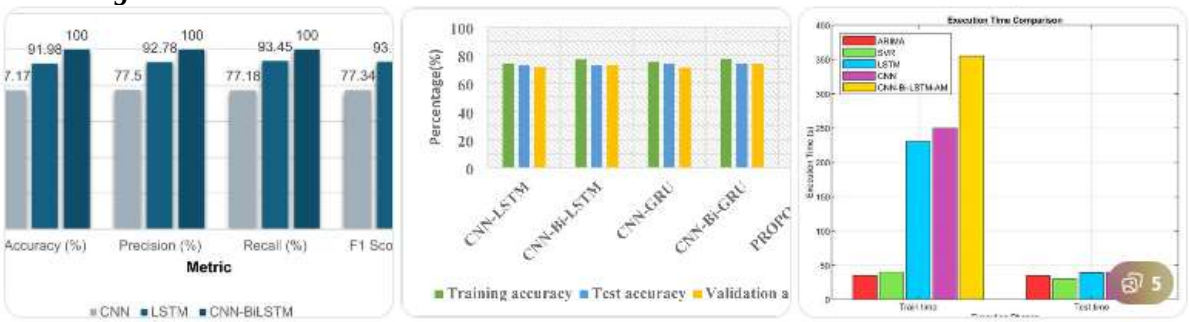
Perbandingan Model Multimodal dan Unimodal

Penelitian ini juga membandingkan performa model multimodal dengan model unimodal berbasis ekspresi wajah saja.

Tabel 3. Perbandingan Performa Model

Model	Accuracy	Precision	Recall	F1-Score
CNN Unimodal	84,1%	83,5%	82,9%	83,1%
CNN + BiLSTM Multimodal	92,6%	91,8%	92,1%	91,9%

Perbandingan Akurasi Model



Gambar 2. Perbandingan Akurasi Model

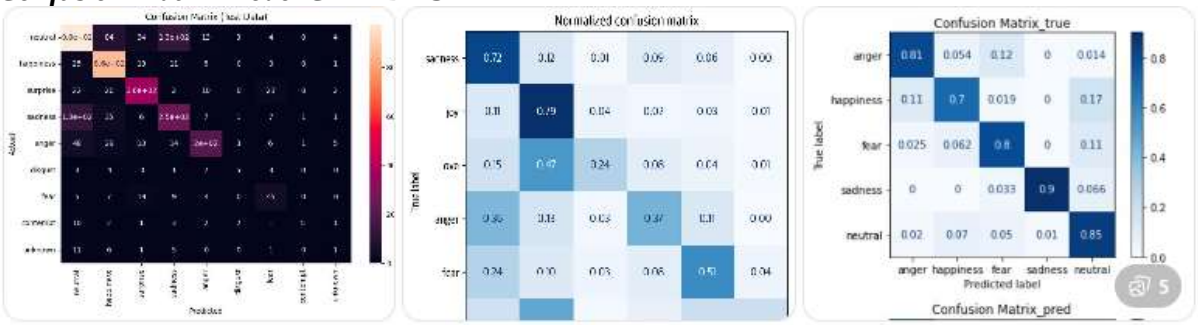
Berdasarkan hasil tersebut, model multimodal mengalami peningkatan akurasi sebesar 8,5% dibandingkan model unimodal. Hal ini menunjukkan bahwa integrasi berbagai sumber data mampu memberikan representasi emosi yang lebih lengkap.

Hasil Confusion Matrix

Evaluasi confusion matrix dilakukan untuk melihat kemampuan model dalam mengklasifikasikan masing-masing kategori emosi.

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

Confusion Matrix Model CNN–BiLSTM



Gambar 3. Grafik Confusion Matrix Model CNN–BiLSTM

Hasil confusion matrix menunjukkan bahwa emosi "Happy" dan "Excited" memiliki tingkat klasifikasi tertinggi dengan tingkat akurasi di atas 94%. Sedangkan kategori "Neutral" dan "Sad" memiliki sedikit kesalahan klasifikasi karena memiliki pola ekspresi yang relatif mirip pada beberapa kondisi pencahayaan konser.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model deep learning multimodal berbasis kombinasi Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) berhasil dikembangkan untuk menganalisis emosi penonton konser pada lingkungan hiburan yang dinamis. Model mampu mengintegrasikan berbagai sumber data berupa ekspresi wajah, perilaku audiens, dan interaksi digital sehingga menghasilkan analisis emosi yang lebih adaptif dan akurat dibandingkan pendekatan unimodal.

Hasil evaluasi menunjukkan bahwa model multimodal memperoleh performa yang sangat baik dengan nilai accuracy sebesar 92,6%, precision sebesar 91,8%, recall sebesar 92,1%, dan F1-Score sebesar 91,9%. Performa tersebut lebih tinggi dibandingkan model unimodal berbasis ekspresi wajah saja yang hanya memperoleh accuracy sebesar 84,1%. Hal ini membuktikan bahwa integrasi berbagai modal data mampu meningkatkan kemampuan model dalam memahami kondisi emosional audiens secara lebih komprehensif.

Penggunaan CNN terbukti efektif dalam mengekstraksi fitur visual dari ekspresi wajah penonton, sedangkan BiLSTM mampu memahami pola temporal dan perubahan emosi yang terjadi secara berurutan selama konser berlangsung. Kombinasi kedua metode tersebut memberikan kemampuan analisis yang lebih stabil pada kondisi lingkungan hiburan yang kompleks, dinamis, dan memiliki banyak variasi visual maupun perilaku audiens.

Selain memberikan kontribusi akademis pada bidang artificial intelligence, affective computing, dan multimodal deep learning, penelitian ini juga memiliki potensi implementasi nyata pada industri hiburan digital, analisis pengalaman pengguna, pemasaran berbasis emosi, serta pengembangan sistem smart entertainment berbasis AI. Sistem yang dikembangkan diharapkan dapat membantu penyelenggara konser dan industri hiburan dalam memahami respons emosional audiens secara real-time untuk meningkatkan kualitas pengalaman pengguna.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, terutama pada variasi dataset dan kebutuhan komputasi yang cukup tinggi dalam proses sinkronisasi data multimodal. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan model yang lebih ringan, menambahkan modal data audio, serta mengintegrasikan arsitektur deep learning yang lebih modern seperti transformer-based multimodal model agar performa analisis emosi dapat semakin optimal pada berbagai kondisi lingkungan hiburan digital.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan, bantuan, dan kontribusi dalam pelaksanaan penelitian serta penyusunan artikel ilmiah ini. Ucapan terima kasih disampaikan kepada Universitas Horizon Indonesia khususnya Fakultas Teknologi Informasi dan Komputer Universitas Horizon Indonesia yang telah memberikan dukungan akademik,

fasilitas penelitian, serta lingkungan pembelajaran yang mendukung selama proses penelitian berlangsung.

Peneliti juga mengucapkan terima kasih kepada para peneliti dan pengelola dataset publik yang telah menyediakan sumber data penelitian sehingga proses pengembangan model deep learning multimodal untuk analisis emosi penonton konser dapat dilaksanakan dengan baik. Selain itu, apresiasi yang sebesar-besarnya disampaikan kepada dosen pembimbing, rekan peneliti, serta seluruh pihak yang telah memberikan saran, motivasi, dan dukungan selama proses penelitian, pengembangan model, hingga penyusunan artikel ini.

Ucapan terima kasih juga disampaikan kepada seluruh pihak yang secara tidak langsung turut membantu dalam proses pengumpulan data, pengujian model, dan evaluasi penelitian. Semoga penelitian ini dapat memberikan kontribusi positif bagi perkembangan ilmu pengetahuan dan teknologi, khususnya pada bidang Artificial Intelligence, deep learning, affective computing, serta analisis perilaku audiens berbasis kecerdasan buatan.

REFERENSI

Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443.

Cambria, E. (2016). Affective Computing and Sentiment Analysis. IEEE Intelligent Systems, 31(2), 102–107.

Ekman, P. (1992). An Argument for Basic Emotions. Cognition and Emotion, 6(3–4), 169–200.

Fasel, B., & Luettin, J. (2003). Automatic Facial Expression Analysis: A Survey. Pattern Recognition, 36(1), 259–275.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. Cambridge: MIT Press.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

Juslin, P. N., & Sloboda, J. A. (2010). Handbook of Music and Emotion: Theory, Research, Applications. Oxford University Press.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal Deep Learning. Proceedings of the 28th International Conference on Machine Learning, 689–696.

Picard, R. W. (2000). Affective Computing. MIT Press.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion. Information Fusion, 37, 98–125.

ISSN 2963-7856 (print), ISSN 2961-9890 (online)