OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

A Systematic Study of Multimodal Emotion Detection in Customer System

Rifqi Favian Hibatullah[1], Rachmad Zildjian[2], Anas Suharman [3], Riddy Bachtiar [4], Mohammad Givi Efgivia[5]

[1]University of Muhammadiyah Prof. Dr. HAMKA, School of Tehcnology Industry and Informatics ,
Tanah Merdeka Road, East Jakarta 13830, Indonesia
[2]University of Muhammadiyah Prof. Dr. HAMKA, School of Tehcnology Industry and Informatics ,
Tanah Merdeka Road, East Jakarta 13830, Indonesia
[3]University of Muhammadiyah Prof. Dr. HAMKA, School of Tehcnology Industry and Informatics ,
Tanah Merdeka Road, East Jakarta 13830, Indonesia
[4]University of Muhammadiyah Prof. Dr. HAMKA, School of Tehcnology Industry and Informatics ,
Tanah Merdeka Road, East Jakarta 13830, Indonesia
[5]University of Muhammadiyah Prof. Dr. HAMKA, School of Tehcnology Industry and Informatics ,
Tanah Merdeka Road, East Jakarta 13830, Indonesia

Abstract
Customer emotion detection is a critical aspect in improving customer experience and happiness in the digital era. This study examines several experiments on the Customer Emotion Detecting Device (CEDD), an AI-based system that uses multimodal sensors—such as facial expressions, vocal tone, and physiological signals—for real-time emotion recognition. CEDD increases accuracy and contextual relevance by combining Convolutional Neural Networks (CNN) with Recurrent Neural Networks (RNN). Its applications are many, including retail, healthcare, and education. However, issues like as data privacy, algorithmic bias, and high implementation costs remain important to its ethical and effective application.

Keywords: *emotion detection, artificial intelligence (ai), customer experience, real-time analysis, physiological sensors.*

Date of Submission: 05-06-2025 Date of acceptance: 17-06-2025

I. Introduction

Artificial Intelligence (AI) has significantly reshaped the landscape of entrepreneurship and business by enhancing operational effectiveness, supporting data-driven decision-making, and elevating customer experience. As AI technology continues to evolve rapidly, its utilization becomes essential for increasing productivity, maintaining competitiveness, and fostering innovation in numerous industry sectors. Despite these advantages, several challenges accompany AI adoption, including algorithmic bias, workforce displacement, and potential regulatory violations (Md Firoz & Md Atiqur, 2025). Beyond its role in mobile devices, AI is steadily becoming a strategic asset in the business domain, adopted by both large corporations and small to medium-sized enterprises (SMEs). It is now widely implemented in modern business environments, continuously developed by IT experts and business professionals. Nevertheless, in countries such as Ukraine, many businesses still encounter significant barriers to accessing and deploying intelligent technologies, often perceiving them as costly or complex (Kraus et al., 2022).

In the academic literature, efforts have been made to categorize emotional concepts such as emotion, affect, feeling, and mood:

- Emotion is described as an immediate and intense reaction to stimuli.
- Affect refers to the outcome of ongoing emotional and social interactions.
- Feelings are subjective emotional experiences linked to specific events or memories.
- Mood denotes a more subtle, prolonged emotional state that influences general disposition.

With the ongoing progress in smart technologies and industry, there is a growing demand for systems that can interpret human emotional states. Automated Emotion Evaluation (AEE) has become a valuable component in fields like robotics, advertising, education, and entertainment (Dzedzickis et al., 2020) The strategy of emotion-based marketing seeks to create emotional resonance with consumers, strengthening their connection with a brand and ultimately boosting sales performance.

Although notable progress has been achieved in emotion recognition via multimodal inputs such as facial cues, speech tone, and bodily gestures, emotion detection from textual sources remains a complex challenge. High-performing text-based emotion recognition systems still face limitations, especially in detecting emotion without relying on the text's length while also accounting for context and expression variability. Generally, there are four primary methods for identifying emotions in text: 1) keyword-based techniques, 2) machine learning approaches, 3) lexical affinity models, and 4) hybrid frameworks. Each has its own strengths and limitations. Among them, hybrid models show great promise due to their ability to merge multiple approaches, though selecting the optimal combination remains a key challenge (Ramalingam et al., 2018).

The task of emotion recognition through images is another area of research with promising applications in social communication but also considerable complexity. Deep learning (DL)-based techniques have demonstrated superior outcomes compared to conventional image processing approaches (Palash & Bhargava, n.d.). This paper introduces an AI-powered system designed to identify human emotions using facial expressions. The method includes three primary phases: detecting facial regions, extracting facial features, and classifying the corresponding emotion. The architecture is based on Convolutional Neural Networks (CNNs), a deep learning model tailored for image analysis tasks. Its effectiveness is evaluated using two datasets: the Facial Emotion Recognition Challenge (FERC-2013) and the Japanese Female Facial Expression (JAFFE) dataset. Results show that the system achieved an accuracy of 70.14% on FERC-2013 and 98.65% on JAFFE. (2020 International Conference for Emerging Technology (INCET), 2020).

In this regard, the Customer Emotion Detecting Device (CEDD) stands out as a technological innovation aimed at delivering real-time emotional analysis of consumers with high precision while embedding ethical safeguards to mitigate risks related to privacy and algorithmic bias (Bui, 2021). A practical response to these issues is the integration of Emotional Intelligence (EI) into AI systems, enabling them to recognize emotions through multimodal machine learning methods like CNN and Recurrent Neural Networks (RNN). Nonetheless, challenges remain due to cultural differences, complex data handling, and ethical concerns involving potential biases and privacy threats (Narimisaei et al., 2024).

II. Method

This study applies a systematic framework to explore the potential use of a Customer Emotion Detecting Device (CEDD) in diverse industrial sectors. The CEDD system is developed using Artificial Intelligence (AI) technologies, focusing on deep learning methods, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). CNN is a type of neural network engineered for pattern recognition and is highly efficient for analyzing visual data (Transformer, 2023). Its structure consists of several core layers: an input layer to receive raw imagery, convolutional layers to extract features through filter operations, activation layers for introducing non-linearity, pooling layers to reduce dimensionality and computational load, and fully connected layers that output predictions in vector form. CNNs are widely utilized in areas such as visual surveillance, object detection, and image-based classification. The accuracy of feature extraction and decision-making can be enhanced by integrating CNN with approaches like YOLO and Gaussian Mixture Models (GMM) (Jogin et al., 2018). Meanwhile, RNNs are designed to process sequential data by forming directed cycles in their architecture, enabling them to maintain context over time. Unlike standard feedforward networks, RNNs can store and reuse information from earlier steps in a sequence, making them particularly effective for tasks such as continuous speech and handwriting recognition (Murugan, 2018). By combining CNN and RNN, the CEDD system is able to conduct more precise and context-aware emotion analysis. The upcoming sections will describe the system's architecture, data flow, and model design in greater depth.

III. Results

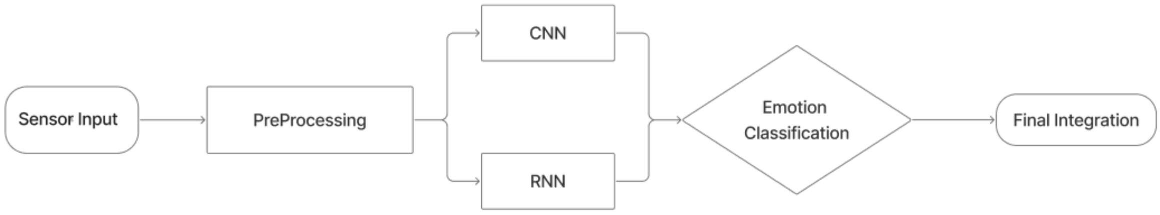
3.1 How CEDD Works

The Customer Emotion Detecting Device (CEDD) operates by identifying and analyzing users' emotional states in real time, utilizing multiple sensors to capture emotional cues from customer interactions (Almulla, 2024). The process initiates with the data acquisition phase, during which tools like cameras (for facial expression detection), microphones (for analyzing voice tone), and physiological sensors (such as heart rate monitors or galvanic skin response detectors) gather raw data from the user. Once the input is collected, it proceeds to the pre-processing stage, where raw signals are cleaned, normalized, and transformed for analysis. For instance, facial images are cropped and resized, and audio inputs are filtered to remove ambient noise—ensuring that only emotion-relevant features remain. The refined data is then processed through two core pathways within the AI model: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

CNNs are responsible for analyzing static visual elements of the face, such as smiling patterns, forehead muscle tension, or wrinkles around the eyes (Ouyang, 2023). These networks automatically extract spatial features that correlate with basic emotional expressions like happiness, anger, or surprise (Lian et al., 2023). Meanwhile, RNNs handle temporal or sequential data, such as changes in voice pitch or speech flow. Their design allows the system to interpret emotional context over time by following the sequence of vocal and facial changes.

The final stage involves classification and emotion inference, where outputs from the CNN and RNN models are processed either independently or in a hybrid structure (Kim et al., 2023). The system calculates a confidence score to determine the most probable emotional state of the user—such as joy, sadness, fear, or neutrality (Pan & Wu, 2023). These insights can then be applied in practical domains such as customized customer service, intelligent product recommendation systems, or automated decision-making processes (Ullah et al., 2023).

CEDD Flow



3.2 The Potential of CEDD

The Customer Emotion Detecting Device (CEDD) holds significant potential across various industries due to its ability to assess and interpret human emotions in real-time using advanced sensors and artificial intelligence algorithms (Rajapakshe et al., n.d.). This technology enables businesses to improve consumer interactions, tailor offerings, and provide more personalized experiences.

1. Retail Industry

Emotion detection technology has revolutionized customer experience, particularly in the retail sector. By utilizing a combination of machine learning algorithms, sensors, and biometric analysis—based on facial expressions, voice intonation, and other physiological cues—these systems can accurately identify human emotions. This not only improves direct customer interactions but also contributes to increased sales and long-term strategic planning in the market (Van Thanh, n.d.).

2. Health Sector

In the healthcare industry, emotion recognition technologies have gained substantial attention, particularly in three areas: application scenarios, multimodal approaches, and clinical applications. Key implications include:

(i) Healthcare professionals can now detect and manage emotions remotely, thanks to the advancement of emotion identification technologies.

(ii) The use of a multimodal approach based on objective physiological signals is expected to improve the accuracy of medical diagnoses, replacing the traditional reliance on subjective emotional assessments.

(iii) Real-time emotion monitoring has proven especially valuable in clinical practice, given the increasing recognition of the connection between emotional states and health conditions during diagnosis, treatment, and recovery (Guo et al., 2024).

3. Academy

For instance, Khuntia and Kale (2024) developed a system capable of recognizing and converting facial expressions into emojis to offer immediate feedback in online learning environments. The system employs image processing algorithms and deep neural networks to accurately classify facial expressions, which are then used to personalize learning content in real-time. This approach highlights how incorporating CEDD technology into e-learning platforms can enhance student interaction and engagement by providing timely and meaningful emotional feedback (Khuntia & Kale, 2024).

IV. Discussion

While the Customer Emotion Detecting Device (CEDD) holds vast potential across various industries, its implementation faces several significant challenges (Geng et al., 2023). A primary concern is the accuracy of emotion detection, which can be affected by factors such as environmental conditions, individual differences, and the quality of input data, such as facial expressions or vocal tone (Latif et al., 2022). These variables can make emotion recognition systems prone to errors, especially in complex, dynamic real-world situations.

In addition, the cost of implementing CEDD—particularly those involving advanced sensors and AI-driven processing infrastructure—remains high, posing a barrier to widespread adoption, particularly in areas with limited resources (Häuselmann et al., n.d.). Ethical concerns also play a major role, especially regarding user privacy. The collection of biometric data, including facial features, voice, and gestures, raises serious privacy risks, especially if collected without clear consent or adequate data protection measures (Mattioli & Cabitza, 2024).

Moreover, ethical concerns regarding privacy remain a critical issue. The gathering of sensitive biometric data, such as facial recognition, voice recordings, and gestures, can lead to potential privacy breaches if done without explicit consent or proper safeguards (Latif et al., 2022). Consequently, there is a pressing need for stringent regulations and a transparent, participatory approach in the development and deployment of these technologies to ensure their fair and ethical use across various industries and society as a whole.

CEDD is a deep learning-driven innovation that uses a contextual and multimodal approach to detect human emotions more accurately and flexibly. By combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), CEDD can simultaneously process facial expressions, voice cues, and conversational context, making it a promising tool for industries such as retail, healthcare, education, and entertainment.

This cross-industry potential highlights that emotions are not merely subjective elements; they can be seen as valuable strategic data for enhancing service personalization, predicting user behavior, and enabling more human-centered decision-making. However, issues like privacy concerns, algorithmic bias, and the need for ethical oversight remain essential challenges that must be addressed before CEDD can be widely and responsibly adopted. With continued development and an inclusive approach, CEDD has the potential to become a pivotal technology in driving future digital transformation through empathy and a deeper understanding of human emotions.

Acknowledgments

This study was supported by a research grant from the Ministry of Education, Culture, Research, and Technology of Indonesia. We would like to express our sincere gratitude to our colleagues in the Department of Informatics, Universitas Muhammadiyah Prof. Dr. HAMKA, for their valuable feedback and assistance during data collection and manuscript refinement. The authors contributed equally to the conception, analysis, and writing of this study. We also extend our heartfelt thanks to Dr. Ir. Mohammad Givi Efgivia., M.Kom for her dedicated help in the preparation and formatting of the manuscript.

References

[1]. Md Firoz, K., & Md Atiqur, R. (2025). The Impact of Artificial Intelligence on Business Transformation: Enhancing Decision-Making, Operational Efficiency, and Customer Experience. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/IJSAT.v16.i1.2421

[2]. Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. In Sensors (Switzerland) (Vol. 20, Issue 3). MDPI AG. https://doi.org/10.3390/s20030592

[3]. Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology — a systematic literature review. In BMC Psychology (Vol. 12, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s40359-024-01581-4

[4]. Häuselmann, A., Fosch-Villaronga, E., Sears, A. M., & Zard, L. (n.d.). EU law and emotion data.

[5]. Jogin, M., Mohana, Madhulika, M. S., Divya, G. D., Meghana, R. K., & Apoorva, S. (2018). Feature extraction using convolution neural networks (CNN) and deep learning. 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018 - Proceedings, 2319–2323. https://doi.org/10.1109/RTEICT42901.2018.9012507

[6]. Khuntia, A., & Kale, S. (2024). Real Time Emotion Analysis Using Deep Learning for Education, Entertainment, and Beyond. http://arxiv.org/abs/2407.04560

[7]. Kraus, K., Kraus, N., Hryhorkiv, M., Kuzmuk, I., & Shtepa, O. (2022). Artificial Intelligence in Established of Industry 4.0. WSEAS Transactions on Business and Economics, 19, 1884–1900. https://doi.org/10.37394/23207.2022.19.170

[8]. Latif, S., Ali, H. S., Usama, M., Rana, R., Schuller, B., & Qadir, J. (2022). AI-Based Emotion Recognition: Promise, Peril, and Prescriptions for Prosocial Path. http://arxiv.org/abs/2211.07290

[9]. Mattioli, M., & Cabitza, F. (2024). Not in My Face: Challenges and Ethical Considerations in Automatic Face Emotion Recognition Technology. In Machine Learning and Knowledge Extraction (Vol. 6, Issue 4, pp. 2201–2231). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/make6040109

[10]. Murugan, P. (2018). Learning The Sequential Temporal Information with Recurrent Neural Networks. http://arxiv.org/abs/1807.02857

[11]. Narimisaei, J., Naeim, M., Imannezhad, S., Samian, P., & Sobhani, M. (2024). Exploring emotional intelligence in artificial intelligence systems: a comprehensive analysis of emotion recognition and response mechanisms. Annals of Medicine & Surgery,

86(8), 4657–4663. https://doi.org/10.1097/ms9.0000000000002315

[12]. Ramalingam, V. V., Pandian, A., Jaiswal, A., & Bhatia, N. (2018). Emotion detection from text. Journal of Physics: Conference Series, 1000(1). https://doi.org/10.1088/1742-6596/1000/1/012027

[13]. Van Thanh, N. (n.d.). Emotion Recognition Systems in Retail A Detailed Analysis of Their Role in Enhancing Customer Interactions, Driving Sales, and Predicting Trends. In Journal of Computational Social Dynamics Research Article: Journal of Computational Social Dynamics.

[14]. Almulla, M. A. (2024). A multimodal emotion recognition system using deep convolution neural networks. Journal of Engineering Research (Kuwait), March. https://doi.org/10.1016/j.jer.2024.03.021

[15]. Bui, H. (2021). Facial Expression Recognition with CNN-LSTM Facial Expression Recognition with CNN-LSTM. January. https://doi.org/10.1007/978-981-15-7527-3

[16]. Geng, L., Fu, H., Tao, H., Lu, Y., Guo, X. & Zhao, L. (2023). Speech Emotion Recognition Based on Dynamic Convolution Recurrent Neural Network. Jisuanji Gongcheng/Computer Engineering, 49(4). https://doi.org/10.19678/j.issn.1000-3428.0064054

[17]. Kim, D. H., Son, W. H., Kwak, S. S., Yun, T. H., Park, J. H. & Lee, J. D. (2023). A Hybrid Deep Learning Emotion Classification System Using Multimodal Data. Sensors, 23(23). https://doi.org/10.3390/s23239333

[18]. Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C. & Zong, Y. (2023). Recognition : Speech , Text , and Face. 1–33.

[19]. Ouyang, Q. (2023). Speech emotion detection based on MFCC and CNN-LSTM architecture. Applied and Computational Engineering, 5(1), 243–249. https://doi.org/10.54254/2755-2721/5/20230570

[20]. Palash, M. & Bhargava, B. (n.d.). EMERSK -Explainable Multimodal Emotion Recognition with Situational Knowledge. 1–11.

[21]. Pan, S. T. & Wu, H. J. (2023). Performance Improvement of Speech Emotion Recognition Systems by Combining 1D CNN and LSTM with Data Augmentation. Electronics (Switzerland), 12(11). https://doi.org/10.3390/electronics12112436

[22]. Rajapakshe, T., Rana, R., Khalifa, S. & Sisman, B. (n.d.). Enhancing Speech Emotion Recognition Through Differentiable Architecture Search. 1–5.

[23]. Transformer, V. (2023). Facial Micro-Expression Recognition Enhanced by Score Fusion and a Hybrid Model from Convolutional LSTM and Vision Transformer.

[24]. Ullah, R., Asif, M., Shah, W. A., Anjam, F., Ullah, I., Khurshaid, T., Wuttisittikulkij, L., Shah, S., Ali, S. M. & Alibakhshikenari, M. (2023). Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. Sensors, 23(13), 1–20. https://doi.org/10.3390/s23136212