JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

MULTI VIEW FEATURE FUSION FOR INDUSTRIAL ANOMALY DETECTION USING 1D-CNN

Daniel Fernando Nainggolan[1], Puguh Hiskiawan [1*]
[1]Data Science Department, Universitas Bunda Mulia
email: *phiskiawan@bundamulia.ac.id

Abstract: Anomalous sound detection is essential for industrial predictive maintenance, as machine failures often originate from subtle acoustic changes during operation. However, high background noise and limitations of conventional Convolutional Neural Networks (CNN) reduce detection reliability. This study proposes a 1D-CNN-based anomaly detection framework with multi-view feature fusion and temporal segmentation to enhance detection performance. The approach combines MFCC, Log-Mel Spectrogram, and Chroma STFT features, while temporal segmentation divides audio signals into 5-second segments to better capture transient anomalies. Experiments on the MIMII dataset under varying Signal-to-Noise Ratio (SNR) conditions show that MFCC and Log-Mel fusion achieves the best performance, with 97.90% accuracy and ROC-AUC of 0.9789. The model maintains accuracy above 90% at −6 dB, demonstrating strong robustness in noisy industrial environments.

Keywords: industrial anomaly detection; 1D-CNN; multi-view feature fusion; temporal segmentation; MIMII dataset.

Abstrak: Deteksi anomali suara merupakan komponen penting dalam sistem pemeliharaan prediktif industri, karena kegagalan mesin sering diawali oleh perubahan akustik yang bersifat halus selama proses operasi. Namun, tingkat kebisingan yang tinggi serta keterbatasan arsitektur Convolutional Neural Network (CNN) konvensional dapat menurunkan keandalan deteksi. Penelitian ini bertujuan mengusulkan kerangka deteksi anomali berbasis 1D-CNN yang mengintegrasikan strategi fusi fitur multi-view dan segmentasi temporal untuk meningkatkan kinerja deteksi. Pendekatan yang digunakan menggabungkan fitur MFCC, Log-Mel Spectrogram dan Chroma STFT, sementara teknik temporal splitting membagi sinyal audio menjadi segmen berdurasi 5 detik untuk menangkap anomali yang bersifat sementara. Eksperimen menggunakan dataset MIMII pada berbagai kondisi Signal-to-Noise Ratio (SNR) menunjukkan bahwa kombinasi MFCC dan Log-Mel Spectrogram menghasilkan kinerja terbaik dengan akurasi 97,90% dan ROC-AUC sebesar 0,9789. Model juga mempertahankan akurasi di atas 90% pada kondisi kebisingan ekstrem (−6 dB) yang menunjukkan ketahanan yang baik dalam lingkungan industri yang bising.

Kata kunci: deteksi anomali industri; 1D-CNN; fusi fitur multi-view; segmentasi temporal; dataset MIMII

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Sustainable industrial development depends heavily on the reliability and performance of production machinery operating in sectors such as manufacturing, energy, and transportation [1], [2]. Machine failures can disrupt system operations and lead to significant financial losses, increased maintenance costs, and safety risks [3], [4]. To address these challenges, maintenance strategies have evolved from reactive and preventive approaches toward predictive maintenance, which enables early detection of abnormal machine behavior and supports timely intervention to minimize operational disruptions [5]

Acoustic signal analysis has emerged as an effective approach for predictive maintenance due to its non-destructive nature, cost efficiency, and capability for real-time monitoring [8], [9]. Machine operations produce characteristic acoustic signatures under normal conditions, while deviations from these patterns may indicate early-stage faults [10]. However, industrial environments typically contain high levels of background noise, and raw audio signals exhibit high dimensionality and complex temporal structures. In addition, short-duration anomalies may be obscured when long audio recordings are processed as a single input, making reliable detection more challenging [11].

Various audio feature extraction techniques have been widely adopted to address these challenges, including Mel-Frequency Cepstral Coefficients (MFCC), Log-Mel Spectrogram, and Chroma Short-Time Fourier Transform (Chroma STFT), which capture complementary acoustic characteristics such as timbre, spectral energy, and harmonic content [12], [13]. Nevertheless, previous studies report inconsistent performance across different feature representations, and no consensus has been established regarding the most effective approach for industrial anomaly detection, particularly under noisy conditions [14]. Furthermore, limited attention has been given to integrating multiple feature representations with temporal segmentation strategies to enhance the detection of transient anomalies.

This study proposes a robust industrial audio anomaly detection framework based on a 1D-Convolutional Neural Network (1D-CNN) combined with multi-view feature fusion and temporal splitting. The proposed approach integrates MFCC, Log-Mel Spectrogram, and Chroma STFT features to capture complementary information, while temporal splitting divides audio signals into shorter segments to improve sensitivity to transient anomalies. Experiments are conducted using the MIMII dataset with pump machine recordings under multiple noise conditions [15]. The contributions of this study include a systematic evaluation of multi-feature representations, improved anomaly detection performance under noisy environments, and the development of a practical framework for industrial predictive maintenance applications [16].

METHOD

Research Design

This research adopts a comparative experimental design to evaluate the effectiveness of different audio signal transformation techniques for industrial machine anomaly detection [17]. The experiment examines preprocessing strategies and feature extraction methods as independent variables, while anomaly detection performance is treated as the dependent variable. The proposed ap-

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

proach, which integrates Feature Fusion and Temporal Splitting, is compared with a baseline model using quantitative evaluation metrics [18]. As illustrated in Figure 1, the research workflow begins with data acquisition from the MIMII dataset, followed by temporal segmentation, feature extraction using MFCC, Log-Mel Spectrogram, and Chroma STFT, and classification using a 1D-CNN model. The final stage evaluates model performance under varying Signal-to-Noise Ratio (SNR) conditions to assess robustness in noisy industrial environments [19].

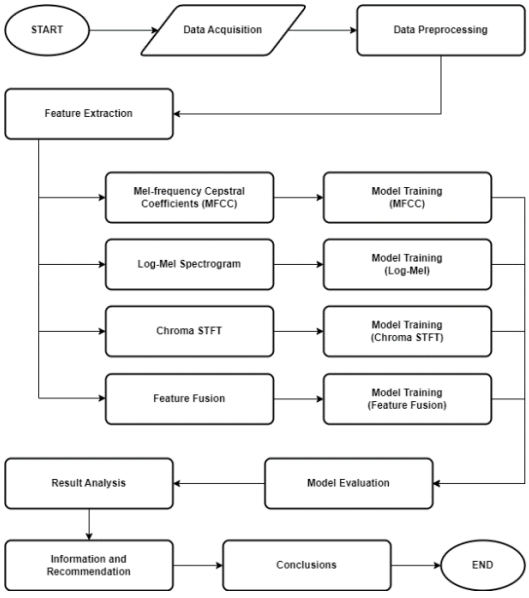

Image 1. Research Workflow of the Proposed Anomaly Detection System

Data Acquisition

This study utilizes audio recordings from the publicly available MIMII (Malfunctioning Industrial Machine Investigation and Inspection) dataset, which contains machine sound data collected under both normal and malfunctioning operating conditions in real industrial environments [15], [20]. Audio signals are recorded using microphones positioned near operating machines to capture realistic acoustic characteristics. The experiment focuses on pump machine recordings (machine ID 00), including both normal and anomalous conditions to enable the model to learn distinct operational patterns. To evaluate robustness, the dataset is tested under varying noise conditions with Signal-to-Noise Ratio (SNR) levels of 6 dB, 0 dB, and −6 dB, simulating realistic industrial environments affected by background noise [21].

Data Preprocessing

Data preprocessing prepares raw audio recordings for feature extraction and model training. In the MIMII dataset, each audio sample has a duration of 10 seconds, which may obscure short-duration anomalies. To address this, temporal splitting is applied to divide each recording into 5-second segments, increasing the number of training samples and improving the model's ability to capture transient anomaly patterns [22]. After segmentation, the dataset is partitioned into training, validation, and testing subsets using an 80:20 split between training and testing data, with a portion of the training set allocated for validation. The final distribution consists of 64% training, 16% validation, and 20% testing data, ensuring balanced evaluation and reducing the risk of overfitting [23], [24].

Feature Extraction and Feature Fusion

Feature extraction transforms segmented audio signals into structured numerical representations suitable for machine learning models. This study employs three complementary feature extraction techniques, namely Mel-Frequency Cepstral Coefficients (MFCC), Log-Mel Spectrogram, and Chroma Short-Time Fourier Transform

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

(Chroma STFT), which capture timbral characteristics, time–frequency energy distribution, and harmonic pitch information, respectively [17], [25]. To leverage the complementary nature of these features, feature fusion is implemented using a vector concatenation strategy, where feature representations are aligned, flattened and combined into a unified feature vector. Multiple feature combinations are evaluated, including MFCC + Log-Mel, MFCC + Chroma STFT and Log-Mel + Chroma STFT. Prior to fusion, all features are normalized using Standard Scaler to ensure consistent scaling, improve model stability, and prevent dominance of features with larger numerical ranges [21].

Model Architecture and Training

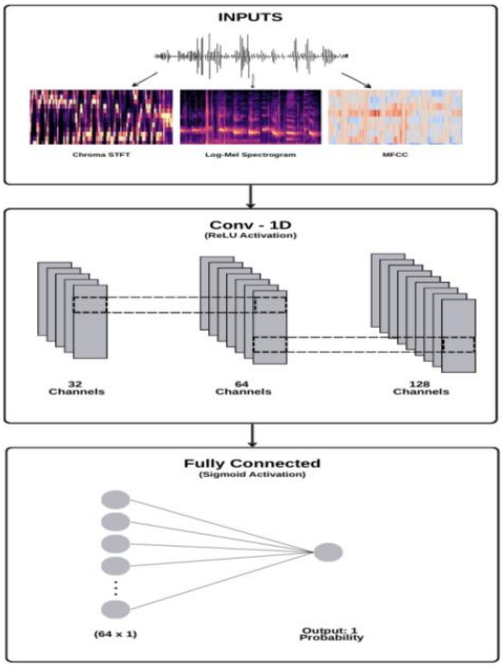

Image 2. Architecture 1D-CNN Model for industrial machine anomaly detection

Anomaly detection in this study is performed using a one-dimensional Convolutional Neural Network (1D-CNN) designed to process sequential acoustic features derived from MFCC, Log-Mel Spectrogram, Chroma STFT, and their fusion combinations [26], [27]. As illustrated in Figure 2, the proposed model consists of three convolutional blocks with 32, 64, and 128 filters, each followed by Batch Normalization and Max-Pooling layers to stabilize learning and reduce feature dimensionality. Dropout regularization is applied to improve generalization, with rates of 0.2 in the first block and 0.3 in subsequent blocks [28]. The extracted feature maps are then flattened and passed to a fully connected Dense layer with 64 neurons using ReLU activation, followed by a Sigmoid output layer that produces a probability score for binary classification of normal and anomalous machine conditions.

Model Evaluation

Model performance is evaluated using quantitative metrics to assess the ability of the proposed model to distinguish between normal and anomalous machine conditions. The evaluation employs Accuracy, Receiver Operating Characteristic–Area Under Curve (ROC-AUC), and Matthews Correlation Coefficient (MCC). Accuracy measures the proportion of correctly classified samples, ROC-AUC evaluates class separability across different classification thresholds, and MCC provides a balanced metric by considering all elements of the confusion matrix, making it suitable for binary classification tasks. In addition, model robustness is assessed under varying environmental noise conditions using Signal-to-Noise Ratio (SNR) levels of 6 dB, 0 dB, and −6 dB, enabling evaluation of model stability in realistic industrial environments affected by background noise.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

RESULT AND DISCUSSION

Comparative Performance of Feature Extraction Methods

Feature fusion consistently outperforms single-feature representations for industrial anomaly detection using the proposed 1D-CNN, as shown in Table 1. The combination of MFCC and Log-Mel Spectrogram achieves the best performance (97.90% accuracy, MCC 0.9051, ROC-AUC 0.9879), while MFCC performs best among individual features (97.10%). In contrast, Chroma STFT shows the lowest performance (93.77%) due to its focus on harmonic structures, which are less relevant for noise-like industrial signals. These results confirm that feature fusion improves detection performance, particularly in noisy environments. These results are competitive with recent studies in industrial audio anomaly detection.

Table 1. Comparative Performance of Feature Extraction Methods

Feature Method	Accuracy (%)	MCC	ROC-AUC
MFCC	97.10	0.8711	0.9801
Log-Mel Spectrogram	95.65	0.8200	0.9746
Chroma STFT	93.77	0.7047	0.9372
MFCC + Log-Mel	**97.90**	**0.9051**	**0.9879**
MFCC + Chroma STFT	96.88	0.8584	0.9904
Log-Mel + Chroma STFT	95.14	0.8080	0.9902

Impact of Temporal Splitting Detection Performance

Temporal segmentation improves anomaly detection performance by enhancing sensitivity to short-duration acoustic irregularities. As shown in Figure 3, models using 5-second segments consistently outperform those using 10-second inputs, with the best accuracy of 97.90% achieved by MFCC and Log-Mel feature fusion. This improvement occurs because shorter segments preserve transient anomaly patterns that may be diluted in longer audio windows.

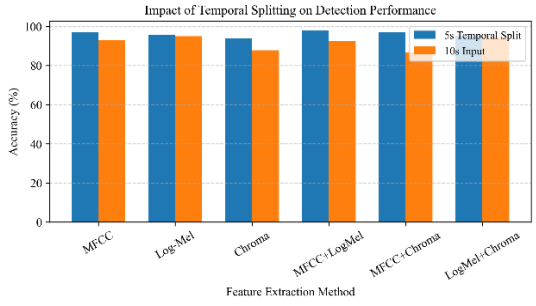

Image 3. Impact of Temporal Splitting on Detection Performance

Robustness Analysis under Different Noise Conditions

Industrial environments often contain background noise that affects anomaly detection performance. As shown in Figure 4, the proposed 1D-CNN maintains high accuracy across SNR levels of 6 dB, 0 dB, and −6 dB, with MFCC and Log-Mel fusion achieving the best results. Although performance decreases at −6 dB, accuracy remains above 90%, indicating strong robustness. These findings show that feature fusion improves stability and enables effective anomaly detection in real-world industrial environments.

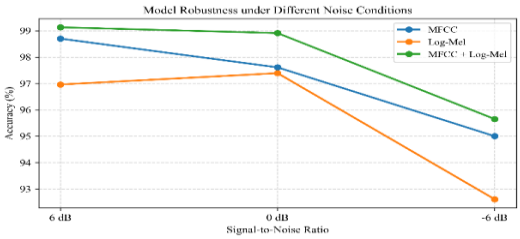

Image 4. Robustness Under Different Signal-to-Noise Ratio Conditions

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 2, Maret 2026, hlm. 391 – 398 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i2.4501
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

CONCLUSION

This research investigates multiple audio feature extraction techniques for industrial anomaly detection using the MIMII dataset. The proposed 1D-CNN with feature fusion and temporal segmentation significantly improves performance, with MFCC and Log-Mel Spectrogram achieving the best results (97.90% accuracy, ROC-AUC 0.9879). The model also maintains accuracy above 90% under noisy conditions (−6 dB), demonstrating strong robustness. Overall, the proposed framework provides an effective solution for predictive maintenance and real-world machine monitoring.

ACKNOWLEGMENTS

The authors gratefully acknowledge Universitas Bunda Mulia for the financial support provided through the Directorate of Research and Community Service, which enabled the completion of this research.

BIBLIOGRAPHY

[1] Unido, "Industrial Development Report 2024 - Turning Challenges Into Sustainable Solutions: The New Era of Industrial Policy," p. 35, 2024.

[2] S. I. Monye, S. A. Afolalu, S. L. Lawal, O. A. Oluwatoyin, and A. G. Adeyemi, "Overview and Impact of Maintenance Process in 4th Industrial Revolution," *E3S Web of Conferences*, vol. 430, pp. 1–12, 2023, doi: 10.1051/e3sconf/202343001220.

[3] M. E. Del Giudice, M. Sharafkhani, M. Di Nardo, T. Murino, and M. C. Leva, "Exploring Safety of Machineries and Training: An Overview of Current Literature Applied to Manufacturing Environments," *Processes*, vol. 12, no. 4, 2024, doi: 10.3390/pr12040684.

[4] I. Rojek, M. Jasiulewicz-Kaczmarek, M. Piechowski, and D. Mikołajewski, "An Artificial Intelligence Approach for Improving Maintenance to Supervise Machine Failures and Support Their Repair," *Applied Sciences (Switzerland)*, vol. 13, no. 8, 2023, doi: 10.3390/app13084971.

[5] M. Molęda, B. Małysiak-Mrozek, W. Ding, V. Sunderam, and D. Mrozek, "From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry," *Sensors*, vol. 23, no. 13, 2023, doi: 10.3390/s23135970.

[6] N. F. M. Hafiz, S. Mashohor, M. H. S. E. M. A. Shazril, A. M. Ali, and M. F. A. Rasid, "Machine Learning Framework for Industrial Machine Sound Classification in Predictive Maintenance," *IEEE Access*, vol. 13, no. August, pp. 154960–154975, 2025, doi: 10.1109/ACCESS.2025.3601999.

[7] A. Senanayaka, P. Lee, N. Lee, C. Dickerson, A. Netchaev, and S. Mun, "Enhancing the accuracy of machinery fault diagnosis through fault source isolation of complex mixture of industrial sound signals," *International Journal of Advanced Manufacturing Technology*, vol. 133, no. 11–12, pp. 5627–5642, 2024, doi: 10.1007/s00170-024-14080-y.

[8] F. A. ERDOĞAN, A. KÜÇÜKMANİSA, and Z. H. KİLİMCİ, "Detection of Fault from Acoustic

Signals in Automobile Engines using Deep Learning Techniques," *Kocaeli Journal of Science and Engineering*, vol. 6, no. 2, pp. 148–154, 2023, doi: 10.34088/kojose.1225591.

[9] M. Romanssini, P. C. C. de Aguirre, L. Compassi-Severo, and A. G. Girardi, "A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery," *Eng*, vol. 4, no. 3, pp. 1797–1817, 2023, doi: 10.3390/eng4030102.

[10] S. Ding, S. Zhang, and C. Yang, "Machine tool fault classification diagnosis based on audio parameters," *Results in Engineering*, vol. 19, no. July, p. 101308, 2023, doi: 10.1016/j.rineng.2023.101308.

[11] M. K. Gourisaria, R. Agrawal, M. Sahni, and P. K. Singh, "Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques," *Discover Internet of Things*, vol. 4, no. 1, 2024, doi: 10.1007/s43926-023-00049-y.

[12] T. T. H. Le, A. A. Adiputra, J. Yun, and H. Kim, "Anomaly Detection in Industrial Machine Sounds Using High-Frequency Features and Gate Recurrent Unit Networks," *IEEE Access*, vol. 13, no. May, pp. 77165–77186, 2025, doi: 10.1109/ACCESS.2025.3565812.

[13] F. Joanda Kaunang, A. Pramana Thenata, B. Hakim, D. Fernando Nainggolan, P. Hiskiawan, and Ranny, "Sound Engine Based In-Situ Environment Leveraging Neural Network Classification Algorithm," in *2025 IEEE International Conference on Artificial Intelligence for Learning and Optimization (ICoAILO)*, 2025, pp. 352–358. doi: 10.1109/ICoAILO66760.2025.11156048.

[14] P. Hiskiawan, S. A. Yasodhara, and D. Alexander, "Mel-Frequency Cepstral Coefficients and Neural Networks for Indonesian Traditional Music Recognition," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2025, pp. 1707–1712.

[15] M. T. Htun, "Compact and Robust MFCC-based Space-Saving Audio Fingerprint Extraction for Efficient Music Identification on FM Broadcast Monitoring," *Journal of ICT Research and Applications*, vol. 16, no. 3, pp. 226–242, Dec. 2022, doi: 10.5614/itbj.ict.res.appl.2022.16.3.3.

[16] R. Artikel *et al.*, "Dilated-Convolutional Recurrent Neural Network untuk Klasifikasi Genre Musik Creative Commons," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, 2024, doi: 10.28932/jutisi.v10i3.9347.

[17] G. Yoo, S. Hong, and H. Kim, "Emotion Recognition and Multiclass Classification in Music with MFCC and Machine Learning," *International Journal on Adavanced Science Engineering Information Technology*, vol. 14, no. 3, 2024, [Online]. Available: https://www.kaggle.com/

[18] F. Mahardhika, M. L. Haryanti, and P. Hiskiawan, "Performance Evaluation of Speech Emotion Recognition Using Hybrid Feature Selection and Machine Learning," in *2025 4th International Conference on Creative Communication*

and Innovative Technology (ICCIT), 2025, pp. 1–7. doi: 10.1109/ICCIT65724.2025.11166879.

[19] A. Alamsyah, F. Ardiansyah, and A. Kholiq, "Music Genre Classification Using Mel Frequency Cepstral Coefficients and Artificial Neural Networks: A Novel Approach," *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 937–948, Dec. 2024, doi: 10.15294/sji.v11i4.13660.

[20] R. Refianti and F. Mahardi, "Comparison of Music Genre Classification Results Using Multilayer Perceptron With Chroma Feature and Mel Frequency Cepstral Coefficients Extraction Features," *International Journal of Engineering, Science and Information Technology*, vol. 3, no. 2, pp. 53–59, 2023, doi: 10.52088/ijesty.v1i4.444.

[21] P. Hiskiawan, J. William, L. Feliepe, and T. Jansel, "A Hybrid Data Science Framework for Forecasting Bitcoin Prices using Traditional and AI Models," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2089–2101, 2025.

[22] M. R. Shadi, H. Mirshekali, and H. R. Shaker, "Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects," Jul. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.rser.2025.115668.

[23] P. Hiskiawan, C. Chih, C. Zheng, and K. Ye, "Processing of electrical resistivity tomography data using convolutional neural network in ERT - NET architectures," *Arabian Journal of Geosciences*, pp. 1–14, 2023, doi: 10.1007/s12517-023-11690-w.

[24] L. Deng, T. Yin, Z. Li, and Q. Ge, "Analysis of the Effectiveness of CNN-LSTM Models Incorporating Bert and Attention Mechanisms in Sentiment Analysis of Data Reviews," *ICBDIE*, pp. 821–829, 2023, doi: 10.2991/978-94-6463-238-5_106.