

Integrating ELECTRA and BERT models in transformer-based mental healthcare chatbot

Junta Zeniarja^{1,2}, Cinantya Paramita^{1,2}, Egia Rosi Subhiyanto^{1,2}, Sindhu Rakasiwi^{1,2},
Guruh Fajar Shidik^{1,2}, Pulung Nurtantio Andono^{1,2}, Anamarija Jurcev Savicevic³

¹Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

²Research Center for Intelligent Distributed Surveillance and Security (IDSS), Universitas Dian Nuswantoro, Semarang, Indonesia

³Department of Health Studies, Public Health Institute of Split-Dalmatia, School of Medicine, University of Split, Split, Croatia

Article Info

Article history:

Received May 2, 2024

Revised Sep 9, 2024

Accepted Sep 29, 2024

Keywords:

BERT

Chatbot

ELECTRA

Mental health

Transformer

ABSTRACT

Over the last decade, the surge in mental health disorders has necessitated innovative support methods, notably artificial intelligent (AI) chatbots. These chatbots provide prompt, tailored conversations, becoming crucial in mental health support. This article delves into the use of sophisticated models like convolutional neural network (CNN), long-short term memory (LSTM), efficiently learning an encoder that classifies token replacements accurately (ELECTRA), and bidirectional encoder representation of transformers (BERT) in developing effective mental health chatbots. Despite their importance for emotional assistance, these chatbots struggle with precise and relevant responses to complex mental health issues. BERT, while strong in contextual understanding, lacks in response generation. Conversely, ELECTRA shows promise in text creation but is not fully exploited in mental health contexts. The article investigates merging ELECTRA and BERT to improve chatbot efficiency in mental health situations. By leveraging an extensive mental health dialogue dataset, this integration substantially enhanced chatbot precision, surpassing 99% accuracy in mental health responses. This development is a significant stride in advancing AI chatbot interactions and their contribution to mental health support.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Cinantya Paramita

Faculty of Computer Science, Universitas Dian Nuswantoro

Semarang, Indonesia

Email: cinantya.paramita@dsn.dinus.ac.id

1. INTRODUCTION

In recent years, the critical importance of mental health care has been increasingly recognized, reflecting both its profound impact on individuals and its substantial cost to the global economy. Studies reveal that a significant number of individuals who succumbed to suicide had interacted with healthcare providers shortly before their death, underscoring the challenges in detecting and addressing suicidal ideation, particularly in those who view such discussions as taboo [1]. Globally, mental illnesses, including common conditions like anxiety and depression as well as more severe disorders, affect close to a billion people, degrading their quality of life and societal contribution. This results in an estimated economic burden of USD 2.5 trillion annually, projected to rise to USD 6 trillion by 2030 [2]. Despite the growing prevalence of psychological distress, a vast majority of those suffering do not receive adequate treatment due to limited access to mental health counseling. Health coaching has emerged as a promising method for supporting those with chronic conditions, but the scarcity of practitioners limits its reach. Meanwhile, the expanding volume

of clinical information in healthcare presents new opportunities for leveraging technology and computer science to improve mental healthcare accessibility and effectiveness.

Artificial intelligence (AI) represents a groundbreaking shift in healthcare, brought about by the convergence of expanding health data and advanced analytical techniques. This technology is pivotal in mirroring human cognitive processes and marks a significant transformation in healthcare dynamics. Furthermore, AI offers unique solutions to overcome social stigma, a notable barrier in accessing healthcare. Among its diverse applications, AI-driven health chatbots have been instrumental in promoting physical activity, healthy eating habits, and broadening patient access to essential health information. These chatbots, sophisticated computer programs designed for simulating interactive human conversation via the internet, are emerging as valuable tools in enhancing patient engagement and healthcare delivery.

Recent research in AI primarily focuses on developing models and algorithms, specifically artificial neural networks [3], deep neural network [4], convolutional neural network (CNN) [5], long short term memory (LSTM) [6] and other deep learning algorithms [7], [8]. AI refers to the capability of machines to perform tasks that typically require human intelligence, such as voice recognition, natural language understanding, and decision-making [9]. Rebelo *et al.* [10] have shown the effects of AI on mental health tasks. They involve several mental healthcare workers such as psychologists, therapists, and psychiatrists, including the tasks they carry out such as assessment, therapy, medical prescriptions, documentation, and monitoring. The ability to comprehend natural language is not limited to human communication but extends to communication between humans and machines, bridged by natural language processing (NLP). The technology employed by NLP has significantly advanced in recent years, enhancing both voice and text-based interactions to assist humans. Some researchers are dedicated to finding models with higher accuracy, capitalizing on breakthroughs in neural networks that bring novel ideas to the fields of AI and NLP, resulting in the emergence of various techniques and methods. Fu *et al.* [11] presented representative works and categorized existing dialogue models into three types, analysing trends in open-domain dialogue development and summarizing objectives in two aspects: informative and controlled. Mah *et al.* [12] utilized text content as part of NLP and AI to assess the connection of the internet of things (IoTs) with human needs.

Chatbots, also known as artificial conversation entities, can be developed using machine learning and pattern matching, with three types based on response generation methods: rule-based, retrieval-based, and generative [13]. Pandey and Sharma [14] compared retrieval-based and generative-based chatbots for mental health, finding that generative models with encoder-decoder architecture achieved the highest accuracy at 94.45%. Simarmata *et al.* [15] explored various transformer models for intent classification, noting bidirectional encoder representation of transformers (BERT) with LSTM's 94.47% accuracy. Yu *et al.* [16] developed a BERT-based chatbot for financial customer service capable of managing complex queries and performing automatic spelling correction. Fatima *et al.* [17] used a COVID-19 rumors dataset for veracity, stance, and sentiment analysis, achieving high accuracy with LSTM+CNN models. The study also highlights the fusion of efficiently learning an encoder that classifies token replacements accurately (ELECTRA) and LSTM models for question classification, combining ELECTRA's language understanding with LSTM's sequential learning abilities [18], [19]. Central to this research is a comprehensive dataset for training chatbots in mental health conversations, employing AI models like BERT, CNN, LSTM, and ELECTRA to enhance the chatbot's ability to understand and respond to human emotions and linguistic subtleties [20], [21].

A critical challenge for mental healthcare chatbots is achieving high accuracy in understanding and responding to user inputs. Despite advancements with models like ELECTRA and BERT, their integration into the Transformer framework for nuanced mental health dialogues remains suboptimal. This accuracy is vital for effective support but faces issues such as data privacy, diverse datasets, and culturally sensitive responses. Evaluated using metrics like accuracy, precision, recall, and F1 score, transformer models outperform others but still require ethical considerations and responsible regulation. Balcombe [22] highlights AI chatbots' decade-long role in improving mental health care, though they lack personal empathy. Human-AI collaboration can integrate human values into AI, impacting various sectors like depression, anxiety, and ADHD support. However, further research is needed for crisis support quality. Caldarini *et al.* [23] reviews the complexities of evaluating chatbots due to diverse conversational goals and domain-specific data, emphasizing the gap between industry practices and academic advancements and the need for a common evaluation framework and better datasets. Rahali and Akhloufi [24] explores Transformer-based models in NLP, noting their significant impact and suggesting improvements through intermediate layers. Khan *et al.* [25] research on a transformer-based model for politeness prediction in conversations highlights its superior performance but notes its limitation to English datasets, indicating a need for future research on multilingual or code-mixed data. The authors suggest that the proposed model is effective for politeness prediction in English texts but lacks important features. Further research will incorporate these features and explore its adaptation to multi-lingual and codemixed data. The model has not undergone evaluation for the detection of hate speech and offensive content.

Chatbots like ChatGPT are AI-powered programs designed to replicate human conversation and perform tasks such as answering questions and controlling smart home devices. A study by Sudheesh *et al.* [26] and a team from Universidad Internacional analyzed sentiments and topics in ChatGPT-based tweets using latent dirichlet allocation (LDA) for topic modeling and the BERT model for sentiment analysis, achieving 96.49% accuracy. The research highlights ChatGPT's ability to enhance user experience through natural language processing, with BERT efficiently categorizing emotions in tweets. Despite the majority of sentiments being positive, some criticisms exist. The chatbot industry is projected to reach \$3.62 billion by 2030, with potential educational benefits. The study emphasizes the importance of public perception, indicating future research will expand data collection from social media to understand sentiments better and develop machine learning approaches for responsible deployment of ChatGPT.

Alruqi and Alzahrani [27] research in "Evaluation of an Arabic chatbot based on extractive question-answering transfer learning and language transformers" demonstrates the AraElectra-SQuAD model's superior performance in improving the accuracy and relevance of Arabic chatbots. Using various Arabic QA datasets and transfer learning techniques, the study addresses the complexity of Arabic chatbots and highlights the need for better computational resources and preprocessing methods. The research emphasizes the importance of confidence metrics and suggests future work should focus on improving coherence, quality, and automated processes. Additionally, Fang and *et al.* [28] propose an effective ELECTRA-based pipeline model for sentiment analysis of tourist attraction reviews, showcasing high efficiency and advocating deep learning models for enhanced analysis. Evaluation metrics such as precision, recall, and F1-score highlight the model's success, suggesting its application to other domains. In "Combating fake news with transformers," Kasnesis *et al.* [29] discusses using language models to assess the subjectivity and stance of social media posts, with RoBERTa outperforming ELECTRA in stance detection. The study highlights the challenges of verifying social media information and suggests future research on tree-based structures, freezing model parameters, and graph neural networks, concluding that misinformation is often linked to subjective language.

This study investigated the integration of advanced NLP models like ELECTRA and BERT into mental healthcare chatbots. While earlier studies have explored the development of AI-driven chatbots for mental health and their potential in handling mental health dialogues, they have not explicitly addressed the challenges related to cultural sensitivity, ethical considerations, and the application of these models to multilingual or code-mixed datasets. The integration of advanced NLP models like ELECTRA and BERT into mental healthcare chatbots is a rapidly developing area of research that promises to enhance the effectiveness and efficiency of these systems. The literature reveals a clear trajectory from simple rule-based systems to complex AI-driven solutions, underscoring the potential of these models in handling the nuances of mental health dialogues. As these technologies evolve, it is imperative to balance their technical advancements with ethical considerations, ensuring they serve as a beneficial tool in mental healthcare. Future research should continue to explore this integration, focusing on improving the emotional intelligence and responsiveness of chatbots to better address the diverse needs of mental healthcare.

2. METHOD

This research followed a structured approach, as depicted in Figure 1, beginning with data collection and progressing through data splitting for training, testing, and validation phases. The modeling phase utilized five different models: BERT, CNN, LSTM, ELECTRA, and a hybrid of ELECTRA and BERT. Evaluation of these models was performed using metrics like accuracy, precision, recall, and F1 score, ultimately determining the classification accuracy. Figure 2, is explained in sections 2.1 through 2.4 of the discussion.

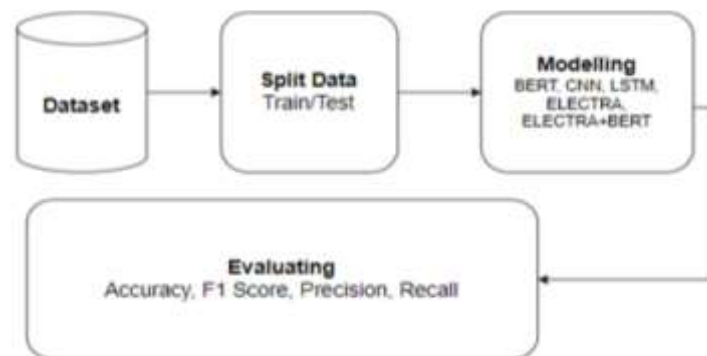


Figure 1. Shows the flowchart of the AI-based models and experimental methods applied

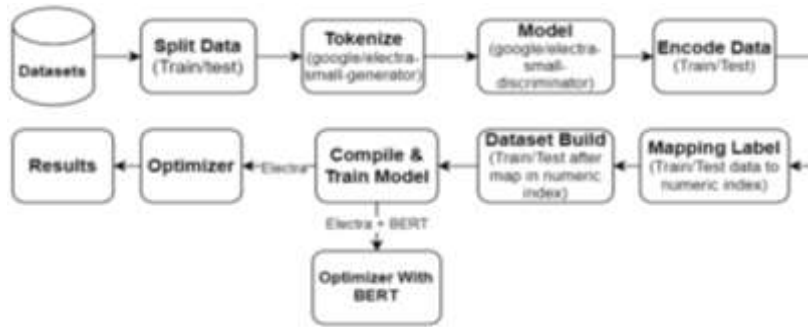


Figure 2. ELECTRA and BERT in transformer model

2.1. Experiment setup

To conduct our training, we leverage Google Colab Pro, equipped with a single NVIDIA P100/T4 GPU with a capacity of up to 15 GB and RAM reaching 8 GB. The pre-trained language models utilized encompass ELECTRA base model and BERT employing the Transformer algorithm, along with the CNN. Subsequently, we apply the ELECTRA base model with LSTM as the algorithm, ELECTRA+BERT optimization involves the use of the AdamW optimizer, setting the learning rate to $1r=5e-5$.

2.2. Dataset

This research utilizes secondary data, comprising two types of Intents.json datasets. The first dataset, Intents.json (Eng), is sourced publicly and includes attributes such as tag(), patterns(), and responses(). Conversely, the second dataset, Intents.json (Ind), possesses similar attributes but is acquired through private sources. These attributes play a crucial role in both datasets. The tag() function is designed to classify the intent or purpose of a question or statement. Patterns(), consisting of text examples or phrases, are crafted to represent the intent associated with a specific tag and are used as training inputs for the NLP model. This allows the model to recognize and accurately map similar inputs to the corresponding tag. On the other hand, response() is the output provided by the chatbot when a pattern is identified with a particular tag. This response is typically predefined by the chatbot's creator and is utilized to deliver an appropriate reply to the user.

2.3. Split data

Following this process, data splitting is executed, wherein the data is methodically divided into distinct sets, typically comprising training, testing, and validation sets. In the provided context, this procedure is specifically applied to the 'Intents.Eng' and 'Intents.Ind' datasets. Such segmentation of data is of paramount importance for the effective training of models, as it allows them to be fine-tuned on diverse data samples and ensures their robust performance on new, unseen data, thereby enhancing the overall efficacy of the modeling process.

2.4. Modeling

Modeling is a stage where deep learning models are implemented to achieve the research objective, namely the accuracy of the dataset for intents data. This experiment involves five predictive algorithms: BERT, CNN, LSTM, ELECTRA, and a combination of ELECTRA and BERT. These five models can handle cases of sentiment analysis, text summarization, or question understanding, where accuracy is measured by comparing the model's predictions with actual answers in the form of text data.

2.4.1. Bidirectional encoder representation from transformers

BERT employs a pre-training method called a masked language model (MLM), in which a specific fraction of input tokens is obscured. The objective of the model is to make predictions regarding these concealed tokens by considering their contextual information. In mathematical terms, when provided with an input sequence X , the loss L is computed as $-\log P(x_i | X_{-i}; \theta)$, with θ representing the model's parameters. Figure 3 shows the architecture of BERT.

2.4.2. Convolutional neural network

The development of a CNN model kicks off with meticulous data preprocessing, where tokenization and padding are crucial for standardizing input lengths. Using TensorFlow's Tftokenizer, the data is

tokenized, and numerical sequences are padded to ensure consistency. Labels are converted to numerical form and encoded into binary matrices via one-hot encoding using a label encoder. The dataset is then divided into 80% for training and 20% for testing, ensuring a balanced approach to model training and validation. The CNN architecture is composed of multiple layers: an embedding layer for vector transformation, convolutional and pooling layers for feature extraction, dense layers for feature connection, and an output layer to generate intent probabilities. The model is trained with the Adam optimizer, employing categorical crossentropy as the loss function and accuracy as the performance metric. This well-structured methodology ensures the CNN model processes and classifies data with high precision and efficiency.

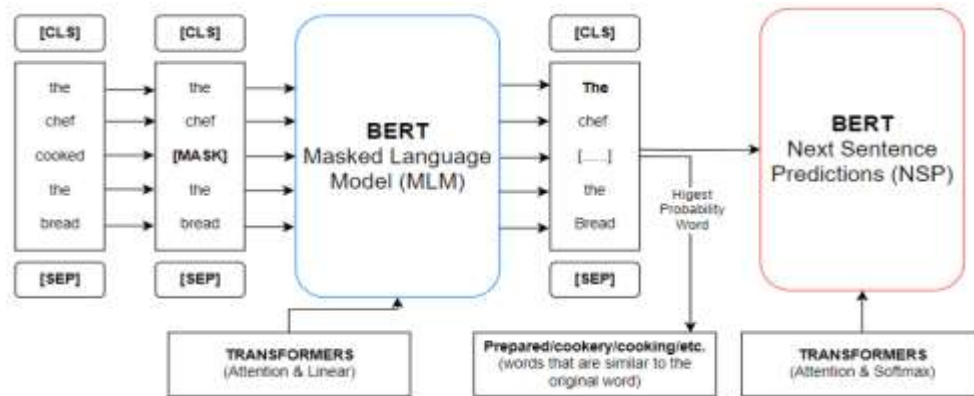


Figure 3. BERT model architecture

2.4.3. Long short-term memory

LSTM is a sophisticated type of artificial recurrent neural network (RNN) architecture used in deep learning, notable for its feedback connections that set it apart from traditional neural networks. An LSTM cell is composed of several key components: the Cell State (C_t), Input Gate (i_t), Forget Gate (f_t), and Output Gate (o_t) [30]. After the embedding process, LSTM networks extract sequential dependencies from the input text. Given their ability to process sequences, LSTMs are particularly adept at handling questions that typically start with "wh" words like "who," "what," or "where," and end with a subject or object. Recognizing this structure helps in gaining deeper insights into the questions' underlying patterns.

2.4.4. Efficiently learning an encoder that classifies token replacements accurately

ELECTRA employs a discriminative training method where the model learns to distinguish between "real" and "fake" tokens within a sentence. Given an input $X=[x_1, x_2, \dots, x_n]$, a generator G proposes substitutes x_i for masked tokens, while a discriminator D evaluates the probability $P(D(x_i) = 1 | X)$ that each token is genuine. The objective is to minimize $-\log(D(x_i))$ for real tokens and $-\log(1 - D(x_i))$ for fake tokens, refining the model's ability to identify authentic versus inauthentic elements in the text. Figure 4 shows the architecture of ELECTRA.

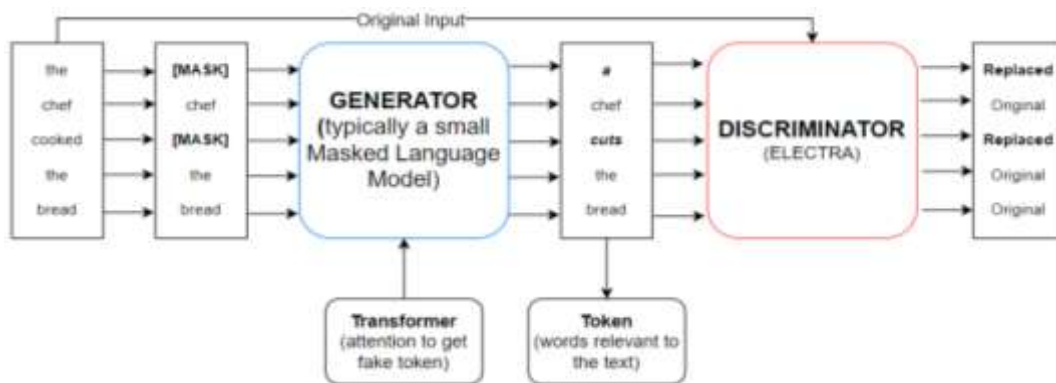


Figure 4. ELECTRA model architecture

2.4.5. Combination of ELECTRA and BERT

In the realm of natural language processing, transformer models like ELECTRA and BERT are pivotal for analyzing intent datasets composed of tags, patterns, and responses. The process starts by organizing labels and sentence patterns into lists, with tags acting as labels and patterns storing sentence structures. Using Sklearn, the data is divided into 80% for training and 20% for testing. ELECTRA's tokenizer, particularly google/electra-small-generator, creates text resembling the original, which is then evaluated by Google/ELECTRA-small-discriminator, leveraging transformers for efficient and accurate modeling. The generator masks and alters words to produce contextually relevant sentences, which are encoded to aid the model's understanding, with adjustments for sentence length and padding. The ELECTRA discriminator trains on this data, assessing if sentences are generated or original, updating the mask generator accordingly. During training, the AdamW optimizer fine-tunes the model using BERT principles. BERT employs MASK and next sentence predictions (NSP) to understand inter-sentence relationships, focusing on word relevance and meaning to decide if sentences need regeneration based on the discriminator's judgment. This sophisticated method enhances the model's natural language understanding and response capabilities.

2.5. Evaluation

After splitting and fine-tuning the model, the evaluation of its performance will be conducted through the analysis of output results. This assessment will involve the utilization of a Confusion Matrix and various metrics such as accuracy, precision, recall, and F1-score, aiming to comprehensively gauge the model's effectiveness across all tasks.

Accuracy: calculates the proportion of accurately predicted samples out of the entire sample set.

$$Accuracy = ((TP + TN))/((TP + TN + FP + FN)) \quad (1)$$

Precision: represents the portion of positive identifications that were accurately made.

$$Precision = TP/((TP + FP)) \quad (2)$$

Recall: demonstrates the proportion of true positives that were correctly recognized.

$$Recall = TP/((TP + FN)) \quad (3)$$

F1 score: this balanced metric functions as the harmonic mean between precision and recall.

$$F1\ Score = 2 \times (Precision \times Recall)/((Precision + Recall)) \quad (4)$$

Where TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative.

3. RESULTS AND DISCUSSION

The Table 1 presents the performance of various deep learning models for intent classification on English and Indonesian datasets. The metrics include accuracy (Acc), F1-score (F1), precision (Pre), and recall (Rec). The models evaluated are BERT, CNN, LSTM, ELECTRA, and a combination of ELECTRA and BERT.

Table 1. Results of accuracy performance of deep learning models for intent English, and intent Indonesian datasets

Models	Dataset	Acc (%)	F1 (%)	Pre (%)	Rec (%)
BERT	Intent English	76.92	77.15	84.48	76.92
	Intent Indonesian	96.79	96.55	97.38	96.79
CNN	Intent English	71.79	69.87	79.57	71.79
	Intent Indonesian	89.94	89.66	93.15	89.94
LSTM	Intent English	64.10	60.91	73.68	64.10
	Intent Indonesian	90.55	90.17	93.07	90.55
ELECTRA	Intent English	98.92	53.61	55.31	57.44
	Intent Indonesian	99.66	48.23	49.12	50.81
ELECTRA+BERT	Intent English	99.46	43.90	42.19	48.93
	Intent Indonesian	99.66	48.74	49.66	51.63

The performance results of various deep learning models for intent classification show that BERT demonstrates strong and balanced performance, achieving 76.92% accuracy for English intents and 96.79% for Indonesian intents, with high F1-scores, precision, and recall, particularly excelling in the Indonesian dataset with an F1-score of 96.55%, precision of 97.38%, and recall of 96.79%. CNN, while performing reasonably well, falls short of BERT's effectiveness, achieving 71.79% accuracy for English intents and 89.94% for Indonesian intents, with slightly lower F1-scores, precision, and recall. LSTM has the lowest performance for English intents at 64.10% accuracy but improves significantly for Indonesian intents with 90.55% accuracy and balanced metrics. ELECTRA shows very high accuracy for both English (98.92%) and Indonesian (99.66%) intents, but with notably lower F1-scores, precision, and recall, particularly for English intents (F1-score of 53.61%, precision of 55.31%, and recall of 57.44%), indicating potential issues like class imbalance or misclassification. The combination of ELECTRA and BERT achieves the highest accuracy for both English (99.46%) and Indonesian (99.66%) intents, but with low F1-scores, precision, and recall, suggesting possible overfitting or other issues. This underscores the need to consider all performance metrics, not just accuracy, when evaluating model performance.

Based on the Table 2, the performance of deep learning models for intent classification shows significant improvement across epochs for both English and Indonesian datasets. At 50 epochs, the model exhibits low accuracy and F1-scores, with English intents at 34.59% accuracy and 10.02% F1-score, and Indonesian intents at 21.38% accuracy and 4.35% F1-score. These results suggest that the model is undertrained and unable to effectively classify intents at this stage.

Table 2. Results of accuracy performance of deep learning models for intent English, and intent Indonesian datasets with different Epoch

Epoch	Dataset	Acc (%)	F1 (%)	Pre (%)	Rec (%)
50	Intent English	34.59	10.02	9.34	12.76
	Intent Indonesian	21.38	4.35	3.82	8.69
100	Intent English	49.19	13.09	11.87	17.02
	Intent Indonesian	85.43	30.32	29.37	35.05
150	Intent English	81.62	34.62	37.58	36.17
	Intent Indonesian	98.03	43.29	44.33	45.38
200	Intent English	93.51	38.15	40.14	42.55
	Intent Indonesian	99.18	45.18	45.24	48.09
250	Intent English	97.84	44.32	43.97	48.93
	Intent Indonesian	99.52	49.17	50.73	51.35
300	Intent English	99.46	43.90	42.19	48.93
	Intent Indonesian	99.66	49.94	50.02	52.71

By 100 epochs, the model shows noticeable improvement, with the Indonesian dataset achieving 85.43% accuracy and a 30.32% F1-score, and the English dataset achieving 49.19% accuracy and 13.09% F1-score. At 150 epochs, performance further increases, with the English dataset reaching 81.62% accuracy and a 34.62% F1-score, and the Indonesian dataset achieving 98.03% accuracy and a 43.29% F1-score. This trend continues at 200 epochs, with the English dataset at 93.51% accuracy and a 38.15% F1-score, and the Indonesian dataset at 99.18% accuracy and a 45.18% F1-score. By 250 epochs, the model reaches high performance, with the English dataset at 97.84% accuracy and a 44.32% F1-score, and the Indonesian dataset at 99.52% accuracy and a 49.17% F1-score. At 300 epochs, the highest performance is observed, with the English dataset achieving 99.46% accuracy and a 43.90% F1-score, and the Indonesian dataset achieving 99.66% accuracy and a 49.94% F1-score. These results indicate that increasing the number of training epochs significantly enhances the model's performance for intent classification, with the most substantial gains observed up to 250 epochs, beyond which accuracy improvements are marginal.

The results validate the initial hypothesis that increasing the number of training epochs enhances model performance significantly. At 50 epochs, accuracy was 34.59% for English and 21.38% for Indonesian intents, but by 300 epochs, accuracy improved to 99.46% for English and 99.66% for Indonesian. This confirms substantial gains up to 250 epochs. The hypothesis that transformer-based models like BERT and ELECTRA would outperform others was partly confirmed; ELECTRA achieved high accuracy (98.92% for English and 99.66% for Indonesian) but had lower F1-scores, precision, and recall, indicating possible balance issues. BERT showed strong performance, especially with Indonesian intents (96.79% accuracy and 96.55% F1-score). The combination of ELECTRA and BERT achieved the highest accuracy but with lower F1-scores, suggesting potential overfitting. The findings also emphasize that relying solely on accuracy can be misleading, and multiple metrics are crucial for a balanced evaluation.

4. CONCLUSION

The results of this research highlight significant advancements in the health sector and related scientific communities by demonstrating that chatbots using the ELECTRA and BERT transformer methods can provide highly accurate, effective, and empathetic support for mental health issues. With an impressive accuracy of 99.66% across English and Indonesian datasets, these chatbots offer valuable 24/7 support and timely responses, enhancing accessibility to mental health resources and potentially reducing the burden on human professionals. This advancement is particularly beneficial in emergency situations, where immediate assistance is critical. Moreover, the study underscores the potential of integrating cutting-edge AI technologies into mental health care, contributing to a deeper understanding of their practical applications and fostering further research in the field. The findings suggest that these AI models can complement traditional mental health services by providing preliminary support and directing users to professional help when needed, thereby enriching the overall mental health care landscape.

This study demonstrated the effectiveness of using ELECTRA and BERT transformer methods in mental health chatbots, achieving high accuracy and providing valuable support in English and Indonesian. However, further research is needed to address limitations such as the models' performance across other languages and culturally diverse contexts. Additionally, ethical concerns, including data privacy and the absence of human empathy in AI solutions, were not discussed. Future research could explore the adaptation of these models to additional languages and culturally diverse contexts to broaden their applicability. Additionally, investigating ways to enhance the ethical aspects of AI-driven solutions, such as improving data privacy and integrating empathetic responses, could address current limitations.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to the Research and Community Service Institute (LPPM) at Universitas Dian Nuswantoro for their invaluable support. This research was funded by the Basic Research Grant Scheme under contract no. 109/A.38-04/UDN-09/XI/2023.





REFERENCES

- [1] A. Li, X. Huang, D. Jiao, B. O'Dea, T. Zhu, and H. Christensen, "An analysis of stigma and suicide literacy in responses to suicides broadcast on social media," *Asia-Pacific Psychiatry*, vol. 10, no. 1, Mar. 2018, doi: 10.1111/appy.12314.
- [2] "Mental health action plan 2013-2020," *World Health Organization*. [Online]. Available: <http://www.who.int/iris/handle/10665/89966>.
- [3] J. M. Górriz *et al.*, "Computational approaches to explainable artificial intelligence: advances in theory, applications and trends," *Information Fusion*, vol. 100, p. 101945, Dec. 2023, doi: 10.1016/j.inffus.2023.101945.
- [4] M. K. Rusia and D. K. Singh, "A color-texture-based deep neural network technique to detect face spoofing attacks," *Cybernetics and Information Technologies*, vol. 22, no. 3, pp. 127–145, Sep. 2022, doi: 10.2478/cait-2022-0032.
- [5] B. P. Babu and S. J. Narayanan, "One-vs-all convolutional neural networks for synthetic aperture radar target recognition," *Cybernetics and Information Technologies*, vol. 22, no. 3, pp. 179–197, Sep. 2022, doi: 10.2478/cait-2022-0035.
- [6] V. Barot and V. Kapadia, "Long short term memory neural network-based model construction and fine-tuning for air quality parameters prediction," *Cybernetics and Information Technologies*, vol. 22, no. 1, pp. 171–189, Mar. 2022, doi: 10.2478/cait-2022-0011.
- [7] L. Kumar and D. K. Singh, "Hardware response and performance analysis of multicore computing systems for deep learning algorithms," *Cybernetics and Information Technologies*, vol. 22, no. 3, pp. 68–81, Sep. 2022, doi: 10.2478/cait-2022-0028.
- [8] S. Vandhana and J. Anuradha, "Spatial and temporal variations on air quality prediction using deep learning techniques," *Cybernetics and Information Technologies*, vol. 23, no. 4, pp. 213–232, Nov. 2023, doi: 10.2478/cait-2023-0045.
- [9] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cognitive Robotics*, vol. 3, pp. 54–70, 2023, doi: 10.1016/j.cogr.2023.04.001.
- [10] A. D. Rebelo, D. E. Verboom, N. R. dos Santos, and J. W. de Graaf, "The impact of artificial intelligence on the tasks of mental healthcare workers: A scoping review," *Computers in Human Behavior: Artificial Humans*, vol. 1, no. 2, p. 100008, Aug. 2023, doi: 10.1016/j.chbah.2023.100008.
- [11] T. Fu, S. Gao, X. Zhao, J. Wen, and R. Yan, "Learning towards conversational AI: a survey," *AI Open*, vol. 3, pp. 14–28, 2022, doi: 10.1016/j.aiopen.2022.02.001.
- [12] P. M. Mah, I. Skalna, and J. Muzam, "Natural language processing and artificial intelligence for enterprise management in the era of Industry 4.0," *Applied Sciences*, vol. 12, no. 18, p. 9207, Sep. 2022, doi: 10.3390/app12189207.
- [13] E. Adamopoulou and L. Moussiades, "Chatbots: history, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.mlwa.2020.100006.
- [14] S. Pandey and S. Sharma, "A comparative study of retrieval-based and generative-based chatbots using deep learning and machine learning," *Healthcare Analytics*, vol. 3, p. 100198, Nov. 2023, doi: 10.1016/j.health.2023.100198.
- [15] R. Simarmata, J. Kristanto, and A. Chowanda, "Utilizing transformer-based deep learning for intent classification on text," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 13, pp. 5078–5084, 2023.
- [16] S. Yu, Y. Chen, and H. Zaidi, "AVA: a financial service chatbot based on deep bidirectional transformers," *Frontiers in Applied Mathematics and Statistics*, vol. 7, Aug. 2021, doi: 10.3389/fams.2021.604842.
- [17] R. Fatima *et al.*, "A natural language processing (NLP) evaluation on COVID-19 rumour dataset using deep learning techniques," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–17, Sep. 2022, doi: 10.1155/2022/6561622.





- [18] K. Clark, M. T. Luong, Q. V. Le, and C. D. Manning, "Electra: pre-training text encoders as discriminators rather than generators," *8th International Conference on Learning Representations, ICLR 2020*, 2020.
- [19] S. Aburass, O. Dorgham, and M. A. Rumman, "An ensemble approach to question classification: integrating electra transformer, GloVe, and LSTM," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 507–514, 2024, doi: 10.14569/IJACSA.2024.0150148.
- [20] S. Sarkar, M. Gaur, L. K. Chen, M. Garg, and B. Srivastava, "A review of the explainability and safety of conversational agents for mental health to identify avenues for improvement," *Frontiers in Artificial Intelligence*, vol. 6, Oct. 2023, doi: 10.3389/frai.2023.1229805.
- [21] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [22] L. Balcombe, "AI chatbots in digital mental health," *Informatics*, vol. 10, no. 4, p. 82, Oct. 2023, doi: 10.3390/informatics10040082.
- [23] G. Caldarini, S. Jaf, and K. McGarry, "A literature survey of recent advances in chatbots," *Information*, vol. 13, no. 1, p. 41, Jan. 2022, doi: 10.3390/info13010041.
- [24] A. Rahali and M. A. Akhloufi, "End-to-end transformer-based models in textual-based NLP," *AI*, vol. 4, no. 1, pp. 54–110, Jan. 2023, doi: 10.3390/ai4010004.
- [25] S. Khan *et al.*, "Transformer architecture-based transfer learning for politeness prediction in conversation," *Sustainability*, vol. 15, no. 14, p. 10828, Jul. 2023, doi: 10.3390/su151410828.
- [26] R. Sudheesh *et al.*, "Analyzing sentiments regarding ChatGPT using novel BERT: a machine learning approach," *Information*, vol. 14, no. 9, p. 474, Aug. 2023, doi: 10.3390/info14090474.
- [27] T. N. Alruqi and S. M. Alzahrani, "Evaluation of an Arabic chatbot based on extractive question-answering transfer learning and language transformers," *AI*, vol. 4, no. 3, pp. 667–691, Aug. 2023, doi: 10.3390/ai4030035.
- [28] H. Fang, G. Xu, Y. Long, and W. Tang, "An effective ELECTRA-based pipeline for sentiment analysis of tourist attraction reviews," *Applied Sciences (Switzerland)*, vol. 12, no. 21, p. 10881, Oct. 2022, doi: 10.3390/app122110881.
- [29] P. Kasnesis, L. Toumanidis, and C. Z. Patrikakis, "Combating fake news with transformers: a comparative analysis of stance detection and subjectivity analysis," *Information (Switzerland)*, vol. 12, no. 10, p. 409, Oct. 2021, doi: 10.3390/info12100409.
- [30] H. R. Sayegh, W. Dong, and A. M. Al-madani, "Enhanced intrusion detection with LSTM-based model, feature selection, and SMOTE for imbalanced data," *Applied Sciences (Switzerland)*, vol. 14, no. 2, p. 479, Jan. 2024, doi: 10.3390/app14020479.

BIOGRAPHIES OF AUTHORS







Junta Zeniarja     is a permanent lecturer at the Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia. He received two master's degrees in Informatics Engineering from Universiti Teknikal Malaysia Melaka (UTeM) in Malaysia and Universitas Dian Nuswantoro in Indonesia. The main research interests are data mining, machine learning, deep learning, information retrieval, and geographic information systems. His current research covers a wide range of Data Mining and Data Science applications, such as predicting student graduation, search engines for kids, thesis document classification, sentiment analysis, and geographic information systems based on information retrieval (Geographic Information Retrieval). He can be contacted at email: junta@dsn.dinus.ac.id.







Cinantya Paramita     is a permanent lecturer at the Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia. She received the bachelor's degree in computer science from Universitas Dian Nuswantoro, Semarang, Indonesia, in 2011, the master's degree in Electrical and Computer Engineering from South China University of Technology, Guangzhou China, in 2014. The main research interests are IT essentials, cybersecurity, data mining, machine learning, and deep learning. She contributing as a trainer for various academies under Ministry of Communications and Informatics, covering IT essentials, cybersecurity, and data analyst. She can be contacted at email: cinantya.paramita@dsn.dinus.ac.id.







Egia Rosi Subhiyakto     completed his Bachelor's degree in Computer Science and Engineering in 2011 at the Universitas Komputer Indonesia, Bandung, West Java. He pursued her Master's degree in Computer Science at Dian Nuswantoro University and Universiti Teknikal Malaysia Melaka (UTeM) in 2014. Since 2014, he has been working as a lecturer at Dian Nuswantoro University, Semarang, in the Computer Science program. Her Google Scholar H-Index is 12, i10-Index is 15, and Scopus H-Index is 3. He can be contacted at email: egia@dsn.dinus.ac.id.







Sindhu Rakasiwi     completed her Bachelor's degree in Computer Science and Engineering in 2010 at Universitas Dian Nuswantoro, Semarang, Central Java. She pursued her Master's degree in Computer Science at Universitas Dian Nuswantoro, in 2012. From 2013 to 2023, she worked as a lecturer at the Universitas Sains dan Teknologi Komputer, Semarang. Since 2023, she has been serving as a lecturer at Dian Nuswantoro University, Semarang, in the Computer Science program. Her Google Scholar H-Index is 5, and her i10-Index is 4. She can be contacted at email: sindhu.rakasiwi@dsn.dinus.ac.id.







Guruh Fajar Shidik     (Member, IEEE) was born in West Borneo, Indonesia, in February 1987. He received the bachelor's degree in computer science from Universitas Dian Nuswantoro, Semarang, Indonesia, in 2009, the master's degree in computer science from Technical Malaysia Melaka University, Malaysia, in 2011, and the Ph.D. degree from Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2016. He is currently an Associate Professor with Universitas Dian Nuswantoro. His research interests include cloud computing, wireless communication, machine learning, and artificial intelligence. He can be contacted at email: guruh.fajar@research.dinus.ac.id.



Pulung Nurtantio Andono     received the bachelor's degree from Trisakti University, Jakarta, Indonesia, in 2006, the master's degree from Dian Nuswantoro University, Semarang, Indonesia, in 2009, and the Ph.D. degree from the Institut Teknologi Sepuluh November (ITS), Surabaya, Indonesia, in 2016. He currently works as a Lecturer and a Researcher at the Faculty of Computer Science, Dian Nuswantoro University. His research interests include image security, computer vision, and 3D image reconstruction. He has authored or coauthored more than 43 refereed journals and conference papers indexed by Scopus. He can be contacted at email: pulung@dsn.dinus.ac.id.



Anamarija Jurcev Savicevic     is medical doctor, specialist of epidemiology, primarius. She has been working at Teaching Public Health Institute of Split and Dalmatia County as a head of Scientific Unit and head of Department for Epidemiology of Respiratory Infections. As a field epidemiologist, she has been deeply involved in different outbreak control activities. At university level, she is experienced teacher, researcher, mentor and reviewer, employed at School of Medicine University of Split and Department of Health Studies University of Split where she runs Department for Preventive Medicine. The main research interests are disease prevention and control as well as health promotion, including lifestyle medicine. Author of numerous papers, editor of university textbooks and authors of several books and publications for health education. Recently, she has been interested in advantages of artificial intelligence in the health service. Full member of Croatian Academy of Medical Sciences. She can be contacted at email: anamarija.jurcev.savicevic@nzjz-split.hr.