OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
journals.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 10, Issue. 1, August 2025. pp. 54-72
doi:10.21108/indojc.v10i1.9707

Visualizing Functional Emotions: Mapping Counseling Responses from Text to Virtual Facial Expressions

Rifki Padilah[1], Rifki Wijaya[2], Shaufiah[3]

School of Computing, Telkom University
Telecommunication Street No. 1, Bandung 40257, Indonesia

[1] rifkipdh@student.telkomuniversity.ac.id
[2] rifkiwijaya@telkomuniversity.ac.id
[3] shaufiah@telkomuniversity.ac.id

Abstract

This research develops an innovative virtual counseling system by integrating text-based emotion classification with visual representation to address the problem of early marriage in Lombok. The system leverages the sophisticated IndoRoBERTa model to accurately classify conselor responses into five functional emotion categories relevant to the counseling context: Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary. The limitations of conventional counseling services in rural areas serve as the primary justification for developing this responsive and accessible technological solution. Evaluation results demonstrate that the IndoRoBERTa model achieves a highly competitive accuracy rate of 89% after being trained on an expanded dataset, an achievement that significantly surpasses previous architectures. In conclusion, this IndoRoBERTa-based system is not only technically viable but also effective as a tool for providing initial empathetic support. Its capability to translate textual emotions into non-verbal visual cues makes it a promising technological solution to bridge the gap in current counseling services.

Keywords: IndoRoBERTa, Emotion Classification, Virtual Counseling, Early Marriage, Facial Expression Mapping, Natural Language Processing (NLP)

Abstrak

Penelitian ini mengembangkan sebuah sistem konseling virtual inovatif dengan mengintegrasikan klasifikasi emosi berbasis teks dan representasi visual untuk mengatasi masalah pernikahan dini di Lombok. Sistem ini memanfaatkan kecanggihan model IndoRoBERTa untuk secara akurat mengklasifikasikan respons pengguna ke dalam lima kategori emosi fungsional yang relevan dalam konteks konseling: Antusiasme, Kelembutan, Analitis, Inspirasional, dan Peringatan. Keterbatasan layanan konseling konvensional di daerah pedesaan menjadi justifikasi utama pengembangan solusi teknologi yang responsif dan mudah diakses ini. Hasil evaluasi menunjukkan model IndoRoBERTa mencapai tingkat akurasi yang sangat kompetitif sebesar 89% setelah dilatih pada dataset yang diperluas , sebuah pencapaian yang secara signifikan melampaui arsitektur sebelumnya. Kesimpulannya, sistem berbasis IndoRoBERTa ini tidak hanya layak secara teknis tetapi juga efektif sebagai alat untuk menyediakan dukungan awal yang empatik. Kemampuannya menerjemahkan emosi tekstual menjadi isyarat visual non-verbal menjadikannya solusi teknologi yang menjanjikan untuk menjembatani kesenjangan dalam layanan konseling saat ini.

Kata Kunci: IndoRoBERTa, Klasifikasi Emosi, Konseling Virtual, Pernikahan Dini, Pemetaan Ekspresi Wajah, Pemrosesan Bahasa Alami

Received on January, 2025. Accepted on August, 2025

I. Introduction

EARLY marriage remains a pressing socio-cultural challenge in Indonesia, particularly in the region of Lombok, which recorded a prevalence rate of 16.59% in 2021, marking it as the nation's second-highest area for this practice [1]. This custom is profoundly embedded within the cultural and traditional norms of the local Sasak community, where it is sometimes endorsed by influential leaders as a perceived safeguard against premarital sexual relations [1]. This normalization often overlooks the severe, long-term consequences for young women, particularly concerning their reproductive health, educational attainment, and overall socio-economic well-being.

As a critical intervention, counseling services are designed to equip adolescents with a thorough understanding of the adverse effects of early marriage, thereby empowering them to make more informed life choices [2], the effectiveness of conventional counseling is significantly hampered, particularly in rural locales, by obstacles including poor infrastructure, prohibitive costs, and inconsistent outreach [3]. These barriers result in a significant portion of youth being deprived of essential guidance and education.

To address these limitations, this paper proposes the use of AI-driven virtual agents, a solution increasingly validated for a wide range of mental health applications through Natural Language Processing (NLP) [4]. A key focus within this domain is the development of systems capable of delivering empathetic support, a capability that has also been systematically examined in recent literature [5].These systems offer distinct advantages, such as 24/7 accessibility and user anonymity, which can foster more open communication [6], [7]. A key innovation explored in this research is the integration of facial expression mapping [8], a technique that translates detected text-based emotions into visual cues, aiming to create a more anthropomorphic and engaging user interaction. However, the effectiveness of such an engaging system is critically dependent on its core ability to accurately classify the user's emotions from text, a task where previous models have shown certain limitations.

In the field of emotion classification for Indonesian text, prior models such as Convolutional Neural Networks (CNN) and K-Nearest Neighbors (KNN) have demonstrated certain limitations. For instance, CNN-based approaches have achieved accuracies around 71.6% but often require complex hyperparameter tuning [9], while KNN has struggled with multi-label contexts [10]. Consequently, models based on the Transformer architecture have become the state-of-the-art. Foundational advancements such as the Robustly Optimized BERT Pretraining Approach (RoBERTa) [11] have introduced more effective training strategies. For the Indonesian language, specific versions like IndoRoBERTa have been developed and optimized, showing strong performance in multi-class sentiment and emotion classification tasks on social media data [12]. Despite this, the potential of this optimized architecture remains underexplored for the specialized task of classifying functional emotions in counseling dialogues for subsequent facial expression mapping.

Therefore, based on the existing research gap, this study specifically aims to evaluate the effectiveness of the IndoRoBERTa model in classifying five functional emotions Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary within the context of early marriage counseling dialogues. The evaluation will also compare the model's performance when trained on an initial balanced dataset versus an expanded, imbalanced dataset to test its robustness. Furthermore, this research aims to develop a conceptual framework for mapping the classified emotion labels onto contextually appropriate virtual facial expressions. The primary hypothesis posited is that the fine-tuned IndoRoBERTa model can not only achieve high and reliable accuracy but also that the proposed visual mapping framework can serve as a valid foundation for the development of more empathetic virtual agents.

To achieve these objectives and test the hypothesis, this research implements and evaluates the IndoRoBERTa model. The output from this classification subsequently serves as the basis for mapping to virtual facial expressions. The primary contribution of this research is demonstrating the efficacy of the IndoRoBERTa architecture in this specialized domain and establishing an initial framework for more advanced, multimodal counseling systems in Indonesia.

II. Literature Review

Emotion classification from text is a Natural Language Processing (NLP) task aimed at identifying and categorizing specific emotions contained within a piece of writing. This task is more complex than standard sentiment analysis, which only categorizes text into positive, negative, or neutral polarities. Emotion classification seeks to recognize a broader and more nuanced spectrum of feelings, such as joy, sadness, anger, or, in the context of this study, functional emotions like enthusiasm and gentleness. In mental health applications, the ability to accurately recognize emotions from user text is crucial. A system capable of understanding a user's emotional state can provide more empathetic and relevant responses, thereby increasing user satisfaction and trust in virtual counseling agents [6], [7].

In the context of mental health applications, the ability to accurately recognize emotions from user text is paramount. The functional emotion taxonomy employed in this study Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary was selected to directly represent verbal communication techniques proven effective in building the therapeutic alliance [13]. The therapeutic alliance, defined as the collaborative bond between a counselor and a client, is consistently identified as a robust predictor of successful psychotherapy outcomes across various modalities.

The selection of these categories is substantiated by recent empirical evidence. For instance, the Gentleness, Enthusiasm, and Inspirational categories in our research mirror findings that supportive statements, such as expressions of encouragement and validation, are significantly correlated with client symptom improvement [14]. Meanwhile, the Analytical category aligns with the importance of exploratory statements, such as the use of open-ended questions, which are crucial for fostering collaboration and trust, particularly in the initial counseling sessions [15]. Finally, the Cautionary category is designed to convey potential risks constructively and supportively, thereby avoiding the challenging or controlling communication styles that are correlated with poorer therapeutic outcomes [14].

A. Previous Approaches in Indonesian Emotion Classification

Initial research in emotion classification for Indonesian text has explored various methods, ranging from conventional machine learning approaches to early deep learning models. Classic approaches such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes have been widely applied. For instance, studies have utilized SVM and Naïve Bayes for sentiment analysis on Indonesian social media data related to political events [16]. While foundational, these methods are highly dependent on manual feature engineering and often struggle to capture the complex contextual meaning of a sentence, as also noted in studies using KNN [10].

As the field progressed, deep learning models such as Convolutional Neural Networks (CNN) were applied. CNNs offer the advantage of automatic feature extraction from text, but their architecture, designed for grid-like data (such as images), has limitations in understanding long-range word dependencies and sequential order in sentences [9]. These limitations have driven researchers to shift towards more advanced architectures capable of handling long-term dependencies in text data.

B. Advancements in Transformer Architecture for Language Understanding

The emergence of the Transformer architecture has revolutionized the field of NLP. With its self-attention mechanism, a Transformer-based model can weigh the importance of each word in a sentence and understand complex contextual relationships, even between distant words. One of the most significant developments from this architecture is RoBERTa (A Robustly Optimized BERT Pretraining Approach), which is an optimization

of the BERT model [11]. RoBERTa introduces more robust pre-training strategies, such as the use of dynamic masking and a much larger training dataset, which theoretically allows it to have a deeper understanding of language.

To handle the uniqueness and linguistic richness of the Indonesian language, a model specifically trained on a local corpus is required. IndoRoBERTa is an implementation of the RoBERTa architecture that has been pre-trained on a large volume of Indonesian data. Previous research has shown that IndoRoBERTa has a very strong performance, often outperforming other models in multi-class sentiment and emotion classification tasks on Indonesian social media data [12]. This demonstrates the great potential of IndoRoBERTa for NLP tasks that require a nuanced understanding of the language.

Providing tangible proof of this "great potential," research by Suwarningsih et al. [17] presents a concrete implementation. This study takes the proven capabilities of IndoRoBERTa in understanding the Indonesian language and applies them to build a Question-Answering (QA) system. Thus, their study serves as a valid case study, demonstrating that the advantages of a RoBERTa architecture trained on a local corpus are not merely theoretical but have been successfully leveraged to develop advanced and functional NLP applications for the Indonesian language.

C. Synthesis and Identification of the Research Gap

The literature review confirms that Transformer-based architectures represent a significant advancement over previous methods for text understanding tasks. In the context of the Indonesian language, initial adaptations of this architecture, such as IndoBERT, have already proven effective for emotion classification in general domains like social media [18]. This success validates the feasibility of using pre-trained language models for Indonesian NLP.

Building upon these foundational models, more robustly optimized architectures like RoBERTa which generally yield superior performance due to enhanced pre-training strategies have been developed, leading to the creation of the IndoRoBERTa model [19]. Despite these advancements, recent systematic reviews on NLP applications for mental health interventions indicate a persistent focus on detecting broad clinical conditions, rather than exploring more granular, functional, and clinically relevant emotion taxonomies [20].

Consequently, the application and evaluation of the more advanced IndoRoBERTa model within a highly specific and nuanced domain, such as the classification of functional emotions in counseling dialogues, remains a largely underexplored area. The counseling domain demands a far deeper and more sensitive contextual understanding than standard sentiment analysis. This research, therefore, aims to address this specific gap by implementing and thoroughly analyzing the performance of the IndoRoBERTa model. The objective is to provide new insights into its capabilities for developing more complex and interactive mental health applications.

D. Mapping Emotions to Virtual Facial Expressions

To create empathetic virtual agents, the ability to non-verbally communicate emotional understanding through facial expressions is crucial. Research in this field often utilizes standard frameworks such as the Facial Action Coding System (FACS) to objectively define expressions based on muscular Action Units (AUs). Building on this system, modern research has developed pipelines that effectively translate discrete emotion labels (e.g., 'happy,' 'angry') into specific AU activations on 3D avatars [21]. Furthermore, validation of this basic emotion mapping has shown that prioritized AU combinations can enhance social presence in avatar-mediated communication [22], laying the foundation for the development of more expressive virtual characters.

Nevertheless, a significant challenge and research gap emerges when addressing more complex and nuanced emotions, such as the functional emotions that are the focus of this study: Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary. Unlike basic emotions, this taxonomy lacks a well-established, one-to-one

mapping to facial expressions. The ambiguity in visualizing an "Analytical" expression, which must be distinguished from "confused," or a "Gentle" expression, which must differ from "Enthusiasm," highlights the absence of a standard for the counseling context. Therefore, this research not only aims to accurately classify these functional emotions using IndoRoBERTa but also to pioneer an initial mapping framework that can serve as a basis for developing more emotionally sensitive virtual counseling agents.

III. Research Method

A. Research Design

This study adopts a descriptive quantitative approach using an experimental method. The primary focus is to implement the IndoRoBERTa model for classifying functional emotions expressed in early marriage counseling conversations and to map those classified emotions into virtual facial expressions. Through this approach, the research aims to build an AI-based counseling system capable of responding to users in a more empathetic and natural manner by integrating text analysis with simultaneous visual emotion representation.these functional emotions using IndoRoBERTa but also to pioneer an initial mapping framework that can serve as a basis for developing more emotionally sensitive virtual counseling agents.

B. System Architecture Overview

The proposed system architecture consists of five main components: data collection, data preprocessing, IndoRoBERTa model implementation, model evaluation, and emotion-to-facial expression mapping. The overall system workflow is illustrated in Fig. 1, which depicts the interaction between components and the sequential stages from raw text input to emotional visual output.

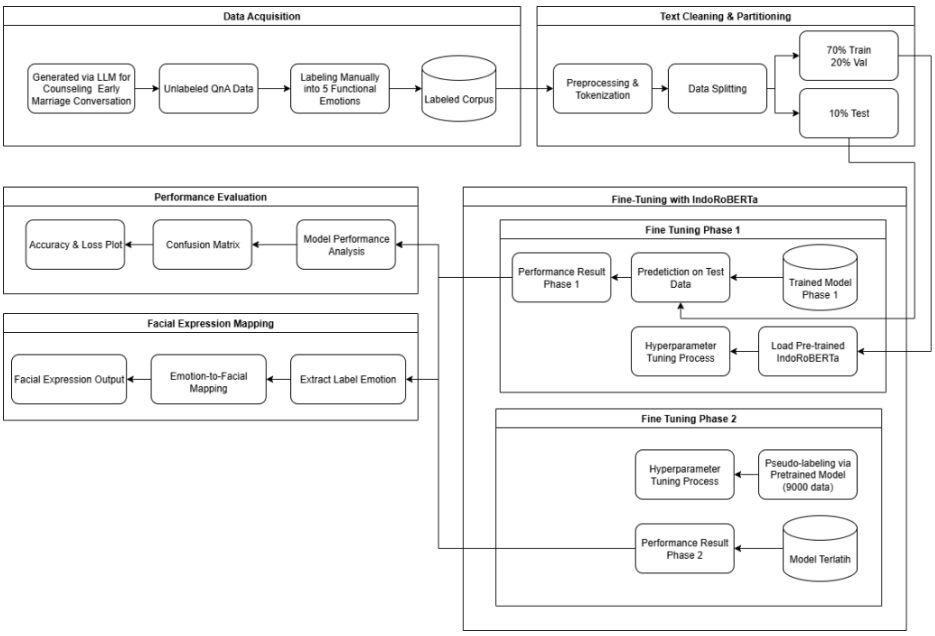

Fig. 1. System Design of IndoRoBERTa-Based Emotion Classification and Mapping

C. Dataset Collecting

The conversation data were generated by leveraging the capabilities of a Large Language Model (LLM) to simulate counseling sessions related to early marriage. The LLM was employed to produce realistic question-

and-answer (QnA) datasets that resemble actual interactions between counselors and participants. To ensure the authenticity and cultural appropriateness of the generated dialogues, especially given the sensitivity of the early marriage topic in Lombok, the LLM-generated data underwent a manual review process. This review was conducted by the research team to verify that each interaction reflected a realistic counseling scenario and respected the local socio-cultural context. Dialogue samples deemed inauthentic, irrelevant, or culturally inappropriate were discarded or substantially revised before proceeding to the labeling stage. This validation process was a crucial step in ensuring the quality and integrity of the data corpus used to train the model. Following this validation, the labeling process categorized each dialogue into one of five functional emotion classes: Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary. A total of 2,000 manually labeled data entries were used as the initial corpus for training the model in the first phase. To provide a clearer understanding of the data's structure and content, Table I presents selected annotated examples from the dataset.

TABLE I
SAMPLE QNA DATA WITH EMOTION LABELS

Input	Response	Emotion Label
Why do some of my friends choose to get married right after high school?	It's common to see peers making that choice, but it's important to understand the long-term impact. Early marriage often limits opportunities for higher education and career growth. Let's talk about your goals and how you can pursue them first.	Cautionary
I feel pressured to marry because my family believes it's the right path for me.	Family expectations can be strong, but it's crucial to make life decisions based on your readiness and aspirations. Building your future doesn't have to start with marriage you deserve time to grow and explore.	Gentleness
Is it wrong if I prioritize college over settling down?	Not at all. Prioritizing education can open up many doors for your future. The more you learn and grow, the more confident and capable you'll become in relationships and beyond.	Inspirational
Will waiting to marry affect my chances of finding a good partner?	Waiting often means becoming a better version of yourself. As you grow, you'll attract people who share your maturity and values. Rushing into marriage doesn't guarantee happiness.	Analytical
I'm scared I'll be left behind if I don't marry soon.	It's natural to feel anxious, but remember that life isn't a race. Focus on building yourself emotionally, mentally, and financially so when the time comes, you're truly ready.	Enthusiasm

The initial dataset used for model training consisted of 2,500 samples, manually annotated to ensure high label quality. The distribution was intentionally balanced, with 500 entries assigned to each of the five emotion categories. This breakdown is shown in Table II.

TABLE II
DISTRIBUTION FOR INITIAL DATASET

Emotion Label	Sample Data
Enthusiasm	400

Gentleness	400
Analytical	400
Inspirational	400
Cautionary	400

Following the semi-automatic data expansion process using the model trained in Phase 1, the dataset was enlarged to a total of 10.000 samples. The final dataset distribution reflects a more realistic imbalance across categories, which poses additional challenges for training in Phase 2. Table III summarizes the final class distribution.

TABLE III
DISTRIBUTION LABEL FOR ALL DATASET

Emotion Label	**Sample Data**
Enthusiasm	1.181
Gentleness	3.105
Analytical	961
Inspirational	4.393
Cautionary	360

In the second phase, the model obtained from phase one was used to automatically label an additional 10.000 previously unlabeled entries. These auto-labeled entries were then used as a training corpus for further fine-tuning the model. The updated model was saved, and its performance was compared to that of the phase one model to observe any improvements in accuracy and generalization.

D. Preprocessing Data

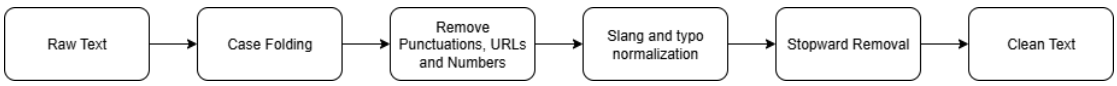

Fig. 2. Preprocessing Data Flow

The following diagram Fig. 2 illustrates the workflow of the text preprocessing steps applied to prepare raw conversation data for machine learning models. Data preprocessing plays a crucial role in ensuring that the text is properly cleaned and structured for further analysis. As shown in Fig. 2, the process starts with Case Folding, which involves converting all the text to lowercase. This ensures uniformity in the text, eliminating distinctions between words like "Marriage" and "marriage" that would otherwise be treated differently by the model.

Following this, Punctuation, URLs, and Numbers are removed, as they are generally irrelevant for emotion classification tasks. For instance, punctuation marks such as exclamation points ("!") and URLs like "www.example.com" do not provide valuable information for emotional content analysis. Their removal helps reduce noise in the data, making the model's task more efficient.

Next, Slang and Typo Normalization is performed, where informal language, such as "u" for "you" and "gr8" for "great," is standardized to its formal equivalents. Similarly, typographical errors like "recieve" are corrected to "receive." This normalization ensures that the model accurately understands the intended meaning of the text, even when informal or incorrect spellings are used.

Stopword Removal is another important step, where common, non-essential words, such as "is," "the," and "and" in English, or "yang," "di," and "dan" in Indonesian, are removed. These stopwords do not contribute significantly to the emotional meaning of the text and are discarded to allow the model to focus on more relevant words.

The outcome of this preprocessing is Clean Text, which is free from irrelevant elements. This clean text is then ready to be used in emotion classification tasks, improving the performance of machine learning models like IndoRoBERTa by ensuring that only meaningful and relevant text is analyzed. The diagram in Figure 2 clearly represents this process, providing a visual overview of the steps involved in data preprocessing.

E. Implementation of the IndoRoBERTa Model

At this stage, the pre-trained IndoRoBERTa model is loaded into the development environment and applied to the processed dataset. IndoRoBERTa is a Transformer-based model optimized for Indonesian text and chosen for its superior ability to handle more complex emotion classification tasks. This model is well-suited for classifying emotions in counseling dialogues due to its strength in processing natural language and recognizing patterns within the text.

The input to the model consists of tokens (T1, T2, ..., Tn), which represent the words or subwords in the conversation text. As shown in Figure 1, these tokens are passed through the model, and each token is associated with an embedding (E1, E2, ..., En). The model processes these embeddings and produces an output that corresponds to the emotional categories (Ent, Gen, Ana, Inp, Cau), which represent different functional emotions such as Enthusiasm, Gentleness, Analytical, Inspirational, and Cautionary. The output of the model is used to classify the emotional tone of the conversation. To visualization the architecture as shown in Fig. 3.

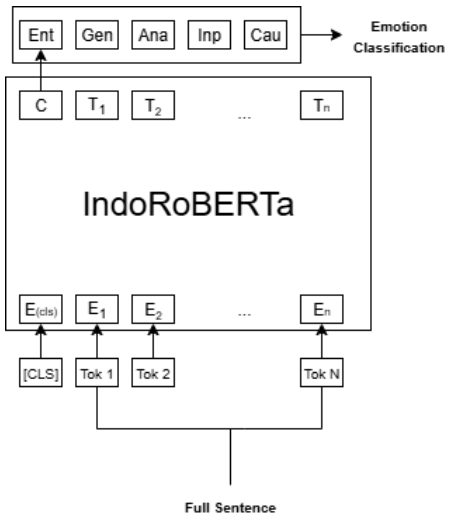

Fig. 3. The architecture of the IndoRoBERTa model

Before using the model for training, hyperparameter tuning is carried out to optimize the classification results. This process involves adjusting key hyperparameters such as the learning rate, batch size, and number of epochs. The choice of these parameters is critical for achieving optimal performance. The hyperparameters used in this study are summarized in Table IV.

TABLE IV
HYPERPARAMETERS USED FOR TRAINING THE INDOROBERTA MODEL

Hyperparameter	**Value**
Base Model	flax-community/indonesian-roberta-base
Optimizer	AdamW
Learning Rate	2e-6
Number of Epochs	10
Batch Size	16
Early Stopping	Patience = 1 epoch

Given that the expanded dataset in Phase 2 exhibited a significant class imbalance (as shown in TableIII), a mitigation strategy was implemented to prevent model bias towards the majority classes. To address this, we applied a class-weighting technique during the Phase 2 fine-tuning process. This technique works by adjusting the loss function to apply a higher penalty to misclassifications of minority classes. The weights are thus calculated in inverse proportion to the frequency of each class, ensuring that the model gives equal consideration to each emotion category during training, even those with fewer samples.

After the hyperparameter tuning, the fine-tuning of the IndoRoBERTa model is performed in two phases. In Phase 1, the model is fine-tuned using the training dataset (70%) to teach the model the emotional patterns in the counseling conversations. In Phase 2, fine-tuning is continued to adjust the model to more specific data, improving the model's accuracy in emotion classification.

This fine-tuning process ensures that the IndoRoBERTa model is adapted to the unique characteristics of the Indonesian language used in early marriage counseling, thus enhancing its ability to classify emotions more effectively.

F. Evaluation Phase

Once the IndoRoBERTa model has been trained and fine-tuned, the next crucial step is to evaluate its performance. Model evaluation helps determine how well the model has learned to classify emotions from the conversation text. This step is essential because it allows us to assess whether the model's predictions align with the expected results and how accurately it can categorize emotions based on the input text.

The evaluation process involves using several performance metrics that provide different insights into how well the model performs in classification tasks. These metrics give a comprehensive understanding of the model's strengths and weaknesses.

a) Accuracy

Accuracy measures the proportion of correctly classified instances relative to the total number of predictions made. This metric provides a clear, high-level overview of the model's overall performance and is particularly useful when the class distribution in the dataset is balanced. As a straightforward yet effective metric, it helps to gauge how well the model performs in general [23], [24]. It is calculated by dividing the sum of true positives and true negatives by the total sum of all predictions.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

b) Precision

Precision evaluates the proportion of true positive predictions among all instances the model classified as positive [25]. This measure becomes critical in applications where the consequences of a false positive error are significant. For instance, in a counseling context, incorrectly classifying a cautionary message as inspirational could have negative repercussions, making it vital to minimize such errors. Therefore, precision serves as a direct indicator of the model's ability to avoid specific misclassification types.

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \tag{2}$$

c) Recall

A model's effectiveness can also be measured by its ability to identify all relevant instances of a class, a capability quantified by the recall metric [25]. This measure is especially important in contexts were failing to identify a positive case carries a significant risk, such as in early marriage counseling. In such scenarios, a failure to recognize a critical emotional signal (a false negative) could severely undermine the system's perceived empathy and utility. Recall is thus used to assess the model's success in minimizing these omissions, ensuring comprehensive detection of all relevant data points.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \tag{3}$$

d) F1-Score

The F1-score balances the inherent trade-off between precision and recall by calculating their harmonic mean[25], [26] . This integrated, single-value metric provides a more holistic assessment of model effectiveness, which is particularly valuable when dealing with imbalanced datasets or when the costs of false positives and false negatives are equally significant. Its formulation ensures that strong performance on one metric does not mask poor performance on the other, leading to a more balanced evaluation of the model's overall capability.

$$\boldsymbol{F1-Score} = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall} \tag{4}$$

e) Accuracy and Loss Plot

The evolution of the model's performance during training was critically assessed using accuracy and loss plots, which respectively measure the rate of correct classifications and prediction errors across each epoch. A decreasing loss function signifies that the model's parameters are converging toward an optimal representation of the data, while an increasing accuracy curve indicates a growing proficiency in mapping inputs to the correct outputs. A key aspect of this analysis is the comparison between training and validation performance to evaluate generalization. While concurrent improvement in both metrics signals effective learning, a significant divergence marked by high training accuracy and low validation accuracy indicates overfitting. In this state, the model has memorized noise from the training set at the expense of its ability to perform on new, unseen data. Therefore, the combined interpretation of these plots is a crucial diagnostic practice for identifying suboptimal learning patterns and ensuring the final model is both accurate and generalizable, especially when considering the function of loss in imbalanced data scenarios [27], [28].

G. Emotion to Facial Expression Mapping

Following the textual emotion classification, this research progresses to a conceptualization phase aimed at translating the categorical outputs into non-verbal communication signals for a virtual agent. In line with advancements in the field of Affective Computing, which focuses on developing systems that can recognize, interpret, and process human affect [29], this study proposes an initial mapping framework. Recent systematic reviews indicate that multimodal fusion particularly audio-visual-text delivers the most reliable affective cues [29]. Nevertheless, the language modality (text) remains a core, low-cost, and ubiquitous componentnt [30].

Therefore, the mapping framework in this research is presented as a proof-of-concept for the text modality. This framework is designed to link the five functional emotion labels to corresponding facial expression representations on an avatar. Given the absence of an established mapping standard for this specific emotion taxonomy, the initial mapping is based on psychological and logical interpretations of each emotional state. Empirical validation through user studies or expert annotation constitutes a crucial future step to refine and confirm the practical effectiveness of this mapping. The specific guidelines for this conceptual mapping are detailed in Table V.

TABLE V
GUIDE FOR MAPPING EMOTION LABELS TO FACIAL EXPRESSIONS

Emotion Label	Facial Expressions	Description
Enthusiasm	A wide smile and sparkling eyes	Creates a positive and enthusiastic atmosphere, showing optimism and active support for the client's thoughts
Gentleness	A sincere smile and calming eyes	Builds a sense of security and shows empathy. This expression communicates that the psychologist is listening with full attention
Analytical	A neutral expression with slightly furrowed brows	Indicates full concentration and deep thought, as if considering information objectively
Inspirational	A confident smile and a reassuring gaze	Shows confidence and hope to empower the client and encourage them to see their future potential
Cautionary	A serious expression with furrowed brows and a sharp gaze	Communicates concern or seriousness about a risk in a firm yet caring manner

IV. Results And Discussion

This chapter presents the primary findings obtained from the implementation and evaluation of the IndoRoBERTa model for the classification of functional emotions within early marriage counseling dialogues. The chapter is organized into two main sections: the presentation of quantitative experimental results and an in-depth discussion of the implications of these findings.

Upon completion of the training process, the model's performance was objectively evaluated using a separate test dataset to obtain an unbiased quantitative assessment. The evaluation results, summarized in the classification report, demonstrated highly satisfactory performance.

A. Implementation Results: Training Phase 1 Fine-Tuning on Initial Data

In the initial phase of the research, the foundation for the emotion classification model was established through the fine-tuning of the IndoRoBERTa linguistic model. This process strategically utilized a focused, high-quality base dataset comprising 2.000 counseling dialogue samples. Each sample within this corpus underwent a meticulous manual annotation process to ensure label accuracy. To prevent inherent bias during the initial learning stage, the dataset was designed to be balanced, with an equal distribution for each of the five defined functional emotion categories: 'Enthusiastic', 'Gentleness', 'Analytical', 'Inspirational', and 'Cautionary'. The fundamental objective of this phase was to construct a solid and reliable base model capable of understanding and differentiating fundamental emotional patterns within the specific domain of early marriage counseling, prior to its exposure to large-scale training.

To understand how the model achieved this level of performance, the dynamics of its learning process were analyzed through the visualization of accuracy and loss curves over ten training epochs. As shown in Table VI.

TABLE VI
CLASSIFICATION REPORT TRAINING PHASE 1

Emotion Label	Precision	Recall	F1-Score	Support
Enthusiam	1.00	0.97	0.99	40
Gentleness	0.97	0.90	0.94	40
Analytical	0.86	0.90	0.88	40
Inspirational	0.91	0.97	0.94	40
Cautionary	0.95	0.93	0.94	40

Overall, the model achieved an impressive classification accuracy of 93.5%, indicating that most of its predictions were correct. Further analysis of the per-class metrics reveals strong performance. The model demonstrated exceptional precision for the 'Antusias' class (1.00), meaning every sample it predicted as enthusiastic was correct. A significant finding was the achievement of the highest recall (0.98) for both the 'Antusias' and 'Inspirational' categories. This implies the model was highly effective at identifying all true instances of these emotions. The balance between precision and recall, reflected in the strong F1-scores across all classes (ranging from 0.88 to 0.99), further affirms the model's holistic effectiveness. To understand how the model achieved this performance, its learning process was analyzed by visualizing the accuracy and loss curves over ten training epochs, as shown in *Fig. 4*.

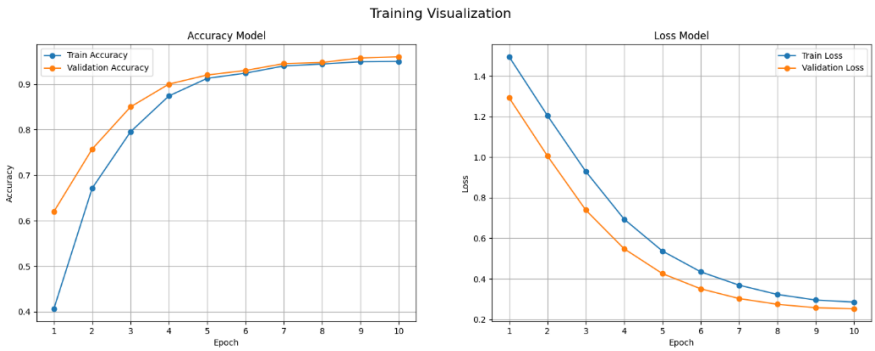

Fig. 4. Training Visualization Plot Phase 1

Observation of the plots reveals a healthy convergence. The accuracy curves for both the training and validation sets show consistent improvement and move in tandem, indicating that the model not only learned effectively from the training data but was also able to generalize its knowledge well to unseen data. Model stability was approached around the sixth epoch, where the performance curve began to plateau, signifying that the model had reached a point of learning saturation. Crucially, no significant divergence was detected between the training and validation curves, providing strong evidence that the model did not suffer from overfitting. Subsequently, to dissect the prediction patterns and specific error types made by the model more deeply, an analysis of the confusion matrix was performed and summrized in *Fig. 5*.

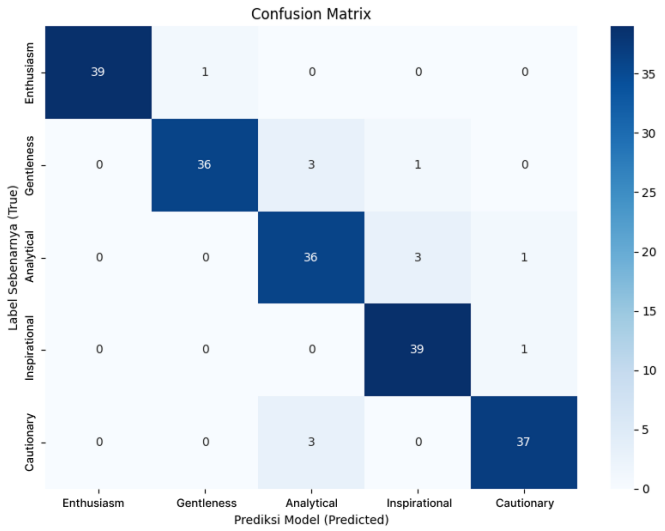

Fig. 5. Training Confusing Matrix Phase 1

An analysis of the confusion matrix reaffirms the model's high accuracy, evidenced by the dominant values concentrated along its main diagonal, which represent correct predictions. A more granular examination of the off-diagonal elements, however, offers insight into the model's specific limitations. The most prominent confusion pattern observed was the misclassification of both 'Lembut' (Gentleness) and 'Peringatan' (Cautionary) samples as 'Analitis' (Analytical), each occurring in three instances. Additionally, the model misclassified 'Analitis' as 'Inspiratif' (Inspirational) in three cases. This phenomenon suggests a degree of semantic overlap within the dataset, where the linguistic characteristics of a gentle or cautionary statement may contain features that the model interprets as analytical. Similarly, an analytical response might be delivered with an inspirational tone, leading to confusion. Despite these specific instances of error, it is crucial to note that the number of misclassifications is minimal when compared to the volume of correct predictions, underscoring the model's overall robustness.

In conclusion, the synthesis of these three analyses the classification report, learning curves, and confusion matrix convincingly confirms that the fine-tuned IndoRoBERTa model possesses a highly reliable capability to distinguish complex emotional nuances in the researched domain. With its demonstrably robust and stable performance, this model serves as a solid and well-validated foundation for the subsequent, larger-scale data expansion phase.

B. Implementation Results: Phase 2 - Fine-Tuning on the Complete Dataset

On the second training phase, the IndoRoBERTa model was retrained utilizing an expanded dataset of 10,000 samples. This step aimed to enhance the model's generalization capability and robustness by exposing it to a larger and more imbalanced corpus, which more accurately reflects real-world data characteristics. The training process commenced from the best-performing checkpoint of Phase 1 to ensure an efficient continuation of learning. The final evaluation on a separate test set demonstrated that the model was able to maintain a highly competitive performance. As detailed in the classification report shown in Table VII, the model achieved an overall accuracy of 88.80%.

TABLE VII
CLASSIFICATION REPORT FOR TRAINING PHASE 2

Emotion Label	Precision	Recall	F1-Score
Enthusiam	0.79	0.85	0.82
Gentleness	0.88	0.86	0.82
Analytical	0.99	0.89	0.90
Inspirational	0.91	0.96	0.93
Cautionary	0.91	0.89	0.82

Despite a slight decrease from the first phase, this accuracy figure still indicates a high level of capability, especially considering the challenge posed by the significant class imbalance in the training data. The classification report shows strong performance on the majority class, 'Inspiratif' (F1-score 0.93). However, a notable trade-off is observed for the minority class 'Analitis', which recorded an exceptionally high precision of 0.99 but a much lower recall of 0.71. This indicates that while the model's 'Analitis' predictions were highly reliable, its ability to identify all instances of this class was impacted by the dominance of other classes in the training data. The learning dynamics during this second phase, which spanned six epochs, are visualized in the training curves shown in *Fig. 6*.

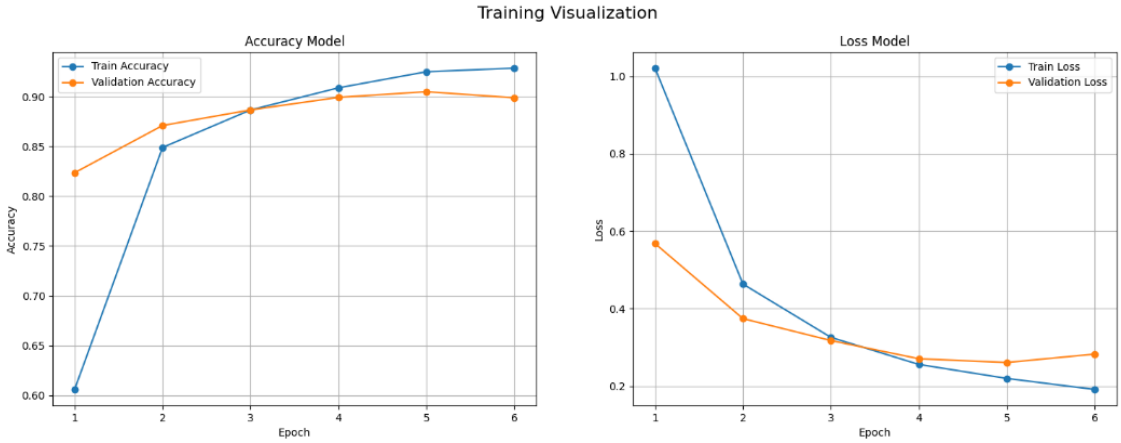

Fig. 6. Training Visualization for Phase 2

The plots show that the validation accuracy curve closely follows the trend of the training accuracy, signifying a stable learning process. However, the validation curve began to show signs of stagnation after the fifth epoch, confirming that terminating the training at the sixth epoch was an appropriate decision to prevent potential overfitting and to secure the model at its optimal performance point. For a more granular analysis of error patterns on the imbalanced data, a confusion matrix was employed, as illustrated in *Fig. 7*.

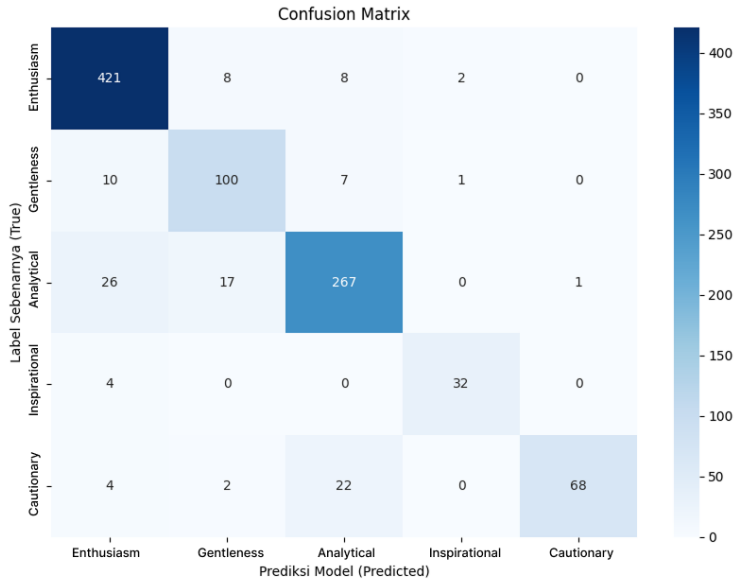

Fig. 7. Confusion Matrix for Phase 2

Overall, the matrix in *Fig. 7* visually confirms the model's strong performance on the majority class, 'Inspiratif', marked by a high diagonal value of 421 correct predictions. However, the matrix also clearly reveals the primary sources of the model's confusion, which are exacerbated by the data imbalance. A significant semantic overlap is evident where the model struggles with less frequent classes. For instance, 26 'Lembut' (Gentleness) samples were misclassified as the more dominant 'Inspiratif (Inspirational)' class, and another 17 were misclassified as 'Antusias'. Furthermore, 22 'Analitis' samples were erroneously classified as 'Lembut'. This confusion pattern indicates that when faced with linguistic ambiguity, the model exhibits a tendency to predict a more dominant class from the training data.

Overall, the results from Phase 2 validate the effectiveness of the two-stage training methodology. Despite being confronted with imbalanced data, the model demonstrated remarkable resilience by maintaining an accuracy of nearly 89%. Training on the larger corpus successfully enhanced the model's generalization ability, while the detailed error analysis highlighted specific areas where semantic overlap between classes presented a more pronounced challenge, providing clear direction for future refinement.e results from Phase 2 validate the effectiveness of the two-stage training methodolog.

C. Mapping Facial Expression

The final stage of the system's framework involves translating the model's emotion classification output into corresponding visual representations for implementation within a virtual counseling agent. The objective of this mapping is to create a more humanistic and empathetic interaction, wherein the counselor's response is not only conveyed textually but is also reinforced by relevant non-verbal cues. This process utilizes a direct mapping approach, where each functional emotion label is specifically associated with a set of facial expression parameters, which are psychologically interpreted within the counseling context. The visual implementation of the mapping for each emotion category is illustrated on *Fig. 8.*

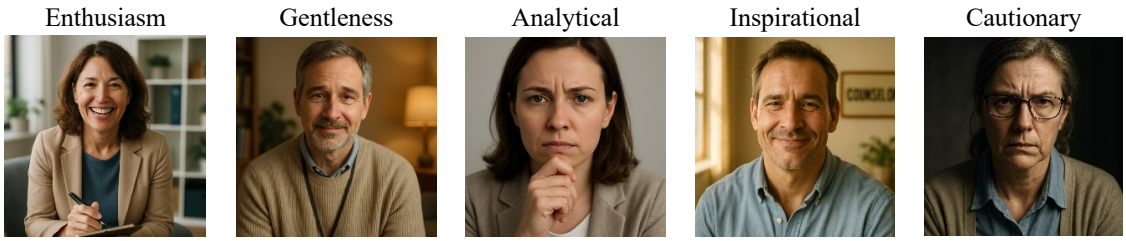

Fig. 8. Visualization of Mapping Emotion Labels to Facial Expressions. The illustrative images were generated by the author using a generative AI model.

Each visual representation serves a specific therapeutic function. The Enthusiasm expression, represented by a wide smile and sparkling eyes, functions to create a positive and motivating interactional atmosphere. This cue actively communicates optimism and support for the client. Meanwhile, the Gentleness expression is manifested through a more subtle smile accompanied by a warm and calming gaze, aiming to build a sense of security and demonstrate profound empathy, thereby encouraging openness.

For moments requiring deep thought, the Analytical expression is displayed through a face that is generally neutral yet focused, with slightly furrowed brows. This representation signals to the client that their information is being considered thoroughly and objectively. In contrast, the Inspirational expression is used to instill hope and empower the client, characterized by a confident smile and a reassuring gaze to encourage them to see future potential. Finally, in situations requiring emphasis on a particular risk, the Cautionary expression is manifested through a more serious countenance and a sharp, yet still caring, gaze. This cue is essential for communicating the gravity of an issue in a firm but non-intimidating manner, while still maintaining the therapeutic relationship.

D. Contextualizing Performance Against Previous Architectures

The performance of the IndoRoBERTa model, which achieved an accuracy of 89% on the five-class functional emotion task in this study, represents a significant advancement. For comparison, a prior study evaluating a six-category emotion classification task reported an accuracy of 71.6% for a Convolutional Neural Network (CNN) [9]. Meanwhile, another study using K-Nearest Neighbors (KNN) for a two-category sentiment classification task (positive/negative) achieved an accuracy of 79% [10].

It is noteworthy that the task undertaken by the IndoRoBERTa model in this research is more nuanced than binary sentiment classification. Nevertheless, its result clearly surpasses these baseline architectures. The superior performance of IndoRoBERTa can be attributed to its ability to understand deep linguistic context and long-range dependencies, a hallmark of the Transformer architecture. This capability is particularly crucial for the domain of counseling dialogue, which is replete with nuance and ambiguity. These results empirically validate that for tasks demanding profound semantic understanding, a Transformer-based architecture such as IndoRoBERTa offers a distinct advantage.

V. Conclusion

This study successfully met its objectives by developing and evaluating an IndoRoBERTa-based emotion classification model for early marriage counseling dialogues. In line with the hypothesis, the model demonstrated high performance, achieving a testing accuracy of 93.5% on the balanced dataset (Phase 1) and 89% on the expanded, imbalanced dataset (Phase 2). This performance was evaluated using standard metrics such as precision, recall, and F1-score, and was further analyzed through accuracy/loss curves and confusion matrices.

Furthermore, the primary contribution of this research is the introduction of a conceptual visual mapping framework, which links each functional emotion category ('Enthusiasm', 'Gentleness', 'Analytical',

'Inspirational', and 'Cautionary') to a corresponding non-verbal expression on a virtual avatar. This framework serves as a conceptual bridge to add expressiveness to the virtual counseling agent. It is important to note that this visual mapping is conceptual and requires future empirical validation through user studies to confirm its practical effectiveness. Thus, the framework successfully establishes a viable foundation for the future development of more dynamic and emotionally intelligent interactions, fulfilling the study's second objective.

References

[1] M. D. H. Rahiem, "COVID-19 and the surge of child marriages: A phenomenon in Nusa Tenggara Barat, Indonesia," *Child Abuse Negl*, vol. 118, p. 105168, Aug. 2021, doi: 10.1016/j.chiabu.2021.105168.

[2] M. E. Greene, M. Siddiqi, and T. F. Abularrage, "Systematic scoping review of interventions to prevent and respond to child marriage across Africa: progress, gaps and priorities," *BMJ Open*, vol. 13, no. 5, p. e061315, May 2023, doi: 10.1136/bmjopen-2022-061315.

[3] A. Sampurna, H. J. Ritonga, and A. R. Matondang, "Integration of Media Literacy in Religious Counseling for Preventing Early Marriage in Nias Barat," *International Journal of Islamic Education, Research and Multiculturalism (IJIERM)*, vol. 6, no. 3, pp. 1205–1218, Dec. 2024, doi: 10.47006/ijierm. v6i3.392.

[4] Z. Guo, A. Lai, J. H. Thygesen, J. Farrington, T. Keen, and K. Li, "Large Language Models for Mental Health Applications: Systematic Review," *JMIR Ment Health*, vol. 11, p. e57400, Oct. 2024, doi: 10.2196/57400.

[5] V. Sorin *et al.*, "Large Language Models and Empathy: Systematic Review," *J Med Internet Res*, vol. 26, p. e52597, Dec. 2024, doi: 10.2196/52597.

[6] G. Park, J. Chung, and S. Lee, "Effect of AI chatbot emotional disclosure on user satisfaction and reuse intention for mental health counseling: a serial mediation model," *Current Psychology*, vol. 42, no. 32, pp. 28663–28673, Nov. 2023, doi: 10.1007/s12144-022-03932-z.

[7] H. Chin *et al.*, "The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study," *J Med Internet Res*, vol. 25, p. e51712, Oct. 2023, doi: 10.2196/51712.

[8] J. Sun, "Research and Application Analysis of Multimodal Emotion Recognition Methods Based on Speech, Text, And Facial Expressions," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 293–297, Mar. 2024, doi: 10.54097/agvjvq19.

[9] M. Y. Baihaqi, E. Halawa, R. A. S. Syah, A. Nurrahma, and W. Wijaya, "Emotion Classification in Indonesian Language: A CNN Approach with Hyperband Tuning," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 137–146, Oct. 2023, doi: 10.24002/jbi. v14i02.7558.

[10] A. Zamsuri, S. Defit, and G. W. Nurcahyo, "Classification of Multiple Emotions in Indonesian Text Using The K-Nearest Neighbor Method," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 2, pp. 1012–1021, Jun. 2023, doi: 10.37385/jaets. v4i2.1964.

[11] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019.

[12] Y. O. Sihombing, R. Fuad Rachmadi, S. Sumpeno, and Moh. J. Mubarok, "Optimizing IndoRoBERTa Model for Multi-Class Classification of Sentiment & Emotion on Indonesian Twitter," in *2024 IEEE 10th Information Technology International Seminar (ITIS)*, IEEE, Nov. 2024, pp. 12–17. doi: 10.1109/ITIS64716.2024.10845566.

[13] E. Bourke, C. Barker, and M. Fornells-Ambrojo, "Systematic review and meta-analysis of therapeutic alliance, engagement, and outcome in psychological therapies for psychosis," *Psychology and Psychotherapy: Theory, Research and Practice*, vol. 94, no. 3, pp. 822–853, Sep. 2021, doi: 10.1111/papt.12330.

[14] J. Kadur, J. Lüdemann, and S. Andreas, "Effects of the therapist's statements on the patient's outcome and the therapeutic alliance: A systematic review," *Clin Psychol Psychother*, vol. 27, no. 2, pp. 168–178, Mar. 2020, doi: 10.1002/cpp.2416.

[15] L. Del Giacco, M. T. Anguera, and S. Salcuni, "The Action of Verbal and Non-verbal Communication in the Therapeutic Alliance Construction: A Mixed Methods Approach to Assess the Initial Interactions with Depressed Patients," *Front Psychol*, vol. 11, Feb. 2020, doi: 10.3389/fpsyg.2020.00234.

[16] Z. N. Maharani, A. Luthfiarta, and N. Z. Farsya, "Sentiment Analysis of the 2024 Indonesian Presidential Dispute Trial Election using SVM and Naïve Bayes on Platform X," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, pp. 440–449, Jun. 2024, doi: 10.47065/bits. v6i1.5380.

[17] W. Suwarningsih, R. A. Pramata, F. Y. Rahadika, and M. H. A. Purnomo, "RoBERTa: language modelling in building Indonesian question-answering systems," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 6, p. 1248, Dec. 2022, doi: 10.12928/telkomnika. v20i6.24248.

[18] S. William, Kenny, and A. Chowanda, "EMOTION RECOGNITION INDONESIAN LANGUAGE FROM TWITTER USING INDOBERT AND BI-LSTM," *Communications in Mathematical Biology and Neuroscience*, vol. 2024, 2024, doi: 10.28919/cmbn/7858.

[19] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[20] M. Malgaroli, T. D. Hull, J. M. Zech, and T. Althoff, "Natural language processing for mental health interventions: a systematic review and research framework," *Transl Psychiatry*, vol. 13, no. 1, p. 309, Oct. 2023, doi: 10.1038/s41398-023-02592-2.

[21] M. A. Witherow *et al.*, "Customizable Avatars with Dynamic Facial Action Coded Expressions (CADyFACE) for Improved User Engagement," Mar. 2024.
[22] S. Kang, H. Song, B. Yoon, K. Kim, and W. Woo, "The Influence of Emotion-based Prioritized Facial Expressions on Social Presence in Avatar-mediated Remote Communication," in *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, Oct. 2024, pp. 1147–1156. doi: 10.1109/ISMAR62088.2024.00131.
[23] K. Wisnudhanti and F. Candra, "Image Classification of Pandawa Figures Using Convolutional Neural Network on Raspberry Pi 4," *J Phys Conf Ser*, vol. 1655, no. 1, p. 012103, Oct. 2020, doi: 10.1088/1742-6596/1655/1/012103.
[24] D. M. Aprilla, F. Bimantoro, and I. G. P. S. Wijaya, "The Palmprint Recognition Using Xception, VGG16, ResNet50, MobileNet, and EfficientNetB0 Architecture," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 1065, Apr. 2024, doi: 10.30865/mib. v8i2.7577.
[25] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
[26] R. Pereira *et al.*, "Systematic Review of Emotion Detection with Computer Vision and Deep Learning," *Sensors*, vol. 24, no. 11, p. 3484, May 2024, doi: 10.3390/s24113484.
[27] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artif Intell Rev*, vol. 58, no. 7, p. 195, Apr. 2025, doi: 10.1007/s10462-025-11198-7.
[28] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices," *Remote Sens (Basel)*, vol. 16, no. 3, p. 533, Jan. 2024, doi: 10.3390/rs16030533.
[29] Y. Wang *et al.*, "A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances," Mar. 2022.
[30] G. Hu *et al.*, "Recent Trends of Multimodal Affective Computing: A Survey from NLP Perspective," Oct. 2024.