Social Network Analysis and Mining (2024) 14:234
https://doi.org/10.1007/s13278-024-01400-z

ORIGINAL ARTICLE

EmoMBTI-Net: introducing and leveraging a novel emoji dataset for personality profiling with large language models

Akshi Kumar[1] · **Dipika Jain**[2]

Received: 19 July 2024 / Revised: 27 November 2024 / Accepted: 28 November 2024

Abstract
Emojis, integral to digital communication, often encapsulate complex emotional layers that enhance text beyond mere words. This research leverages the expressive power of emojis to predict Myers-Briggs Type Indicator (MBTI) personalities, diverging from conventional text-based approaches. We developed a unique dataset, EmoMBTI, by mapping emojis to specific MBTI traits using diverse posts scraped from Reddit. This dataset enabled the integration of Natural Language Processing (NLP) techniques tailored for emoji analysis. Large Language Models (LLMs) such as FlanT5, BART, and PEGASUS were trained to generate contextual linkages between text and emojis, further correlating these emojis with MBTI personalities. Following the creation of this dataset, these LLMs were applied to understand the context conveyed by emojis and were subsequently fine-tuned. Additionally, transformer models like RoBERTa, DeBERTa, and BART were specifically fine-tuned to predict MBTI personalities based on emoji mappings from MBTI dataset posts. Our methodology significantly enhances the capability of personality assessments, with the fine-tuned BART model achieving an impressive accuracy of 0.875 in predicting MBTI types, which notably exceeds the performances of RoBERTa and DeBERTa, at 0.82 and 0.84 respectively. By leveraging the nuanced communication potential of emojis, this approach not only advances personality profiling techniques but also deepens insights into digital behaviour, highlighting the substantial impact of emotive icons in online interactions.

Keywords Sentiment analysis · Personality · MBTI · Emojis · LLM · Natural language understanding

1 Introduction

Automatic personality detection utilizes machine learning algorithms and natural language processing (NLP) techniques to analyze digital communication, such as texts, social media posts, or emails, inferring individuals' personality traits based on established models like the Myers-Briggs Type Indicator (MBTI) (Keh and Cheng 2019; Hernández et al. 2018; Cui and Qi 2017; Hernandez and Scott 2017; Ismail et al. 2017; Ren et al. 2021; Jain et al. 2021; Kumar et al. 2023). By examining linguistic patterns, word choices, and communication styles, algorithms can categorize individuals into different personality types, such as Introversion (I) vs. Extraversion (E), Sensing (S) vs. Intuition (N), Thinking (T) vs. Feeling (F), and Judging (J) vs. Perceiving (P). This process offers valuable insights into behaviour, enables personalized digital experiences, and contributes to psychological research by studying large-scale datasets. Leveraging the MBTI framework, automatic personality detection algorithms can decode the nuances of digital behaviour, providing actionable insights for personal development, relationship building, targeted marketing, and a deeper understanding of human psychology in the digital age.

User-generated content on social media platforms has become a valuable resource for personality detection, as it encompasses more than just text. Beyond written posts and comments, user-generated content includes images, videos, likes, shares, and other interactions, providing a rich source of data for inferring individuals' personality traits. Thus, a multi-modal approach to personality detection, incorporating textual, visual, and emotive elements, allows a more comprehensive understanding of users' personalities and preferences, facilitating personalized experiences and targeted interventions in various digital contexts. More specifically, emojis, beyond their colourful appearance, are

✉ Akshi Kumar
Akshi.Kumar@gold.ac.uk

[1] Department of Computing, Goldsmiths, University of London, London, UK

[2] Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

Published online: 10 December 2024

integral to digital communication, offering visual cues that enhance the expression of emotions and attitudes alongside text. This symbiosis provides a deeper understanding of individuals' digital self-expression. For example, in a text message conversation, a person might use a series of heart emojis 💖 😍 🥰 to convey affection and excitement about a forthcoming event. These emojis complement the text, adding layers of emotional nuance and conveying the sender's enthusiasm more vividly than words alone. Similarly, a laughing emoji 😂 might accompany a joke or humorous comment, amplifying the sense of amusement and creating a more engaging interaction.

As emojis play a crucial role in enriching digital communication by conveying emotions and attitudes in a concise and visually appealing manner, enhancing the overall expressive capacity of written text. analyzing emoji usage to infer personality types based on MBTI principles merges the realms of digital communication and psychological typology. This integration offers a novel approach to understanding the psychological underpinnings of individuals through their digital expressions.

Building on this foundation, this research aims to refine predictive models that adeptly correlate specific emojis with MBTI traits to discern personality types from textual conversations. For instance, frequent use of emotive emojis like hearts or smileys may suggest a preference for Feeling over Thinking, while the use of more ordered or structured symbols like checkmarks or clocks might indicate a Judging rather than Perceiving inclination. This methodology leverages the inherent expressiveness of emojis to tap into subtle psychological cues that may be less apparent in plain text. The utilization of emoji-based personality detection has practical applications across various domains. In social media, understanding user personality can enhance content personalization and ad targeting, improving user engagement. In educational technologies, it can help tailor learning experiences to suit different personality-driven learning styles, potentially increasing effectiveness and satisfaction. Furthermore, in professional settings, this approach could assist in team formation and leadership strategies by providing insights into team members' preferred communication styles and decision-making processes.

Transformers models and large language models (LLMs) have revolutionized natural language processing (NLP) with their capacity for complex tasks like personality detection from textual data. These models are designed to be fine-tuned on specialized datasets, which enhances their ability to discern and predict personality traits effectively. The inherent flexibility and robustness of transformer architectures enable them to process and generate text that closely mimics human communication, grounded in the nuances learned from extensive training data. The adaptability of transformers and LLMs is a key advantage, particularly for dynamic applications such as personality profiling. They are capable of ongoing learning, which allows them to refine and adapt their predictions as they are exposed to new data and evolving contexts. This capability is essential in environments where communication styles and interactions are constantly changing, ensuring that the models remain relevant and accurate over time.

Building upon the theoretical and technical groundwork, this research specifically aims to leverage emojis as a novel data source for inferring personality traits based on the MBTI model, which will serve to advance our understanding of digital behaviour and its psychological implications. The research objectives are as follows:

- *Develop the Emo-MBTI dataset* Curate and refine a novel dataset by integrating emoji usage with textual data from the MBTI dataset and additional posts scraped from the r/mbti subreddit on Reddit, aiming to map emojis to specific MBTI personality types for enhanced personality analysis.
- *Implement and evaluate advanced LLMs for Emoji-Text Contextual Mapping:* Employ and assess the effectiveness of advanced language models like FlanT5 (Liusie et al. 2024), BART(Lewis et al. 2019), and PEGASUS in understanding and generating the contextual relationships between text and emojis, preparing the groundwork for accurate personality type predictions.
- *Fine-Tune transformers for MBTI personality prediction* Utilize fine-tuned transformer models such as Roberta(Liao et al. 2021), DeBERTa (He et al. 2020), and BART to predict MBTI personality types from emoji usage, leveraging the nuanced comprehension of emoji contexts to enhance personality profiling accuracy.

This paper is structured to methodically explore the interface between emoji usage and personality profiling within the digital communication realm. We begin with a Literature Review that scrutinizes prior research on personality detection. Following this, the section on MBTI Personalities and Emojis discusses the theoretical underpinnings of our study, explaining how emojis can serve as practical indicators of MBTI personality traits. The Emo-MBTI Dataset section describes the meticulous process of dataset assembly—detailing the collection, integration, and preparation of data that combines emoji usage with textual content from diverse sources. In the Emoji-Based MBTI Personality Prediction Model section, we detail the methodologies used for training and fine-tuning Large Language Models (LLMs) to predict MBTI personality types from emojis. The Results and Discussion segment presents the outcomes of our experiments, showcasing how different models perform in the context of emoji-based personality prediction. This section highlights key findings, discusses model efficiencies, and examines the implications of the results in the broader context of digital communication and psychological analysis. Finally, the Conclusion and Future Work section summarizes the contributions of this research and outlines directions for future inquiry. It suggests potential expansions of the current work, including the exploration of additional models, further diversification of data sources, and the practical application of our findings in various domains such as marketing, mental health, and user experience design.

2 Formal problem definition and mathematical formulation

In this research, we address the challenge of predicting Myers-Briggs Type Indicator (MBTI) personality types by analyzing the use of emojis in textual communications. This problem involves interpreting the emotional and contextual nuances conveyed through emojis within text messages, a critical aspect of understanding digital human interactions.

3 Notation and definitions

- *Emojis Set (E)* Let $E = \{e_1, e_2, \ldots, e_n\}$ represent the set of all possible emojis that may appear in text messages.
- *Text Messages (X)* Define $X = \{x_1, x_2, \ldots, x_m\}$ as the collection of input text messages, where each x_i consists of words interspersed with emojis from E.
- *MBTI Types (Y)* Let $Y = \{y_1, y_2, \ldots, y_k\}$ denote the set of MBTI personality types, with each y_j corresponding to one of the 16 distinct personality types recognized by the MBTI framework.

3.1 Problem modelling

The objective is to construct a predictive function $f{:}X \rightarrow Y$ that assigns an MBTI type to a given text message based on its content and the emojis used. This function f is derived from a predictive model trained on a dataset $D \subset X \times Y$, where each pair (x_i, y_j) indicates that message x_i is associated with personality type y_j.

To optimize the performance of our model, we minimize a loss function $L(f(x_i), y_i)$ across the dataset DD. The loss function typically employed is the cross-entropy loss, which is particularly suited for classification tasks.

4 Literature review

The study of personality through textual analysis, particularly English texts, has seen a growing interest over the years. A notable dataset in this field is the publicly available Kaggle_MBTI dataset, which has been instrumental in analyzing personalities across 16 different dimensions. Throughout the years, various researchers have utilized this dataset to deepen our understanding of personality traits. For instance, in Hernandez and Scott (2017) applied a Recurrent Neural Network (RNN) model to this dataset, while Ismail et al. (2017) undertook a questionnaire-based approach to analyze MBTI personality traits. In subsequent years, more advanced models were employed; Ren et al. (2021) used a BERT model in 2020 to achieve 54% accuracy, and Jain et al. (2021) improved upon this in 2022 with a BERT-base model, reaching 69% accuracy. The following year, Kumar et al. (2023) tested the effectiveness of a kernel-based soft voting ensemble model on these personality dimensions.

Research in personality detection has also extended beyond textual data to include audio-visual and multimodal data. The First Impression dataset from Chalearn, which focuses on Big-Five personality traits, is a prominent example. Gürpinar et al. (2016a; b) explored this dataset using a pre-trained DCC model to analyze facial expressions and employed the OpenSMILE tool to extract facial emotions, achieving an accuracy of 91.3%. Following this, Rai (2016) emphasized leveraging visual information through deep networks. Zhang et al. (2016) developed a deep bimodal regression model using the same dataset, achieving an identical accuracy of 91.3%. Over the next few years, various approaches were tested on this dataset, including Kaya et al. (2017) use of a CNN with a KELM model in 2017 which reached an accuracy of 91.7%, and Gucluturk et al.'s (2017) application of a ResNet model in 2018, achieving 91.18% accuracy.

In recent years, there has been an increasing interest in integrating non-verbal cues such as facial expressions and body language into personality analysis. This trend is reflected in the work of Junior et al. (2019) and Aslan (2019) in 2019, who employed spatial–temporal modelling with CNN and RNN models along with pre-trained deep learning models such as ResNet and VGG, achieving an accuracy of 91.16%. Artha Agastya et al. (2019) continued this exploration by utilizing a range of deep learning models like DNN, CNN, RNN, and ResNet.

In Kosan et al. (2022) outlined a study aimed at predicting personality traits using social media data, particularly from Twitter. The research emphasized the importance of understanding individuals' personal characteristics for various sectors such as law enforcement, human resources, and advertising. It highlighted the need for labelled datasets and structural analysis of social media platforms to develop accurate prediction models. The methodology involved creating a personality dataset from Twitter, transforming unstructured data into meaningful formats, and utilizing LSTM-based prediction models. Evaluation was conducted on both the created dataset and an existing benchmark dataset (PAN-2015-EN).

More recently in Kennison et al. (2024) examined the relationship between personality traits, language use, and emoji usage on X (formerly Twitter). Through linguistic analysis and surveys across two samples, they found that higher emoji usage correlated with lower levels of openness to experience, while also revealing associations with specific language patterns in social media posts. These findings underscore the nuanced nature of personality expression and linguistic cues in online communication. Liao et al. (2024) also proposed a comprehensive framework that incorporates audio, video, and non-verbal audio-visual elements for personality analysis using the same benchmark multimodal dataset.

Emojis, like text, images, videos, and audio, also serve as an important means of communication, particularly in conveying emotions, thoughts, and feelings. This aspect of digital communication has not been overlooked in personality research. A recent innovative study by Saeidi (2024) involved analyzing screenshots from WhatsApp conversations that included emojis, providing new insights into personality dimensions through this unique form of expression. This research highlights how even seemingly trivial elements like emojis can offer valuable information about an individual's personality traits. We demonstrate how sophisticated algorithms can leverage emojis and language models to infer personality traits, offering insights into digital behaviour and its psychological underpinnings. Through such advancements, the intricate connection between emojis, personality, and digital communication continues to be explored, unlocking new avenues for understanding human behaviour in the digital age.

5 MBTI personalities and emojis

The use of emojis in digital communication can significantly illuminate aspects of an individual's personality, as evidenced by their choices of specific emojis in various contexts. This is particularly relevant when considering the Myers-Briggs Type Indicator (MBTI), which categorizes personality types based on how individuals perceive the world and make decisions. Emojis, as non-verbal cues, offer a unique window into how personality traits manifest in modern communication methods.

For example, consider a conversation where Person A texts, "I had a challenging day at work 😩," choosing the weary face emoji to express feelings of stress and exhaustion. The emoji here does more than simply emphasize the sentiment; it also gives clues to Person A's emotional state and possibly their personality type. If Person A frequently uses such emojis to express discomfort or distress, they might have a personality that leans towards sensitivity or introversion, where personal and professional struggles are internalized deeply. In response, Person B replies, "I totally get it, let's plan something fun to unwind! 🎉," employing the party popper emoji to communicate enthusiasm and a proactive attitude. This response, marked by a contrasting emoji, suggests a personality that might be more extroverted or feeling-oriented, focusing on emotional support and active engagement in social activities as coping mechanisms. The choice of the party popper emoji, which is vibrant and celebratory, indicates a disposition towards positivity and a preference for interpersonal interaction as a way to alleviate stress.

Analyzing such emoji usage provides profound insights into personality traits, particularly in how different MBTI types might choose to express emotions and interact. Introverts and extroverts, for instance, might consistently use different sets of emojis that align with their typical behaviors—introverts opting for more subdued, reflective emojis (like 😌 or 🌜) and extroverts choosing more expressive, outgoing emojis (like 🥳 or 💃). Moreover, this kind of analysis can also uncover how individuals with a 'thinking' versus a 'feeling' orientation might use emojis. Thinkers might prefer emojis that are less about emotional expression and more about depicting thoughts or activities (like 🧠 or 📖), while feelers are likely to use emojis that convey empathy, affection, or emotional states directly (like ♡ or 😢).

Thus, the use of emojis in digital texts is not just a casual or aesthetic choice but is deeply tied to an individual's personality and communication style. By examining patterns in emoji use, we can gain deeper insights into how people of different MBTI types express themselves and interact with others in a digital context. This understanding not only enhances interpersonal communication but also has potential applications in areas like personalized marketing, mental health assessments, and the development of AI-driven communication tools that are sensitive to individual personality differences.

6 Methodology

As we embark on detailing the methodology of our study, we adopt the OODA (Observe, Orient, Decide, Act) Loop framework, a systematic approach guiding our data collection, preparation, model training, evaluation, decision-making, and iterative refinement, ensuring a comprehensive and rigorous analysis. This process is integral in ensuring the robustness and accuracy of our analysis.

6.1 Data collection and preparation

- *Observation (Observe):* Data was collected from various sources, including social media platforms and online

forums, to observe patterns in textual expressions and emoji usage related to personality traits.
- *Orientation (Orient):* Relevant features were extracted from the observed data, such as text content and emoji sequences, to create inputs for the AI models.

6.2 Model training and evaluation

- *Observation* (*Observe*) Patterns and trends within the data were observed to inform the training of AI models for personality analysis.
- *Orientation* (*Orient*) The performance of AI models was evaluated using metrics such as accuracy, F1 scores, and Rouge scores to orient decisions regarding model selection and optimization.

6.3 Model selection and optimization

- *Decision* (*Decide*) The most suitable AI models for personality analysis were selected based on their performance metrics, computational efficiency, and interpretability.
- *Action* (*Act*) Hyperparameters of the selected models were optimized through techniques such as hyperparameter tuning to improve their performance on personality classification tasks.

6.4 Model deployment and iterative improvement

- *Decision* (*Decide*) Deployed models were monitored for performance, and feedback was collected to inform iterative improvements.
- *Action* (*Act*) Iterative refinement of models was conducted based on new data or insights gained from their usage to enhance their accuracy and applicability.

Figure 1 illustrates the OODA Loop framework, used for this study.

The following sections probe the specific details, shedding light on each phase of our methodology and explicating the systematic approach we employ to conduct our study.

7 Emo-MBTI dataset

The Emo-MBTI Dataset is a specialized resource designed to bridge the gap between digital communication, represented by emoji usage, and psychological profiling through the Myers-Briggs Type Indicator (MBTI). This dataset innovatively combines natural language processing and psychological analysis to explore the intersection of emoji usage and personality traits.

7.1 Core composition of the Emo-MBTI dataset

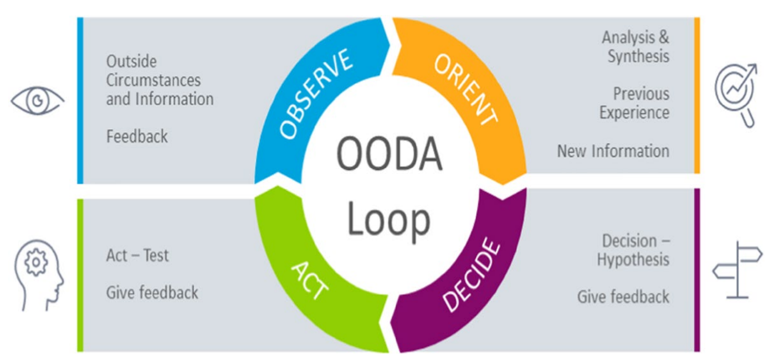
The primary foundation of the Emo-MBTI Dataset is derived from the MBTI Dataset, which includes data from over 8600 individuals. These data snippets, taken from the last 50 posts of each individual on Reddit, are categorized into one of the 16 distinct MBTI personality types. Each post is segmented using "|||" (three pipe characters) to demarcate individual entries, facilitating a structured analysis of textual content related to distinct personality profiles as shown in Fig. 2.

7.2 Data enrichment through advanced scraping

To enrich the foundational MBTI dataset, we employed advanced web scraping techniques targeting the r/mbti subreddit on Reddit. Our focus was on users who frequently use emojis in their posts, a criterion that suggests heightened emotional expression. Specifically, we collected the last 50 posts from users who utilized more than three emojis per post, using these emojis as contextual labels for the text. Using the Python Reddit API Wrapper (PRAW), we meticulously extracted data from users on the r/mbti subreddit. The initial scraping yielded 6280 data points, which were then filtered focus on users demonstrating consistent engagement:

- 6029 data points included users with more than 10 comments.
- 5527 data points were further refined to include only those with more than 50 comments, ensuring a robust representation of active community members.

Fig. 1 OODA loop framework



	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.|||That's another silly misconce...
5	INTJ	'18/37 @.@|||Science is not perfect. No scien...
6	INFJ	'No, I can't draw on my own nails (haha). Thos...
7	INTJ	'I tend to build up a collection of things on ...
8	INFJ	I'm not sure, that's a good question. The dist...
9	INTP	'https://www.youtube.com/watch?v=w8-egj0y8Qs||...

Fig. 2 Snapshot of MBTI dataset

The snapshot of this enriched dataset is shown in Fig. 3

This process not only expanded our dataset but also ensured the relevance and depth of the data collected, focusing on users with significant subreddit participation.

7.3 Integration of advanced NLP models for emoji mapping

Post-collection, the data underwent processing where a sophisticated model was trained to map text to emojis effectively. This model leverages the nuances of sequence-to-sequence generation, using generative AI models, specifically Large Language Models (LLMs), known for their capability in handling complex language tasks, For the task of mapping text to emojis, three sophisticated models were utilized: FlanT5, PEGASUS, and BART, each chosen for their unique strengths in summarization and contextual understanding.

- **FlanT5**: This transformer-based model excels at generating concise, contextually accurate summaries by deeply understanding the intricacies within extensive texts. Its proficiency in handling complex narratives makes it highly effective for detailed summarization tasks.
- **PEGASUS**: Renowned for its abstractive summarization capabilities, PEGASUS employs a unique pre-training strategy tailored specifically for summarization. This approach enables it to filter extensive text into essence-focused summaries, adeptly capturing the core messages.
- **BART**: Utilizing a denoising autoencoder architecture, BART is designed to reconstruct and simplify text, making it particularly suitable for summarizing complex content into more digestible forms, including the translation of text sentiments and themes into corresponding emojis.

These models were chosen for their ability to efficiently process and condense textual information, facilitating accurate and expressive emoji mappings that reflect the nuanced emotional and thematic undertones of the original content. The dataset after mapping is as shown in Fig. 4:

7.4 Dataset analysis and bias mitigation

We conducted a thorough statistical analysis of the emoji usage across different MBTI personality types within our dataset. This analysis revealed significant variations in emoji usage patterns. For example, we found that extroverted personality types tend to use a wider range of emojis more frequently compared to introverted types. These insights are critical as they help identify inherent dataset imbalances that could influence the model's performance.

index	label	text
0	🎭 🤐 🤬	While currently on trial for fraud ||| The Sau...
1	🚗 🚬 🎩	I love how all these analysts and fans contin...
2	🪶 📷 🤮	I immediately burst into tears ||| Look into ...
3	💩 🐀 😮	No ||| sorry ||| I had b rob too it feels b...
4	🥚 🍑 👀	Thanks man ||| I may hit you up I'm in SEPA ...
5	👯 🏞️ 💍	Rock vs Austin with Mankind as special ref |||...
6	🌮 🥤 🍟	That makes total sense ||| Lolz ||| the ide...
7	🌋 🔪 🖊️	Heidi Gardner is *FORTY* ||| Revoke Stefan Kra...
8	💓 🕺 📲	That still seems really low but I don't know w...
9	😈 🚪 🐖	Some books I have used and like are* Suomen m...

Fig. 3 MBTI posts relabelled with emojis

Unnamed: 0	type	posts	emoji
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...	🎥 📺 💻 💻 🎉 🎭 🎥
1	ENTP	'I'm finding the lack of me in these posts ver...	💑 😄 😂
2	INTP	'Good one _____ https://www.youtube.com/wat...	🙌 ❤️ 🎥 💔 ❤️ 😭
3	INTJ	'Dear INTP, I enjoyed our conversation the o...	👤 🌍 🫢 💀 💀 💀
4	ENTJ	'You're fired.|||That's another silly misconce...	😫 💥 ❓ 👭 ✈️ 🌴
...	...	...	...
8670	ISFP	'https://www.youtube.com/watch?v=t8edHB_h908||...	😺 ❤️ 💻 Dutch 📚 😄
8671	ENFP	'So...if this thread already exists someplace ...	💻 📱 💔 🎥 🍿
8672	INTP	'So many questions when i do these things. I ...	🫢 💭 💰 🐴 🐴
8673	INFP	'I am very conflicted right now when it comes ...	😭 💔 👶 📚 🖼️ 🧛
8674	INFP	'It has been too long since I have been on per...	📺 👩‍⚕️ 🌍 🌠 😮 ❤️

Fig. 4 Snapshot of the emoji labelled MBTI dataset

To address the identified biases, we implemented synthetic minority oversampling techniques (SMOTE) to improve the representation of underrepresented personality types in our dataset. This method enhances the dataset's balance by artificially generating new instances for less prevalent classes, thereby equalizing the training conditions across all personality types. Such an approach significantly diminishes the model's inclination to favour classes that are more commonly represented, ensuring a fairer and more accurate analysis across the spectrum of personality traits.

6.5. Final dataset and utility: The final Emo-MBTI Dataset, as visualized in the snapshots, features posts labelled with emojis that correlate with the inferred MBTI types based on the textual analysis provided by the models. This dataset not only facilitates the exploration of how different personality types might prefer certain types of emojis but also serves as a valuable tool for psychological research, digital behaviour analysis, and the development of AI-driven tools that can understand and predict human emotional and psychological states through digital footprints.

8 EmoMBTI-Net: the emoji-based MBTI personality prediction model

In this study, our goal is to predict MBTI personality types based on the usage of emojis in textual communications. As introduced earlier, the integration of emojis within natural language processing frameworks offers a novel approach to understanding complex human behaviours and personality traits. The rationale behind this methodology lies in the richness of emojis as expressive tools that convey emotional and contextual nuances beyond the capacity of plain text. Emojis encapsulate a spectrum of emotions and sentiments, making them valuable in the psychological profiling inherent to personality assessment.

The principal challenge we faced was selecting appropriate models equipped to process emojis as fundamental components of language. Traditional models like BERT (Lin et al. 2021), DistilBERT (Tamburini et al.), and ALBERT (Lan et al. 2019), while robust in general language tasks, do not inherently support emoji interpretation within their tokenization processes. This limitation necessitates the exploration of alternative models that natively incorporate emoji understanding to ensure the accuracy of personality predictions. Consequently, our approach has focused on identifying and leveraging transformer-based models that are inherently capable of processing and interpreting complex, emoji-rich text. These models, such as NLI-RoBERTa, BART, and NLI-DeBERTa, offer sophisticated mechanisms for embedding contextual and emotional nuances, which are critical in capturing the subtleties required for accurate personality profiling from textual communications. Embracing these advanced technologies allows us to explore the depths of emoji semantics, translating seemingly simplistic symbols into profound insights into human psychology and behaviour, thereby enhancing the predictive capabilities of our system. This shift towards models adapted for emoji interpretation not only addresses the shortcomings of conventional NLP tools but also sets a new standard in the nuanced analysis of digital communication.

8.1 NLI-RoBERTa

NLI-RoBERTa, an extension of the RoBERTa (Cui and Qi 2017) (Robustly optimized BERT approach) model, is designed for Natural Language Inference (NLI) tasks. The model employs a transformer architecture known for its efficiency in grasping contextual relationships within text. NLI-RoBERTa is pre-trained on extensive text datasets using a self-supervised learning paradigm, enabling it to capture complex semantic subtleties essential for discerning between entailment, contradiction, and neutrality in texts. Such capabilities make NLI-RoBERTa highly effective for sequence classification, including MBTI personality prediction, where understanding the layered meanings conveyed by emojis is crucial. The strength of NLI-RoBERTa lies in its transformer-based design, incorporating multiple layers of self-attention mechanisms and feedforward neural networks. This architecture is adept at learning contextual embeddings from vast amounts of data, crucial for emoji-based text interpretation. By eliminating the next sentence prediction task and adopting dynamic masking, the model enhances its focus on relevant textual segments, crucial for accurate personality assessment.

8.2 BART

BART is versatile in handling both sequence-to-sequence and sequence classification challenges. Its architecture combines a bidirectional transformer encoder with an auto-regressive decoder, forming a denoising autoencoder that excels in reconstructing noisy inputs to predict original sequences. This innovative training approach enables BART to identify and retain critical textual features, including emojis, making it suitable for detailed personality analysis. The bidirectional encoder ensures comprehensive understanding of context, capturing dependencies in text from both directions, which is vital for interpreting the emotional and contextual layers conveyed by emojis. The auto-regressive decoder enhances the model's ability to generate coherent outputs that are contextually aligned with the input, thereby supporting sophisticated sequence classification tasks like MBTI personality prediction.

8.3 NLI-DeBERTa

NLI-DeBERTa refines the DeBERTa (Hernandez and Scott 2017) model with a focus on Natural Language Inference. Its standout feature is the disentangled attention mechanism, which optimizes focus on different segments of input text independently, crucial for parsing complex emoji-laden communications. This ability to focus selectively on relevant parts of the text facilitates a deeper understanding of long-range dependencies and intricate semantic relationships, essential for accurate personality profiling. The specialized architecture of NLI-DeBERTa, with its nuanced attention to detail in context modelling, makes it an exceptional choice for tasks requiring an advanced grasp of subtle textual interactions. Its efficacy in MBTI classification is enhanced by its capacity to disentangle attention weights during pre-training, thus improving the model's overall ability to interpret and analyze emoji-based communication in the context of personality assessment.

The following Fig. 5 outlines a sophisticated model for processing and predicting personality traits based on textual input, particularly focusing on the Hindi language.

8.4 Workflow summary

The Emo-MBTI Dataset comprises posts labelled with emojis that reflect inferred MBTI types derived from textual analysis by the transformer models. The dataset is processed independently by each model: NLI-RoBERTa leverages its natural language inference capabilities to deduce personality traits from text and emojis; BART re-encodes or transforms text to emphasize emotional or stylistic features for personality prediction; and NLI-DeBERTa uses its decoding-enhanced attention mechanisms for a deeper analysis of emoji-text interaction. Each model undergoes training, tuning, and testing on separate data splits to ensure comparable performance metrics. The outputs from each model are not only the predicted MBTI types but also insights into which emojis correspond to specific personality traits. These results are thoroughly compared to evaluate accuracy, precision, recall, and F1 scores, providing a basis for understanding which model best handles the nuances of emoji usage in textual communications and why.

9 Results and discussion

This section presents the results from our comprehensive evaluation, illuminating the performance of diverse language models across both emoji mapping and MBTI personality classification tasks. As the first research initiative using this dataset, we lack baseline comparisons, making this study foundational in setting initial benchmarks for future explorations. The application of meticulous analysis using Rouge scores and F1 metrics has revealed clear patterns of strengths and weaknesses, offering valuable insights into the nuanced interpretation of textual expressions and the efficacy of different models in capturing the intricacies of human communication.

9.1 Contextual connection

As we set out to explore the predictive capabilities of large language models in interpreting emojis within digital communications, the results presented here underscore our objective to enhance personality profiling through textual

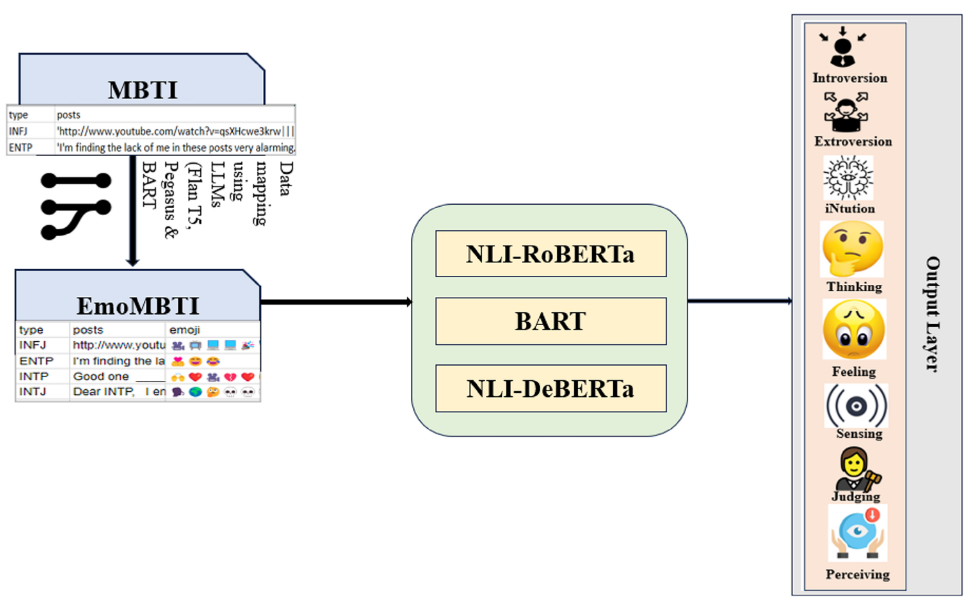


Fig. 5 The MBTIEmoNet model

and emoji analysis. These findings not only affirm our hypotheses but also pave the way for innovative applications in computational linguistics.

9.2 Performance across models

The Rouge scores, often used as a measure of similarity between the expected and the model-generated outcomes, were utilized here to gauge the efficacy of each model in emoji mapping, as shown in Table 1.

The rouge scores reveal several key points about the performance of these models:

- *Higher efficiency in learning representations* BART's exceptionally low rouge score of 0.0250 suggests a superior ability to capture and reproduce the nuances required for emoji mapping. This implies that BART's training process effectively minimized the loss, leading to more accurate emoji predictions.
- *Comparison with other models* Flan-T5, despite being a powerful and versatile model, showed a much higher final training loss in this task. This indicates that while Flan-T5 is generally effective across various tasks, it may not be as optimized as BART for the specific nuances of emoji mapping. PEGASUS, known for its summarization capabilities, performed better than Flan-T5 but still fell short compared to BART. Its middle-range score suggests it has moderate capabilities in emoji mapping, potentially due to its underlying architecture which may prioritize certain aspects of language understanding differently.
- *Choice for MBTI classification* Given BART's outstanding performance in minimizing training loss, it was the logical choice for implementing the final emoji mappings for MBTI personality classification. The low rouge score indicates a high level of precision in BART's emoji mappings, which is crucial for accurately capturing personality traits through emojis.

This graph in Fig. 6 illustrates the final training losses (Rouge Scores) for Flan-T5, PEGASUS, and BART models. BART demonstrates a significantly lower Rouge Score, indicating a superior ability to capture the nuances required for emoji mapping due to minimized loss during training.

Table 1 Emoji mapping rouge scores using LLMs

Model	Final training loss (rouge score)
Flan-T5	44.1280
PEGASUS	11.5288
BART	0.0250

Overall, BART consistently outperformed NLI-RoBERTa and NLI-DeBERTa, achieving the highest overall accuracy of 86.160% as shown in Table 2.

Notably, BART displayed notable strengths in discerning traits related to introversion, intuition, and perceiving. However, the models demonstrated varying degrees of success in understanding traits such as extroversion and sensing.

The F1 scores presented in Table 3 for the models NLI-RoBERTa, NLI-DeBERTa, and BART provide an in-depth look at their performance across various MBTI personality traits. These scores are crucial for understanding how well each model balances precision and recall in their classifications, which is particularly important in personality analysis where both false positives and false negatives can lead to significant misinterpretations.

Here's a detailed discussion of the F1 scores:

9.2.1 Overall performance

- **BART** shows the highest overall F1 score at **0.862**, indicating its superior general performance in classifying MBTI personalities.

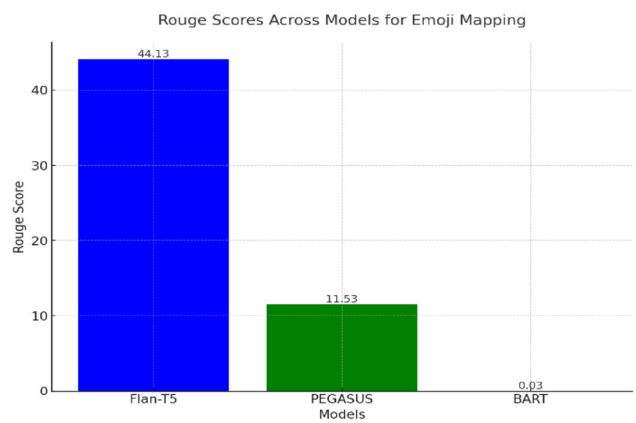


Fig. 6 Rouge scores across models for emoji mapping

Table 2 Accuracy of different models

Accuracy	NLI-RoBERTa	NLI-DeBERTa	BART
Overall	0.78571	0.80654	0.8616
Introversion	0.84824	0.85504	0.88497
Extroversion	0.63736	0.63636	0.76331
Intuition	0.92096	0.92359	0.95000
Sensing	0.14035	0.18181	0.19318
Thinking	0.73437	0.76948	0.83742
Feeling	0.84308	0.83631	0.87172
Judging	0.73818	0.76351	0.81111
Perceiving	0.82897	0.83870	0.88356

Table 3 F1 scores of different models

F1	NLI-RoBERTa	NLI-DeBERTa	BART
Overall	0.786	0.807	0.862
Introversion	0.848	0.855	0.885
Extroversion	0.637	0.636	0.763
Intuition	0.921	0.924	0.950
Sensing	0.140	0.182	0.193
Thinking	0.734	0.769	0.837
Feeling	0.843	0.836	0.872
Judging	0.738	0.764	0.811
Perceiving	0.829	0.839	0.884

- **NLI-DeBERTa** follows with **0.807**, and **NLI-Roberta** has an F1 of **0.786**. This suggests that while all models are relatively effective, BART excels in balancing precision and recall most effectively.

The graph in Fig. 7 showcases the accuracy and F1 scores for MBTI personality classification using EmoMBTI-Net, focusing on how well each model performs in this specific context.

Figure 8 presents the Radar Chart for model performance on MBTI Traits. Each spoke of the radar chart represents a different personality trait, and each model's performance on these traits is plotted as a separate line on the chart. BART (in red) shows superior performance in most traits compared to NLI-RoBERTa (in blue) and NLI-DeBERTa (in green), particularly noticeable in Extroversion and Intuition.

9.2.2 Trait-specific performance

- **Introversion and extroversion**:
  - BART outperforms in both traits, with particularly strong gains in Extroversion (**0.763**), which is notably higher than the scores for NLI-RoBERTa (**0.637**) and NLI-DeBERTa (**0.636**). This indicates BART's better capability in detecting features related to extroverted behaviour through text.
- **Intuition**:
  - All models perform well with Intuition, with BART leading at **0.950**. This suggests that the models are effective at picking up on the abstract and theoretical content often associated with intuitive personalities.
- **Sensing**:
  - All models struggle with Sensing, with F1 scores remaining below **0.200**. This is likely due to the difficulty of capturing concrete and detail-oriented expressions through text, which are less prevalent or explicitly marked compared to intuitive expressions.
- **Thinking and feeling**:
  - BART again leads in these traits, with Thinking at **0.837** and Feeling at **0.872**. These traits involve logical versus empathetic content, indicating BART's stronger ability to distinguish between structured argumentation and emotional expressions.
- **Judging and perceiving**:
  - BART performs best in both Judging (**0.811**) and Perceiving (**0.884**), showing its effectiveness in identifying structured versus spontaneous expressions in text.

The results indicate BART's robustness in nuanced text interpretation, making it particularly suited for tasks

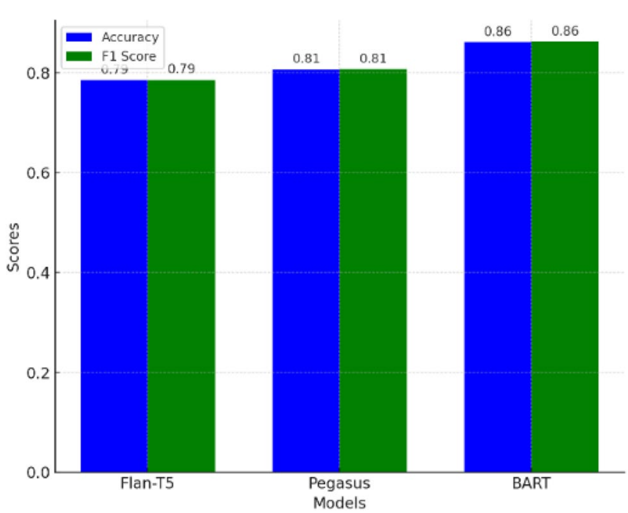


Fig. 7 Overall accuracy and F1 scores across models in EmoMBTI-Net

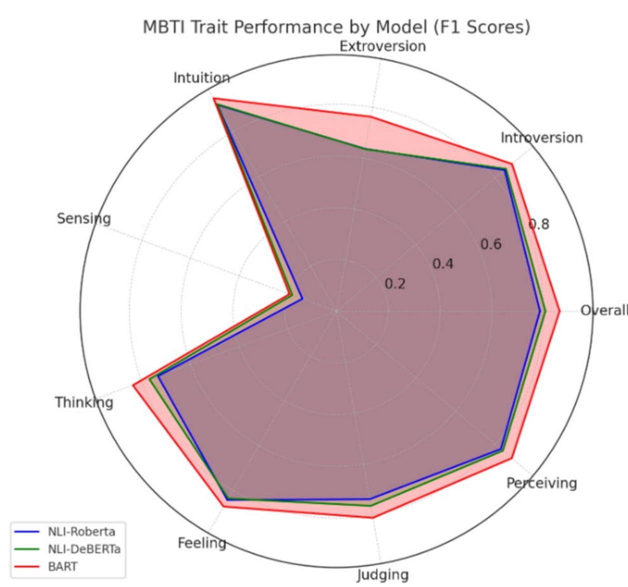


Fig. 8 Radar chart for model performance on MBTI traits

requiring deep semantic understanding, such as personality assessment through text analysis. The superior F1 scores for BART across most traits suggest that its training and underlying model architecture may be more attuned to the subtleties of personality-expressive language.

The lower scores in Sensing across all models highlight a common challenge in natural language processing—capturing concrete and sensory-related information, which tends to be less explicitly expressed. This suggests a potential area for further model refinement and training, possibly by incorporating more sensory-specific training data or employing techniques that better capture these aspects.

In conclusion, the F1 scores reveal the strengths and weaknesses of each model in handling different aspects of personality classification. BART's leading performance across multiple traits recommends it as the preferable model for applications in personality analysis, offering more reliable and nuanced insights into individual personality types based on their textual expressions.

The choice of pre-trained language model and hyperparameters significantly influenced the performance of both emoji mapping and MBTI classification tasks. For emoji mapping, the hyperparameters such as learning rate and epochs were tuned for each LLM, optimizing their performance. Similarly, in MBTI classification, different learning rates and epochs were applied to NLI-RoBERTa, NLI-DeBERTa, and BART as shown in Table 4. An iterative tuning process was crucial in selecting appropriate learning rates and epochs, which were dynamically adjusted based on model performance during validation to prevent overfitting while maximizing accuracy.

Overall, the study highlights the effectiveness of large language models in leveraging emojis for personality detection tasks, with BART demonstrating superior performance in both emoji mapping and MBTI classification. The findings underscore the importance of model selection and hyperparameter tuning in optimizing the performance of such tasks and provide valuable insights into the nuances of interpreting personality traits conveyed through emojis.

Table 4 Hyperparameters used

Model	Learning rate	Epochs
Emoji mapping		
Flan-T5	5e−5	3
PEGASUS	3e−4	3
BART	1e−5	3
MBTI classification		
NLI-RoBERTa	3e−5	10
NLI-DeBERTa	3e−6	10
BART	1e−5	10

9.3 Limitations and ethical considerations

While our study provides valuable insights into the performance of language models in emoji mapping and MBTI personality classification tasks, several limitations must be acknowledged.

9.3.1 Limitations

- *Dataset scope and diversity* The size and diversity of the dataset used for training and evaluation may affect the generalizability of our findings. This limitation underscores the need for larger, multilingual, and culturally diverse datasets that can account for variations in emoji usage and linguistic expressions across demographic and cultural groups.
- *Contextual nuances* Emojis can carry context-specific meanings that vary based on individual user intent, cultural norms, or platform-specific interpretations. Our current methodology may not fully capture these nuances, which could impact the accuracy of personality inferences. Future work could explore incorporating context-aware models to address this challenge.
- *Domain-specific biases* Training data often reflect biases present in user-generated content, potentially skewing model predictions. Domain-specific biases in personality traits associated with emojis or MBTI categories could hinder general applicability. Techniques such as bias quantification, fairness-aware learning, and targeted dataset augmentation should be further explored.
- *Scalability and computational constraints* Training large-scale language models demands significant computational resources, which can limit scalability. Optimizing model architectures and adopting efficient training techniques (e.g., knowledge distillation or low-rank adaptations) could help reduce resource demands without sacrificing performance.
- *Evaluation metrics* While we utilized Rouge scores and F1 metrics, these may not fully capture the intricacies of personality classification. Future research could explore alternative evaluation metrics such as interpretability measures, psychological validity checks, or human-in-the-loop validation to offer a holistic assessment of model performance.

9.3.2 Ethical considerations

- *Privacy and consent* The collection and analysis of sensitive textual data raise privacy concerns. Strict adherence to ethical data collection practices, including obtaining explicit user consent, anonymizing datasets, and ensuring compliance with data protection laws (e.g., GDPR), is imperative.

- *Algorithmic bias and fairness* Bias in training data or algorithmic processes could lead to skewed or discriminatory outcomes. For example, biased personality predictions in hiring or mental health applications might disproportionately affect certain groups. Conducting regular bias audits and implementing debiasing strategies are essential to mitigate such risks.
- *Transparency and accountability* The use of opaque AI systems for personality analysis may erode trust among stakeholders. Transparent model architectures, interpretability tools, and user-facing explanations are critical to ensuring accountability in high-stakes applications. Mechanisms for recourse, such as allowing users to contest or correct AI-generated personality assessments, should also be incorporated.
- *Misuse and overgeneralization* Personality analysis, when misapplied, may lead to overgeneralized conclusions about individuals, influencing decisions inappropriately in areas like hiring, marketing, or psychological evaluations. Clear boundaries for ethical use and robust regulatory oversight are necessary to prevent misuse.
- *Psychological impact* Providing users with AI-generated personality insights without appropriate context or explanation could lead to unintended psychological effects, such as anxiety or self-stereotyping. Involving domain experts like psychologists in model validation and developing user-centric feedback mechanisms could alleviate these concerns.
- *Access and Inclusivity* Access to the benefits of such AI systems should be equitable. Researchers and policymakers must ensure that the development and deployment of personality analysis systems do not exacerbate existing digital divides or marginalize underrepresented populations.

By addressing these limitations and ethical concerns, future research can refine the methodologies for personality classification while ensuring that such advancements contribute positively to society. Adhering to best practices in AI ethics, transparency, and inclusivity will help maximize the societal benefits of this innovative approach to personality profiling.

9.4 Practical implications

The findings of our study have significant practical implications for real-world applications, particularly in the development of personalized recommendation systems, chatbots, and mental health assessment tools. By leveraging insights gained from analyzing personality traits through text, developers can enhance user experiences, tailor content recommendations, and provide more personalized interactions. For personalized recommendation systems, understanding users' personality traits can facilitate the delivery of content and product recommendations tailored to their individual preferences and characteristics. By integrating personality-based profiling into recommendation algorithms, platforms can enhance user engagement, satisfaction, and retention rates.

Similarly, in the context of chatbots and virtual assistants, incorporating personality-aware dialogue generation techniques can enhance the conversational experience and foster more meaningful interactions. Chatbots capable of adapting their language and tone to match users' personality traits can improve communication effectiveness and user satisfaction. Furthermore, in the field of mental health assessment and therapy, analyzing personality traits through text can provide valuable insights for psychological research and clinical practice. By identifying linguistic markers associated with various personality dimensions, clinicians can gain a deeper understanding of patients' psychological profiles and tailor interventions accordingly.

10 Conclusion and future work

The study offers a comprehensive comparison of DeBERTa and BART models in the emerging field of emoji-based MBTI personality trait classification, introducing a novel approach that integrates emojis as an expressive and psychological dimension of personality profiling. Among the tested models, BART consistently demonstrates superior accuracy, achieving F1 scores of 0.885 for introversion, 0.950 for intuition, and 0.884 for perceiving. These results highlight BART's ability to effectively capture nuanced and abstract patterns within textual and emoji-based data, positioning it as the most reliable model for this task. However, challenges remain in classifying traits like extroversion and sensing, where lower F1 scores across all models indicate difficulty in recognizing concrete and socially explicit cues. In these areas, NLI-DeBERTa outperforms NLI-RoBERTa slightly, showcasing the value of disentangled attention mechanisms in handling complex interactions between text and emojis.

The study also introduces the innovative Emo-MBTI dataset, which integrates emoji usage with textual content from Reddit, mapped to MBTI personality traits. This dataset forms the foundation for the analysis and highlights the potential of emojis as psychological indicators that enrich digital communication and personality profiling. The findings emphasize the critical need for tailoring model selection based on specific strengths, optimizing their performance for distinct personality dimensions. While the integration of emojis into personality analysis offers a fresh perspective on understanding human behavior in digital communication, several challenges require further attention. These include addressing dataset diversity, mitigating domain-specific biases, and resolving the inherent ambiguity of emoji interpretation.

Moreover, the study underscores the ethical implications of this research, emphasizing the importance of safeguarding user data privacy, addressing biases in model predictions, and ensuring transparency in deploying such systems.

Future research will prioritize several key directions to build on these findings. First, systematic ablation studies will be conducted to dissect the influence of specific components of the models, such as attention mechanisms and contextual embeddings, on their performance. This granular analysis will help refine model configurations, reduce overfitting, and better understand the interplay of features that drive accurate predictions. Second, expanding the dataset through synthetic augmentation and the inclusion of multilingual and culturally diverse data will test and improve the generalizability of the models. This approach aims to address biases and enhance the robustness of personality predictions across varied linguistic and demographic contexts. Furthermore, exploring multimodal learning techniques will form an integral part of future work. By integrating textual data with emojis, and potentially audio or visual cues, researchers can create a more holistic representation of communication dynamics. This integration could significantly enhance the accuracy of personality predictions, providing a richer, multi-faceted dataset for model training. Temporal analysis of emoji usage patterns will also be explored to understand how personality traits may manifest and evolve over time, offering insights into dynamic aspects of personality profiling.

Lastly, collaborating with experts in psychology to validate the models' predictions against established psychological assessments will be critical. Such interdisciplinary validation will not only ensure the scientific credibility of the findings but also open avenues for practical applications in mental health, education, marketing, and personalized user experience design. These advancements aim to transform the emoji-based approach from a promising concept into a widely applicable tool for understanding and predicting human behaviour in the digital age.

Author Contribution AK conceptualized the study, designed the methodology, led the data curation process, and wrote the original draft of the manuscript. AK also supervised the project and handled the project administration. DJ was responsible for the development and implementation of the computational models, conducted the data analysis, and prepared all figures and tables for the manuscript. DJ also contributed to the literature review and assisted in writing the manuscript. Both authors AK and DJ contributed to the final version of the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

Artha Agastya IMA, Dwi Handayani DO, Mantoro T (2019) A systematic literature review of deep learning algorithms for personality trait recognition. In: 2019 5th international conference on computing engineering and design (ICCED), pp 1–6, https://doi.org/10.1109/ICCED46541.2019.9161107

Aslan S (2019) Multimodal video-based personality recognition using Long Short-Term Memory and convolutional neural networks. arXiv: Computer vision and pattern recognition

Cui B, Qi C (2017) Survey analysis of machine learning methods for natural language processing for MBTI personality type prediction. Final Report Stanford University

Gucluturk Y, Guclu U, Perez M, Jair Escalante H, Baro X, Guyon I, Andujar C, Van Lier R (2017) Visualizing apparent personality analysis with deep residual networks. In: Proceedings of the IEEE international conference on computer vision workshops, pp 3101–3109

Gürpinar F, Kaya H, Salah AA (2016) Multimodal fusion of audio, scene, and face features for first impression estimation. In 2016 23rd international conference on pattern recognition (ICPR), 2016 Dec. IEEE, pp 43–48

Gürpınar F, Kaya H, Salah AA (2016) Combining deep facial and ambient features for first impression estimation. In: Computer vision—ECCV 2016 workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, proceedings, Part III 14. Springer, Berlin, pp 372–385

He P, Liu X, Gao J, Chen W (2020). Deberta: decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654

Hernandez RK, Scott I (2017) Predicting Myers–Briggs type indicator with text. In: 31st Conference on neural information processing systems (NIPS 2017)

Hernández Y, Peña CA, Martínez A (2018) Model for personality detection based on text analysis. In: Batyrshin I, Martínez-Villaseñor M, Ponce Espinosa H (eds) Advances in computational intelligence. MICAI 2018. Lecture notes in computer science, vol 11289. Springer, Cham. https://doi.org/10.1007/978-3-030-04497-8_17

Ismail S, Babak B, Ismail S (2017) Significant of MBTI personality model on decision making in university program selection. In: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 62–67. IEEE

Jain D, Kumar A, Beniwal R (2021) Personality BERT: a transformer-based model for personality detection from textual data. In: Bashir AK, Fortino G, Khanna A, Gupta D (eds) Proceedings of international conference on computing and communication networks.

Lecture notes in networks and systems, vol 394. Springer, Singapore. https://doi.org/10.1007/978-981-19-0604-6_48
Junior JCJ, Güçlütürk Y, Pérez M, Güçlü U, Andujar C, Baró X, Escalera S (2019) First impressions: a survey on vision-based apparent personality trait analysis. IEEE Trans Affect Comput 13(1):75–95
Kaya H, Gurpinar F, Ali Salah A (2017). Multi-modal score fusion and decision trees for explainable automatic job candidate screening from video CVS. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1–9
Keh SS, Cheng I (2019) Myers–Briggs personality classification and personality specific language generation using pre-trained language models. arXiv:1907.06333
Kennison SM, Fritz K, Hurtado Morales MA, Chan-Tin E (2024) Emoji use in social media posts: relationships with personality traits and word usage. Front Psychol 15:1343022
Kosan MA, Karacan H, Urgen BA (2022) Predicting personality traits with semantic structures and LSTM-based neural networks. Alexandria Eng J 61(10):8007–8025
Kumar A, Beniwal R, Jain D (2023) Personality detection using kernel-based ensemble model for leveraging social psychology in online networks. ACM Trans Asian Low-Resour Lang Inf Process ACM (TALLIP) 22:1–20. https://doi.org/10.1145/3571584
Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R (2019) Albert: a lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2019) Bart: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461
Liao W, Zeng B, Yin X, Wei P (2021) An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Appl Intell 51:3522–3533
Liao R, Song S, Gunes H (2024) An open-source benchmark of deep learning models for audio-visual apparent and self-reported personality recognition. IEEE Trans Affect Comput 15:1590–1607. https://doi.org/10.1109/TAFFC.2024.3363710
Lin J, Liu Y, Zeng Q, Jiang M, Cleland-Huang J (2021) Traceability transformed: Generating more accurate links with pre-trained Bert models. In: 2021 IEEE/ACM 43rd international conference on software engineering (ICSE). IEEE, 2021, May, pp 324–335
Liusie A, Manakul P, Gales M (2024) LLM comparative assessment: zero-shot NLG evaluation through pairwise comparisons using large language models. In: Proceedings of the 18th conference of the European chapter of the association for computational linguistics, 2024 Mar, volume 1: long papers, pp 139–151
Rai N (2016) Bi-modal regression for apparent personality trait recognition. In: 2016 23rd international conference on pattern recognition (ICPR). https://doi.org/10.1109/icpr.2016.7899607
Ren Z, Shen Q, Diao X, Xu H (2021) A sentiment-aware deep learning approach for personality detection from text. Inf Process Manag 58(3):102532
Saeidi S (2024) Identifying personality traits of WhatsApp users based on frequently used emojis using deep learning. Multimedia Tools Appl 83(5):13873–13886
Tamburini F, Cimiano P, Preite S (XXXX) Deep question answering: a new teacher for DistilBERT
Zhang CL, Zhang H, Wei XS, Wu J (2016) Deep bimodal regression for apparent personality analysis. In: European conference on computer vision, Oct 2016. Springer, Cham, pp 311–324