An Intelligent Fusion-based Behavioral Trait Prediction for Autistic Spectrum Disorder with Artificial Intelligence

Monalin Pal*[1], Rubini P. [2]
[1] CMR University (CMRU), Bangalore, India
[2] CMR University (CMRU), Bangalore, India
Emails: monalin.19cphd@cmr.edu.in; rubini.p@cmr.edu.in

Abstract

Autism spectrum disorder (ASD) is a neurological and developmental condition impacting individuals' interactions with others, communication, learning, and behavior. While autism can be identified at any point in life, it is characterized as a "developmental disorder" due to the typical onset of symptoms within the initial two years of life. As individuals with ASD transition from childhood to adolescence and young adulthood, they might face challenges in establishing and having friendships, communicating with both peers and adults, and understanding the expected behaviors in education or work. The current study introduces a novel approach for suggesting the right behavioral strategy to assist Autistic Spectrum Disorder with the help of supervised BERT (Bidirectional Encoder Representations from Transformers). Our model achieved an accuracy of 88% with the help of BERT to predict the right behavioral trait. This research demonstrates cost-effectiveness and efficiency in offering recommendations for ASD, making it suitable for applications requiring near real-time outcomes.

Keywords: Deep Learning; Machine Learning; Autism Spectrum Disorder; BERT (Bidirectional Encoder Representations from Transformers); Fusion Processing; Information Fusion; Neural networks; Social Media; Applied Behavioral Analysis

1. Introduction

The prevalence of Autism stands at 1 in 100 children, with a noticeable increase in recent times as per WHO [1]. Autism spectrum disorder (ASD) is a developmental disability arising from brain differences, with some individuals having identifiable causes like genetic conditions while others have unknown origins. People with ASD may exhibit unique behaviors, communication styles, social interactions, and learning approaches, setting them apart from the majority. Physical appearance usually doesn't distinguish them from others. The range of abilities in individuals with ASD is diverse, with some having advanced conversational skills and others being nonverbal. Support needs also vary, from those requiring substantial assistance in daily life to those who can function independently. ASD typically starts before the age of 3 and can persist throughout life, although symptoms may increase over time. As individuals with ASD transition to adolescence and young adulthood, challenges may arise in forming and maintaining friendships, communicating with peers and adults, and grasping expected behaviors in educational or work settings. Additional conditions like anxiety, depression, or attention-deficit/hyperactivity disorder may also be present, occurring more frequently in individuals with ASD compared to those without ASD.

Topic models provide a statistical framework that effectively addresses the issue, especially with the availability of digital archives containing scholarly literature. This framework aligns closely with the natural writing process, where authors typically select a few topics and proceed to write about them. In the probabilistic framework of topic modeling, a probability distribution over words is linked to each topic, and a probability distribution over topics is associated with each document. The document creation process involves iteratively selecting a topic from the document-topic distribution and drawing a word from the topic-word distribution. In various domains, including digital humanities, linguistics, and cognition, topic models have been employed for this specific purpose and have demonstrated their applicability [2], [3], [4].

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

2. Related Work

LDA operates as a parametric model, defined by specifying the number of topics. Determining the optimal set of topics that best describes the data involves fitting multiple models and subsequently engaging in model selection, which is computationally challenging. The Hierarchical Dirichlet Process Mixture Model (HDPMM) [5] tackles this limitation by using a non-parametric prior on the parameter space of the mixture components. This approach allows the data to determine the model's complexity, automatically inferring the appropriate number of topics. In research [6], the authors proposed a study titled "Modeling social support in the autism community on social media". The information model theory presented in this research is applied to evaluate social support based on theoretical frameworks from social science, linguistics, and social network theory. The research showed that the autism blogger community offers members significant social support through blogs and Twitter.

Given potential misunderstandings about Applied Behavior Analysis (ABA), examining the content shared online and the sentiment expressed is crucial. As individuals with Autism Spectrum Disorder (ASD) and their family members engage in online communities [7], researching online content becomes evident to understand the experiences and perspectives of adults with ASD and other members of neurodiverse communities [8, 9]. This study yields valuable insights into factors influencing decision-making regarding ABA-based interventions. Consequently, it could enhance the ability of healthcare professionals to guide families in making informed choices, with a comprehensive understanding of the advantages and limitations of available options (10). This study marks an initial step toward understanding the landscape surrounding ABA-based interventions for children with ASD within social media platforms.

Several analyses have been conducted to explore content and linguistic attributes within social media platforms, specifically Reddit. Techniques such as topic modeling and Linguistic Inquiry and Word Count (LIWC) have been used for large corpora of data. Topic modeling utilizes text-mining algorithms to understand patterns within the data [11], examining how words cluster together in their usage. LIWC, on the other hand, is an automatic text analysis algorithm that calculates the percentage of words in a text matching various emotional, cognitive, structural, and process dimensions, using a predefined dictionary [12]. In the study [13], authors have used cluster analysis to understand conversational involvement, emotion, and informational support in the subreddit r/Aspergers, revealing it to be a supportive community. A study [6] explored the social support provided by popular ASD bloggers active on blogs and Twitter through LIWC analysis, indicating significant social support within the ASD community on both platforms. Additionally, research proposed in [14] focused on patterns and themes of ASD-related tweet content on Twitter.

Authors in [16] introduced a novel learning approach for SBERT models, incorporating keyword information to enhance semantic vector expression. They devised a method to generate training data containing both positive and negative keywords using sentence n-grams. Subsequently, they utilized this data to train an SBERT model, resulting in the creation of keyword pairs. Their study demonstrated that an SBERT model trained on keyword-inclusive data exhibited a performance improvement of 2.74% compared to the conventional SBERT model. In another study [17], authors investigated sentiment classification using five-word embedding techniques and three machine-learning classification models. They used CounterVectorizer, TfidfVectorizer, Word2vec (CBOW), Word2vec (Skip-gram), and Pretrain_Word2vec for word embeddings, coupled with Decision Tree, RandomForest, and Logistic Regression models for classification. Their research revealed that CounterVectorizer outperformed TfidfVectorizer in sentiment classification tasks. Additionally, they conducted an accuracy analysis based on the number of emotion categories, finding that using two emotion categories (positive and negative) yielded over a 30% accuracy improvement compared to using seven emotion categories. Furthermore, they determined that subdivided emotion categories were unsuitable for emotion classification using embedding techniques.

In the current study, we focus on creating a classification model with the help of BERT to infer the right behavioral trait. 10 behavioral trait classes are identified for the current study. The emphasis of the study lies in assessing the reliability and reproducibility of outcomes, with potential applications in real-time ASD therapy. Our model outperforms the previous research with an accuracy of 88% to detect the right behavioral trait.

3. Data Analysis

The present study utilizes the Text-based Early Autism Spectrum Disorder Detection Dataset for Toddlers (TASD) dataset [15], The TASD-Dataset, designed for the early detection of Autism Spectrum Disorder (ASD) in toddlers, encompasses text sequences describing the situations of both ASD and non-ASD toddlers. It incorporates crucial ASD assessment features, including Attention Response, Change Reaction, Word Repetition, Eye Contact,

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

Emotional Empathy, Finger Movements, Focused Attention, Follow Pointing, Repetitive Behavior, and Toy Arranging. Each feature is linked to specific toddler behaviors, and these discussions provide comprehensive parental observations, offering insights into the understanding and expression of these behaviors. The dataset facilitates the development of machine learning models, allowing a thorough exploration of behavioral indicators essential for identifying ASD risk in early childhood. Each behavioral trait is explained in detail in Table 1.

Table 1: TASD Dataset Behavioral Traits and Features

Behavioral Trait	Features
Attention Response	This pertains to a child's reaction or response to external stimuli, such as sounds, movements, or instructions. In ASD assessment, variations in attention response may signal differences in sensory processing or responsiveness
Change Reaction	Examines how a child adapts to changes in their environment or routine. Children with ASD may exhibit resistance or difficulty adjusting to changes, resulting in visible reactions
Word Repetition	Involves repeating words or phrases, often without contextual relevance. This linguistic pattern is observed in children with ASD, showcasing repetitive speech or echolalia
Eye Contact	Indicates the ability or frequency with which a child makes eye contact during social interactions. Limited or atypical eye contact is common in ASD, affecting social communication
Emotional Empathy	Focuses on a child's capacity to understand and respond to others' emotions. Challenges in emotional empathy might manifest as difficulties in recognizing or appropriately responding to others' feelings
Finger Movements	Observe any repetitive or stereotyped finger movements or hand gestures. These movements might manifest as repetitive actions or mannerisms, often seen in children with ASD
Focused Attention	Refers to a child's ability to concentrate on a specific task or activity for an extended period. Children with ASD might face difficulty maintaining focused attention on tasks of interest or importance
Follow Pointing	Assesses a child's ability to follow another person's pointing gesture to focus on or locate an object or event. Difficulty in following pointing gestures might indicate challenges in social communication
Repetitive Behavior	Encompasses a range of repetitive actions, behaviors, or interests that are persistent and often resist change. These behaviors might include repetitive movements, insistence on sameness, or rigid adherence to routines
Toy Arranging	Involves a child's inclination or patterns in arranging toys or objects in a specific or repetitive manner. This behavior might be related to a preference for order or specific patterns often observed in ASD

4. Proposed Methodology

With the help of TASD Dataset, we built a text classification model using BERT. Data Preprocessing is performed on the dataset with natural language processing techniques like stop word removal, stemming, and lemmatization. The dataset has 297 entries, out of which 163 are ASD and 134 are Non ASD. We consider the text data of ASD and split it into 80% train and 20% test for the BERT Classification Model to predict the Behavioral Trait.

A. Bidirectional Encoder Representations from Transformers (BERT)

BERT is an advanced deep learning architecture designed for natural language processing tasks, helping contextual understand text or sentences. Trained on extensive datasets like Wikipedia, BERT can be fine-tuned to enhance its accuracy for specific tasks. BERT leverages transformer-based architecture, wherein every output is intricately linked to each input, with weights dynamically calculated based on input-output connections. Featuring a Masked Language Model (MLM), BERT conceals a word within a sentence during training and predicts the hidden word based on contextual cues. The self-attention mechanism is central to BERT's functionality, achievable through bidirectional transformers. The encoder and decoder are integrated within the transformer as a sequence-to-sequence model. The mathematical formulation for BERT is as in equation (1).

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

$$Attention(A,B,C) = softmax\left(\frac{ArB^T}{\sqrt{d_b}}\right)C \quad (1)$$

A, B, and C denote embedding vectors transformed by a weight matrix within the transformer. Once the weight matrices are learned, the transformer model changes into a language model. The BERT model operates based on a predefined set of protocols for representing input text. The input text is segmented into embedding components, which consist of three embeddings. As illustrated in Fig. 1, token embeddings are generated through the tokenization process of the input text. Segment embedding is essential for distinguishing between two distinct sentences, such as "My dog is cute; he likes playing," allowing the model to comprehend the context better. Additionally, BERT employs positional embedding to separate the positional placement of each word within a sentence. The input representation within the BERT model is derived from the summation of these three embeddings.

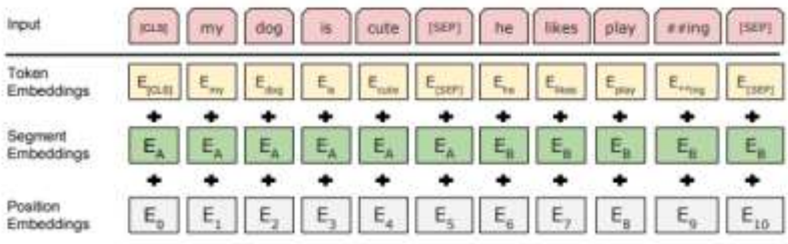

Figure 1: Text Embeddings in BERT Model

B. Architecture

Figure 2 from the Transformer architecture illustrates two main components: the Encoder and the Decoder. Once the input text is transformed into input embeddings, the Encoder commences processing the input embeddings. Within the Encoder block are two layers: the Multi-Head Attention layer and the Feed Forward Neural Network layer. After encoding the data, the Encoder block forwards it to the Decoder block, where the Multi-Head Attention layer and the Feed Forward Neural Network layer are connected, along with an additional layer known as Masked Multi-Head Attention, which handles various masks. To get the behavioral strategy for Autism Spectrum, the autism-related dataset is utilized in conjunction with the BERT model. The proposed system leveraging the BERT model extracts data from the prepared dataset and performs NLP tasks such as tokenization and embedding. The embedded data and labeled data are split into training and testing datasets for the BERT model. The training data is utilized for model training, while the testing data assesses the model's performance.

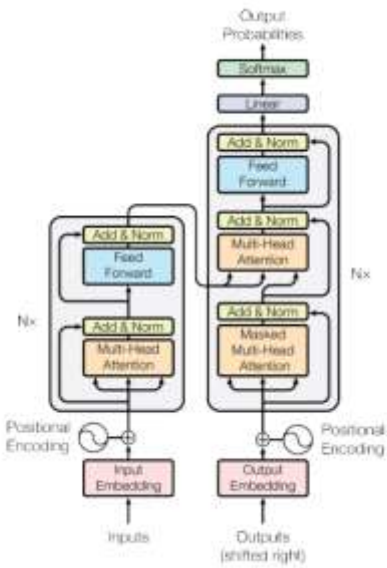

Figure 2: Transformer Architecture

The end-to-end pipeline is shown in Figure 3, where the dataset is pre-processed with Natural Language Processing techniques such as stop-word removal, stemming, and Lemmatization, which is fed to the BERT Model for the classification.

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

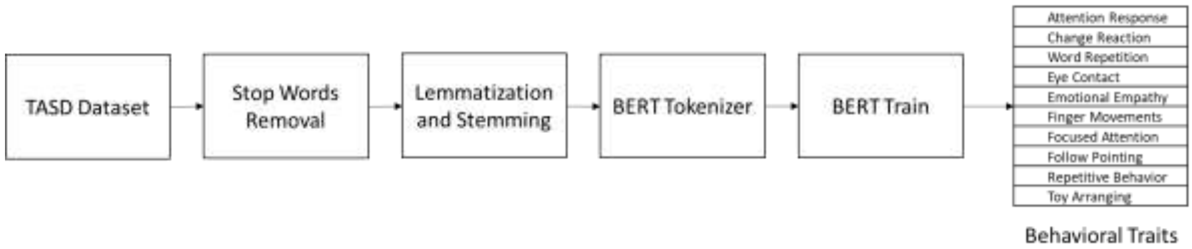

Figure 3: Model Pipeline

C. Pre-training BERT

For Masked LM, a deep bidirectional model would outperform unidirectional (left-to-right) models and basic combinations of left-to-right and right-to-left models. However, traditional language models are limited to being trained in either a left-to-right or right-to-left manner. This limitation arises because allowing bidirectional training would enable each word to "see" itself in the training process. The term "Transformer encoder" is often used to describe the bidirectional model, while the model that only considers the left context is known as a "Transformer decoder." A technique known as "masked LM" (MLM) is helpful to train deep bidirectional models. This involves randomly masking a certain percentage of the input tokens and then predicting these masked tokens. Specifically, we randomly hide 15% of the WordPiece tokens in each input sequence.

Unlike denoising autoencoders, which aim to reconstruct the entire input, the focus is solely on predicting the masked words. This approach enables the development of a pre-trained model that understands context from both directions.

The concept of Next Sentence Prediction (NSP) is helpful for numerous downstream tasks like Question Answering (QA) and Natural Language Inference (NLI). These tasks require understanding the relationship between two sentences, which is not inherently captured by conventional language modeling. Pre-training on a binarized NSP task is performed to address this gap, which can be easily constructed from any monolingual corpus. In this pre-training setup, when selecting pairs of sentences, A and B for each training instance, we ensure that 50% of the time, B is the actual subsequent sentence following A (labeled as IsNext), while the remaining 50% of the time, it is a random sentence drawn from the corpus (labeled as NotNext). This task predicts whether sentence B follows sentence A (NSP). Pre-training is shown in Figure 4.

D. Fine-tuning BERT

Fine-tuning is a self-attention mechanism within the Transformer architecture, enabling BERT to handle a wide range of downstream tasks effectively. Whether these tasks involve single text inputs or pairs of texts, BERT can adapt by simply substituting the appropriate information and outputs. The conventional approach typically involves encoding each text independently before applying bidirectional cross-attention for tasks involving text pairs. However, BERT takes a different approach by utilizing the self-attention mechanism to integrate these stages. By concatenating the text pair and employing self-attention, BERT inherently incorporates bidirectional cross-attention between the two sentences. In the fine-tuning phase, task-specific inputs and outputs into BERT and train all parameters end-to-end. For input, the sentences A and B from pre-training serve as (1) sentence pairs in paraphrasing, (2) premise-hypothesis pairs in entailment, (3) passage-question pairs in question answering, and (4) a text-∅ pair in text classification or sequence tagging.

Regarding output, token representations are directed into an output layer for tasks at the token level, such as sequence tagging or question answering [CLS] representation is directed into an output layer for classification tasks, such as entailment or sentiment analysis. Fine-Tuning is shown in Figure 4.

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

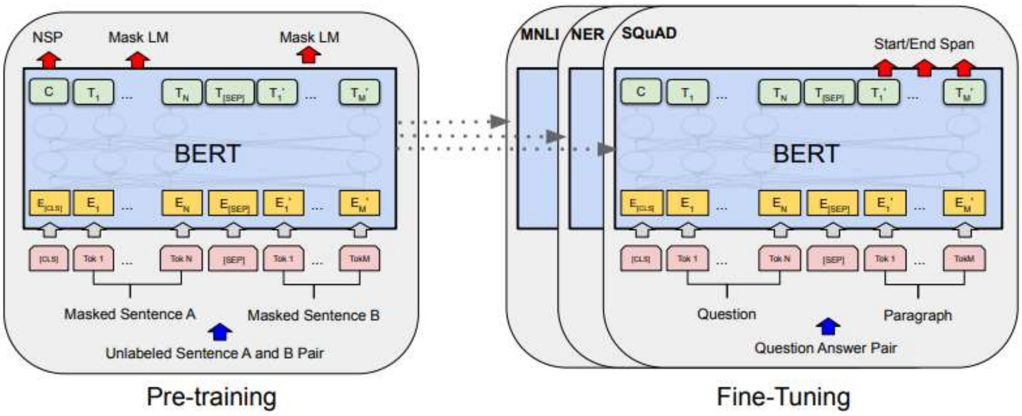

Figure 4: Pre-training and Fine-Tuning for BERT

5. Results and Discussions

Experiments are conducted using Google Colab due to its GPU requirement. Data preprocessing is performed on the TASD dataset with NLP Techniques such as stop words removal, lemmatization and stemming. The performance of the BERT model has been assessed using Accuracy and Loss. In this study, the BERT model serves as the multi-class classifier tasked with identifying 10 types of Behavioral Traits.

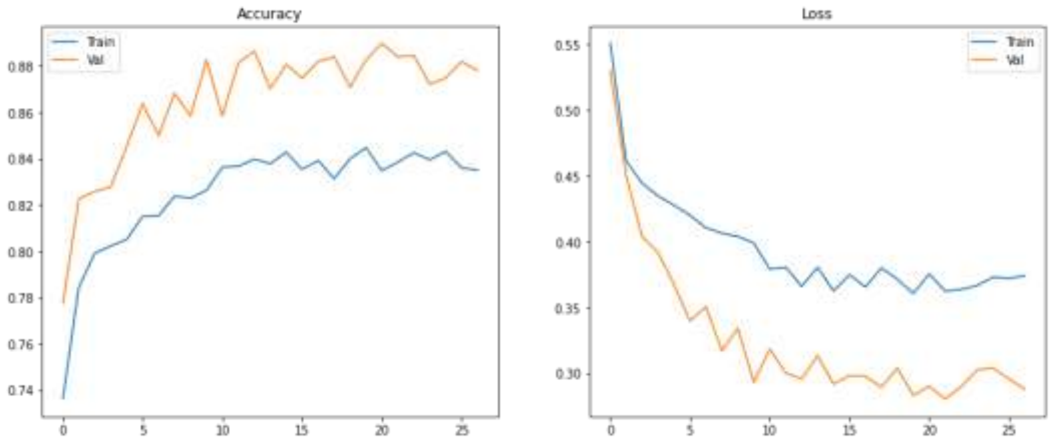

Figure 5: Accuracy and Loss

Accuracy and Loss are shown in Figure 5, and the model summary is shown in Figure 6; we have used a "bert-base-uncased" tokenizer with a sequence length of 55, and the model is trained with Batch Normalization and Relu Activation function. The model achieves an accuracy of 88%, surpassing the outcomes of prior studies and demonstrating enhanced robustness in terms of accuracy. The model's performance is notably stable, making it suitable for real-time applications in ASD Behavioral Traits prediction.

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

```
Layer (type)                Output Shape                 Param #    Connected to
==================================================================================================
input_ids (InputLayer)      [(None, 55)]                 0          []

attention_mask (InputLayer  [(None, 55)]                 0          []
)

tf_bert_model (TFBertModel  TFBaseModelOutputWithPooli   1094822    ['input_ids[0][0]',
)                           ngAndCrossAttentions(last_   40          'attention_mask[0][0]']
                            hidden_state=(None, 55, 76
                            8),
                             pooler_output=(None, 768)
                            , past_key_values=None, hi
                            dden_states=None, attentio
                            ns=None, cross_attentions=
                            None)

tf.__operators__.getitem (  (None, 768)                  0          ['tf_bert_model[0][0]']
SlicingOpLambda)

batch_normalization (Batch  (None, 768)                  3072       ['tf.__operators__.getitem[0][
Normalization)                                                      0]']

dense (Dense)               (None, 128)                  98432      ['batch_normalization[0][0]']

dropout_37 (Dropout)        (None, 128)                  0          ['dense[0][0]']

dense_1 (Dense)             (None, 32)                   4128       ['dropout_37[0][0]']

dense_2 (Dense)             (None, 1)                    33         ['dense_1[0][0]']

==================================================================================================
Total params: 109587905 (418.04 MB)
Trainable params: 109586369 (418.04 MB)
Non-trainable params: 1536 (6.00 KB)
__________________________________________________________________________________________________
```

Figure 6: BERT Model Summary

6. Conclusion

Detecting autism at an early stage is a crucial step to mitigate potential complications. This research demonstrates the efficiency of artificial intelligence (AI) for behavioral trait prediction for autistic spectrum disorder. TASD dataset underwent data preprocessing with NLP techniques which was further fed to the BERT Multi class classification model to predict the behavioral traits. The BERT achieved an impressive accuracy of 88% on the TASD dataset, showcasing the efficiency of early autism detection. Future investigations will incorporate additional datasets to optimize the model and explore a multimodal approach by integrating social media data for enhanced accuracy in Behavioral Trait detection for autism spectrum disorder (ASD).

References

[1] W. H. O., "From: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders", 2023.
[2] D. Hall, D. Jurafsky, and C. D. Manning, "Studying the history of ideas using topic models," in EMNLP. Association for Computational Linguistics, 2008, pp. 363–371.
[3] D. Mimno, "Computational historiography: Data mining in a century of classics journals," JOCCH, vol. 5, no. 1, p. 3, 2012.
[4] U. C. Priva and J. L. Austerweil, "Analyzing the history of cognition using topic models," Cognition, vol. 135, pp. 4–9, 2015.
[5] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," Journal of the american statistical association, vol. 101, no. 476, 2006.
[6] Saha A, Agarwal N (2016) Modeling social support in autism community on social media. Netw Model Anal Health Inform Bioinform 5(1):1–14.
[7] Parsloe SM. Discourses of disability, narratives of community: reclaiming an autistic identity online. J Appl Commun Res. (2015) 43:336–56. doi: 10.1080/00909882.2015.1052829.
[8] Kapp SK, Steward R, Crane L, Elliott D, Elphick C, Pellicano E, et al. "People should be allowed to do what they like": autistic adults' views and experiences of stimming. Autism. (2019) 23:1782–92. doi: 10.1177/13623613198 29628
[9] McGill O, Robinson A. "Recalling hidden harms": autistic experiences of childhood applied behavioural analysis (ABA). Adv Autism. (2020) 7:269– 82. doi: 10.1108/AIA-04-2020-0025.

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024

[10] Elwyn G, Frosch G, Edwards A, Montori VM. Investing in deliberation: a definition and classification of decision support interventions for people facing difficult health decisions. Medical Decision Making. (2010) 30:701– 11. doi: 10.1177/0272989X10386231.

[11] Nunez-Mir GC, Iannone BV, Pijanowski BC, Kong N, Fei S. Automated content analysis: addressing the big literature challenge in ecology and evolution. Methods Ecol Evol. (2016) 7:1262–72. doi: 10.1111/2041-210X.12602.

[12] Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin: University of Texas at Austin. T29G6Z. (2015).

[13] Thin N, Hung N, Venkatesh S, Phung D. Estimating support scores of autism communities in large-scale web information systems. In: International Conference on Web Information Systems Engineering. Cham: Springer (2017). p. 347–55. doi: 10.1007/978-3-319-68783-4_24.

[14] Bellon-Harn M, Ni J, Manchaiah V. Twitter usage about autism spectrum disorder. Autism. (2020) 24:1805–16. doi: 10.1177/1362361320923173.

[15] HACHEMI, Rania; degha, Houssem Eddine (2023), "TASD-Dataset: Text-based Early Autism Spectrum Disorder Detection Dataset for Toddlers", Mendeley Data, V1, doi: 10.17632/87s2br3ptb.1.

[16] SangMin Park, JaeYoon LeeYu Ri Sung, JaeEun Kim. (2022). SBERT-IQ: A SentenceBERT-based embedding model considering keyword information content. Proceedings of the Korean Information Science Society, (), 1058-1060.

[17] HyeJin Yoon, JahWan Koo, and UngMo Kim. "Comparison of Word Embedding Performance for Sentiment Classification." Proceedings of the Annual Meeting of the Korean Information Processing Society 28.2 (2021): 760-763.

[18] S. Rubio-Martín, M. T. García-Ordás, M. Bayón-Gutiérrez, N. Prieto-Fernández and J. A. Benítez-Andrades, "Early Detection of Autism Spectrum Disorder through AI-Powered Analysis of Social Media Texts," 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), L'Aquila, Italy, 2023, pp. 235-240, doi: 10.1109/CBMS58004.2023.00223.

[19] K. Pineda, A. M. Perrotti, F. Poursardar, D. Graber and S. Jayarathna, "Using BERT to Understand TikTok Users' ADHD Discussion," 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI), Bellevue, WA, USA, 2023, pp. 213-214, doi: 10.1109/IRI58017.2023.00043.

Doi: https://doi.org/10.54216/FPA.150110
Received: August 22, 2023 Revised: December 11, 2023 Accepted: February 27, 2024