Using Computational Models to Detect Autistic Tendencies for Children from their Story Book Narratives

Ruihan Sun[1], Jasin Wong[2][0000−0003−0045−6393], Eva E. Chen[3][0000−0003−2194−197X], and Arbee L.P. Chen[4][0000−0003−2872−4484]⋆

[1] Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
ruihansun98@gmail.com

[2] Department of Special Education, National Tsing Hua University, Hsinchu, Taiwan
jswong@bu.edu

[3] College of Education, National Tsing Hua University, Hsinchu, Taiwan
evachen@gapp.nthu.edu.tw

[4] Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan
arbee@asia.edu.tw

Abstract. Diagnosing autism spectrum disorder (ASD) conventionally demands significant time and resources. Language deficits are key markers of ASD, particularly in constructing narratives. This study leverages computational models to analyze story book narratives from seven children with ASD and 16 typically-developing (TD) peers. By transcribing and training models on limited data using augmentation techniques, our best model achieved over 90% accuracy, sensitivity, and specificity—outperforming previous models by 20% in ASD detection. This research showcases the efficacy of our approach in efficiently assessing language abilities and identifying ASD tendencies. The method holds promise for enhancing diagnostic efficiency and providing comprehensive language evaluations to support children with ASD and their caregivers.

Keywords: Computational models · Natural language processing · Deep neural networks · Autism spectrum disorders · Story book narratives.

1 Introduction

1.1 Overview

As a neurodevelopmental disease, autism spectrum disorders (ASD) are characterized by difficulties in social communication and interaction as well as repetitive and restricted behavior patterns [1]. Although these characteristics can be observed in early childhood, ASD is frequently not officially diagnosed until later in life [25]. In 2023, the latest prevalence of people with ASD in Taiwan is reported to be 19,078 (out of a total population being 23.57 million people), with

⋆ Corresponding author

74.6% of these individuals identified with mild ASD [31]. Compared to the global ASD prevalence of 100/10,000 [43], the ASD prevalence in Taiwan is low due to a lack of medical capacity for ASD assessment and diagnosis. In Taiwan, most children with mild ASD are often not diagnosed until they enter elementary school [41].

Diagnosing ASD is a time-consuming process, primarily relying on behavioral signs, leading to a prolonged diagnostic journey [28]. In Taiwan, pediatricians may identify ASD signs during routine check-ups, prompting further evaluation. The child is then referred to psychiatric or pediatric rehabilitation clinics, where a comprehensive assessment is conducted by various professionals. This involves multiple hospital visits, spanning over six months. While ASD screening tools like STAT and M-CHAT are used, their detection rates, with sensitivities ranging from 0.62 to 0.78 and specificities from 0.83 to 0.91, still fall short of ideal standards [13, 40].

In addition, students with ASD undergo an extended evaluation process to qualify for special education support [10]. In Taiwan, an ASD student must obtain a medical diagnosis from specific centers (often located in major cities), undergo detailed assessments to test IQ and learning-related capabilities, and await the final decision from the County Committee for Identification and Placement of Gifted and Disabled Students. These last two steps typically take two to four months, in addition to the six months of hospital assessments. Due to the substantial time and resource investment, students with ASD and their families often face significant challenges in accessing essential resources. In this paper, we propose a cross-disciplinary, non-invasive alternative for detecting ASD in a relatively fast and accessible manner. We employ computational models to detect ASD with reasonable accuracy and efficiency, complementing the traditional screening and diagnostic process.

1.2 Narrative ability in ASD

Among the myriad characteristics that can suggest the presence of ASD in children, language is a key trait. That is, although individuals with ASD exhibit a broad range of verbal-linguistic abilities, a notable aspect of their language profile is a pervasive deficit in *pragmatic language skills* [22]. One of the most frequent pragmatic language problems children with ASD face is their ability to relate a narrative [2]. When examining the narratives produced by children with ASD, researchers found that these stories tend to lack coherence and causal connections, and may include irrelevant or inappropriate components [9]. When telling stories, children with ASD have fewer utterances and less lexical diversity than their typically-developing (TD) peers [4]. In addition, children with ASD may be unable to describe characters' thoughts and feelings because they do not understand the motivations behind the characters' actions in the story [3].

Previous research in ASD have measured narrative styles in children using traditional linguistic approaches, such as semantic analysis, discourse analysis, and grammatical analysis [2, 4, 7, 9, 11, 21]. However, most linguistic analytic methods require extensive time and staffing to transcribe and code the texts and recordings. In the present study, following the approach proposed in [37], we used computational models with natural language processing (NLP) to offer clinicians decision-making support akin to psychological testing tools.

Due to parental concerns about privacy loss and the potential disruption of regular school activities, it is challenging to identify eligible children with ASD to participate in the study. In order to augment the dataset used for training models, we have also employed the technique of data augmentation. Additionally, we incorporate language features from children's narratives into the model for joint learning (details will be provided in the Method and Data Collection sections).

1.3 Goal

Our present study is intended to be a preliminary examination into the validity of NLP as a computer-based method for detecting autistic tendencies from children' narrative discourses. We tested several NLP approaches and selected the ones that performed the best according to the performance matrices: sensitivity, specificity and accuracy. We aimed to:

1) Examine the feasibility of using computational models to detect children with ASD through an automatic text analysis.
2) Explore the linguistic features of the children with ASD and their TD peers.
3) Improve the computational model capability by adding the external knowledge of the language features.

2 Related Work

In this section, we introduced some relevant tools used in our study and previous studies on the narratives produced by autistic individuals.

2.1 The Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) assessment and picture book task

The ADOS-2 is a standardized assessment tool designed to assist in diagnosing ASD for individuals from 12 months to adulthood [18]. The ADOS-2 consists of semi-structured activities (including plays and interviews) to provide examiners the opportunities to observe behaviors that are relevant to the diagnosis of ASD. Examiners may choose from five different modules based on an individual's age and communication levels. The assessment takes approximately 40 to 60 minutes to complete. Below are the five modules of the assessment. Only one module would be chosen for a single individual.

1) Toddler Module – for children aged 12 to 30 months who do not often use phrases to speak.
2) Module 1 – for children 31 months and older who do not use phrases frequently.

3) Module 2 – for children of any age who use phrases to speak but are not fluent in spoken language.
4) Module 3 – for children and teenagers who are fluent in spoken language.
5) Module 4 – for older adolescents and adults who are fluent in spoken language.

In this study, we used the story-telling activity in Module 3 to collect children's narratives. The activity requires the examined child to tell a story based on a picture book without words: *Tuesday* [39].

The story is about the adventures of a group of frogs who float on water lily petals and visit a nearby town. These pictures depict unreal and humorous scenarios, as well as various psychological and emotional states of the characters. Participants were asked to look at pictures and tell stories. The instructions given are as follows: *"Look at this book. It presents a story about frogs. Can you tell me the story as we progress?"* During the data collection, the experimenter tried to avoid getting involved, showing non-verbal expressions of hints or encouragement, or providing guidance. The experimenter's goals were to videotape child's narratives and transcribe the story using the Codes for Human Analysis of Transcripts (CHAT) [20].

This task evaluates the children's capacity to comprehend and discuss a sequential story presented in a picture book and utilize visual cues provided to construct a narrative. Additionally, it evaluates the children's aptitude in recounting a story in a sequential manner.

2.2 Child Language Data Exchange System

Established in 1984 by Brian MacWhinney and Catherine Snow, the Child Language Data Exchange System (CHILDES) is the largest computerized database for child languages [20]. The CHILDES has the following three features:

1) A database currently consists of 230 corpora with 30 languages and includes transcripts of spontaneous language interactions between young children and caregivers, playmates, and teachers.
2) Using the Codes for Human Analysis of Transcripts (CHAT), a standardized, universally-used and multiple-level language coding system, to manually transcribe the language samples.
3) Using the Computerized Language ANalysis (CLAN) program to analyze the transcripts standardized by CHAT, researching various aspects of language usage, such as lexicon, syntax, morphology, phonology, discourse, and narrative.

There are many studies that have examined and compared the narrative abilities of autistic children and their TD peers using the Tuesday picture book and transcribed the recordings to text using the CHAT coding system [15, 26].

2.3 Chinese Language Sample Analysis

To collect and analyze the child language data systematically, we use the Chinese Language Sample Analysis (CLSA) procedure to guide our data collection process . The CLSA procedure [12, 33] was developed based on the CHILDES and CHAT, the two most established ways to record (CHILDES) and analyze (CHAT) naturally occurring conversations [12, 20, 33]. The CLSA further modified the CHILDES and CHAT guidelines to better fit the linguistic features of the Chinese language. As a standardized approach to analyzing language samples, the CLSA has been frequently used to analyze school-aged children's narrative discourses in Taiwan [35, 36]. According to the CLSA, researchers collect children's spontaneous verbal language in natural settings through video and audio recordings. The language samples were then transcribed verbatim. The CLSA provides clear guidelines for sentence and word segmentation in Chinese, allowing an accurate assessment of children's linguistic abilities, including phonology, vocabulary, grammar, and pragmatics. It also exhibits high sensitivity in detecting children's improvement in language development [12]. The following are the CLSA guidelines for sample collection, transcription, and analysis of Chinese-speaking children's language samples.

Sample Collection:

1. Examiners who use CLSA should receive formal test administration training.
2. CLSA only applies to children whose Mean Length of Utterance-characters $(MLU{-}c) \geq 5$.
3. CLSA can only be used for children who have the ability to speak without hearing or visual impairments.
4. Examiners need to collect children's demographic information, including:
   (a) Name
   (b) Gender
   (c) Chronological age
   (d) Family situation:
      i. Home ranking
      ii. Primary caregiver
      iii. Primary caregiver's education level
      iv. The primary language used in the family
      v. The occupational name of the main source of income
      vi. The child has the disability identified or not
      vii. Contact information
      viii. Other special matters
   (e) Video camera or recording equipment operates normally
   (f) A well-lit and quiet environment
   (g) Only the examiner and the child are in the room
   (h) Interact with the child for a few minutes before formal recording to reduce fear, discomfort, or resistance
   (i) There is no time limit for sample collection (in principle, each data collection should last 20-30 minutes)
   (j) Examiners encourage the child to talk by facilitating and supporting utterances when the child is unresponsive or responds infrequently

Transcription:

1. Utterance segmentation rules:
   (a) A child speaks a passage of speech apparently in one breath without a pause, often expressing a pause with the terminal intonation, called an utterance.
   (b) If the pause time exceeds two seconds, it is an indicator of a clearly cut utterance.
   (c) When children use "then," "and" and other related words to link sentences, each sentence should be divided into an independent utterance.
   (d) When the pause, tone change, deep breath, etc. appear in the sentence, the sentence should be split.
2. Word segmentation rules:
   Use Chinese Word Segmentation System developed by Academia Sinica [29], and manually corrected the segmentation according to the word segmentation rules of the Taiwan Corpus of Child Mandarin (TCCM) [34].

Analysis:

After the audio recordings are transcribed into text, the CHAT and CLAN will be used for language sample analysis. CLSA assesses several language characteristics to determine children's linguistic ability.

1. Mean length of utterance (MLU)
   MLU is calculated by the total number of characters or words in the selected valid utterances divided by the total number of selected utterances. Autistic children often show differences in MLU compared to their typically developing peers.
2. Mean length of the five longest utterances (MLU5)
   In addition to MLU, MLU5 can offer supplementary insights into child language development. MLU5 is calculated by the total number of characters or words in the five longest sentences in the language sample, then divides the total number by five to take the average. In terms of Chinese, [44] indicates that MLU5 can reflect the highest performance level of children in the complexity of sentences.
3. Vocabulary Diversity (VOCD)
   Vocabulary diversity is often used to evaluate children's vocabulary ability in spontaneous language. The larger the VOCD value is, the higher the lexical diversity will be.
   In Chinese, sentences are made up of words, and words are made up of characters. [42] found that calculating MLU with words (MLU-w) as a unit has higher test-retest reliability than calculating MLU with characters (MLU-c) as a unit. Therefore, in our study we used MLU-c, MLU-w, MLU5-c, MLU5-w, VOCD-c, VOCD-w as measures for analysis.
4. Word analysis
   According to grammatical functions, Chinese words can be divided into two

categories: notional words and function words. Notional words are meaningful words, including noun, verb, adjective, numeral, measure word, pronoun, and adverb. Function words are words that can not stand alone as a meaningful segment but have grammatical meaning or carry out specific language functions, such as preposition, conjunction, auxiliary word, and interjection. Word analysis can be used to assess children's vocabulary usage for each part of speech.

2.4 Computational models for analyzing narration

Previous research on children's narrative skills mainly uses manual analytic methods to investigate language characteristics across different developmental stages and populations with various language abilities. Only a few studies used NLP tools or neural networks to investigate the narrative performance differences between autistic individuals and TD peers.

To quantitatively characterize narrative performance in autistic individuals, sentiment and language abstraction analyses based on the Linguistic Category Model were employed [5]. Their experiments included 50 Polish-speaking children (25 with ASD and 25 controls with TD, aged 7-25). Language samples were generated during two standardized tasks from the ADOS-2: Telling a Story from a Book and Description of a Picture. The results supported the sentiment and language abstraction analyses as a valuable tool. Specifically, autistic individuals demonstrated lower levels of language abstraction compared to TD individuals and exhibited difficulties expressing emotions.

In Chojnicka and Wawer's work, they collected 50 Polish-speaking individual's narratives in picture book (25 autstic participants and 25 TD controls, aged 7-25). The aim of the research was to identify the narratives produced by autistic individuals. Two text encoders, Embeddings from Language Models (ELMo) and Universal Sentence Encoder (USE), were utilized, along with three classification algorithms: XGBoost, support vector machines, and a dense neural network layer [37]. Their experiments, employing deep neural network text representation models, demonstrated higher sensitivity, specificity, positive predictive values, and negative predictive values compared to human raters. However, these values were lower when compared to the two standardized instruments used in the study: ADOS-2 and Social Communication Questionnaire (SCQ) [27]. The SCQ consists of a series of 40 questions, which are answered by a parent, caregiver, or teacher who knows the individual well. The questions cover various aspects of social communication, such as social interaction, verbal and nonverbal communication, and restricted and repetitive behaviors commonly associated with autism. Each item in the SCQ is scored based on the presence or absence of specific behaviors related to ASD. The total score provides an indication of the likelihood that the individual may have autism, with higher scores suggesting a higher likelihood of the condition.

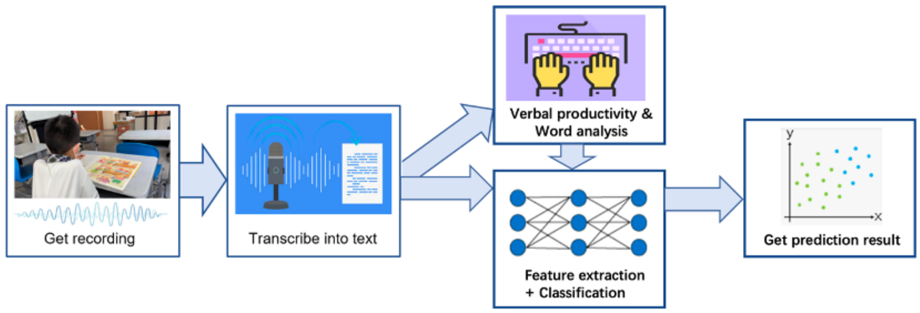

Fig. 1. The flowchart of our approach.

3 Data Collection

As shown in Figure 1, the whole approach can be divided into 5 steps: collect language samples by audio recording, transcribe the recordings into text, conduct verbal productivity and word analysis, perform feature extraction and classification, and acquire the prediction result. In this section, we introduce how we obtained recordings and transcribed them into text files. We also explain how we conducted verbal productivity and word analysis. In the next section, we describe several techniques we applied to feature extraction and classification.

3.1 Participants

A total of 23 school-aged children were enrolled, including 7 children diagnosed with ASD (five mild, one moderate and one severe diagnosis) in the ASD group and 16 age-matched peers in the TD group. All of the participants with ASD in our research were boys (Table 1). The gender proportion of our participants reflected the fact that boys are more frequently diagnosed with ASD than girls [6]. We exhausted all the students with ASD at the school. Due to the cultural expectation differences among genders, the male-to-female ratio of ASD in Taiwan is higher than the international ratios (5:1 vs. 3:1) [17, 31]. Most students completed reading the two story books within 20 minutes. There is no significant group difference between children with ASD (Mean/ SD= 879.91/ 340.22) and TD (Mean/ SD= 1053.00/ 316.68) in reading time (t = -1.18, p= .25, Cohen's d = 0.53).

3.2 Procedure

Participants were recruited from the main local ASD association and a public elementary school catering to both typical and special education needs students in Hsinchu, a mid-sized city in Taiwan. To be included, children must have been: (1) aged between 6 and 12 years, (2) capable of communicating in Mandarin Chinese as their primary language, and (3) not diagnosed with hearing or visual

Table 1. Group descriptive characteristics

Background variables	ASD (n=7)	TD (n=16)	Statistical test
Age (mean & standard deviation)	10.5 (0.85)	10.28 (1.37)	$t = .49$
Gender			Fisher's exact = .003**
Male	7 (100%)	5 (31.25%)	
Female	0	11 (68.75%)	
Education of the main caregiver			$\chi^2 = 2.65$
Trade license/Certificate/High school diploma/GED or Below	3 (42.86%)	6 (37.5%)	
Bachelor's degree	3 (42.86%)	10 (62.5%)	
Doctorate or master's degree	1 (14.29%)	0	
Occupation of the main caregiver			$\chi^2 = 7.49$
Engineer	0	5 (31.25%)	
Administrator/Manager	1 (14.29%)	3 (18.75%)	
Service industry	1 (14.29%)	2 (12.5%)	
Freelance work	5 (71.43%)	3 (18.75%)	
Others (Designer/Typist)	0	3 (18.75%)	

ASD: autism spectrum disorder; TD: typically developing.
$**p<.01$.

impairments. Participants were divided into two groups, ASD and TD. Those in the ASD group all received an official medical diagnosis or disability identification. Those in the TD group were children who did not (1) have a personal or family history of ASD, (2) have a personal history of developmental disorders, and (3) have neurological or psychiatric disorders or suspected genetic syndromes and developmental issues.

The administrators at the association and elementary school screened the potential participants and offered research flyers to their parents. Afterwards, the first author visited the school to explain the study in detail to the parents and students. Parents' written consent and students' verbal consent were obtained before the data collection process. The research procedure was reviewed and approved by the Central Regional Research Ethics Committee at China Medical University [BLOCK FOR REVIEW].

3.3 Materials

In this study, we collected language samples using two distinct story books with varied styles and cultural backgrounds. The following two story books were selected: *Tuesday*, a book from ADOS-2, and the Chinese book *Spit the Seeds* [16].

Tuesday, a 29-page illustrated book about the night-time adventures of a colony of frogs, has been utilized to collect children's language samples in multiple past studies [15, 26, 37]. There is no text in the storybook, with the narrative implied through the illustrations. *Spit the Seeds*, a 40-page illustrated book about a piglet eating papaya too quickly, was selected as a local Taiwanese comparison to *Tuesday*. The story book received the prestigious Hsin-Yi Children Literature

Award in Taiwan, and is often purchased by local schools for their classrooms. There is some Chinese text in the book.

3.4 Get recording and transcription

To transcribe and analyze children's naturally occurring languages in Chinese, we use the CLSA procedure to guide our data collection process [12]. According to the CLSA's guidelines, the study was carried out in a well-lit, quiet classroom with only the experimenter and the child. Prior to formal data collection, the experimenter interacted with the tested child for two minutes to build rapport, establishing a shared focus of attention with the child. Once the study began, the experimenter asked the child to choose their first book to read. When the child did not respond to the story book or responded in one-word utterances, the experimenter provides more guidance to encourage the child to share more, such as *"What is happening here?"*

All the collected language samples were then transcribed based on the CHAT transcription manual. Segments of utterances, words, pauses, and repetitions were determined based on the CLSA's guidance. Afterwards, the results were manually checked and corrected based on the word segmentation rules stated in the TCCM [34].

3.5 Data Analysis

Analysis of Verbal Productivity

Using the transcriptions, we then compared the narrative skills between the children from the ASD and TD groups. We examined children's verbal productivity by calculating the number of utterances, number of characters, and number of words. At the sentence level, we calculated the MLU and MLU5 in characters and words [12]. We also measured the lexical diversity by calculating the VOCD in characters and words [12]. Finally, we investigated whether children in different groups have different preferences between the two story books.

Table 2 shows the comparisons of verbal productivity between the ASD and TD groups. First, children with ASD tend to narrate less than the TD children and have fewer utterances, characters, and words when telling stories. This is especially evident in the following five measures (c: characters; w: words): MLU-c, MLU5-c, VOCD-c, MLU-w, and MLU5-w. The t-test for independent samples reveals that the ASD and TD groups differ significantly in MLU-c ($t = 2.33$, $p = .03$, Cohen's $d= 0.89$), MLU5-c($t = 2.34$, $p = .03$, Cohen's $d= 0.90$), VOCD-c ($t = 2.23$, $p = .037$, Cohen's $d= 0.86$), MLU-w ($t = 2.42$, $p = .025$, Cohen's $d= 0.94$), and MLU5-w ($t = 2.63$, $p = .016$, Cohen's $d= 1.03$), while the differences between VOCD-w ($t = 1.94$, $p = .065$, Cohen's $d= 0.85$) and book preference (Fisher's exact = 1.65, $p = .66$, Odds ratio=1.61) are not significant. The findings suggest that children with ASD tend to use a limited range of characters and words when speaking. However, paradoxically, they show an adept use of lexical diversity.

Table 2. Narrative performance on verbal productivity measures

Verbal Productivity Measures	**ASD Mean(SD)**	**TD Mean(SD)**
Number of utterances	69.29 (44.48)	83.94 (25.67)
Number of characters	452.43 (354.76)	643 (235.68)
Number of words	277.71 (220.28)	409.13 (149.38)
MLU-c	5.69 (2.57)*	7.47 (1.17)
MLU5-c	12.11 (7.27)*	17.16 (3.27)
VOCD-c	47.01 (29.43)*	66.92 (14)
MLU-w	3.54 (1.62)*	4.75 (0.82)
MLU5-w	7.43 (4.46)*	11.09 (2.29)
VOCD-w	69.14 (58.05)	115.06 (49.55)
Preference	3:4	5:11

SD: standard deviation; ASD: autism spectrum disorder; TD: typically developing; MLU-c: mean length of utterance-characters; MLU5-c: mean length of the five longest utterances-characters; VOCD-c: vocabulary diversity-characters; MLU-w: mean length of utterance-words; MLU5-w: mean length of the five longest utterances-words; VOCD-w: vocabulary diversity-words; Preference: Which book do children prefer? (*Tuesday*: *Spit the Seeds*).
$*p<.05$; $**p<.01$.

Word Analysis

In addition to the macroscopic statistics of verbal productivity, we conducted independent sample t-tests to examine if there was a group difference in narrative skills between the ASD and TD groups. We classified all transcribed words into two main categories: *notional words* (including nouns, verbs, adjectives, numerals, measure words, pronouns, and adverbs) and *function words* (including prepositions, conjunctions, auxiliary words, and interjections).

Table 3 shows the results of word analysis in notional and function words. Overall, the proportion of notional words used by children with ASD is higher than that of TD children (t = 2.48, p = .022, Cohen's d= 0.86), and the use of function words is lower than that of TD children (t = 2.49, p = .021, Cohen's d=0.92). Children with ASD are unable to fluently use measure words and auxiliary words compared to the TD children (measure word: t = 3.07, p = .006, Cohen's d=1.22; auxiliary word: t = 3.21, p = .004, Cohen's d=1.46). This finding may be attributed to the prevalent utilization of narrative discourses among children with ASD primarily for declarative statements or imperative expressions, resulting in limited usage of measure words and auxiliary words. However, children with ASD used adjectives more often than TD children in their narratives (t = 2.24, p = .036, Cohen's d=0.77). This represents a unique pattern, as compared to TD children. Although children with ASD often show deficits in narrative expression, they show relative proficiency in specific areas of adjective use.

Table 3. Word analysis

Parts of Speech	ASD Mean%(SD)	TD Mean%(SD)
Noun	34.52% (0.23)	22.64% (0.06)
Verb	19.80% (0.07)	20.62% (0.03)
Adjective	9% (0.09)*	4.04% (0.01)
Numeral	2% (0.02)	2% (0.01)
Measure word	2% (0.02)**	3.93% (0.01)
Pronoun	9% (0.07)	11.35% (0.03)
Adverb	9.62% (0.07)	13.67% (0.03)
Notional word	85% (0.1)*	78.65% (0.03)
Preposition	3% (0.02)	4.91% (0.02)
Conjunction	7% (0.05)	7% (0.04)
Auxiliary word	4% (0.03)**	7.72% (0.02)
Interjection	0% (0)	1% (0.01)
Function word	14% (0.1)*	20.82% (0.03)

SD: standard deviation; ASD: autism spectrum disorder; TD: typically developing.
$*p<.05$; $**p<.01$.

4 Method

In this section, we describe several techniques we applied to feature extraction and classification.

4.1 Data augmentation

As we only recruited seven autistic children and 16 TD children, which is far from enough data for the subsequent training of the model. In order to increse the size of data, we used data augmentation techniques proposed by [38], i.e. for a given sentence in the training set, we randomly chose and performed one of the operations to do data augmentation.

To balance the amount of data between ASD and TD children, we amplified the data for ASD children by 5,10,15,20 times, and 2,4,6,8 times for TD children. Figure 2 shows the visualized result of the augmented ASD and TD datasets for picture book *Tuesday*. We applied Term Frequency-Inverse Document Frequency (TF-IDF) [30] to get feature vectors, k-means [19] for clustering, Principal Component Analysis (PCA) [23] for dimension reduction, and Matplotlib to plot the 2-D latent space representations as shown in Figure 2. We found that the resulting latent space representations for the augmented sentences closely surrounded the original sentences, which were divided into 7 clusters (ASD dataset) and 16 clusters (TD dataset) in the figure. This showed that in most cases, augmented sentences conserved the labels of the original sentences.

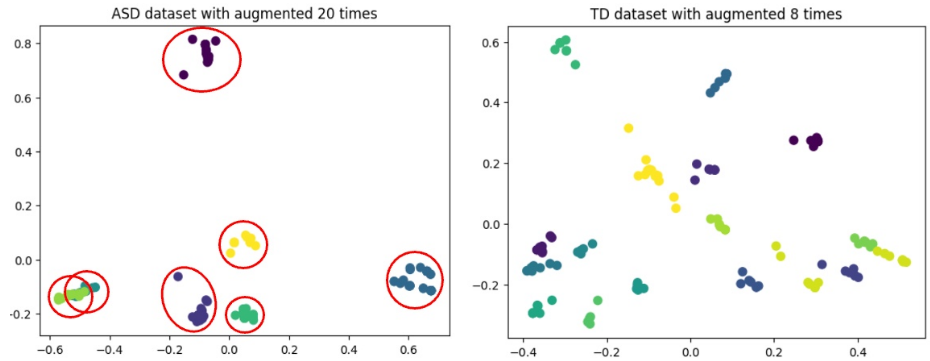

Fig. 2. Visualization of the augmented datasets for the picture book *Tuesday*

4.2 Neural networks in detecting ASD

In our study, we compared several methods for classifying participants with ASD and TD. The methods we tested differed not only in the various ways they represented utterances, but also in how they handled verbal productivity and word analysis data for the classification algorithms.

To evaluate the predictive ability of the model with limited data samples, we employed k-fold cross-validation. In each iteration, one of the k folds was used as the validation set, while the remaining k-1 folds were used as the training set. After each iteration, the performance metrics, such as accuracy, were recorded. The final performance of the model was reported as the average of these k iterations. In our experiments we used 5-fold cross-validation.

Model 1: TF-IDF with Support Vector Machines (SVM)

The first approach we applied to represent utterances is TF-IDF [30]. TF-IDF combines term frequency and inverse document frequency to identify significant words. SVM [24] finds a decision boundary to separate data classes.

In our experiment set-up, we initially employed data augmentation techniques to enhance the transcribed text dataset along with its corresponding labels. Subsequently, TF-IDF was applied to extract informative textual features from the augmented data. To classify the resulting feature vectors, we employed the SVM algorithm. Finally, a 5-fold cross-validation strategy was employed to assess the performance of the classification model.

Model 2: Bidirectional Encoder Representations from Transformers (BERT)

BERT [8] is a significant language model renowned for capturing contextual information. Its training involves pre-training on unlabeled text, followed by

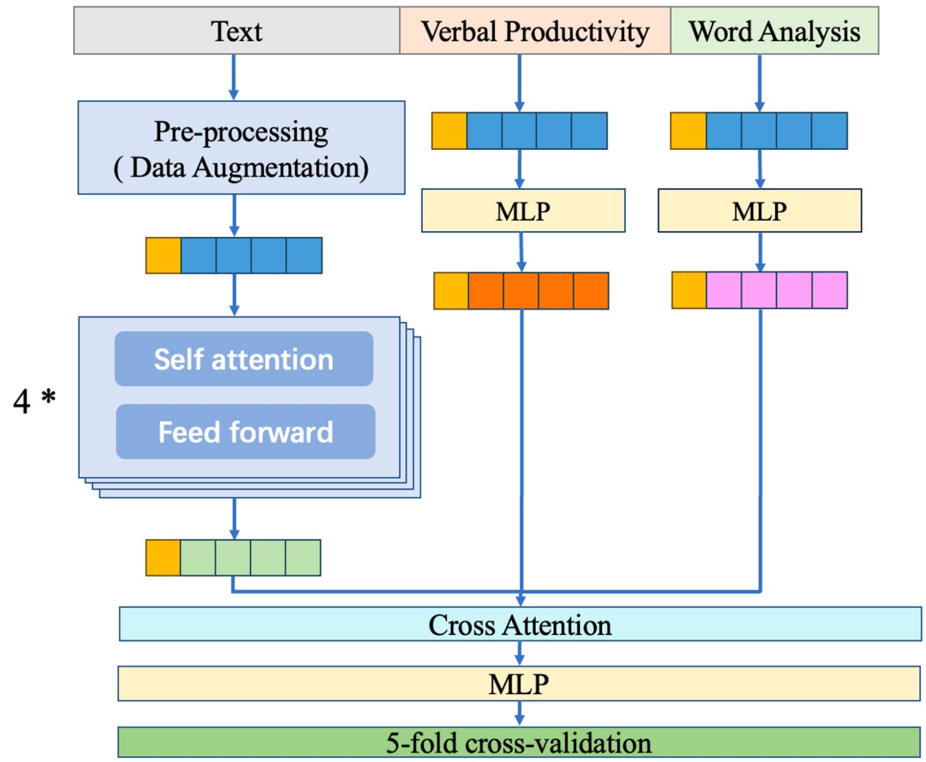

Fig. 3. Visualization of augmented datasets for picture book *Tuesday*

fine-tuning on labeled data for specific tasks. In this study, we used the Tiny-Bert_4L_zh [14] as our pre-trained model, which is a compact and efficient version of the BERT model specifically designed for Chinese language understanding tasks. It is especially suitable for our small datasets. In addition to simply using TinyBERT to train children's narratives, we also added verbal productivity and word analysis data as additional information to enhance the learning ability of the model.A simplified schema of encoding and classifying utterances using TinyBERT is illustrated in Figure 3.

First, we amplified the transcribed documents with their corresponding labels, and then input the augmented data into TinyBERT of the 4-layer transformer architecture. At the same time, the corresponding verbal productivity and word analysis data were regularized and passed into the multilayer perceptron (MLP) with one linear layer. By concatenating the output of the MLP with the TinyBERT representation, we subsequently performed cross-attention and applied an additional layer of MLP for the final classification. Furthermore, we added the 5-fold cross-validation to assess the model performance.

5 Experiments

5.1 Data preprocessing

We divided the collected narratives (from two different story books) into six different datasets for the experiments, as follows:

1. *Combine*: Narratives of the two story books were combined for each participant, resulting in a total of 23 narrative documents.
2. *Separate*: Narratives of the two story books were kept separate for each participant, resulting in a total of 46 narrative documents.
3. *Combine_DA*: The Combine dataset with data augmentation applied.
4. *Separate_DA*: The Separate dataset with data augmentation applied.
5. *Book 1_DA*: The 23 narratives for the story book *Tuesday* with data augmentation applied.
6. *Book 2_DA*: The 23 narratives for story book *Spit the Seeds* with data augmentation applied.

To make the model training more powerful, we used StandardScaler from the Scikit-Learn package to standardize all features of the verbal productivity and word analysis so that the mean of the data was 0 and the variance was 1.

5.2 Model Settings

In accordance to [38], we used α=0.05 (with α being the proportion of changed words in each utterance) to do data augmentation. Since [38] focuses on enhancing English corpus data, we employed the Chinese stop word vocabulary [32] developed by Harbin Institute of Technology and synonyms toolkit for the replacements. We used the TinyBert_4L_zh as our pretrained model with the number of transformer layers M=4, the hidden size d'=312, and the head number h=12. For the training, we used 20 EPOCH, 5e-5 LR, 16 BATCH_SIZE, and 200 Embedding_dim.

5.3 How much augmentation is appropriate for model training?

As mentioned in Section 4.1, we amplified the data for ASD children by 5, 10, 15, 20 times, and corresponding 2, 4, 6, 8 times for TD. In order to select the most appropriate amplification factor, we conducted experiments with the data of different amplification factors under the two models.

In the TF-IDF with SVM model (Model 1), as illustrated in Figures 4 (a) to (d), the experiment results conducted on the four distinct datasets indicate that the accuracy exhibits an upward trend as the amplification factor increases. Notably, in the Book 1, Book 2, and Combine datasets the model achieves their best performance when the ASD amplification factor is 15 times and the TD amplification factor 6 times. Subsequently, the accuracy stabilizes or even declines. However, in the Separate dataset, the best performance is observed when the ASD amplification factor is 20 times and the TD amplification factor 8 times.

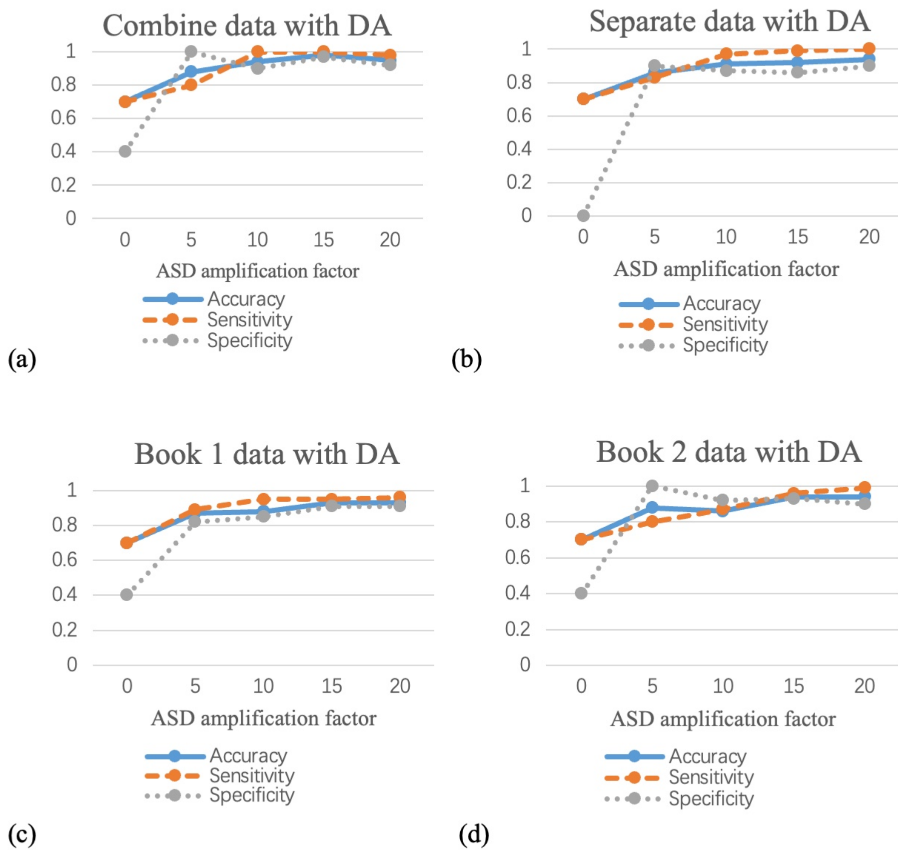

Fig. 4. Performance on Model 1 with different amplification factors.
(X-axis represents the ASD amplification factors; Y-axis represents the model performances.)

A similar trend is observed in sensitivity (detection rate of children with ASD), wherein the Book 1, Book 2, and Separate datasets achieve their best performance when the ASD amplification factor is 20 times. Conversely, in the Combine dataset, the best performance is attained when the ASD amplification factor is increased by a factor of 10. Finally, the specificity (the number of children detected as TD as a proportion of the total number of children with TD) is also on the rise, but unlike the accuracy and sensitivity, a smaller amplification factor can achieve higher performance.

In the TinyBERT model (Model 2), as illustrated in Figures 5 (a) to (d), the empirical findings derived from the four datasets demonstrate that the accuracy reaches its pinnacle when the amplification factor for ASD is set at 10 times and that for TD 4 times. Subsequently, the accuracy is either stabilized or shows a decline. Similar patterns are observed in terms of sensitivity for the Book 2,

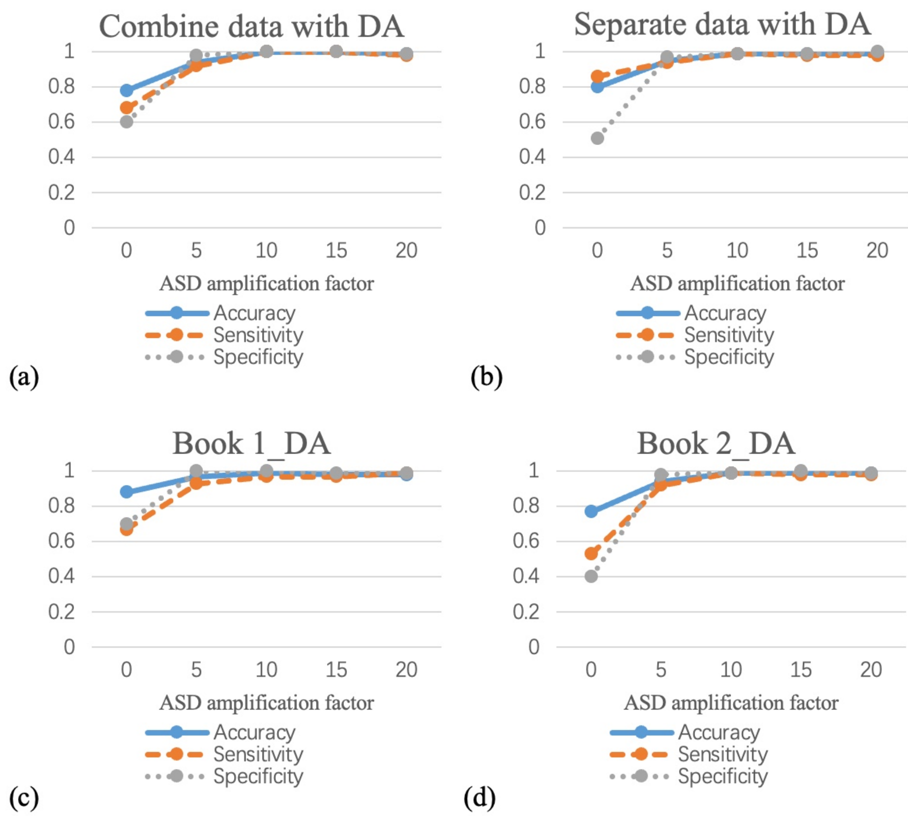

Fig. 5. Performance on Model 2 with different amplification factors.
(X-axis represents the ASD amplification factors; Y-axis represents the model performances.)

Combine, and Separate datasets. However, the best performance in the Book 1 dataset is achieved when the ASD amplification factor is increased by a factor of 20. The distribution of the best specificity values varies across the datasets, although the overall discrepancy is not substantial.

Based on the above observations, we found the most suitable data amplification factor for the two models: 15 times for ASD and 6 times for TD for the TF-IDF with SVM model, and 10 times for ASD and 4 times for TD for the TinyBERT model.

5.4 Does the choice of story books influence the classification performance of the proposed methods?

To examine whether or not the narratives collected from different story books influence the results of the classification, we used two story books as training

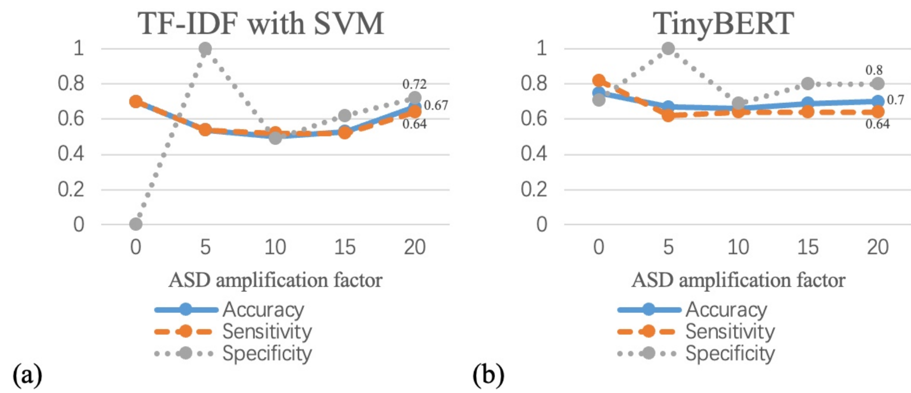

Fig. 6. Two models' performance on Book 1 for training and Book 2 for testing (X-axis represents the ASD amplification factors; Y-axis represents the model performances.)

sets to train the proposed models respectively and compared the classification performance between the proposed models. We first used the Book 1 dataset as the training data, and the Book 2 dataset as the testing data. Then we used the Book 2 dataset as the training data, and the Book 1 dataset as the testing data.

Figure 6 presents an analysis of the two models' performance on Book 1 for training and Book 2 for testing, with different data augmentation factors. The results indicate that the accuracy and sensitivity of the two models are very similar with data augmentation, showing an upward trend. However, the specificity shows large fluctuations when the amplification factor is less than 10. That is, when the amount of data is too small, the model training can be insufficient to classify TD and ASD groups. By comparison, TinyBERT has a relatively stable and better performance than TF-IDF with SVM, likely because TinyBERT's language expression is very powerful and can capture semantic differences and word order well, even with a small amount of data.

Figure 7 presents an analysis of the two models' performance on Book 2 for training and Book 1 for testing with different data augmentation factors. The results indicate that the trends of the accuracy and sensitivity of the TF-IDF with SVM model are very similar with data augmentation. However, when the amplification factor is less than 10, the accuracy and sensitivity show a declining trend, and the specificity remains zero. This is likely due to the small data volume (for example, without data augmentation, we have only 23 narrative documents). However, when the amplification factor is more than 10, the performance measures show an upward trend. For the TinyBERT model, the trends of the accuracy, sensitivity and specificity are very similar with different amplification factors. TinyBERT's performance measures are also similar with TF-IDF with SVM when the amplification factor is 20.

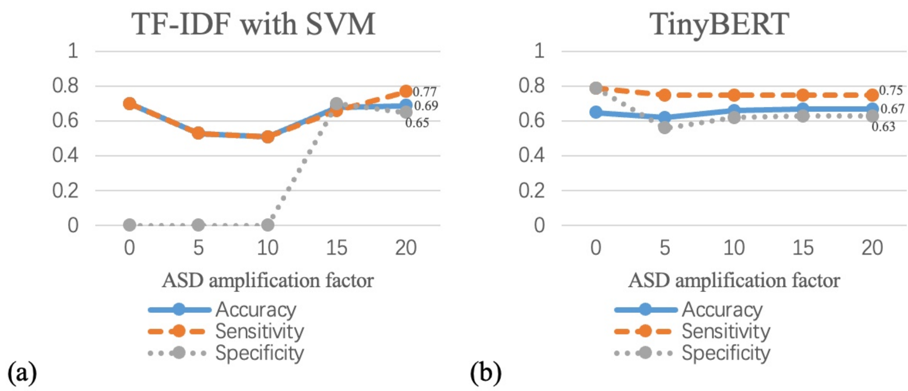

Fig. 7. Two models' performance on Book 2 for training and Book 1 for testing. (X-axis represents the ASD amplification factors; Y-axis represents the model performances.)

From Figures 6 and 7, we can conclude that both TF-IDF with SVM and TinyBERT can learn certain ASD characteristics to reach a sensitivity of 0.64 to 0.77 in the two testing sets. The results indicate that the narratives from different story books do not differ in classifying ASD and TD groups. However, because the TF-IDF with SVM is easily limited by the amount of data, we used TinyBERT in the following experiment and examined if adding information of linguistic characteristics improves the model performance.

5.5 Does adding verbal productivity and word analysis information as external knowledge to the computational models improve the classification performance?

To further improve the model performance in specificity, we added the results of verbal productivity and word analysis to the model building as supplementary information. The renewed model is called *TinyBERT with external knowledge*. We used 4 datasets (Combine, Separate, Book 1, and Book 2) that had not undergone data augmentation for the experiments. As Table 4 shows, both TinyBERT and *TinyBERT with external knowledge* achieve the same accuracy for the Combine dataset, but the sensitivity and specificity are improved by 3% and 7%. In the Separate dataset, the accuracy is increased by 3%, the specificity is increased by 22%, and the sensitivity is decreased by 3%. In the task of selecting which book as the training set, the accuracy, sensitivity, and specificity are improved compared with TinyBERT without external knowledge. Especially when Book 2 is used as the training set, the sensitivity is increased by 5%, the specificity is increased by 13%, and the accuracy is increased by 9%.

Table 4. Performance comparison by adding external knowledge from verbal productivity and word analysis

		Sensitivity	**Specificity**	**Accuracy**
TinyBERT	Combine	0.76	0.4	0.78
	Separate	0.86	0.51	0.8
	Book 1	0.82	0.71	0.75
	Book 2	0.79	0.79	0.76
TinyBERT with external knowledge	Combine	0.79	0.47	0.78
	Separate	0.83	0.73	0.83
	Book 1	0.82	0.84	0.81
	Book 2	0.84	0.92	0.85

5.6 Does our proposed methods perform better than existing approaches?

In Table 5, we compare our three examined methods (*TF-IDF with SVM*, *TinyBERT*, and *TinyBERT with external knowledge*) with the results from ADOS-2, SCQ and [37]. The three methods demonstrate evident enhancements in terms of sensitivity, specificity, and accuracy. Particularly, the TF-IDF with SVM model exhibits a remarkable improvement of 7% to 8% in both sensitivity and accuracy, along with a substantial 16% increase in specificity compared to [37]. Furthermore, TinyBERT outperforms [37] by 14% to 21% across all three performance metrics. Notably, the integration of external knowledge from verbal productivity and word analysis into TinyBERT yields exceptional results, surpassing 90% for sensitivity, specificity, and accuracy. Notice that the performance of TinyBERT with external knowledge is close to that of ADOS-2 and SCQ, the two standardized assessment tools that help diagnose ASD in children and adults. These findings reaffirm the efficacy of incorporating external knowledge through verbal productivity and word analysis, facilitating superior learning within the model.

Table 5. Comparisons among the standardized ASD-screening tools (ADOS-2 & SCQ), the results of the previous research, and the proposed three methods.

	Sensitivity	**Specificity**	**Accuracy**
ADOS-2	1.00	0.92	0.96
SCQ	0.92	0.96	0.94
Results from Wawer and Chojnicka (2022) [37]	0.72	0.68	0.7
TF-IDF with SVM	0.79	0.84	0.78
TinyBERT	0.86	0.89	0.84
TinyBERT with external knowledge	0.93	0.91	0.92

6 Conclusion

We aimed to use computational models and language ability analysis from children's story book narratives for autistic tendency detection. Thanks to the continuous progress of computational models in recent years, we can use high performance models to address the problem. It is important to acknowledge that, thus far, the majority of research on ASD detection through speech has relied on assessment sheets or rudimentary language analysis tools, with only a limited number of researchers exploring computer-based methods.

Our proposed methods showed better performance in different tasks. First, we conducted experiments on TF-IDF with SVM and TinyBERT under different data amplification factors. The experiment results showed that when the ASD amplification factor was 15 and 10 respectively, the two models had achieved relatively stable performance. The accuracy, sensitivity and specificity can reach more than 90%, among which TinyBERT can reach 99% and above.

In the second experiment, we alternately used two different story books as the training set in the experiments. The experiment results showed that both the TF-IDF with SVM and TinyBERT can learn certain ASD characteristics, and reached a sensitivity of 0.64-0.77 in the two testing sets. TinyBERT was more stable than the TF-IDF with SVM which was easily limited by the amount of data.

In the third experiment, we examined whether verbal productivity and word analysis can be used to improve the performance of the model. In this experiment, we did not use the dataset that had undergone data augmentation, but only used the collected original data. We conducted ablation experiments on 4 different datasets (Combine, Separate, Book 1, and Book 2) and the results showed that verbal productivity and word analysis can improve the accuracy, sensitivity and specificity to a certain extent, in the case of a very small amount of data, which also proved the importance of the use of additional knowledge to improve the performance of the model.

In the last experiment, we compared the three methods we proposed with the most relevant study of [37]. In order to have the same amount of data, we amplified the original data to be the same amount as [37] (25ASD, 25TD). It can be seen from Table 5 that the three methods we proposed surpassed [37]. Among the three methods, TinyBERT with the external knowledge from verbal productivity and word analysis achieves the best performance, with the sensitivity, specificity, and accuracy all exceeding 90%.

These experiment results hold promise for facilitating more accessible and expedient services in autistic tendency detection in the future.

7 Limitations and Future Work

It is undeniable that there are many limitations in our study. First, it is important to note that the available dataset contains a limited number of samples, exclusively comprising males within the collected ASD group. Moreover, the gender distribution within the TD dataset is imbalanced. This inherent imbalance has the potential to introduce bias during subsequent analyses of language characteristics. The small number of the sample size may be a reflection of the difficulties in recognizing ASD children with an official diagnosis. Our methods can be the first step to increase the ASD screen rate and lower the barriers to detecting children with mild ASD in school settings. Second, owing to the challenges associated with collecting personal data, we are unable to obtain information pertaining to the verbal IQ and non-verbal IQ of the participants. Consequently, the impact of the IQ variability on the results cannot be determined. Third, it is noteworthy that our experiment solely relies on audio recordings, without accompanying video recordings that can capture children's nonverbal and behavioral expressions. As a result, the absence of visual records limits our ability to analyze the nuances of children's nonverbal cues and gestures. Finally, we utilize speech-to-text transcription for subsequent analysis, which hinders the inclusion of children's voice intonation in the recorded data. It is worth mentioning that achieving high-quality speech-to-text transcription poses a significant challenge. In our research, manual transcription is employed, which introduces a possibility of transcription errors and potential deviations during the transcription process.

Autistic tendency analysis through children's narratives remains a difficult challenge, with many factors to consider. Future endeavors should incorporate additional language assessment methods, such as the analyses of sentence structure composition, narrative coherence, pitch variations, and speech fluency. In order to achieve a more nuanced understanding of the language characteristics specific to children with ASD, it may be advantageous to include a comparative analysis involving a cohort of language-impaired children. This comparative approach can more effectively help delineate the distinctive features associated with ASD.

Acknowledgements

We extend our heartfelt gratitude to the children and their families who took part in this study, generously contributing to our dataset. We are also grateful to the Zhulian Elementary School in East District of Hsinchu City and the Hsinchu Autism Association for their cooperation in recruiting the participants for the study and providing the experiment venue.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ruihan Sun. The first draft of the manuscript was written by Ruihan Sun and all other authors commented on previous versions of the manuscript. All authors read and approved the final manuscript

Funding

This research was funded by [109-2221-E-468-014-MY3] (awarded to Dr. Arbee L.P. Chen) from National Science and Technology Council, Republic of China.

Statements and Declarations

Competing Interests

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Consent to Participate

We obtained the written consent from the legal guardian of each participant.

Ethical Approval

This study was reviewed and approved by the Institutional Review Board at China Medical University (IRB number: CRREC-112-002).

References

1. Association, A.P., et al.: Diagnostic and statistical manual of mental disorders, 5th edn) edn. American Psy-chiatric Association., Retrieved from http://psychiatryonline. org/doi/book/10.1176/appi. books. 9780890425596 (2013)
2. Baixauli, I., Colomer, C., Roselló, B., Miranda, A.: Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. Research in Developmental Disabilities **59**, 234–254 (2016)
3. Baron-Cohen, S., Leslie, A.M., Frith, U.: Does the autistic child have a "theory of mind"? Cognition **21**(1), 37–46 (1985)
4. Capps, L., Losh, M., Thurber, C.: "the frog ate the bug and made his mouth sad": Narrative competence in children with autism. Journal of abnormal child psychology **28**, 193–204 (2000)
5. Chojnicka, I., Wawer, A.: Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction. PLoS One **15**(3), e0229985 (2020)
6. Christensen, D.L.: Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2012. MMWR. Surveillance summaries **65** (2016)
7. Colle, L., Baron-Cohen, S., Wheelwright, S., Van Der Lely, H.K.: Narrative discourse in adults with high-functioning autism or asperger syndrome. Journal of autism and developmental disorders **38**, 28–40 (2008)

8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
9. Diehl, J.J., Bennetto, L., Young, E.C.: Story recall and narrative coherence of high-functioning children with autism spectrum disorders. Journal of abnormal child psychology **34**, 83–98 (2006)
10. Education, M.o.: Handbook of identification methods for students with disabilities and gifted talents. https://spe-girc.ntnu.edu.tw/wp-content/uploads/2021/08/jianding201608.pdf (2014)
11. Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J.R.M., Weed, E.: Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in american and danish autistic children. Autism Research **15**(4), 653–664 (2022)
12. Huang, R., W.S.T.I.H.T..Z.Z.: Chinese Language Sample Analysis Guide. Psychological Publishing (2016)
13. Hung, Y.C., C.I..W.C.: Detection of autism spectrum disorder in different settings: Accuracy of the modified checklist for autism in toddlers. Bulletin of Special Education **44**(3), 33–61 (2019)
14. Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., Liu, Q.: Tinybert: Distilling bert for natural language understanding. arXiv preprint arXiv:1909.10351 (2019)
15. Kuijper, S.J., Hartman, C.A., Bogaerds-Hazenberg, S., Hendriks, P.: Narrative production in children with autism spectrum disorder (asd) and children with attention-deficit/hyperactivity disorder (adhd): Similarities and differences. Journal of abnormal psychology **126**(1), 63 (2017)
16. Lee, C.: Spit the Seeds. Hsin-yi (1993)
17. Loomes, R., Hull, L., Mandy, W.P.L.: What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry **56**(6), 466–474 (2017)
18. Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., Bishop, S., et al.: Autism diagnostic observation schedule–2nd edition (ados-2). Los Angeles, CA: Western Psychological Corporation **284** (2012)
19. MacQueen, J.: Classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. pp. 281–297 (1967)
20. MacWhinney, B., Snow, C.: The child language data exchange system: An update. Journal of child language **17**(2), 457–472 (1990)
21. Marini, A., Ozbič, M., Magni, R., Valeri, G.: Toward a definition of the linguistic profile of children with autism spectrum disorder. Frontiers in psychology **11**, 808 (2020)
22. Parsons, L., Cordier, R., Munro, N., Joosten, A., Speyer, R.: A systematic review of pragmatic language interventions for children with autism spectrum disorder. PloS one **12**(4), e0172242 (2017)
23. Pearson, K.: Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science **2**(11), 559–572 (1901)
24. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. the Journal of machine Learning research **12**, 2825–2830 (2011)
25. Prevention, C.f.D.C.a.: Screening and diagnosis of autism spectrum disorder. https://www.cdc.gov/ncbddd/autism/screening.html (2022)

26. Rumpf, A.L., Kamp-Becker, I., Becker, K., Kauschke, C.: Narrative competence and internal state language of children with asperger syndrome and adhd. Research in Developmental Disabilities **33**(5), 1395–1407 (2012)
27. Rutter, M., B.A., Lord, C.: The Social Communication Questionnaire: Manual. Western Psychological Services (2003)
28. Schaaf, C.P., Betancur, C., Yuen, R.K., Parr, J.R., Skuse, D.H., Gallagher, L., Bernier, R.A., Buchanan, J.A., Buxbaum, J.D., Chen, C.A., et al.: A framework for an evidence-based gene list relevant to autism spectrum disorder. Nature Reviews Genetics **21**(6), 367–376 (2020)
29. Sinica, A.: The chinese word segmentation system. http://ckipsvr.iis.sinica.edu.tw/
30. Sparck Jones, K.: A statistical interpretation of term specificity and its application in retrieval. Journal of documentation **28**(1), 11–21 (1972)
31. of Social Welfare in Taiwan, N.D.: The annual statistics of people with disabilities in 2023. https://dosw.gov.taipei/cp.aspx?n=BA5B128CF7454DDC (2023)
32. of Technology, H.I.: stopwords. https://github.com/goto456/stopwords/blob/master
33. Tsai, I.F.: A study of chinese language sample analysis for 3-5 years old children., unpublished Manuscript
34. Tsay, J.S.: Taiwan child language corpus: data collection and annotation. In: Proceedings of the Fifth Workshop on Asian Language Resources (ALR-05) and First Symposium on Asian Language Resources Network (ALRN) (2005)
35. Tseng, Y.H., .L.H.: Examining performance of expository and conversational discourse in mandarin-speaking children with language impairment. Journal of Special Education **46**, 1–30 (2017)
36. Tseng, Y.H., .L.H.: Investigating performance of expository discourse in mandarin-speaking children with language impairment: Language sampling analysis. Bulletin of Special Education **48**(1), 31–60 (2023)
37. Wawer, A., Chojnicka, I.: Detecting autism from picture book narratives using deep neural utterance embeddings. International Journal of Language & Communication Disorders **57**(5), 948–962 (2022)
38. Wei, J., Zou, K.: Eda: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196 (2019)
39. Wiesner, D.: Tuesday. Houghton Mifflin Harcourt (1991)
40. Wong, Y.S.: Utility of the screening tool for autism in two-year-olds (stat) and the autism diagnostic observation schedule (ados) for detecting autism spectrum disorder in toddlers under aged 24 months: A follow-up study., unpublished Manuscript
41. Wu, C.Y.: Urban and rural differences in age at initial diagnosis and healthcare utilization among pre-school children with autism, unpublished Manuscript
42. Wu, Q.: Exploring the reliability of oral language proficiency indicators in children with communication disorders, unpublished Manuscript
43. Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M.S., Saxena, S., Yusuf, A., Shih, A., Elsabbagh, M.: Global prevalence of autism: A systematic review update. Autism research **15**(5), 778–790 (2022)
44. Zhou, J.: Research on Chinese Children's Language Development: Application and Development of International Children's Corpus Research Methods. Educational Science Press (2009)