Jurnal Teknologi Pendidikan, April 2025, 27 (1), 137-149
DOI: http://dx.doi.org/10.21009/JTP2001.6
p-ISSN: 1411-2744 e-ISSN: 2620-3081
Accredited by Directorate General of Strengthening for Research and Development
Available online at http://journal.unj.ac.id/unj/index.php/jtp

Automated Framework for Communication Development in Autism Spectrum Disorder Using Whisper ASR and GPT-4o LLM

Naela Fauzul Muna[1(*)], Mukhammad Andri Setiawan[2]
[1,2]Universitas Islam Indonesia, Yogyakarta, Indonesia

Received : March 21, 2025
Revised : March 30, 2025
Accepted : April 29, 2025

Abstract

Autism Spectrum Disorder (ASD) is a developmental condition impacting communication, social interaction, and behavior. Communication assessments for children with ASD are often conducted manually, making the process time-consuming, which can lead to delays in developing educational programs and a lack of standardization due to subjective evaluations. This study introduces an automated framework using Whisper and GPT-4o to enhance the efficiency and accuracy of evaluating communication abilities and language patterns in children with ASD. The research adopts a Research and Development (RnD) approach with the ASET model (Analyze, System Design, Execution, Testing), engaging children with mild and moderate verbal ASD and teachers from four autism schools in Daerah Istimewa Yogyakarta, Indonesia. Data were collected through interviews, classroom observations, audio recordings, and a matrix-based evaluation. Whisper was employed for automated transcription, integrated with GPT-4o for speaker diarization and communication analysis. Results showed an 89.1% reduction in analysis time compared to manual methods. Whisper achieved a low Word Error Rate (WER) for mild autism (average 5%) and a higher rate for moderate autism (average 23%). GPT-4o contributed to the process with high speaker diarization accuracy (93.9% for mild autism and 89.2% for moderate autism). The framework identified detailed communication improvements through the matrix-based evaluation, including verbal, pragmatic, semantic, sentence structure, and echolalia aspects. It provided insights previously undetected by teachers, such as specific developmental patterns in each aspect. The future research should integrate intonation and emotional analysis, refine diarization accuracy, and validate the approach across diverse populations.

Keywords: Autism Spectrum Disorder, Communication Assessment, Artificial Intelligence, Whisper Automatic Speech Recognition, GPT Large Language Model

(*) Corresponding Author: naelafauzulm@gmail.com

How to Cite: Muna, N. F., & Setiawan, M. A. (2025). Automated Framework for Communication Development in Autism Spectrum Disorder Using Whisper ASR and GPT-4o LLM. *JTP - Jurnal Teknologi Pendidikan*, *27*(1), 137–149. https://doi.org/10.21009/jtp.v27i1.54243

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that impacts communication, social interaction, and behavior due to abnormal brain development (Reddy Ananthula Akash, 2023; Hodis et al., 2025; World Health Organization, 2023). The diagnosis of ASD is based on the observation of behavioral characteristics, focusing on communication, social interaction, and repetitive patterns of behavior (Harrison et al., 2021; Grabrucker, 2021). Globally, the prevalence of autism has been increasing. The World Health Organization

estimates that 1 in 100 individuals is diagnosed with ASD (Zeidan et al., 2022; Talantseva et al., 2023). In Indonesia, the National Statistik Center, Badan Pusat Statistik, estimates there are 3.2 million children with ASD out of a total population of 270.2 million (Rahmawati, 2024; Dwi Pratiwi et al., 2023).

The increasing prevalence of ASD highlights the need to address key challenges, particularly in communication and social (Fuller & Kaiser, 2020). Ideally, communication assessments for children with ASD should be rapid to enable timely interventions that significantly enhance developmental outcomes and quality of life (Okoye et al., 2023; Desideri et al., 2021), as well as accurate and standardized to ensure consistency and reliability (Kasari et al., 2013; Trembath et al., 2019). These assessments should comprehensively analyze linguistic features, as children with ASD often struggle with pragmatics, grammar, semantics, syntax, phonology, and morphology (Vogindroukas et al., 2022; Schaeffer et al., 2023).

However, the current reality reveals that the assessment of communication is still dependent on manual transcription and analysis by trained professionals, which is both time-consuming and resource-intensive (O'Sullivan et al., 2023). Moreover, interviews and observations with teachers at specialized autism schools indicate that progress evaluations are frequently based on subjective judgments, such as intuitive impressions, resulting in inconsistencies and imprecise tracking of developmental progress. This significant gap, between the need for an efficient, objective assessment system and the prevailing slow, subjective manual methods creates an urgent need for research to develop automated solutions that can expedite interventions and improve communication support for children with ASD.

To overcome these challenges, we have explored various approaches to identify and leverage the most effective methods or technologies for improving the assessment and analysis of communication in individuals with ASD. One approach utilized BERT and ChatGPT-3 (text-davinci-003) models to detect autism by analyzing text dialogues between parents and children through sentiment analysis. The proposed system shows promise for affordable and accessible ASD detection through text analysis. However, a gap exists in ChatGPT-3, where the model performs well in understanding text but is less efficient in handling natural language responses that involve data aggregation, such as performing calculations or analyzing data for tasks like cosine similarity (Mukherjee et al., 2023).

Another approach involved the development of LENA Autism Screening, which uses audio recordings and statistical algorithms to differentiate children's vocalizations from environmental sounds, enabling early autism detection. This approach is valuable for identifying signs of ASD from a child's speech. However, a room for improvement exists in the need for more comprehensive and standardized data collection from children diagnosed with ASD, and it requires tools that come with significant costs (Richards et al., 2010).

More recently, a study implemented OpenAI's Whisper model for ASR, demonstrating its ability to analyze linguistic features relevant to ASD diagnosis and providing outputs that align with the Vineland Adaptive Behavior Scales-II (VABS-II). However, a gap in this research is the manual speaker diarization, which is necessary to distinguish between the child and the therapist or parent in conversations. Another limitation in this research is the limited scope of linguistic features that have been tested, such as the number of communication units (c-units)

per minute, the speech repetition (echolalia), and morphological complexity (O'Sullivan et al., 2023).

Despite recent advancements, several challenges remain in current approaches to speech analysis for ASD: 1) Speaker diarization issues, requiring manual identification of speakers (e.g., child, therapist, parent) in recorded conversations (O'Sullivan et al., 2023); 2) Narrow linguistic feature analysis, which overlooks important ASD characteristics like communication patterns, sentence complexity, echolalia, and other relevant markers (O'Sullivan et al., 2023); and 3) Limited capacity for processing natural language responses, as seen in systems like ChatGPT-3, which struggle with speech repetition analysis and multi-marker data aggregation (Mukherjee et al., 2023).

To address these limitations, this research proposes an integrated solution combining OpenAI's Whisper model with GPT-4o to develop an automated framework for transcribing speech and analyzing key linguistic features, such as sentence complexity, echolalia, and pragmatics. Additionally, GPT-4o will enable automatic speaker diarization, distinguishing between the child and the therapist or parent. The data will be collected from children officially diagnosed with ASD to ensure the framework is based on real-world scenarios. To overcome the subjective assessments currently used by teachers, we introduce a Matrix-Based Evaluation framework, developed from the literature review and validated through feedback from teachers at specialized autism schools.

The primary purpose of this research is to develop an efficient and automated framework for assessing the communication abilities of children with Autism Spectrum Disorder (ASD), integrating OpenAI's Whisper model for transcription and GPT-4o for speaker diarization and linguistic analysis. This study aims to address the inefficiencies and subjectivity of manual assessment methods by creating a standardized, technology-driven solution that supports teachers in specialized autism schools in monitoring and enhancing communication progress. Specifically, the research expects to achieve a significant reduction in analysis time, deliver detailed and objective insights into linguistic features such as pragmatics, semantics, and echolalia, and provide educators with actionable data to design tailored interventions, ultimately improving developmental outcomes for children with ASD.

METHODS

The objective of this research is to develop an efficient and automated framework integrating OpenAI's Whisper and GPT-4o to assess communication abilities in children with ASD, addressing inefficiencies and subjectivity in manual methods while enhancing progress monitoring for educators.

This research adopts a Research and Development (R&D) approach using the ASET model (Analysis, System Design, Execution, Testing), a custom model designed to systematically develop and evaluate an automated communication assessment tool for children with Autism Spectrum Disorder (ASD). The ASET model comprises four phases: Analysis, which involves identifying the need for a rapid, objective assessment tool through interviews and observations of current

manual methods; System Design, which entails structuring the framework with Whisper for transcription, GPT-4o for speaker diarization and analysis, and a matrix-based evaluation for communication assessment; Execution, which focuses on developing and implementing the system using real-world data from ASD classrooms; and Testing, which evaluates the framework's performance in terms of technical accuracy, time efficiency, and practical utility. This model was selected for its linear, product-focused approach, aligning with the study's emphasis on creating and testing a functional technology solution tailored to the specific needs of ASD educators and students.

The study took place from April to December 2024, to provide a diverse range of data across various levels of ASD severity, we conducted research in several specialized autism schools in Daerah Istimewa Yogyakarta, including SLB Bina Anggita, SLB Dian Amanah, SLB Fredofios, and SLB Fajar Nugraha. The population of this study included two groups: children diagnosed with ASD who have verbal communication abilities, and teachers actively involved in educating these children in specialized autism schools.

The research employed a combination of instruments to gather data, we utilized semi-structured interviews, direct classroom observations, audio recordings of conversations, and a matrix-based evaluation framework. We conducted Semi-structured interviews to explore key communication characteristics of children with ASD and the common language patterns used by them. Classroom observations captured real-time communication behaviors during regular class activities, providing a deeper understanding of how children communicate in natural settings. We utilized audio recordings of teacher-student conversations to transcribe and analyze communication progress.

To evaluate communication progress, we developed a matrix-based evaluation based on insights from the literature review. This matrix was then validated through teacher interviews regarding the communication characteristics experienced by children with autism as identified in the literature review, as well as through classroom observations. It focuses on five critical aspects of communication: verbal ability, pragmatics, semantics, sentence structure, and echolalia. Additionally, reliability was assessed by checking whether the matrix provides consistent results when used at different times for the same child (test-retest reliability), ensuring stability in the evaluation process over time. The final version of the matrix, including descriptions and defined indicators for each aspect

Data Collection

We conducted semi-structured interviews with teachers and direct classroom observations to explore the communication behaviors of children with ASD. We designed interview questions based on a literature review to align with the research objectives, while observations focused on real-time behaviors, particularly vocal characteristics (intonation, volume, response speed), language structures (syntax, grammar, word usage), and communication content (response relevance and echolalia patterns).

In addition to the interviews, we conducted classroom observations with all students with autism from the selected schools to examine how they communicate

during regular classroom activities. These observations focused on various aspects of communication, including vocal characteristics, language structure, and communication content.

To capture authentic communication patterns in natural settings, we recorded conversations among teachers and students between November and Desember 2024 during one-month learning activities. We collected audio recordings from a specialized autism school, involving two students with mild verbal autism and their teacher, as well as two students with moderate verbal autism and their teacher. They transcribed and analyzed these recordings using a matrix-based evaluation framework that focuses on five critical communication aspects.

Data were analyzed quantitatively (e.g., percentage of relevant responses, echolalia frequency) and qualitatively (e.g., textual summaries per aspect) using GPT-4o and the matrix-based evaluation, with visualizations (bar charts, line graphs) to track progress over time.

Matrix Based Evaluation

The matrix-based evaluation framework, developed from the literature review, interviews, and observations, focused on five key communication aspects: verbal ability, pragmatics, semantics, sentence structure, and echolalia. This matrix is used to evaluate the communication progress of each child by analyzing their verbal ability, pragmatic skills, sentence structure, and frequency of echolalia. Each aspect is assessed based on the child's response in natural settings, where teachers' and students' interactions provide the data for analysis.

The matrix was validated using expert judgment from six experienced teachers at specialized schools for autism. The validation process involved a thorough review of the matrix indicators to confirm their alignment with real classroom practices and the communication challenges typically encountered by children with ASD. The final version of the matrix, including descriptions and defined indicators for each aspect, is detailed in Table 1.

Table 1. Matrix Based Evaluation

Assessment Aspect	Description	Indicator
Verbal	Ability to speak and produce sounds	Producing sounds, uttering meaningful words.
Pragmatics	Social interaction ability	Greeting, responding when called, engaging in two-way conversations, saying words like "thank you" and "please."
Semantics	Understandi-ng word meanings and instructions	Understanding simple instructions, everyday vocabulary, creating and understanding question sentences, expanding vocabulary.
Sentence Structure	Sentence formation and grammar	Forming simple sentences, adding affixes, using conjunctions.
Echolalia	Frequency of repeating words or phrases heard	Number and frequency of echolalia, repeating the ongoing topic or echoing the speaker's questions.

Automated Framework Flow

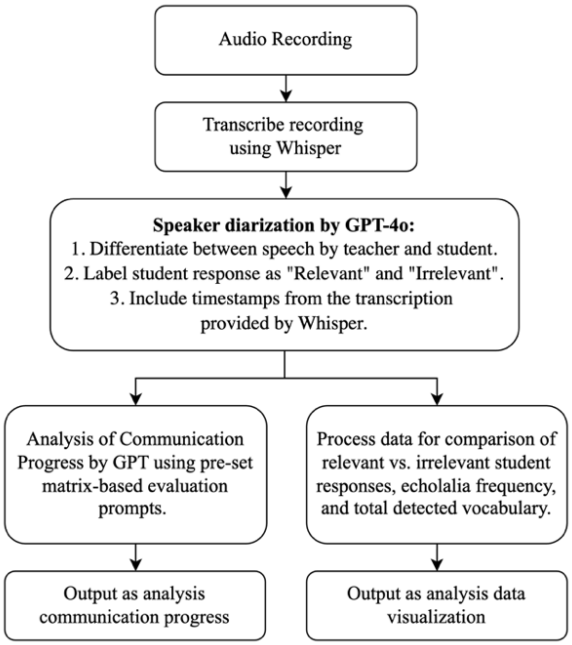

Figure 1. General System Flow

The analysis process, as outlined in Figure 1, begins with recording the conversation between the student and the teacher. The audio is then transcribed using Whisper, producing a textual output. Following the transcription, speaker diarization is performed using GPT to differentiate between the student and teacher's speech and to label the student's responses as "Relevant" or "Irrelevant". Timestamps are also added to each segment of the conversation.

The structured data is then subjected to communication progress analysis using a matrix-based evaluation, as detailed in Table 1. Specifically, the communication progress analysis aims to evaluate the communication abilities of children with ASD over time by comparing the current diarized transcription with previous versions, using GPT-4o to track developmental changes. This analysis employs the matrix to assess five key aspects: verbal, pragmatics, semantics, sentence structure, and echolalia, assigning scores or labels based on predefined criteria.

GPT-4o identifies patterns, such as an increase in relevant responses or a reduction in echolalia, providing a comprehensive overview of communication progress. The resulting data is analyzed both quantitatively and qualitatively. Quantitative analysis includes metrics such as the percentage of relevant responses and the frequency of echolalia, while qualitative analysis provides textual summaries for each aspect: verbal ability, pragmatics, semantics, sentence structure, and echolalia, as well as an overall summary (e.g., "The student enhanced sentence structure by consistently forming complete sentences"). To enhance interpretability, data visualizations are generated, including bar charts depicting scores for each aspect and line graphs illustrating trends over time. The outcomes of this analysis are presented as both detailed textual summaries and visual

representations, enabling educators to effectively monitor developmental progress and tailor interventions to the specific needs of children with ASD.

Automatic Speech Recognition Using Whisper

This study employs Whisper, an advanced automatic speech recognition (ASR) model developed by OpenAI and released in September 2022. Trained on a diverse multilingual and multitask dataset comprising 680,000 hours of audio, Whisper demonstrates robust capabilities in recognizing various accents, mitigating background noise, and interpreting technical terminology with high accuracy. For the purposes of this research, the Whisper v2-large model, accessed through the OpenAI API under the identifier "whisper-1," was utilized due to its superior performance in achieving precise speech recognition across a range of contextual settings.

The implementation of Whisper in this study entails a systematic process that commences with efficient audio management, utilizing the MPEG4 AAC (M4A) format, which is selected for its ability to deliver high-quality audio while maintaining a reduced file size compared to alternatives such as WAV. Following audio preparation, the files are submitted to the Whisper model for transcription, whereby the system identifies spoken content and converts it into text with exceptional fidelity, providing a critical foundation for subsequent analytical procedures within the research framework.

Speaker Diarization Using GPT-4o

After generating transcriptions with Whisper, the subsequent phase involves speaker diarization to determine the identities of speakers, such as teachers or students, within the conversation. This process commences with the transformation of Whisper's verbose JSON transcription output into a condensed data model, which reduces data volume and facilitates expedited processing by GPT-4o while circumventing constraints on input length. The transcription is then segmented into 60-second intervals, creating manageable units that enhance the efficiency and effectiveness of GPT-4o's analysis, with each segment subsequently forwarded to GPT-4o for diarization. In this step, GPT-4o employs a structured prompt comprising persona, task, instructions, examples, output format specifications, and user requests, to identify conversational patterns and distinguish between speakers, a method that has been demonstrated to markedly improve the performance of GPT-based models. This technique allows GPT-4o to differentiate teacher and student speech patterns, annotate relevant and irrelevant student responses, and append timestamps to each segment for straightforward temporal tracking.

Communication Progress Analysis Using GPT-4o

After the transcription is processed by GPT-4o and yields a new version with speaker diarization formatting, the next step is to analyze the communication progress of children with ASD in comparison to the previous transcription. This analysis is performed with the assistance of GPT-4o, enabling a detailed comparison and evaluation of the communication progress over time. The analysis is conducted based on a matrix-based evaluation, as outlined in Table 1, which is incorporated into the prompt to guide the evaluation process.

Framework Evaluation and Validation

To assess and validate the framework's performance, a multifaceted evaluation strategy was adopted to confirm its efficacy and precision in analyzing conversations involving children with Autism Spectrum Disorder (ASD). This process entails a comparative analysis of the time efficiency, wherein the duration required by the framework to complete transcription, analysis, and reporting is juxtaposed with that of conventional manual methods, thereby quantifying improvements in processing speed. Furthermore, the framework's analytical outputs are systematically compared with those produced through manual assessments by teachers, allowing for an assessment of its accuracy and the richness of insights generated, thus demonstrating its capacity to deliver detailed and reliable evaluations of communication patterns in children with ASD.

RESULTS & DISCUSSION

Result of Whisper

The performance of the Whisper model in transcribing conversations of children with Autism Spectrum Disorder (ASD) was evaluated using the Word Error Rate (WER), defined by the equation:

$$WER = \frac{S+D+I}{N} \times 100\% \quad (1)$$

Where (S) represents substitutions, (D) represents deletions, (I) represents insertions, and (N) is the total number of words in the reference transcription. The results indicate that the average WER for children with mild autism is relatively low, ranging between 0.56% and 10%, with a median value of approximately 5%. The narrow distribution of WER values reflects the consistency of the transcription results, highlighting the ability of the model to accurately capture structured and clear speech.

Conversely, the average WER for children with moderate autism is significantly higher, ranging from 13% to 46%, with a median value of approximately 23%. The wide spread of WER values for this group indicates a higher degree of complexity in their communication patterns.

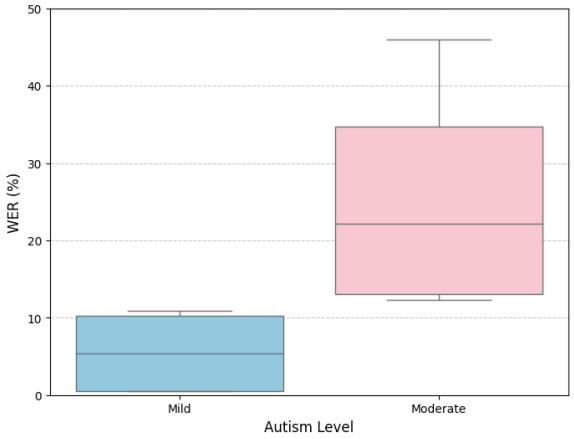

Figure 2. Word Error Rate (WER) of Whisper

Result of Speaker Diarization Using GPT-4o

Building on the transcription results, speaker diarization accuracy, enabled by GPT-4o, was evaluated to measure the system's ability to distinguish between speakers. The accuracy was determined using:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Predictions} \quad (2)$$

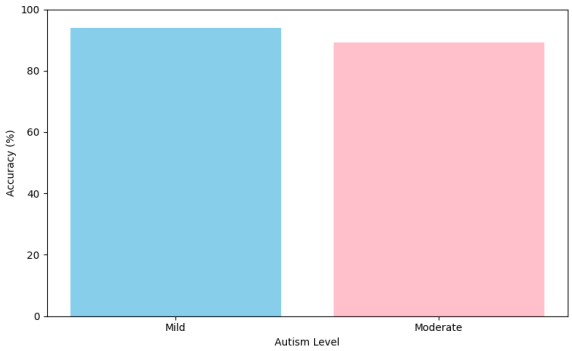

Figure 3. Accuracy of Speaker Diarization by GPT-4o

Figure 3 illustrates the performance of the system across the two autism categories. For mild autism, the system achieved an accuracy of 93.9%, indicating its high performance in identifying speakers within this category. In contrast, for moderate autism, the system's accuracy was slightly lower at 89.2%, reflecting the challenges in processing complex speech patterns typical of this group. The comparison reveals a performance gap of 4.7% between the two categories, suggesting that the system is more effective when dealing with structured and clear speech patterns.

Result of Speaker Diarization Using GPT-4o

The automated framework, integrating Whisper and GPT-4o, not only delivers robust transcription and diarization results but also significantly boosts efficiency compared to manual methods. A detailed comparison of analysis time is presented in Table 2.

Table 2. Comparison of Analysis Time Between Manual and Automated Methods

Analysis Stage	Manual Analysis (Minutes)	Automated Framework (Minutes)	Time Reduction
Transcription	Not Conducted	0,2	-
Speaker Diarization	Not Conducted	0,7	-
Communication Analysis	3	0,3	90%
Report Writing	8	0	100%
Total Time	11	1,2	89.1%

Table 2 presents a comparison between the proposed automated framework and manual methods. In the manual approach, transcription and speaker diarization were not conducted. The automated system efficiently completed transcription in 0.2 minutes and speaker diarization in 0.7 minutes. For communication analysis, the manual process required 3 minutes, while the automated framework reduced this to 0.3 minutes, achieving a 90% time reduction. Likewise, report writing, which took an average of 8 minutes in the manual approach, was eliminated entirely (0 minutes) by generating structured results that are ready for use, resulting in a 100% time reduction.

Discussion

The findings of this study reveal that the verbal communication abilities of children with Autism Spectrum Disorder (ASD) play a pivotal role in shaping the performance of automated tools like the Whisper Automatic Speech Recognition (ASR) system and the GPT-4o-powered speaker diarization system. For children with mild autism, whose speech is typically clearer and more structured, Whisper delivered optimal transcription performance, reflected in low and stable Word Error Rate (WER) values. This outcome aligns with prior research, such as that noted in (O'Sullivan et al., 2023), which suggests that ASR systems thrive when processing predictable and articulate speech patterns. Conversely, children with moderate autism, who often exhibit more intricate communication challenges like echolalia, articulation difficulties, and less predictable speech, yielded higher WER values. This discrepancy highlights a key limitation in current ASR technology: its struggle to adapt to the atypical speech patterns prevalent across varying autism severity levels.

This pattern of performance extends to speaker diarization, where GPT-4o demonstrated notable strengths and limitations. For children with mild autism, the system achieved an impressive accuracy of 93.9%, benefiting from structured, consistent two-way interactions that facilitate clear speaker differentiation. In contrast, for those with moderate autism, accuracy dipped to 89.2%. This decline stems from conversational complexities, repetitive speech, limited engagement, and frequent question repetition by teachers, that introduce ambiguities and challenge the model's ability to distinguish speakers. Despite this 4.7% performance gap, GPT-4o marks a significant leap forward from manual diarization methods, which, as (O'Sullivan et al., 2023) observed, were labor-intensive and susceptible to human error. By automating this process, the system offers a faster, more reliable, and scalable alternative, though it still necessitates teacher involvement, particularly for moderate autism cases, to ensure accuracy in real-world settings.

Beyond accuracy, the integrated framework combining Whisper and GPT-4o dramatically enhances efficiency and analytical depth. The total processing time plummeted from 11 minutes with manual methods to just 1.2 minutes, an 89.1% reduction, streamlining transcription, diarization, communication analysis, and report generation. This time-saving is invaluable in educational and clinical contexts, where swift, consistent evaluations are critical for customizing interventions. More importantly, the automated analysis surpasses manual methods by delivering richer, more detailed insights. While previous studies, such as (O'Sullivan et al., 2023), pointed to constraints in analyzing linguistic features like communication units per minute, echolalia, and morphological complexity, this framework overcomes those barriers. It provides a nuanced examination of verbal, pragmatic, semantic, sentence structure, and echolalia patterns, revealing insights previously unnoticed by teachers. For example, it identified developmental stagnation in some children while spotlighting vocabulary growth and improved social responses in others, offering a window into individual progress that can guide personalized therapeutic strategies.

The framework's real-world relevance was affirmed by teachers from four specialized autism schools in Daerah Istimewa Yogyakarta, Indonesia, whose feedback validated its alignment with their observational assessments. However, its performance was less robust for moderate autism cases, where greater speech variability exposed areas for improvement. This suggests that while the system excels with structured communication, adaptations, such as training on autism-specific datasets or fine-tuning for atypical speech. could broaden its effectiveness across the spectrum. In essence, this study advances autism communication analysis by quantifying the interplay between speech complexity and automated tool performance, bridging gaps in prior research, and delivering a scalable, efficient solution that enriches our understanding of ASD communication profiles while highlighting avenues for future refinement.

CONCLUSION

This study successfully develops an framework integrating OpenAI's Whisper model and GPT-4o to address the challenges in assessing communication abilities in children with Autism Spectrum Disorder (ASD). By automating transcription, speaker diarization, and the analysis of linguistic features such as verbal ability, pragmatics, semantics, sentence structure, and echolalia, the framework offers significant advancements over manual methods. The results indicate that the framework provides accurate and nuanced insights into communication progress. For children with mild ASD, it captures improvements in pronunciation, sentence complexity, and reduced echolalia with contextual relevance. For children with moderate ASD, it identifies stagnation and highlights challenges in verbal and semantic development. These findings underline the compatibility of GPT-4o for text-based analysis and its capacity to enhance ASD communication evaluations with detailed and objective assessments. Compared to manual methods, the framework achieves an 89.1% reduction in analysis time while offering deeper insights into communication progress, addressing gaps noted in prior studies, such as limited linguistic feature analysis and reliance on manual speaker diarization.

For comparison, a similar study by (O'Sullivan et al., 2023) utilized OpenAI's Whisper model for automatic speech recognition to analyze linguistic features in ASD, aligning outputs with the Vineland Adaptive Behavior Scales-II (VABS-II). Their approach demonstrated Whisper's efficacy in transcription but relied on manual speaker diarization and focused on a narrower set of features, such as communication units per minute and echolalia, without fully addressing pragmatics or sentence complexity. In contrast, our research introduces a novel integration of Whisper with GPT-4o, enabling automated speaker diarization and a comprehensive matrix-based evaluation of five critical communication aspects. This novelty lies in the combination of advanced ASR with large language model capabilities, providing a scalable, efficient, and holistic analysis that captures a broader spectrum of ASD communication patterns while eliminating manual diarization bottlenecks.

Despite these contributions, the framework has limitations. It is unable to analyze intonation, speech fluency, and emotional cues, which are essential for understanding social and emotional nuances in communication. Furthermore, errors in speaker diarization and difficulties in processing fragmented or ambiguous speech, especially in moderate ASD cases, require further attention. In conclusion, this study bridges key gaps in ASD communication assessment by providing a robust and automated solution with significant improvements over existing methods. Future research should focus on incorporating features such as intonation and emotional analysis, improving diarization accuracy, and validating the framework across diverse populations. This approach has the potential to empower educators and therapists in tailoring interventions and enhancing developmental outcomes for children with ASD, setting a foundation for more advanced, technology-driven ASD assessment tools.

REFERENCES

Desideri, L., Pérez-Fuster, P., & Herrera, G. (2021). Information and communication technologies to support early screening of autism spectrum disorder: A systematic review. *Children*, *8*(2). https://doi.org/10.3390/children8020093

Dwi Pratiwi, R., Dwi Pranata, A., Ayuningtyas, G., Azzahra, P., Program Studi, D. S., Widya Dharma Husada Tangerang, Stik., & Program Studi, M. S. (2023).

DETERMINAN KEJADIAN ANAK AUTIS BASED ON SYSTEMATIC REVIEW. In *Nursing Science Journal (NSJ)* (Vol. 4, Issue 2).

Fuller, E. A., & Kaiser, A. P. (2020). The Effects of Early Intervention on Social Communication Outcomes for Children with Autism Spectrum Disorder: A Meta-analysis. *Journal of Autism and Developmental Disorders*, *50*(5), 1683–1700. https://doi.org/10.1007/s10803-019-03927-z

Grabrucker, A. M. (2021). *Autism Spectrum Disorders*. Exon Publications, Brisbane, Australia. https://doi.org/https://doi.org/10.36255/exonpublications.autismspectrumdisorders.2021

Harrison, J. E., Weber, S., Jakob, R., & Chute, C. G. (2021). ICD-11: an international classification of diseases for the twenty-first century. In *BMC Medical Informatics and Decision Making* (Vol. 21). BioMed Central Ltd. https://doi.org/10.1186/s12911-021-01534-6

Hodis, B., Mughal, S., & Saadabadi, A. (2025). *Autism Spectrum Disorder*.

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. In *Autism Research* (Vol. 6, Issue 6, pp. 479–493). https://doi.org/10.1002/aur.1334

Mukherjee, P., Gokul, R. S., Sadhukhan, S., Godse, M., & Chakraborty, B. (2023). Detection of Autism Spectrum Disorder (ASD) from Natural Language Text using BERT and ChatGPT Models. *IJACSA) International Journal of Advanced Computer Science and Applications*, *14*(10). https://doi.org/10.14569/IJACSA.2023.0141041

Okoye, C., Obialo-Ibeawuchi, C. M., Obajeun, O. A., Sarwar, S., Tawfik, C., Waleed, M. S., Wasim, A. U., Mohamoud, I., Afolayan, A. Y., & Mbaezue, R. N. (2023). Early Diagnosis of Autism Spectrum Disorder: A Review and Analysis of the Risks and Benefits. *Cureus*. https://doi.org/10.7759/cureus.43226

O'Sullivan, J., Bogaarts, G., Kosek, M., Ullmann, R., Schoenenberger, P., Chatham, C., Nobbs, D., Murtagh, L., Lindemann, M., Parish-Morris, J., Liberman, M., Aponte, E., Dorn, J., & Lipsmeier, F. (2023). Automatic Speech Recognition for ASD Using the Open-Source Whisper Model from OpenAI. *International Society for Autism Research*.

Rahmawati, S. (2024). Optimalisasi Fokus: "Strategi Pembelajaran untuk Meningkatkan Konsentrasi pada Anak dengan Gangguan Spektrum Autisme (GSA)." In *Jurnal Kependidikan* (Vol. 13, Issue 2). https://jurnaldidaktika.org

Reddy Ananthula Akash, T. (2023). Comprehensive Review of Autism Spectrum Disorder: Etiology, Early Signs, and Diagnostic Assessment. *International Journal of Science and Research (IJSR)*, *12*(8), 480–485. https://doi.org/10.21275/sr23803085129

Richards, J. A., Xu, D., & Gilkerson, J. (2010). *Development and Performance of the LENA Automatic Autism Screen*.

Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., Hendriks, P., Kissine, M., Manenti, M., Marinis, T., Meir, N., Novogrodsky, R., Perovic, A., Panzeri, F., Silleresi, S., Sukenik, N., Vicente, A., Zebib, R., Prévost, P., & Tuller, L. (2023). Language in autism: domains, profiles and co-occurring conditions. *Journal of Neural Transmission*, *130*(3), 433–457. https://doi.org/10.1007/s00702-023-02592-y

Talantseva, O. I., Romanova, R. S., Shurdova, E. M., Dolgorukova, T. A., Sologub, P. S., Titova, O. S., Kleeva, D. F., & Grigorenko, E. L. (2023). The global prevalence of autism spectrum disorder: A three-level meta-analysis. In *Frontiers in Psychiatry* (Vol. 14). Frontiers Media S.A. https://doi.org/10.3389/fpsyt.2023.1071181

Trembath, D., Paynter, J., Sutherland, R., & Tager-Flusberg, H. (2019). Assessing Communication in Children with Autism Spectrum Disorder Who Are Minimally

Verbal. In *Current Developmental Disorders Reports* (Vol. 6, Issue 3, pp. 103–110). Springer. https://doi.org/10.1007/s40474-019-00171-z

Vogindroukas, I., Stankova, M., Chelas, E. N., & Proedrou, A. (2022). Language and Speech Characteristics in Autism. In *Neuropsychiatric Disease and Treatment* (Vol. 18, pp. 2367–2377). Dove Medical Press Ltd. https://doi.org/10.2147/NDT.S331987

World Health Organization. (2023). *Autism*.

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. In *Autism Research* (Vol. 15, Issue 5, pp. 778–790). John Wiley and Sons Inc. https://doi.org/10.1002/aur.2696