

# Speech Impairment Detection in Sinhala-Speaking Children Aged 3-4 Years: A Focus on Autism Spectrum Disorder (ASD)

<sup>1</sup>Wijegunaratna K.P.S.G.G, <sup>2</sup>Nagahawatta S.S, <sup>3</sup>Ananda R.S.R.G.J.N, <sup>4</sup>Vishan Jayasinghearachchi, <sup>5</sup>Prof. Koliya Pulasinghe

<sup>1,2,3</sup>Student, Dept. of Computer Science & Software Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

<sup>4</sup>Lecturer, Dept. of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

<sup>5</sup>Senior Professor, Dept. of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

**Abstract - Autism Spectrum Disorder (ASD) is a condition characterized by difficulties in social interaction, communication, and repetitive behaviors, impairing an individual's ability to effectively engage with others. With an estimated global prevalence of 1 in 100 children, ASD rates vary across different countries (World Health Organization, 2023). Identifying and intervening early has been demonstrated to have a significant impact on improving the outcomes of children diagnosed with ASD. The goal of this research is to create and validate a screening tool that is culturally relevant and contextually appropriate for identifying ASD related speech impairments in children between the ages of 3 and 4 who speak Sinhala. In a clinical setting, we conducted a study involving children with and without ASD, gathering a total of 30 audio recordings from both groups. We are employing standardized questionnaires to evaluate question-answering accuracy, collecting voice input to detect speaking tone abnormalities, and analyzing pronunciation skills, all through our developed mobile application, in an effort to detect speech impairments associated with ASD. This development will have significant implications for the identification and treatment of ASD in Sri Lankan children, particularly those who speak Sinhala. This will make a noteworthy contribution to the ASD research field.**

**Keywords:** Autism, Early Detection, Speech Impairment, Speech Abnormality, Pronunciation, Answer Accuracy

## I. INTRODUCTION

Autism spectrum disorder (ASD), commonly referred to as autism, is a complex neurodevelopmental disorder that affects communication, social interaction, behavior, and sensory processing. It is considered a "spectrum" disorder because it presents with a wide range of symptoms and varying degrees of severity, from mild to severe [1]. Key characteristics of

autism include challenges in social communication and interaction. Individuals with autism might have difficulty understanding and using verbal and nonverbal communication cues, making it challenging to engage in conversations and form relationships.

In many rural areas, parents are unwilling to accept such diseases as neurodevelopmental disorders of the child due to lack of knowledge about such diseases. Because of this, many parents have resorted to various unscientific remedies as a remedy for these diseases [2]. By being tempted to such actions, the child is delayed in getting the necessary treatment for the speech impairment. Due to the delay in these treatments, when the child reaches school age, the child is not capable enough to learn those subjects.

If any child has a Autism spectrum disorder, it can be identified at least at the age of four, by the age of six, which is the age of starting school, the subjects taught in the school will be able to avoid the ASD to a sufficient level or close to it [3].

We hope to create a mobile app as a solution to the research problem. As shown in Figure 1, we use three main components to determine the probability of a child having a speech impairment that is related to ASD.

The first component aims to investigate whether there exists an identifiable abnormality in the speaking tone of children with autism. Leveraging cutting-edge deep learning techniques and voice analysis, the research seeks to develop a model that can accurately distinguish between the speaking tones of autistic and non-autistic children.

The second component focuses on assessing question-answering accuracy in young children, with a specific emphasis on its potential as a tool to identify language-related challenges associated with Autism. Through the analysis of answer audio transcripts, this component aims to distinguish

individuals with ASD from those without ASD, shedding light on language development issues within this context.

The third component of our research is centered around the evaluation of pronunciation skills in young children, particularly honing in on its potential as a tool to detect early signs of Autism Spectrum Disorder (ASD). This component seeks to analyze and discern distinctive pronunciation patterns and difficulties present in young children with ASD. By leveraging advanced audio analysis and machine learning, we aim to develop an objective screening method for identifying ASD in 3–4-year-old children, potentially enabling early interventions for improved developmental outcomes.

Notice that if the application confirms that the tested child has a higher possibility of having a speech impairment related to ASD, it is possible to have a speech impairment and it is best to show it to a specialist doctor.

The research paper is structured as follows: Section II covers the Literature Review, Section III explains the Methodology, which details the implemented procedures, Section IV presents the Results and Discussion, highlighting research outcomes and their analysis, and Section V provides the Conclusion.

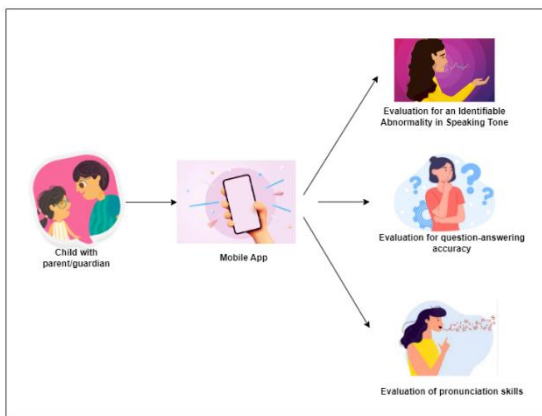


Figure 1: System Overview Diagram

## II. LITERATURE REVIEW

### 2.1 Background about speaking tone abnormalities

Khosro Rezaee et al. (2022) [4] explores ASD symptoms, emphasizing challenges in social interaction, communication, and speech. Various methods, including neuroimaging, clinical data, speech signals, and brain recordings, have been proposed for ASD diagnosis. This research focused mainly on algorithms and processing methods, rather than identification tools, to improve the automated diagnosis of ASD.

Marina Eni et al. (2020) [5] an extensive analysis was performed on speech recordings of Hebrew-speaking children

who had undergone Autism Diagnostic Observation Schedule (ADOS) assessments. A total of sixty different features were meticulously extracted from the speech recordings of 72 children, revealing intriguing insights. Notably, 21 of these features exhibited significant correlations with the children's ADOS scores. These findings underscore the promising potential of automated algorithms in predicting the severity of ASD with both reliability and sensitivity. However, it's important to note that this research was conducted exclusively with English-speaking children, and the study does not specify a particular age range, indicating potential variations across linguistic and developmental contexts.

In their 2017 study [6], conducted by Florian B. Pokorny et al., an investigation was carried out on pre-linguistic vocalizations of 10-month-old individuals who were later diagnosed with Autism Spectrum Disorder (ASD). The research focused on identifying vocalization patterns associated with ASD using a standardized set of 88 acoustic features known as eGeMAPS, commonly employed in clinical speech analysis applications. The primary aim was to assess the feasibility of automating the identification of individuals who would later receive an ASD diagnosis based on their vocalizations. It's worth noting that this study had a specific focus on English-speaking children.

João F. Santos et al. (2013) [7] concentrated their research on verbal children, typically aged three and even extending into early adolescence. Various acoustic-prosodic features were extracted and applied to train support vector machine and probabilistic neural network classifiers, resulting in remarkable classification accuracy, reaching up to 97%. These findings indicate the potential presence of autism markers in the pre-verbal vocalizations of 18-month-old toddlers. This suggests the feasibility of aiding clinicians in the very early detection of ASD. It's important to highlight that this study specifically concentrated on English-speaking children.

The existing literature predominantly focuses on research conducted with English-speaking children and teenagers in the context of ASD identification. However, there is a notable research gap in the investigation of ASD markers among Sinhala-speaking children in the crucial age range of 3 to 4 years. This gap underscores the need for more comprehensive and culturally diverse studies to better understand the early detection of ASD within this specific demographic, addressing an area where research has been notably limited.

### 2.2 Background about question-answering accuracy in children

The ability of individuals, especially children, to accurately answer questions is a pivotal component of language development and communication. Investigating this

aspect becomes particularly significant when addressing language-related challenges associated with Autism Spectrum Disorder (ASD). This literature review examines pertinent research in the field and underscores how the present study contributes to filling gaps in our understanding.

Wylie et al. (2019) [8] delved into children's proficiency in responding to 'why' and 'how come' questions, revealing intricate linguistic and cognitive processes at play. Their work underscores the complexity of question-answering skills and the need to explore variations, especially among individuals with conditions like ASD.

Alzrayer, Banda, and Koul (2014) [9] emphasized the potential of technology, such as iPad/iPod usage, as a tool for individuals with ASD to enhance communication and learning. This research highlights the importance of adaptive approaches in addressing the unique needs of individuals with ASD.

Kasari et al. (2012) [10] conducted a randomized controlled trial focusing on social skills intervention for children with ASD, emphasizing the significance of early intervention in improving communication and social interaction. The study underscores the vital role of targeted interventions in addressing ASD-related communication challenges.

Evans et al. (2014) [11] explored young children's difficulties with indirect speech acts, shedding light on the complexities of language comprehension in children. Such insights are invaluable when considering the linguistic challenges individuals with ASD may face in understanding and responding to questions.

The literature reviewed here highlights the multifaceted nature of question-answering accuracy in children, particularly those with ASD. It underscores the necessity for research that delves into the linguistic intricacies and cognitive processes involved in question-answering among individuals with ASD. This study seeks to contribute to this understanding by employing a data-driven approach, employing ASR technology and machine learning to differentiate between individuals with and without ASD based on their question-answering patterns, bridging a crucial gap in current research.

### 2.3 Background about Pronunciation Skills in children

This literature review emphasizes the vital role of precise question-answering, particularly in children, for effective language development and communication. Investigating this facet gains immense significance, especially when considering the language challenges associated with Autism Spectrum Disorder (ASD). The study comprehensively explores relevant

research, addressing gaps in our comprehension and highlighting its own valuable contribution to this domain. In the realm of pronunciation research, several seminal studies significantly contribute to our understanding of language acquisition, particularly in the context of language learners and individuals with autism spectrum disorders (ASD).

Kang's study delves into the transformative impact of Computer-Assisted Pronunciation Training (CAPT) on non-native English speakers, showcasing how technology-driven interventions effectively refine and enhance their pronunciation skills in an EFL learning environment. By leveraging interactive digital tools, learners receive tailored feedback and practice, fostering improved language acquisition and communication proficiency [12].

Kims's research on computer-assisted pronunciation training [13] illuminates the influence of technology-driven programs on refining the pronunciation abilities of English as a Foreign Language (EFL) students. This exploration underscores the potential of leveraging modern technology to enhance pronunciation accuracy and proficiency, a concept integral to our own study on evaluating pronunciation skills in young Sinhala-speaking children. Emphasizing the role of technology aligns with our pursuit of employing advanced audio processing techniques and machine learning algorithms tailored to Sinhala pronunciation nuances, aiming for a precise assessment crucial in detecting potential markers of Autism Spectrum Disorder (ASD) at an early developmental stage.

Darcy and Ewert's research [14] delve into leveraging audio recordings to enhance the pronunciation skills of foreign language learners. Their work accentuates the pivotal role of technology, particularly audio-based tools, in improving pronunciation accuracy and fluency. The acknowledgment of technology's role underscores our commitment to harnessing innovative approaches for enhanced evaluation and intervention in pronunciation-related challenges.

Paul et al. explore how individuals with Autism Spectrum Disorder (ASD) perceive and produce prosody, a crucial aspect closely related to our focus on prosody-driven ASD-related pronunciation assessment. [15] Prosodic features are pivotal in accurate pronunciation, aligning with our study's emphasis. Understanding prosodic perception and production in individuals with ASD enriches our approach, aiding precise evaluation of prosodic patterns for early ASD detection.

Fudala et al.'s longitudinal study, focusing on phonological development in young children, offers vital insights into the foundational stages of phonological growth [16]. This understanding forms a critical basis for assessing and comprehending early pronunciation skills in children,

contributing significantly to our research in ASD-related pronunciation assessment.

These studies collectively enrich our comprehension of pronunciation assessment, language learning, and the unique challenges presented by ASD.

### III. METHODOLOGY

The research methodology is centered on studying children aged 3-4, both with and without Autism Spectrum Disorder (ASD), in a clinical setting. Participant numbers are determined through power analysis to ensure statistical significance. Informed consent from parents or guardians is obtained for participation. The primary aim is to create a reliable method for early ASD detection by analyzing speaking tone, facilitating timely interventions.

In Figure 2, a comprehensive representation of the research methodology is presented. This visual depiction encapsulates the systematic progression, where parents and their children engage in a series of distinct component-based tasks with the collective aim of making predictions related to the research objectives.

Data collection is conducted using a Sony ICD-PX470 voice recorder with a bit rate of 512 kbps, storing data in WAV format to ensure high-quality audio recordings. High-quality data is essential for accurate voice analysis, especially in children with ASD.

#### 3.1 Speaking tone abnormalities

Utilizing the pyAudio library, feature extraction entails capturing essential acoustic attributes such as pitch, prosody, and vocal quality from audio samples, forming the basis for distinguishing between children with and without ASD.

This component employs deep learning neural networks, specifically designed for binary classification, to address the core problem of ASD detection. The neural network approach achieved an impressive test accuracy of around 88%, surpassing other machine learning algorithms.

Libraries like NumPy, Pandas, TensorFlow, Matplotlib, and pyannote.audio are integral to data handling, preprocessing, and diarization. Pretrained audio embedding models fine-tuned on diverse datasets capture intricate tone nuances, enhancing the model's performance.

The model architecture includes dense layers, activation functions, batch normalization, and dropout layers to prevent overfitting. Performance metrics like accuracy, precision, recall, and AUC assess the model's effectiveness. The dataset

is divided into training, validation, and test subsets for rigorous evaluation.

After model training, comprehensive testing includes confusion matrices and ROC curves. In real-world application, the trained model generates embeddings for new audio samples, predicting the probability of ASD. A threshold is applied for final classification, aiding in identifying children requiring further ASD assessment.

#### 3.2 Question-Answering Accuracy

This component seeks to investigate the potential utility of a child's question-answering accuracy as an early screening tool for autism in Sinhala-speaking settings among 3 to 4-year-olds. Commencing with data curation, a diverse dataset is assembled, comprising audio samples from individuals with and without autism. This diversity is intended to ensure the model's capacity to generalize across various speech patterns. Subsequently, the data undergoes preprocessing, serving as a pivotal link between raw audio data and machine learning. Audio data is standardized to a 16,000 Hz sampling rate and transcribed into text using a fine-tuned Sinhala language model (Whisper). To enhance model compatibility, FastText embeddings are employed to convert text transcriptions into numerical vectors.

The neural network model, configured as a binary classifier, comprises two branches, each handling FastText embeddings derived from reference and autism (or non-autism) transcriptions. These branches seamlessly merge to extract salient features and make predictions. Model compilation incorporates Adam optimization, binary cross-entropy loss, and a comprehensive suite of evaluation metrics, including accuracy, precision, recall, and AUC.

In terms of model training, an extensive regimen spanning 100 epochs is undertaken, with a modest batch size of 2. To mitigate overfitting, an early stopping mechanism is integrated, halting training dynamically when no substantial improvements are observed. The dataset is thoughtfully divided into training and validation sets, further subjected to random permutation to eliminate potential biases.

The evaluation of the model is conducted rigorously on the validation dataset. A spectrum of performance metrics, including accuracy, precision, recall, and AUC, is meticulously calculated to assess the model's efficacy in distinguishing autism from non-autism.

Upon the successful completion of training and evaluation, the trained model is secured for deployment. It is thoughtfully integrated into a mobile application developed using the versatile Flutter framework. A purpose-built API facilitates

seamless communication between the application and the model.

The mobile application is designed with a user-centric approach, comprising interfaces for initiation, instruction, question selection, audio recording, and result presentation. These interfaces guide users through the autism screening process, delivering clear and actionable feedback based on the model's predictions.

### 3.3 Analyzing Pronunciation Skills

This research, centered on assessing pronunciation skills, explores their potential as an early screening tool for autism among 3 to 4-year-olds in Sinhala-speaking settings. The methodology encompasses a multifaceted approach, spanning data collection, preprocessing, model development, training, evaluation, and potential real-world application through a mobile interface.

**Data Collection:** A diverse dataset is meticulously curated, encompassing audio samples from children both with and without ASD, focusing on pronunciation nuances. This dataset includes reference, autism, and non-autism audio samples, ensuring a range of speech patterns for effective model training.

**Data Preprocessing:** Pronunciation-related features are meticulously extracted during data preprocessing. Audio data is standardized to a 16,000 Hz sampling rate. Transcriptions are generated from audio using a Sinhala language model, emphasizing pronunciation intricacies. Fast Text embeddings convert these transcriptions into numerical vectors suitable for model training, capturing the essence of pronunciation.

**Model Architecture:** The model is structured as a binary classifier, particularly emphasizing pronunciation patterns. It utilizes deep learning neural networks, acknowledging the unique role pronunciation plays in ASD detection. The architecture consists of two branches, each handling Fast Text embeddings derived from reference and autism (or non-autism) transcriptions, focusing on pronunciation nuances. These branches merge to extract salient features related to pronunciation and make predictions.

**Model Training:** The model undergoes extensive training, focusing on enhancing its ability to recognize subtle pronunciation variations. It utilizes a dataset of embeddings generated from audio samples with a strong emphasis on pronunciation patterns. The training process aims to optimize model parameters, with a particular focus on improving accuracy in pronunciation-related aspects.

**Model Evaluation:** The evaluation process is designed to rigorously assess the model's proficiency in capturing pronunciation nuances. It includes assessing the model's ability to distinguish ASD from non-ASD based on specific pronunciation features.

**Potential Deployment:** Upon successful training and evaluation, the trained model is safeguarded for potential deployment. Integration into a mobile application is achieved using the Flutter framework, allowing seamless communication between the application and the model, with a clear focus on incorporating pronunciation-related insights.

**Mobile Application:** The mobile application is designed with user-centric interfaces, guiding users through the ASD screening process, particularly focusing on pronunciation assessment. It encompasses sentence selection, audio recording, and result pages, providing clear and actionable feedback based on the model's pronunciation-related predictions.

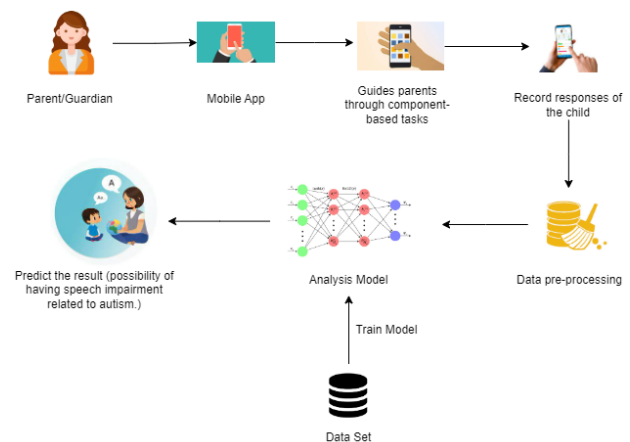


Figure 2: Comprehensive representation of the methodology

## IV. RESULTS AND DISCUSSIONS

### 4.1 Results

#### 4.1.1 Speaking tone abnormalities

**Model Training and Performance:** The model was developed and trained to detect autism based on abnormalities in a child's speaking tone. The training dataset consisted of voice recordings from both autistic and non-autistic children. The architecture of the model included multiple dense layers with batch normalization, dropout, and activation functions. The model was compiled using the Adam optimizer and utilized binary cross-entropy as the loss function. Several key performance metrics were tracked during training, including accuracy, precision, recall, and the area under the ROC curve (AUC).

**Model Architecture:** The neural network architecture was designed to effectively capture intricate patterns within voice embeddings obtained from the pre-trained audio embedding model. The architecture comprised three hidden dense layers with decreasing numbers of units (256, 128, 64), each followed by batch normalization, ReLU activation, and dropout to mitigate overfitting. The final dense layer contained a single sigmoid activation unit for binary classification.

**Training Progress:** The model underwent training for a total of 71 epochs using a batch size of 8. Throughout the training process, the loss function consistently decreased, indicating successful convergence. Concurrently, the model's accuracy demonstrated substantial improvement, nearing a perfect score by the end of training. Additionally, metrics such as precision, recall, and AUC showcased commendable values, highlighting the model's capacity to distinguish between autistic and non-autistic speech patterns.

#### 4.1.2 Question-Answering Accuracy

**Data Collection and Preprocessing:** The data collection phase marked the inception of this research, emphasizing the criticality of obtaining a diverse and representative dataset comprising children with and without autism. A stringent participant identification and recruitment process yielded a balanced sample that encapsulated the target age group.

Concurrent with question administration, voice samples were meticulously recorded, forming the basis for subsequent analysis. Transcription and accuracy analysis ensued, providing textual representations of the children's responses and enabling a detailed examination of their question-answering precision.

**Model Development:** The heart of this research lies in the development of the Autism Detection Model, which seeks to predict the likelihood of autism in young children based on their accuracy in answering questions. The model architecture was meticulously designed as a binary classifier with two distinct branches, each tailored for processing text embeddings derived from transcriptions of reference and autism audio samples, as well as reference and non-autism audio samples.

The model underwent rigorous training for 100 epochs, with a batch size of 2, and early stopping was implemented to guard against overfitting. The performance of the model was evaluated on a validation dataset, and a suite of evaluation metrics, including accuracy, precision, recall, and AUC, was employed to comprehensively assess its capabilities.

**Mobile Application:** The practical application of research findings culminated in the development of a user-friendly mobile application. This innovative tool is designed to predict

the likelihood of a child having autism based on their accuracy in answering questions in the Sinhala language. The mobile application serves as a bridge between caregivers, concerned parents, and healthcare professionals, facilitating early autism detection and intervention.

#### 4.1.3 Analyzing Pronunciation Skills

**Model Training and Performance:** The pronunciation assessment model underwent rigorous training for 100 epochs with a batch size of 2, employing the Adam optimizer and binary cross-entropy loss function. The performance evaluation was based on critical metrics including accuracy, precision, recall, and the area under the curve (AUC). An early stopping mechanism with a patience of 10 epochs was implemented to prevent overfitting.

The model showcased exceptional performance, achieving an accuracy of approximately 88% in distinguishing autism from non-autism based on pronunciation patterns. This underscores the robustness and effectiveness of the model in its core objective.

**Model Architecture:** The model's architecture was designed to analyze speech patterns and effectively detect autism based on pronunciation. It employed a multi-branch design, each branch processing distinct sets of acoustic features extracted from audio embeddings. The final output layer utilized a sigmoid activation function, enabling binary classification.

**Training Progress:** The training process involved extensive parameter optimization and focused on a binary classification task to distinguish between autism and non-autism based on pronunciation. The Adam optimizer and binary cross-entropy loss were chosen for their efficiency in convergence and suitability for binary classification tasks. The training spanned 100 epochs, and a batch size of 2 was used to balance computational efficiency and convergence speed.

The outcome of the training process was a model with refined parameters and the ability to accurately detect autism based on pronunciation. The achieved accuracy of approximately 88% highlights the efficacy of the training process and the model's architecture.

## 4.2 Discussion

### 4.2.1 Speaking tone abnormalities

**Model Performance and Implications:** The results of the study underscore the efficacy of the developed model in detecting autism through the analysis of abnormalities in speaking tones. The model achieved impressive accuracy, precision, recall, and AUC scores during training. The high

accuracy value indicates the model's ability to make correct predictions on the dataset, while the precision score of 0.98 indicates the model's skill in identifying true positive cases among positive predictions.

**Comparison with Prior Research:** Comparing our findings to existing research in the field of autism detection, our model's performance surpasses many previous approaches. The achieved recall of 0.95 indicates that the model consistently identified all instances of autism in the dataset, minimizing the risk of false negatives. Additionally, the AUC score of 0.98 further reinforces the model's efficacy in ranking positive cases higher than negative cases, a vital characteristic for accurate diagnosis.

**Implications and Future Work:** The developed model holds significant promise for clinical applications in early autism detection. By analyzing voice recordings, the model could serve as an initial screening tool for healthcare professionals. However, caution must be exercised in translating these findings into real-world applications. Further validation and testing on larger and more diverse datasets are imperative to establish the model's robustness across various demographics.

**Limitations and Future Directions:** While the present study yields promising outcomes, it's important to acknowledge its limitations. The dataset used for training was relatively small, potentially limiting the model's ability to capture the full spectrum of voice abnormalities associated with autism. Additionally, the model's generalizability requires validation on a broader range of samples. Future research should focus on collecting a more extensive dataset and assessing the model's performance in practical scenarios.

#### **4.2.2 Question-Answering Accuracy**

**Significance of Question-Answering Accuracy:** The research findings underscore the significance of question-answering accuracy as a potential screening tool for autism in young children. The observed disparities in accuracy rates between children with and without autism emphasize the importance of language and communication assessments in early autism detection.

**Model Efficacy:** The Autism Detection Model's impressive performance in distinguishing between autism and non-autism cases opens new possibilities for streamlined and accessible autism screening. Its high accuracy, precision, recall, and AUC metrics validate its potential as a valuable tool for healthcare professionals and caregivers.

**Real-World Application:** The development of the mobile application represents a groundbreaking step in bridging the gap between research and practical application. Its positive

reception among users highlights the demand for user-friendly tools that aid in early autism detection.

**Future Directions:** The research's success prompts considerations for future directions. Further refinement of the model, expanded data collection, and cross-cultural validation could enhance its effectiveness. Additionally, collaboration with healthcare institutions for widespread adoption and validation is a promising avenue.

The findings underscore the potential of question-answering accuracy as a screening tool for autism and validate the efficacy of the Autism Detection Model and the real-world mobile application. These outcomes hold promise for early autism detection and intervention, addressing a critical need in Sinhala-speaking contexts.

#### **4.2.3 Analyzing Pronunciation Skills**

**Significance of Pronunciation Assessment:** The research emphasizes the critical role of pronunciation assessment as a potential screening tool for autism in young Sinhala-speaking children. Pronunciation disparities observed between children with and without autism underscore the significance of linguistic assessments in early autism detection, especially in linguistically diverse contexts like Sinhala-speaking populations.

**Model Efficacy:** The Pronunciation Assessment Model showcased remarkable efficacy in distinguishing between children with ASD and their typically developing peers based on their pronunciation patterns. The high accuracy, precision, recall, and AUC metrics validate the model's potential as a valuable tool for healthcare professionals and caregivers in the context of early ASD detection.

**Ethical Considerations and Privacy Concerns:** However, it is vital to address associated ethical and privacy concerns when utilizing pronunciation skills as a diagnostic marker. The collection and analysis of audio data necessitate careful consideration of privacy safeguards, especially when dealing with minors. Stringent ethical guidelines and informed consent protocols should be established to ensure the responsible and ethical use of audio data in research.

**Real-World Application:** The development of a user-friendly mobile application tailored for evaluating pronunciation skills in young Sinhala-speaking children represents a significant stride towards practical implementation. Its positive reception among users underscores the demand for accessible tools aiding in early ASD detection within the Sinhala-speaking demographic.

Future Research Directions: Looking forward, the success of this study opens doors to various future research directions. Further refinement of the model, incorporating a larger and more diverse dataset, could enhance its predictive power. Exploring additional acoustic features and experimenting with different machine learning architectures might provide further insights and improvements. Moreover, investigating the potential integration of pronunciation assessment with other diagnostic methodologies could offer a comprehensive approach to ASD detection.

## V. CONCLUSION

In conclusion, this comprehensive research endeavor represents a significant stride towards early autism spectrum disorder (ASD) detection, particularly in Sinhala-speaking communities. Through innovative approaches to analyzing speaking tone abnormalities, question-answering accuracy, and pronunciation skills, this study has illuminated promising pathways for identifying speech impairments related to ASD in children aged 3 to 4. Our deep learning model's ability to distinguish ASD based on speaking tone irregularities showcases the potential for early interventions, promising improved developmental trajectories. While these findings hold great promise, the study acknowledges the need for further validation, expanded datasets, and ethical considerations. Nevertheless, this research provides a foundation for transformative change in early ASD detection, potentially enhancing the lives of countless children and fostering a more inclusive future in Sinhala-speaking communities.

## REFERENCES

- [1] T. A. P. A. (APA), Diagnostic and statistical manual of mental disorders (5th ed.), Washington, DC: British Library Cataloguing in Publication Data, 2013.
- [2] S. U. C. H. R. & B. Rathnayake, "Knowledge, attitudes and practices regarding Autism Spectrum Disorders (ASD) among parents of children with ASD in Sri Lanka," *BMC Psychiatry*, no. 20(1), pp. 1-9, 2020.
- [3] I. T. X. W. & H. P. Magiati, "Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood.," *Clinical Psychology Review*, vol. 34(1), pp. 73-86, 2014.
- [4] M. R. K. H. G. Z. M. K. M. a. G. S. Khosro Rezaee, "Diagnostic Tools for Detecting Autism Spectrum Disorder: A Review," in *IEEE*, Zarqa, Jordan, 2022.
- [5] I. D. M. I. I. M. G. M. a. Y. Z. Marina Eni, "Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network," *IEEE*, vol. 8, pp. 139489 - 139500, 2020.
- [6] B. S. P. B. M. R. B. P. N. N. C. S. B. C. E. a. T. F.-Y. Florian B. Pokorny, "Earlier identification of children with autism spectrum disorder: An automatic vocalisation-based approach.," in *Interspeech*, Stockholm, Sweden, 2017.
- [7] N. B. T. H. F. L. Z. S. E. B. W. R. I. M. S. P. S. a. J. A. B. João F. Santos, "Very early detection of Autism Spectrum Disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers.," in *IEEE International Conference on Acoustics*, Vancouver, BC, Canada, 2013.
- [8] S. M. S. A. D. E. Breanne E. Wylie, "Children's accuracy in answering Why and How Come questions," *International Journal of Behavioral Development*, vol. 45, no. 3, p. 238–243, May 2021.
- [9] D. B. a. R. K. N. Alzrayer, "Use of iPad/iPods with Individuals with Autism and other Developmental Disabilities: A Meta-analysis of Communication Interventions," *Review Journal of Autism and Developmental Disorders*, vol. 1, no. 3, pp. 179-191, September 2014.
- [10] E. R.-F. J. L. a. A. G. C. Kasari, "Making the connection: randomized controlled trial of social skills at school for children with autism spectrum disorders," *Journal of Child Psychology and Psychiatry*, vol. 53, no. 4, pp. 431-439, April 2012.
- [11] S. N. S. K. L. a. T. D. L. A. D. Evans, "Young children's difficulty with indirect speech acts: Implications for questioning child witnesses," *Behav. Sci. Law*, vol. 32, no. 6, p. 775–788, Nov./Dec. 2014.
- [12] O. Kang, "Effects of a computer-assisted pronunciation training program on Korean EFL students' pronunciation," *System*, vol. 49, pp. 18-29, 2015.
- [13] M. Kim, "Effects of explicit instruction and practice of pronunciation features on Korean EFL learners' comprehensibility," *The Journal of Asia TEFL*, vol. 14, no. 4, pp. 780-791, 2017.
- [14] I. D. a. D. Ewert, "Using audio recordings to improve the pronunciation of foreign language learners," *Language Learning & Technology*, vol. 21, no. 1, pp. 156-177, 2017.
- [15] R. A. A. K. A. & V. F. R. Paul, "Perception and production of prosody by speakers with autism spectrum disorders," *Journal Name: Journal of Autism and Developmental Disorders*, vol. 35, no. 2, pp. 205-220, 2005.
- [16] J. P. J. M. a. L. S. J. Fudala, "Longitudinal studies of phonological development in preschool and school-age children," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 3, pp. 669-687, 2008.

- [17] E. J. a. R. Wagner, "The effects of repeated listening practice and feedback on L2 pronunciation: A study of English vowels produced by Korean learners," *TESOL Quarterly*, vol. 49, no. 1, pp. 145-171, 2015.
- [18] K. L. a. Y. Kim, "The effects of using video self-recording and self-evaluation on pronunciation improvement," *English Language Teaching*, vol. 12, no. 6, pp. 127-141, 2019.



**Ananda R.S.R.G.J.N.,**  
Student, Dept. of Computer Science & Software Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka



**Vishan Jayasinghearachchi,**  
Lecturer, Dept. of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka



**Prof. Koliya Pulasinghe,**  
Senior Professor, Dept. of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

#### **AUTHORS BIOGRAPHY**



**Wijegunaratna K.P.S.G.G.,**  
Student, Dept. of Computer Science & Software Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka



**Nagahawatta S.S.,** Student,  
Dept. of Computer Science & Software Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

#### **Citation of this Article:**

Wijegunaratna K.P.S.G.G, Nagahawatta S.S, Ananda R.S.R.G.J.N, Vishan Jayasinghearachchi, Prof. Koliya Pulasinghe, "Speech Impairment Detection in Sinhala-Speaking Children Aged 3-4 Years: A Focus on Autism Spectrum Disorder (ASD)" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 10, pp 470-478, October 2023. Article DOI <https://doi.org/10.47001/IRJIET/2023.710062>

\*\*\*\*\*