Interspeech 2024
1-5 September 2024, Kos, Greece

NeuRO: An Application for Code-Switched Autism Detection in Children

Mohd Mujtaba Akhtar[1], *Girish** [1], *Orchid Chetia Phukan*[2], *Muskaan Singh*[3]

[1]School of Computer Science, UPES, Dehradun, India, [2]IIIT-Delhi, India,
[3]CARL, ISRC, SCEI, Ulster University
*equal contribution

mmakhtar.research@gmail.com, girish.research.pr@gmail.com, orchidp@iiitd.ac.in, m.singh@ulster.ac.uk

Abstract

Code-switching is a common communication phenomenon where individuals alternate between two or more languages or linguistic styles within a single conversation.Autism Spectrum Disorder(ASD) is a developmental disorder posing challenges in social interaction, communication, and repetitive behaviors.Detecting ASD in individuals with code-switch scenario presents unique challenges. In this paper, we address this problem by building an application **NeuRO** which aims to detect potential signs of autism in code-switched conversations, facilitating early intervention and support for individuals with ASD.

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

Autism Spectrum Disorder (ASD), is a complex neurodevelopmental disorder due to synaptic dysfunction and altered brain connectivity.It affects an individual's daily life with aspects such as social interaction, communication, and behavior. These aspects emerge in early childhood as early as 3 years and vary in degrees and their symptoms range from mild to severe. Children with ASD often struggle with basic social cues, such as making eye contact, responding to their names, sharing interests, and moving with others. Early diagnosis and intervention are crucial for children with autism, as they can receive speech and language therapy, occupational therapy, and social skills training to support them and their families. Previous studies focusing on the diagnosis of audio focused on low-level acoustic characteristics such as pitch, intensity, and rhythmic pattern as potential markers [1]. In addition to these characteristics of vocal outburst and utterance level, which contribute to understanding of audios [2, 3], some studies investigated the incorporation of para-lingual features [4], such as voice quality, articulation and resonance, improving the overall performance of automatic autism detection. In this work, we explore the potential of incorporating linguistic and paralinguistic features for improved code-switched ASD detection. By experimenting with linguistic and paralinguistic features, we aim to capture more comprehensive and discriminative representations of speech for autism detection and the topmost performance is achieved by CNN classifier with paralinguistic features.

2. System Design

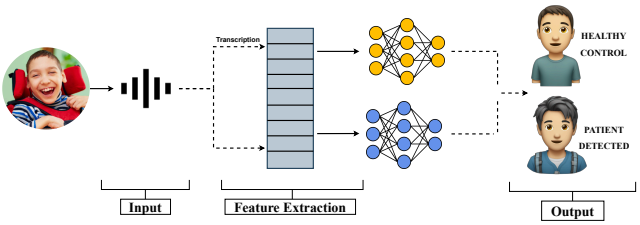


Figure 1: *System Design*

The modeling system design behind our application, **NeuRO** is shown in Figure 1 for autism detection. For the dataset to be used for modeling, raw audio samples from two distinct groups: children potentially diagnosed with autism spectrum disorder (PT) and healthy controls (HC).These raw audio are further preprocessed by resampling to 16 kHz. *For linguistic feature extraction*, we employ the Whisper [5] model for speech transcription and use the Natural Language Toolkit (NLTK) to extract linguistic characteristics from code-mixed transcripts. These extracted features cover the average lengths of words and sentences, the rate of speech, and the language distribution, providing information on code mixing and language switching patterns between English and Hindi. On the other hand, we extract paralinguistic features using TRILLsson [6].

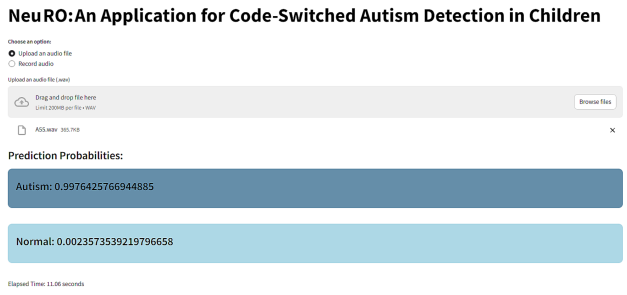
We explore classical ML techniques such as SVM, RF and also RNN, CNN, Transformer encoder for experimentation with individual linguistic and paralinguistic features. For SVM and RF, we use the default parameters. For RNN, we use 50 hidden states and the output layer with sigmoid function. We use CNN with 64 filters with size 3, then succeeded by hidden layer with 128 neurons followed by the output. We use transformer with number of heads as 4 with one hidden layer of 128 neurons followed by the output layer with sigmoid function. We also lead an investigation into combining linguistic and paralinguistic features for autism detection, we use only the CNN and transformer model as it has shown the top most performance with individual models. We train the neural models (RNN, CNN, Transformer) for 50 epochs. We employ 5-fold cross-validation with 4 folds used for training and one fold for testing.

We present our results of the models used in our system design in Table 1. All of our results are computed on the internally curated dataset of children in code-mixing scenarios alternating between English and Hindi. The data collection comprises interactive sessions soliciting responses from children aged 3 to 13. The total of 159.75 m in 61 participants was equally divided into 105.6 m of 30 ASD and 53.99 m of 31 non-ASD.

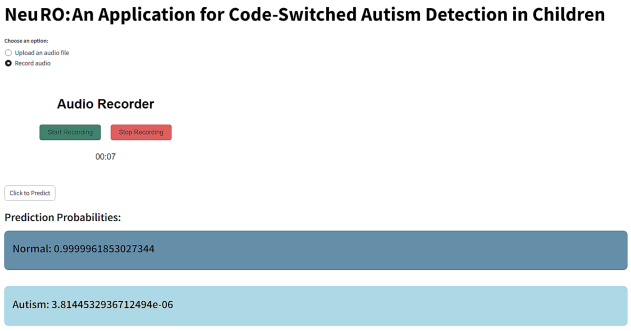
Table 1: *Evaluation Scores; Scores are average of 5-folds; F1 stands macro-average f1-score*

Model	Accuracy	F1
Linguistic Representation Modeling		
SVM	86.98	81.44
RF	85.12	80.80
RNN	85.12	78.90
Transformer	88.37	85.14
CNN	87.44	80.49
Paralinguistic Representation Modeling		
SVM	95.81	94.67
RF	94.77	93.50
RNN	83.72	76.82
Transformer	97.38	96.84
CNN	**98.13**	**97.37**
Fusion with Linguistic+Paralinguistic		
Transformer	97.81	**97.37**
CNN	96.27	95.16

Our data set is diverse and strong, ready for detailed analysis and training classification models, as evident from our results in Table 1.CNN with paralinguistic features obtained the best performance. We were limited in our experimental setup due to the scarcity of data; to the best of our knowledge, there is no existing code-mixed audio corpus to extend the generalizability of our system design.



((a))



((b))

Figure 2: *NeuRO Application Interfaces: (a) Audio from Device (b) Real-time Audio Recording*

3. NeuRO Interface

To enhance the accessibility and user-friendliness of our autism detection tool, we developed and integrated a web-based user interface as in Figure 2. The user interface (UI) was designed to be user-friendly, allowing even non-technical users to easily navigate the platform. After the user uploads a voice recording, the back-end system analyzes the file and extracts embeddings, as described in our approach. Subsequently, the extracted features are then fed into our trained models for classification into PT or HC. This made it easy for data to move from the user interface to the models.

4. Conclusion

In this paper, we present a novel real-world application, **NeuRO**, for the detection of autism from code-switched speech. The model at the core of the system is CNN model with paralinguistic features. This model is seamlessly integrated into the web-based UI to provide prediction of healthy control or PT from a given audio. In addition, users can intuitively interact with the system through the system interface provided. Overall, we believe that our proposed system has the potential for widespread usability in diverse fields, such as healthcare, education, and training.

5. References

[1] H. Tanaka, S. Sakti, G. Neubig, T. Toda, and S. Nakamura, "Linguistic and acoustic features for automatic identification of autism spectrum disorders in children's narrative," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsych@ACL 2014, Baltimore, Maryland, USA, June 27, 2004*, P. Resnik, R. Resnik, and M. Mitchell, Eds. Association for Computational Linguistics, 2014, pp. 88–96. [Online]. Available: https://doi.org/10.3115/v1/W14-3211

[2] S. Cho, M. Liberman, N. Ryant, M. Cola, R. T. Schultz, and J. Parish-Morris, "Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations," in *Proc. Interspeech 2019*, 2019, pp. 2513–2517.

[3] M. Eni, I. Dinstein, and Y. Zigel, "Comparing Hand-Crafted Features to Spectrograms for Autism Severity Estimation," in *Proc. INTERSPEECH 2023*, 2023, pp. 4154–4158.

[4] M. A. Shahin, B. Ahmed, D. V. Smith, A. Duenser, and J. Epps, "Automatic screening of children with speech sound disorders using paralinguistic features," in *29th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2019, Pittsburgh, PA, USA, October 13-16, 2019*. IEEE, 2019, pp. 1–5. [Online]. Available: https://doi.org/10.1109/MLSP.2019.8918725

[5] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *International Conference on Machine Learning*. PMLR, 2023, pp. 28 492–28 518.

[6] J. Shor and S. Venugopalan, "TRILLsson: Distilled Universal Paralinguistic Speech Representations," in *Proc. Interspeech 2022*, 2022, pp. 356–360.