



# Automatic Cry Analysis: Deep Learning for Screening of Autism Spectrum Disorder in Early Childhood

Ana Laguna<sup>1</sup> · Sandra Pusil<sup>1</sup> · Anna Lucia Paltrinieri<sup>2,3</sup> · Silvia Orlandi<sup>4,5</sup>

Accepted: 18 March 2025  
© The Author(s) 2025

## Abstract

**Purpose** The objective of this study is to identify the acoustic characteristics of cries of Typically Developing (TD) and Autism Spectrum Disorder (ASD) children via Deep Learning (DL) techniques to support clinicians in the early detection of ASD.

**Methods** We used an existing cry dataset that included 31 children with ASD and 31 TD children aged between 18 and 54 months. Statistical analysis was applied to find differences between groups for different voice acoustic features such as jitter, shimmer and harmonics-to-noise ratio (HNR). A DL model based on Recursive Convolutional Neural Networks (R-CNN) was developed to classify cries of ASD and TD children.

**Results** We found a statistical significant increase in jitter and shimmer for ASD cries compared to TD, as well as a decrease in HNR for ASD cries. Additionally, the DL algorithm achieved an accuracy of 90.28% in differentiating ASD cries from TD.

**Conclusion** Empowering clinicians with automatic non-invasive Artificial Intelligence (AI) tools based on cry vocal biomarkers holds considerable promise in advancing early detection and intervention initiatives for children at risk of ASD, thereby improving their developmental trajectories.

**Keywords** Cry analysis · Early childhood · Autism spectrum disorder · Deep learning

## Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition affecting 1 in 36 children in the U.S. (CDC, 2022; Mughal et al., 2024). ASD encompasses a wide range

of cognitive, social, and communicative challenges, which have a significant impact on the child, their family and society in general, including areas such as education, the health-care system, and employment.

Early intervention during the initial years of life emerges as a pivotal factor in the management of ASD, owing to its potential for favorable outcomes facilitated by neuroplasticity (Dawson, 2008). Unfortunately, in a majority of cases, ASD is not detected until after the age of 4 (Baio et al., 2018), leading to a missed critical window for early stimulation. The subtle nature of early signs of autism presents challenges for both clinicians and caregivers. Therefore, any behavioral cues that raise concerns should be diligently combined with clinical assessments to enable the early identification of atypical development. This proactive approach enables the timely implementation of intervention strategies, thereby improving the child's long-term prognosis and substantially elevating their overall quality of life.

Over the last decade, numerous studies have leveraged Artificial Intelligence (AI) to analyze early signs of ASD in children using various neurophysiological signals, such as eye-tracking (Jones et al., 2023), electroencephalograms

---

✉ Ana Laguna  
ana.laguna@zoundream.com

<sup>1</sup> Novartis Campus - SIP Basel Area AG, Lichtstrasse 35, Basel 4056, Switzerland

<sup>2</sup> Neonatology Department, Barcelona Center for Maternal-Fetal and Neonatal Medicine (BCNatal), Hospital Clínic, Universitat de Barcelona, Barcelona, Spain

<sup>3</sup> Department de Cirurgia I Especialitats Mèdico-quirúrgiques, Universitat de Barcelona, Barcelona 08036, Spain

<sup>4</sup> Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" – (DEI), University of Bologna; Health Sciences and Technologies, Interdepartmental Center for Industrial Research (CIRI-SDV), University of Bologna, Bologna, Italy

<sup>5</sup> IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy

(Gabard-Durnam et al., 2019), magnetic resonance imaging (Shen et al., 2022), and functional near-infrared spectroscopy (Conti et al., 2022), to detect potential indicators of ASD before behavioral symptoms appear.

However, the techniques mentioned above are considered reliable but involve complex clinical procedures, implying substantial hospital expenditures (Okoye et al., 2023). This fact combined with the recent success of top Deep Learning (DL) techniques applied to much more accessible and low-cost unstructured data signals, has shown promising potential in the domain of early ASD detection (Kim et al., 2023; Kojovic et al., 2021; Manigault et al., 2023). In this context, cry analysis has emerged as a compelling approach for early ASD detection due to its accessibility, non-invasive nature, cost-effectiveness, and ease of recording in both clinical and home settings. It allows for longitudinal assessment, enabling researchers to track developmental changes over time, and is strongly associated with neurodevelopmental conditions (Esposito et al., 2017; Oren et al., 2016). Infant cries provide a unique window into the neurological and physiological state of the infant (Laguna et al., 2023; Laguna, Pusil, Bazán, Laguna et al., 2023a, b; Orlandi et al., 2012), offering the potential to identify early markers of ASD through acoustic features. Research has identified atypical cry patterns in toddlers with ASD under 18 months, with differences observed in acoustic features such as jitter, shimmer, harmonic-to-noise ratio (HNR), and fundamental frequency (F0) (Orlandi, Manfredi, Orlandi et al., 2012a, b; Santos et al., 2013; Sheinkopf et al., 2012; Unwin et al., 2017). Machine learning (ML) algorithms have also been applied to classify these cries, revealing its promising predictive value as a vocal early indicator of ASD (Khozaei et al., 2020; Manigault et al., 2023).

In this study, our primary objective is to determine the distinctive acoustic characteristics present in cries from children aged 18 to 54 months, comparing those with TD to those with ASD. Moreover, we aim to assess the potential application of cry analysis using DL techniques to support clinicians in the early detection of ASD. Empowering clinicians with automatic, non-invasive AI-driven cry vocal biomarkers tools presents a compelling avenue for enhancing early detection endeavors. This has the potential to significantly advance early intervention strategies, leading to more timely and precisely targeted support for children at risk of ASD, thereby enhancing their developmental trajectories.

## Methods

### Participants

The study participants were drawn from the cry dataset from (Khozaei et al., 2020), which encompassed a total of 62 individuals aged between 18 and 54 months. This cohort was divided into two distinct groups: 31 individuals diagnosed with ASD and 31 TD individuals. Within each group, there were 24 boys and 7 girls. The average ages for the ASD and TD groups were 35.6 and 30.8 months, respectively. The autism diagnosis procedure started with the Gilliam Autism Rating Scale-Second Edition (GARS-2) questionnaire (Samadi & McConkey, 2014) which was answered by the parents. Then the caregivers were interviewed, based on the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) (Wiggins et al., 2019), while the participants were evaluated and observed by two Ph.D. degree child clinical psychologists. In addition, the diagnosis of ASD was separately confirmed by at least a child psychiatrist in a different setting. It is important to note that the official version of Autism Diagnostic Observational Schedule (ADOS) is not available for Farsi. Thus, there are different approaches taken to evaluate participants in Iran (Samadi & McConkey, 2014).

### Data Collection

As explained by (Khozaei et al., 2020), data was recorded using high-quality devices (74.2%, Sony UX560 and UX512F voice recorders) and smartphones with a custom voice-recording application (25.80%). Recordings were made in WAV format (16-bit, 44.1 kHz) to ensure consistency across devices. A variety of devices and recording locations such as homes (12.90% ASD sample, 45.16% TD sample), autism centers (87.10% ASD sample) and health centers (54.84% TD sample) were used to avoid bias and increase generalizability. Parents and trained voice collectors were instructed to record in quiet environments with the devices held approximately 25 cm from the participant's mouth. Recordings not meeting these conditions, as well as cries associated with pain, were excluded. The reasons for crying differed between the ASD and TD groups, reflecting distinct behavioral and emotional triggers. In both groups, the most common causes were related to complaining or discomfort (74.20% for ASD group and 67.75% for the TD group), while other factors such as sleepiness, hunger, and anxiety (25.80% for the ASD group and 32.25% for the TD group) also contributed to the crying episodes. In the pre-processing phase, only pure crying sounds were retained, with segments containing other vocalizations or non-empty mouth cries eliminated. Finally, the average number of

cry instances per infant per group includes for the ASD group:  $6.10 \pm 5.05$  instances, while the TD group showed  $5.39 \pm 3.66$  instances. For more details, see Supplementary Material Table 1 S.

### Ethical Considerations

The study protocol received approval from the ethics committee at Shahid Beheshti University of Medical Sciences and Health Services, Tehran (Iran). Prior to enrollment, comprehensive informed consent was acquired from the parents or legal guardians of the participants. This ensured they were well-informed about the study's objectives, procedures, and possible advantages and risks.

### Procedures

For this publicly available dataset, we extracted a range of pitch-based audio features for each cry pattern using Praat software (Boersma, 2002). The total number of frames depends on both the audio duration and the time step used. Perturbation vocal metrics, including jitter, shimmer, and HNR, were chosen due to their widespread use in clinical contexts (Meghashree & Nataraja, 2019; Teixeira et al., 2013). Jitter measures pitch variability over time by calculating the mean absolute difference between consecutive pitch periods, using a period range from 0.0001 to 0.02 s and a pitch floor of 1.3 to eliminate low-frequency noise. Shimmer quantifies the amplitude variation between successive pitch periods, applying the same parameters as jitter but with an added amplitude ratio of 1.6 to capture shimmer across each audio frame. HNR quantifies the periodicity of the voice signal, with higher values indicating more periodic (voiced) sounds and lower values suggesting noisier (aperiodic) signals. HNR was calculated using an autocorrelation-based method in Praat, analyzing the signal in short, overlapping frames. For more details, see Supplementary Material.

### Statistical Analysis

Following the extraction of quantitative features, an exploratory analysis was conducted to discern statistically significant differences between the studied groups for each feature. P-values were computed using the U Mann-Whitney test designed for independent samples. Results are reported in terms of mean  $\pm$  standard error mean and statistically significant p-values are coded as follows: \*\*\* $p < 0.001$ , \*\* $p < 0.01$  and \* $p < 0.05$ .

### Deep Learning Classification Analysis

To demonstrate the potential of DL techniques for automated classification of cry patterns into ASD and TD, we trained a Recurrent Convolutional Neural Network (R-CNN) from scratch. This hybrid architecture combines the strengths of Convolutional Neural Networks (CNNs), which excel at extracting spatial features (Krizhevsky et al., 2017; LeCun et al., 2015) and Long Short-Term Memory (LSTM) recurrent networks, which capture extended temporal relationships in sequential data (Bahdanau et al., 2016; Graves, 2014). Hybrid CNN-LSTM models have been successfully applied in tasks with a sequential component like video and speech recognition (Donahue et al., 2016).

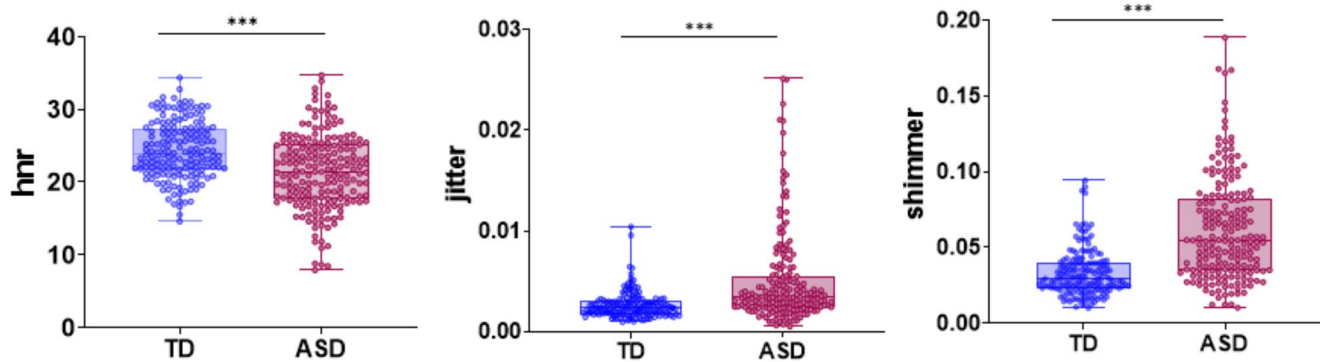
The R-CNN architecture used the extracted spectrographic information as image representations, which serves as input to the model. Input images of size  $128 \times 128$  with a single channel (grayscale) are processed through a CNN block with a kernel size of 3 and 32 output channels, capturing spatial features.

This architecture consists of a CNN with three convolutional layers (comprising 32, 64 and 64 filters, respectively) and two LSTM layers with 288 neurons, followed by one fully-connected layer with 128 units. For the purpose of this study, 80% of the available dataset was used for model training, while the remaining 20% was set aside to assess the model performance and derive validation metrics. The data distribution includes 140 samples for TD and 147 samples for ASD in the training set, while the testing set comprises 28 samples for TD and 44 samples for ASD.

To initialize the weights of our R-CNN model, we employed the Kaiming uniform initialization (He et al., 2015) technique, known to promote stable convergence during training. To mitigate the risk of overfitting, we implemented multiple regularization techniques. Specifically, we employed Dropout (Srivastava et al., 2014) with a rate of 0.2 in each of the LSTM layers. In addition, to enhance generalization, we integrated two data augmentation methods, namely Frequency Masking and Time Masking (Park et al., 2019), which were randomly applied to the training split.

The Adam optimizer (Kingma & Ba, 2017) was utilized for the gradient descent, coupled with a Cyclic learning rate (lr) scheduler (Smith, 2017). The base lr was set to  $10e-6$ , and a maximum lr reached to  $10e-5$ . To strike a balance between computational efficiency and model convergence, we established a batch size of 16. Throughout the training process, we monitored both Binary Cross Entropy (BCE) loss and accuracy for both the training and validation sets.

The training process spanned 2000 epochs, with the model exhibiting superior validation accuracy being designated as the final trained classifier. All aspects of the training procedure, including data preprocessing and model



**Fig. 1** Statistically significant differences between the ASD and TD group. \*\* represents a statistically highly significant difference  $p < 0.01$  from pairwise comparisons. \*\*\* indicates a statistically highly significant difference  $p < 0.001$  for all the pairwise comparisons

**Table 1** Metrics obtained in the DL classification for the ASD and TD acoustic cry patterns. Performance metrics are reported in terms of accuracy, sensitivity (true positives, number of ASD instances), specificity (true negatives, number of TD instances), and F-score

Performance Metrics	Accuracy	Sensitivity	Specificity	Precision	F-score
Model	90.28%	88.64%	88.64%	95.12%	91.75%
Acoustic Cry Patterns					
R-CNN					

optimization, were implemented using the PyTorch v.2.0.1 framework.

DL performance metrics were reported in terms of accuracy (the proportion of true results -both true positives and true negatives- among the total number of cases examined), sensitivity (number of ASD instances, the proportion of true positives correctly identified by the model) and specificity (number of TD instances, the proportion of true negatives correctly identified by the model). Additional DL methods were tested, for more details referred to the Supplementary Material.

## Results

### Vocal Perturbation Measures in Autism

We performed a statistical exploratory analysis of cry acoustic features to assess and characterize the differences between ASD and TD cries. Figure 1 illustrates the statistically significant features identified in this analysis for each group. Notably, the ASD group exhibited higher levels of jitter, and shimmer compared to the TD group ( $p < 0.01$ ). Furthermore, a significant reduction in HNR was observed in the ASD group relative to the TD group ( $p < 0.0001$ ).

## Deep Learning Classification of Autistic Cry Patterns

The classification analysis aimed to evaluate the efficacy of audio features in distinguishing between cries of children with ASD and TD. The R-CNN algorithm exhibited remarkable performance, achieving a validation accuracy of 90.28% (AUC 90.83%). Furthermore, the model exhibited a specificity of 88.64%, a recall of 88.64%, precision of 95.12%, and an F-score of 91.75% (Table 1).

## Discussion

The current study delved into the potential of DL techniques to differentiate between cries of children diagnosed with ASD and those TD, within an age range of 18 to 54 months. First, we aimed to objectively identify characteristic differences in cry patterns between the two groups through the analysis of audio features, including attributes such as jitter, shimmer, and HNR. The second aim was to establish the feasibility of leveraging an AI-based automatic system to accurately identify these characteristic differences in the spectrogram's cry patterns.

Our study revealed notable differences between the ASD and TD groups in various frequency-based cry features. Precisely, the ASD group exhibited increased levels of jitter and shimmer coupled with a reduced HNR when contrasted with the TD group. Our results are consistent with previous research (Santos et al., 2013), which also identified increased levels of jitter and shimmer, along with reduced HNR, as distinguishing acoustic features in children with ASD integrated into a ML model to classify ASD and TD groups. The study further emphasizes that these vocal quality differences, linked to breathiness, hoarseness, and roughness, can serve as early biomarkers for ASD or even other disorders or pathologies (Meghashree & Nataraja, 2019; Santos et al., 2013; Teixeira & Fernandes, 2015).

Consequently, these findings indicate distinctive cry characteristics associated with ASD children, pointing to the potential value of these features in effectively discerning between the two groups during early-stage ASD screening and assessment.

Regarding the automatic classification of ASD and TD cries, previous research (Khozaei et al., 2020; Motlagh et al., 2013) with the same dataset used ML pattern recognition algorithms to distinguish between ASD and TD cry patterns within the ages range of 2 to 3 years. They extracted various audio features such as temporal features, energy features, harmonic features, perceptual and spectral features. Their SVM model showed an average accuracy of 89.3% accuracy including both genders (Khozaei et al., 2020). Notwithstanding, our study is pioneering on using DL for classification of ASD and TD cries without gender differentiation being able to predict if a cry belongs to an autistic or neurotypical child with a precision of 90.28% even in a very reduced dataset.

To translate this approach into clinical practice, the proposed tool could be integrated as a complementary aid in pediatric evaluations or as an at-home screening resource for caregivers. Its non-invasive nature and ease of use would allow for continuous remote monitoring, potentially enhancing early detection and facilitating timely interventions. However, successful implementation would require clinician training and addressing ethical considerations. Further research should focus on real-world testing and integration strategies to maximize clinical utility and acceptance.

## Limitations

While our findings highlight the potential of using cry acoustic features and DL for early autism identification, the study generalizability is limited by the sample size, demographic variation, and the specific age range of participants (18 to 54 months). Future research should include larger, more diverse populations across different age groups and cultural backgrounds to validate the model performance in varying developmental stages and contexts. Additionally, cry characteristics may evolve as children grow, potentially influencing model accuracy, making it essential to explore age-specific models. Another limitation is that autism diagnosis in this study was conducted using the GARS-2 parental questionnaire and a DSM-5-based parental interview by child clinical psychologists, with independent confirmation by a child psychiatrist. However, the widely used ADOS tool was not administered due to the lack of an official Farsi translation, which is less common in Iran. Expanding the dataset and refining diagnostic methodologies will be crucial for improving the robustness and applicability of this approach in broader populations.

## Conclusion

Early signs of infant atypical development are frequently challenging for clinicians and caregivers. Often, clinical procedures are invasive and non-accessible. Thus, doctors lack signs raising the alarm during the first months of life. Nevertheless, early screening and detection of autism are crucial for initiating timely interventions and support, enhancing the overall outcomes and well-being of affected children and their families. The statistical trends and AI techniques show encouraging results suggesting the potential utility of the automatic cry analysis as a reliable non-invasive and objective tool for identifying early markers of autism using cry as a vocal biomarker in early childhood.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10803-025-06811-1>.

**Author Contributions** AL: Conceptualization, Funding acquisition, Methodology, Data processing, Formal analysis, Manuscript– original draft preparation. SP: Conceptualization, Methodology, Data processing, Formal analysis, Manuscript– original draft preparation and preparing figures. ALP: Conceptualization, Advice on Neonatology, Manuscript– review & editing. SO: Advice on Data analysis pipeline and Result discussion, Manuscript– review & editing. All present authors contributed to the article and approved the submitted version.

**Funding** The research has been funded by Zoundream AG together with the Swiss State Secretariat for Education, Research and Innovation (SERI), the Swiss Innovation Agency (Innosuisse), the Future of Health Grant, and the NEOTEC grant (SNEO-20211305) received from the Spanish Center for Technological Development and Innovation (CDTI).

## Declarations

**Competing Interests** The authors declare competing interests (Funding, Employment or Confidentiality interests) in relation to the work described herein. Ana Laguna and Sandra Pusil are employed by Zoundream AG. Ana Laguna is also co-founder of the company and owns stock in Zoundream AG. Silvia Orlandi received compensation for the collaboration as members of the scientific advisory board of Zoundream AG. Anna Lucia Paltrinieri declared no potential conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (No. arXiv:1409.0473). arXiv. <https://doi.org/10.48550/arXiv.1409.0473>
- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L. C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., & Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, 67(6), 1–23. <https://doi.org/10.15585/mmwr.ss6706a1>
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5.
- CDC. (2022, August 4). *Research on Autism Spectrum Disorder*. Centers for Disease Control and Prevention. <https://www.cdc.gov/ncbddd/autism/seed.html>
- Conti, E., Scaffei, E., Bosetti, C., Marchi, V., Costanzo, V., Dell'Oste, V., Mazziotti, R., Dell'Osso, L., Carmassi, C., Muratori, F., Baroncelli, L., Calderoni, S., & Battini, R. (2022). Looking for fNIRS signature in autism spectrum: A systematic review starting from preschoolers. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.785993>
- Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, 20(3), 775–803. <https://doi.org/10.1017/S0954579408000370>
- Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2016). *Long-term Recurrent Convolutional Networks for Visual Recognition and Description* (No. arXiv:1411.4389). arXiv. <https://doi.org/10.48550/arXiv.1411.4389>
- Esposito, G., Hiroi, N., & Scattoni, M. L. (2017). Cry, baby, cry: Expression of distress as a biomarker and modulator in autism spectrum disorder. *The International Journal of Neuropsychopharmacology*, 20(6), 498–503. <https://doi.org/10.1093/ijnp/pyx014>
- Gabard-Durnam, L. J., Wilkinson, C., Kapur, K., Tager-Flusberg, H., Levin, A. R., & Nelson, C. A. (2019). Longitudinal EEG power in the first postnatal year differentiates autism outcomes. *Nature Communications*, 10(1), 4188. <https://doi.org/10.1038/s41467-019-12202-9>
- Graves, A. (2014). *Generating Sequences With Recurrent Neural Networks* (No. arXiv:1308.0850). arXiv. <https://doi.org/10.48550/arXiv.1308.0850>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* (No. arXiv:1502.01852). arXiv. <https://doi.org/10.48550/arXiv.1502.01852>
- Jones, W., Klaiman, C., Richardson, S., Lambha, M., Reid, M., Hamner, T., Beacham, C., Lewis, P., Paredes, J., Edwards, L., Marrus, N., Constantino, J. N., Shultz, S., & Klin, A. (2023). Development and replication of objective measurements of social visual engagement to aid in early diagnosis and assessment of autism. *JAMA Network Open*, 6(9), e2330145. <https://doi.org/10.1001/jamanetworkopen.2023.30145>
- Khozaei, A., Moradi, H., Hosseini, R., Pouretamad, H., & Eskandari, B. (2020). Early screening of autism spectrum disorder using cry features. *Plos One*, 15(12), e0241690. <https://doi.org/10.1371/journal.pone.0241690>
- Kim, J. H., Hong, J., Choi, H., Kang, H. G., Yoon, S., Hwang, J. Y., Park, Y. R., & Cheon, K. A. (2023). Development of deep ensembles to screen for autism and symptom severity using retinal photographs. *JAMA Network Open*, 6(12), e2347692. <https://doi.org/10.1001/jamanetworkopen.2023.47692>
- Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (No. arXiv:1412.6980). arXiv. <https://doi.org/10.48550/arXiv.1412.6980>
- Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose Estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-94378-z>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the Acm*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Laguna, A., Pusic, S., Acero-Pousa, I., Zegarra-Valdivia, J. A., Paltrinieri, A. L., Bazán, À., Piras, P., Palomares i Perera, Garcia-Algar, C., O., & Orlandi, S. (2023a). How can cry acoustics associate newborns' distress levels with neurophysiological and behavioral signals? *Frontiers in Neuroscience*, 17. <https://www.frontiersin.org/journals/neuroscience/articles/https://doi.org/10.3389/fnins.2023.1266873>
- Laguna, A., Pusic, S., Bazán, À., Zegarra-Valdivia, J. A., Paltrinieri, A. L., Piras, P., Palomares i Perera, C., Pardos Véglia, A., Garcia-Algar, O., & Orlandi, S. (2023b). Multi-modal analysis of infant cry types characterization: Acoustics, body Language and brain signals. *Computers in Biology and Medicine*, 167, 107626. <https://doi.org/10.1016/j.compbiomed.2023.107626>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Manigault, A. W., Sheinkopf, S. J., Carter, B. S., Check, J., Helderman, J., Hofheimer, J. A., McGowan, E. C., Neal, C. R., O'Shea, M., Pastyrnak, S., Smith, L. M., Everson, T. M., Marsit, C. J., Dansereau, L. M., DellaGrotta, S. A., & Lester, B. M. (2023). Acoustic cry characteristics in preterm infants and developmental and behavioral outcomes at 2 years of age. *JAMA Network Open*, 6(2), e2254151. <https://doi.org/10.1001/jamanetworkopen.2022.54151>
- Meghashree, & Nataraja, D. N. P. (2019). Acoustic features of cry of infants with high risk factors. *International Journal of Health Sciences and Research*, 9(12), 62–67.
- Motlagh, S. H. R. E., Moradi, H., & Pouretamad, H. (2013). Using general sound descriptors for early autism detection. *2013 9th Asian Control Conference (ASCC)*, 1–5. <https://doi.org/10.1109/ASCC.2013.6606386>
- Mughal, S., Faizy, R. M., & Saadabadi, A. (2024). Autism Spectrum Disorder. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK525976/>
- 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*, 15(8), e43226. <https://doi.org/10.7759/cureus.43226>
- Oren, A., Matzliach, A., Cohen, R., & Friedmann, H. (2016). Cry-based detection of developmental disorders in infants. *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 1–5. <https://doi.org/10.1109/ICSEE.2016.7806073>
- Orlandi, S., Bocchi, L., Donzelli, G., & Manfredi, C. (2012a). Central blood oxygen saturation vs crying in preterm newborns. *Biomedical Signal Processing and Control*, 7(1), 88–92. <https://doi.org/10.1016/j.bspc.2011.07.003>

- Orlandi, S., Manfredi, C., Bocchi, L., & Scattoni, M. L. (2012b). Automatic newborn cry analysis: A non-invasive tool to help autism early diagnosis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2012*, 2953–2956. <https://doi.org/10.1109/EMBC.2012.6346583>
- Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. *Interspeech 2019*, 2613–2617. <https://doi.org/10.21437/Interspeech.2019-2680>
- Samadi, S. A., & McConkey, R. (2014). The utility of the Gilliam autism rating scale for identifying Iranian children with autism. *Disability and Rehabilitation*, 36(6), 452–456. <https://doi.org/10.3109/09638288.2013.797514>
- Santos, J. F., Brosh, N., Falk, T. H., Zwaigenbaum, L., Bryson, S. E., Roberts, W., Smith, I. M., Szatmari, P., & Brian, J. A. (2013). Very early detection of Autism Spectrum Disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7567–7571. <https://doi.org/10.1109/ICASSP.2013.6639134>
- Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, 5(5). <https://doi.org/10.1002/aur.1244>
- Shen, M. D., Swanson, M. R., Wolff, J. J., Elison, J. T., Girault, J. B., Kim, S. H., Smith, R. G., Graves, M. M., Weisenfeld, L. A. H., Flake, L., MacIntyre, L., Gross, J. L., Burrows, C. A., Fonov, V. S., Collins, D. L., Evans, A. C., Gerig, G., McKinstry, R. C., Pandey, J., & IBIS Network. (2022). Subcortical brain development in autism and fragile X syndrome: Evidence for dynamic, Age- and Disorder-Specific trajectories in infancy. *The American Journal of Psychiatry*, 179(8), 562–572. <https://doi.org/10.1176/appi.ajp.21090896>
- Smith, L. N. (2017). *Cyclical Learning Rates for Training Neural Networks* (No. arXiv:1506.01186). arXiv. <https://doi.org/10.48550/arXiv.1506.01186>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res*, 15(1), 1929–1958.
- Teixeira, J. P., & Fernandes, P. O. (2015). Acoustic analysis of vocal dysphonia. *Procedia Computer Science*, 64, 466–473. <https://doi.org/10.1016/j.procs.2015.08.544>
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic Analysis–Jitter, shimmer and HNR parameters. *Procedia Technology*, 9, 1112–1122. <https://doi.org/10.1016/j.protcy.2013.12.124>
- Unwin, L. M., Bruz, I., Maybery, M. T., Reynolds, V., Ciccone, N., Dissanayake, C., Hickey, M., & Whitehouse, A. J. O. (2017). Acoustic properties of cries in 12-Month old infants at High-Risk of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(7), 2108–2119. <https://doi.org/10.1007/s10803-017-3119-z>
- Wiggins, L. D., Rice, C. E., Barger, B., Soke, G. N., Lee, L. C., Moody, E., Edmondson-Pretzel, R., & Levy, S. E. (2019). DSM-5 criteria for autism spectrum disorder maximizes diagnostic sensitivity and specificity in preschool children. *Social Psychiatry and Psychiatric Epidemiology*, 54(6), 693–701. <https://doi.org/10.1007/s00127-019-01674-1>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.