Analyzing neonatal vocal expression: Methological approaches to identifying neurological and psychiatric signatures

Original Research

Analyzing neonatal vocal expression: methodological approaches to identifying neurological and psychiatric signatures

Syed Taimoor Hussain Shah[1,*,†], Syed Adil Hussain Shah[1,2,†], Andrea Buccoliero[2,3], Iqra Iqbal Khan[4,5], Syed Baqir Hussain Shah[6], Angelo Di Terlizzi[2], Giacomo Di Benedetto[7]

[1]Politecnico di Torino, Department of Mechanical and Aerospace Engineering, PolitoBioMed Lab, Corso Duca degli Abruzzi 24, Turin I-10129, Italy
[2]GPI SpA, Department of Research and Development (R&D), Via Ragazzi del '99, Trento 38123, Italy
[3]Human Science Department, Università degli Studi di Verona, Lungadige Porta Vittoria, 17, Verona 37129, Italy
[4]Department of Computer Science, Bahauddin Zakariya University, Multan 60800, Pakistan
[5]Department of Computing and Emerging Technologies, Emerson University Multan, Multan 60000, Pakistan
[6]COMSATS University Islamabad (CUI), Wah Campus, Department of Computer Science, Grand Trunk Road, Wah 47040, Pakistan
[7]7HC SRL, Rome 00198, Italy

***Correspondence**: taimoor.shah@polito.it
DOI: https://doi.org/10.56280/1703023560

Received: 2 June 2025 Accepted: 24 June 2025 Online Published: 30 June 2025

Abstract

Analyzing neonatal vocal expression provides invaluable insights into brain function and the emergence of consciousness, as early vocalization patterns reflect neurodevelopmental trajectories and sensory integration processes. Despite progress in neonatal healthcare, identifying reliable neurological and cognitive markers from infant vocal sounds remains challenging, as it requires linking complex, multi-level brain activity with perceptual acoustic features. This paper reviews methodological approaches used to analyze neonatal vocal expressions, with a focus on techniques that bridge data-driven models with clinical applications. We examine computational methods, including signal processing, feature extraction algorithms, and machine learning models designed to capture vocal biomarkers of neurological or psychiatric disorders. Approaches include spectro-temporal analysis to detect atypical acoustic patterns, deep learning models like convolutional neural networks (CNNs) for automated feature learning, and explainable AI techniques that connect model outputs to clinically interpretable vocal features. We also explore multimodal approaches that combine vocal data with physiological and behavioral signals to improve diagnostic accuracy. The review addresses challenges in neonatal vocal analysis, including data scarcity, demographic variability, and the need for generalization across different recording environments. To mitigate these issues, we highlight advances in domain adaptation, transfer learning, and data augmentation, which enable models to generalize across diverse clinical scenarios. We emphasize the need for clinical validation and interdisciplinary collaboration to ensure practical adoption of these models in healthcare. Future research should focus on refining predictive models with larger, more diverse datasets and enabling real-time analysis for continuous neonatal monitoring. By evaluating existing methodologies and proposing future directions, this study aims to advance neonatal vocal analysis and support early diagnosis and intervention in pediatric healthcare.

Keywords: Neonatal vocal expression, neurological and psychiatric signatures, signal processing, machine/deep learning, explainable AI, pediatric healthcare

1. Introduction

Analyzing neonatal vocal expression represents a frontier in developmental neuroscience and pediatric healthcare, offering a unique and noninvasive pathway for understanding the early architecture of the human brain (Andonotopo et al., 2025). Neonatal vocalizations, beginning with the very first cries after, are far more common than primitive reflexive sounds; they are rather complex, biologically orchestrated signals that emerge from the intricate coordination of the neurological, respiratory, and cognitive systems (Romo et al., 2024; Shah et al., 2025). From an evolutionary perspective, these early sounds have played a crucial role in securing caregiver attention, signaling physiological needs, and promoting social bonding. Today, they are increasingly recognized as rich behavioral biomarkers that can offer critical insight into an infant's sensory processing

Journal of Multiscale Neuroscience

capacity, emotional regulatory mechanisms, and early motor control (Filippa & Kuhn, 2024). This multidimensional nature makes them valuable proxies for assessing the integrity of the developing central nervous system, with immense potential to inform early diagnosis of neurodevelopmental and neuropsychiatric conditions.

Over the past decade, rapid advancements in computational neuroscience and biomedical engineering have fueled interest in harnessing infant vocalizations as early warning signals for atypical brain development (Onciul et al., 2025). Researchers are exploring whether subtle deviations in cry acoustics or cooing patterns could serve as preclinical indicators for conditions such as autism spectrum disorder, cerebral palsy, or language impairments well before such conditions manifest in overt behavioral symptoms. Early identification is particularly critical, as timely therapeutic interventions during periods of peak neuroplasticity can significantly improve cognitive and behavioral outcomes later in life. However, despite this exciting promise, transforming these raw vocal outputs into actionable, clinically relevant information poses formidable scientific and technical hurdles (Husain et al., 2025).

One major challenge stem from the inherent variability and fleeting nature of neonatal vocalizations. Unlike adult speech, which follows structured phonetic and linguistic patterns, infant cries are brief, highly context-dependent, and easily influenced by environmental stimuli or physiological states (Kao & Zhang, 2025; Nussbaum et al., 2025). Furthermore, inconsistencies in recording environments, differences in microphone quality, and demographic variability, such as language or cultural factors, add layers of complexity, introducing confounding variables that can degrade model accuracy and generalizability across populations. Extracting meaningful patterns from these signals requires not only robust signal processing techniques but also advanced computational models capable of distinguishing noise from neurologically meaningful variation.

Moreover, bridging the gap between sophisticated computational models and real-world clinical practice necessitates methodological rigor and interpretability. Healthcare professionals need tools that not only achieve high predictive accuracy under controlled laboratory conditions but also maintain robustness and transparency in the noisy, dynamic context of neonatal intensive care units (NICUs) or home-based monitoring settings (Sheikh et al., 2025). This calls for explainable Artificial Intelligence (AI) frameworks that clarify how models reach diagnostic conclusions, fostering trust and acceptance among clinicians and caregivers alike. Additionally, integrating vocal analysis with other complementary data streams, such as physiological signals (e.g., heart rate variability), EEG patterns, or contextual behavioral observations, can significantly enhance diagnostic precision and provide a more holistic understanding of an infant's health status.

This review aims to provide a comprehensive examination of the state-of-the-art methodologies in neonatal vocal analysis, outlining how signal processing, acoustic feature extraction, and Machine Learning intersect to address the unique demands of this sensitive domain. We survey traditional techniques for denoising, segmentation, and feature extraction before delving into more recent advances such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention-based models that automate and refine the learning of complex acoustic patterns. A critical focus is placed on explainable AI tools that illuminate model decision pathways, as well as strategies for domain adaptation, transfer learning, and data augmentation that help mitigate data scarcity and promote generalizability. In recognition of the limitations of single-modality approaches, we further explore emerging multimodal frameworks that combine vocal signals with physiological, behavioral, and environmental data to construct a more comprehensive, context-aware portrait of neonatal neurodevelopment. This reflects a broader trend in pediatric medicine toward systems-level analysis and personalized care pathways. Finally, we address the practical and ethical considerations that accompany the clinical translation of these tools, highlighting the importance of interdisciplinary collaboration, rigorous validation, and attention to data privacy. By synthesizing these diverse threads, this review not only charts the current landscape but also identifies promising avenues for future research and clinical application. Ultimately, by advancing robust, interpretable, and contextually integrated vocal biomarkers, the field moves closer to realizing the vision of early, accessible, and precise neurodevelopmental monitoring, transforming how we detect, understand, and respond to risk in the earliest stages of human life.

To explore this promising domain comprehensively, Section 2 outlines the neurodevelopmental significance

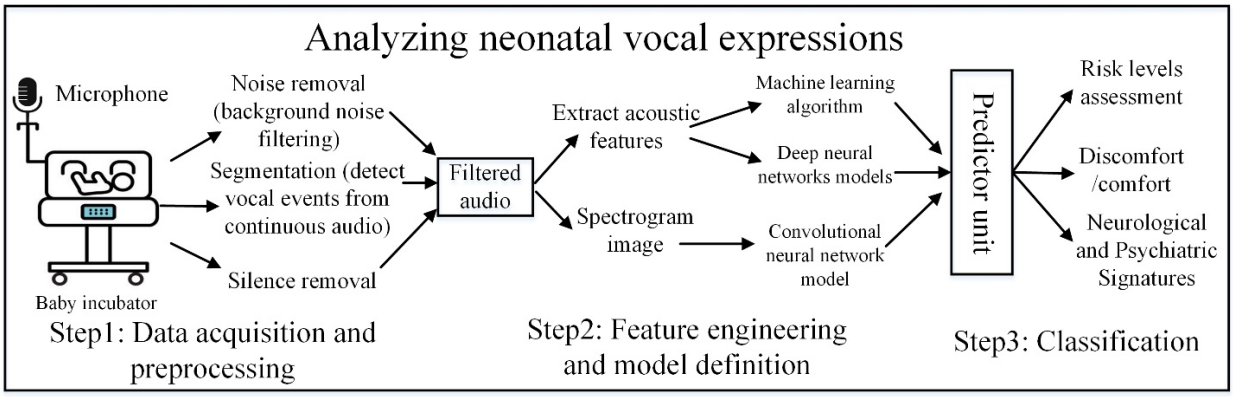

Figure 1 *A general conceptual workflow illustrating the typical stages of neonatal vocal analysis, from data collection and preprocessing to feature extraction, modeling, and clinical interpretation.*

of neonatal vocalizations, while Section 3 discusses thekey challenges inherent in analyzing such delicate data. Section 4 outlines the signal processing and acoustic feature extraction methods, followed by Section 5, which explores machine learning approaches specifically designed for vocal analysis. Section 6 highlights the benefits of multimodal and integrative frameworks, and Section 7 addresses methodological challenges and proposed solutions. Section 8 examines clinical translation and validation pathways, while Section 9 explores future directions for research and application. Finally, Section 10 provides concluding remarks, summarizing key insights and their implications for advancing neonatal neurodevelopmental care.

2 Neurodevelopmental significance of neonatal vocalizations

Neonatal vocalizations serve as one of the earliest and most accessible manifestations of neural activity in the developing brain (Zhang, 2025). These vocal expressions, including cries, coos, and proto-speech sounds, arise from complex coordination between the central nervous system, respiratory system, and vocal tract. Even in the absence of fully developed cognitive or linguistic abilities, infants produce vocal patterns that are shaped by underlying neural circuits involved in motor planning, auditory feedback, and affective regulation (Wang & Song, 2022). The structure, frequency, pitch, and rhythm of these early sounds are not arbitrary; they reflect maturational processes in brain regions such as the brainstem, limbic system, and auditory cortex (Cappelli & Noccetti, 2022). As such, careful analysis of neonatal vocal output provides a noninvasive proxy for assessing neurological development, offering insights into both typical and atypical trajectories of brain maturation.

In the context of recent advances, a typical framework for analyzing neonatal vocal expressions can be conceptually outlined in three core stages: data acquisition and preprocessing, feature engineering with model definition, and final classification or risk assessment, as outlined in **Figure 1**. In this approach, audio signals are first captured in clinical environments such as neonatal intensive care units using microphones placed near incubators. Rigorous preprocessing steps, including background noise removal, silence trimming, and segmentation of relevant vocal events, are employed to ensure that only high-quality, meaningful audio segments are retained. Subsequently, acoustic features are extracted and, where appropriate, spectrogram images are generated to visually represent the temporal and spectral properties of the cry sounds. These features then serve as inputs to various computational models, ranging from traditional ML algorithms to more advanced deep neural networks and convolutional architectures. The processed outputs enable classification tasks that support clinicians in assessing discomfort levels, identifying possible neurological or psychiatric markers, and estimating overall risk. As illustrated in the diagram, this multi-step pipeline reflects best practices commonly reported in the literature, offering a structured perspective on how interdisciplinary tools can be integrated to unlock the diagnostic potential of infant vocalizations.

Early vocalizations are more reflexive acts than true indicators of an infant's ability to interact with and respond to the sensory environment (Filippa & Kuhn,

2024). The emergence of patterned, intentional vocal output signals the gradual development of conscious awareness and the infant's capacity to perceive, process, and react to external stimuli. This process involves the integration of multimodal sensory inputs such as touch, vision, and hearing with internal motor and affective states (Sanna, 2025). The dynamic interplay between sensation and expression facilitates the infant's engagement with caregivers and surroundings, which in turn supports socioemotional bonding and cognitive stimulation (La Rosa et al., 2024). Vocalizations, therefore, represent a bridge between the internal neurological state and external behavioral expression, enabling researchers to track the formation of consciousness and early perceptual motor coordination through acoustic analysis (Rudenko, 2023).

The clinical implications of neonatal vocal expression are profound, particularly in the early detection of neurological and psychiatric disorders. Atypical vocal characteristics such as abnormal pitch, monotonic cries, prolonged silence, or irregular prosody can be early indicators of conditions such as autism spectrum disorder (ASD), cerebral palsy, or perinatal brain injury (Filippa et al., 2021; Marschik et al., 2022a). Numerous studies (Bartl-Pokorny et al., 2022; Long et al., 2023; Marschik et al., 2022b; Wagner et al., 2025) have documented how infants with neurological impairments exhibit vocal patterns distinct from those of their typically developing peers. These differences often precede observable behavioral symptoms, positioning vocal analysis as a valuable tool for early risk assessment. In psychiatric contexts, alterations in vocal expression may signal disruptions in affective processing and social communication, domains commonly affected in disorders such as ASD or early-onset mood disorders (Ding and Zhang, 2023; Kamiloğlu and Sauter, 2021; Ribolsi et al., 2022). As such, neonatal vocal analysis not only supports early diagnosis but also holds promise for tracking developmental progress and treatment outcomes over time.

3 Challenges in analyzing neonatal vocal data

To support robust research on infant vocalization, cry detection, snoring recognition, and pain assessment, this study leverages a diverse collection of datasets, each contributing unique acoustic contexts and annotation standards (see **Table 1** for a comprehensive summary). Among these, AudioSet (Gemmeke et al., 2017) stands out as a massive benchmark of over two million human-annotated YouTube clips, from which we specifically extracted categories relevant to infant cries and snoring events, using both weakly and strongly labeled subsets. The Baby Chillanto Database (BCD) (Reyes-Galaviz et al., 2008) provides a pathology-focused corpus of short infant cry samples categorized into clinically meaningful conditions such as asphyxia, deafness, hunger, normal, and pain cries, facilitating nuanced classification tasks. Donate A Cry (Veres, 2025) complements this by focusing on the emotional and need-driven aspects of baby cries, covering categories such as hunger, burping, belly pain, discomfort, and tiredness, thereby enabling models to map vocal cues to daily care needs.

The environmental and ambient noise contexts are captured through the ESC-50 dataset (Piczak, 2015), which we filter for infant crying and snoring sounds to augment the training data with real-world variability. Building on these public resources, the Infant Cry and Snoring Detection (ICSD) dataset (Liu et al., 2025) integrates samples from eight source datasets, which are systematically cleaned and balanced into weakly labeled, strongly labeled, and synthetic event clips specifically tailored for our detection models. Similarly, the DSPLab Baby Sounds challenge dataset (Alexlinander, 2022) provides additional labeled baby cry clips designed for benchmarking machine learning pipelines via MFCCs and SVMs. For pain detection, the Infant FLACC Pain Level Video Dataset (IFPaLVD) (Kristian et al., 2023) offers carefully annotated audio recordings of infants assessed for both crying and pain severity via the widely accepted FLACC scale, enriching our ability to study the acoustic correlates of distress and discomfort.

In addition to traditional audio datasets, we include rich, ego-centric multimodal corpora that capture infant and child behavior in naturalistic contexts. BV-Home (Long et al., 2024) comprises more than 400 hours of daily life home recordings from 28 families, capturing spontaneous infant vocalizations, interactions, and environmental context alongside parent-reported language measures. BV-Preschool extends this perspective into early education settings, documenting child speech and interactions within a Montessori-inspired preschool environment. Ego-SingleChild, also from the same research initiative, provides longitudinal recordings from a single child via a wearable headband camera, whereas SAYCam (Sullivan et al., 2021) offers a similar head camera view spanning 476 to examine the relationship between early language experience and vocal development.

Table 1 *Comprehensive overview of all datasets employed in this study for infant vocalization, snoring detection, pain assessment, and contextual behavioral analysis.*

Dataset	Description	Categories Used in Research	Total Clips Used	Clip Duration	Additional Notes
AudioSet (Gemmeke et al., 2017)	2 Million+ human-annotated YouTube audio clips with 632 event classes; includes weakly and strongly labeled subsets	Infant Cry, Snoring	Weakly labeled: 1391 (Cry), 1713 (Snoring); Strongly labeled: 424 (Cry), 383 (Snoring)	10 sec	Strong labels include timestamps for events; hierarchical ontology
Baby Chillanto (BCD) (Reyes-Galaviz et al., 2008)	Mexican database for infant cry pathology classification; 5 pathology classes	All 5 infant cry categories	2,268 samples	1 sec	Categories: asphyxia, deaf, hunger, normal, pain
Donate A Cry (Veres, 2025)	Collected from 0–2-year-old babies; designed for infant need recognition	Hunger, Burping, Belly Pain, Discomfort, Tiredness	457 files	7 sec	Cleaned and categorized for infant need recognition
ESC-50 (Piczak, 2015)	2,000 environmental sounds across 50 classes and 5 main categories	Infant Cry, Snoring (extracted only)	Subset extracted	5 sec	Includes 10 classes per category across 5 main sound groups
ICSD Dataset (Liu et al., 2025)	Proposed Infant Cry and Snoring Detection (ICSD) dataset built from 8 unified source datasets; used for event detection task	Infant Cry, Snoring	8,000 (train), 1,000 (validation/test); plus: 1,699 Cry & 1,577 Snoring weakly labeled (train); 338 Cry & 305 Snoring real strongly labeled (train)	10 sec	Includes weakly labeled, real strongly labeled, and synthetic strongly labeled clips; test set excludes weak labels; fully cleaned and standardized
DSPLab: Detecting Baby Sounds (Alexlinander, 2022)	Kaggle competition dataset to classify baby sounds using MFCCs and SVM (or other models)	Baby sound types (not explicitly listed)	3,996 labeled clips (train) + dev set	Not specified	F1 score used for evaluation
IFPaLVD (Infant FLACC Pain Level Video Dataset) (Kristian et al., 2023)	Pain and cry assessment dataset collected at Dr. Soetomo General Hospital using FLACC scale	Cry/No Cry, Pain levels: Neutral, Discomfort, Mild, Moderate, Severe	253 audio recordings	Not specified	23 infants (<1 year); pain labels based on tuple (pain level, cry); 5 pain categories derived from tuple combinations

Table 1 Con't…

Dataset	Description	Categories Used in Research	Total Clips Used	Clip Duration	Additional Notes
BV-Home (Long et al., 2024)	Home recordings of infant-toddler daily life collected from 28 families (avg. child age 11 months); includes ego-centric video, audio, transcripts, and motion data	Infant vocalizations in naturalistic contexts	Not explicitly clips, but ~433 hours of recordings	Varies	Ego-centric, long-form recordings; parent-reported language development; audio, transcript, motion data available
BV-Preschool (Long et al., 2024)	Egocentric preschool recordings in Montessori-like setting	Child vocalizations and interactions	Not explicit clips; ~63 hours	Varies	39 children (2.11–5.11 years); play-based learning
Ego-SingleChild (Long et al., 2024)	Frequent recordings of a single infant with headband camera	Single child daily vocalizations	Not explicit clips; 47 hours	Varies	High continuity; alternative camera; lower resolution
SAYCam (Sullivan et al., 2021)	Longitudinal head-camera recordings of daily infant life	Infant language & visual experience	Not explicit clips; 476 hours	Varies	3 infants; focus on language acquisition context

Despite the breadth and depth of these datasets, analyzing neonatal vocal data poses persistent challenges. First, the limited availability of high-quality, consistently annotated infant recordings remains a bottleneck owing to ethical restrictions, the fragile nature of neonatal subjects, and practical constraints in clinical environments such as NICUs (Keles & Bagci, 2023). Consequently, many studies rely on relatively small or demographically narrow samples, limiting the statistical power and generalizability of the resulting models (Frank, 2020). Furthermore, inconsistent recording protocols, varying microphone placements, and environmental noise, such as hospital equipment hums or household disturbances, introduce significant acoustic variability, which can degrade feature extraction and classification accuracy (Mallegni et al., 2022).

Moreover, neonatal vocalizations are inherently influenced by biological and social factors such as age, gestational maturity, cultural background, and health status (Hou et al., 2024). Premature infants, for example, may vocalize differently than their full-term peers because of developmental differences in respiratory control or neurological function. Without careful inclusion of diverse demographic groups, models risk embedding bias and may fail to detect anomalies accurately across varied populations (Meissen et al., 2024). Finally, even well-tuned models often struggle to maintain their performance when transferred from controlled research conditions to real-world settings, such as busy hospitals or remote home monitoring. Variations in recording devices, background activity, and infant states (e.g., feeding, sleeping, or interacting) can shift the acoustic profile of vocalizations, demanding robust techniques such as domain adaptation and transfer learning to bridge this gap.

4 Signal processing and acoustic feature extraction

The foundation of computational infant cry analysis lies in transforming raw audio into mathematically tractable representations that capture both the spectral (frequency-related) and temporal (time-varying) properties of vocalizations (Fu et al., 2025). A range of signal processing techniques, including cepstral analysis, wavelet transforms, zero-crossing rates, energy measures, and image-based time series encoding, are used to extract distinctive features that encode the subtle dynamics of neonatal cries, coos, and atypical vocal behavior.

Table 2 summarizes how these methods have been systematically applied in recent studies, each employing specialized tools to ensure precise and consistent feature extraction.

4.1 Cepstral features: MFCC and GFCC

Mel-frequency cepstral coefficients (MFCCs) (Ali et al., 2021) are among the most widely adopted acoustic features for cry analysis. They approximate how the human cochlea perceives sound by mapping the power spectrum of short overlapping frames onto the Mel scale, a scale that reflects human auditory sensitivity to frequency. MFCCs are computed by applying a Fourier transform to windowed frames, mapping the result through triangular Mel filter banks, taking the logarithm of the power at each filter, and then performing a discrete cosine transform (DCT) to decorrelate the coefficients.

In Dey et al. (2025), MFCCs were extracted from the Baby Chillanto dataset (1049 normal and 340 asphyxia cries). The audio data were resampled to 8 kHz, segmented into 1-second windows, denoised, balanced with random oversampling, and framed before MFCC extraction. They visualized MFCCs and structured the data as Pandas DataFrames for downstream processing via Librosa, a popular Python library for audio analysis. Similarly, Ozcan and Gungor (Ozcan & Gungor, 2025) used 13-dimensional MFCCs on the Donate Cry dataset, applying robust data augmentation to diversify training data for their structure-tuned artificial neural network. Kumar Nukala et al. (2024) and Hammoud et al. (2024) combined MFCCs with other metrics in a multidomain framework to encode short-term spectral envelopes alongside other cues.

Expanding beyond MFCCs, Zayed et al. (2023) employed gammatone frequency cepstral coefficients (GFCCs). Unlike MFCCs, GFCCs use a gammatone filter bank, which models human auditory filters more accurately than triangular Mel filters do, especially in noisy conditions. GFCCs were extracted via MATLAB scripts, complementing prosodic and image-based features to construct a robust, fused representation.

4.2 Prosodic features: harmonic ratio

The harmonic ratio (HR) (Bellanca et al., 2013) quantifies the proportion of periodic (harmonic) energy relative to total signal energy, serving as a measure of voice periodicity and phonatory control. In infants, HR helps capture breath support and vocal fold vibration stability. Zayed et al. (2023) extracted HR using MATLAB, integrating it with GFCCs and spectrogram-derived features for multidomain fusion.

4.3 Spectrograms and spectro-temporal representations

A spectrogram visualizes how signal energy is distributed over frequency and time, offering a time–frequency representation useful for highlighting transitions, pitch modulations, and noise bursts. It is typically computed via the short-time Fourier transform (STFT), which divides the signal into short overlapping frames and computes a frequency spectrum for each.

Zayed et al. (2023) generated spectrogram images in Python and extracted deep features via a pretrained VGG16 CNN. This allowed the capture of intricate spectral patterns that simpler features might overlook. In multidomain setups, the spectrogram serves as a high-dimensional, image-based descriptor fused with GFCC and HR.

Kumar Nukala et al. (2024) and Hammoud et al. (2024) extended this by using Mel-Spectrograms, which compress frequency bins onto the Mel scale, further aligning them with human perception. This is especially beneficial when analyzing cries, where high-pitched harmonics and formants convey critical diagnostic information.

4.4 Zero-crossing rate (ZCR) and root mean square energy (RMS)

The zero-crossing rate (ZCR) (Joo et al., 2021) is the rate at which the signal waveform crosses the zero-amplitude axis. This indicates the noisiness or tonal quality of a signal: voiced sounds tend to have a lower ZCR, whereas noisy or unvoiced segments have a higher ZCR. The RMS energy quantifies the signal's average power, reflecting the loudness and respiratory effort in cries. Kumar Nukala et al. (2024) and Hammoud et al. (2024) extracted ZCRs and RMSs via Python libraries, integrating them as simple yet informative descriptors of time-domain dynamics.

4.5 Autocorrelation-based Features

Narayanan et al. (2024) used autocorrelation function (ACF) (Hassani et al., 2024)-based features, such as FZCP (fractional zero-crossing periodicity), rmax (maximum autocorrelation value), kmax (lag at maximum autocorrelation), ZCP12 (zero-crossing periodicity at lag 12), and DR (decorrelation ratio). These features capture periodicity and pitch

Table 2 *Datasets and extracted features used for infant cry and community sound classification.*

Study	Dataset	Features Used	Feature Extraction Details
(Xu et al., 2025)	Custom CED Sound Dataset	FF-Orbital Patterns, UTMDWT	Orbital + wavelet band features (5632); INCA selection
(Dey et al., 2025)	Baby Chillanto dataset: 1049 normal + 340 asphyxia cries; audio sampled to 8 kHz, 1-second segments	MFCC	MFCCs extracted per frame (time & frequency domain); preprocessing: noise removal, outlier handling, label encoding, Random Oversampling; MFCC plots and Pandas DataFrame conversion for model input
(Ozcan & Gungor, 2025)	Donate a Cry	MFCC + Data Augmentation	MFCC (13), robust DA, structure-tuned ANN
(Narayanan et al., 2024)	ESC, Donate a Cry, YouTube	ACF	Five ACF-based features (FZCP, rmax, kmax, ZCP12, DR); blockwise detection
(Kumar Nukala et al., 2024)	Donate a Cry	ZCR, RMS, MFCC, Mel-spectrogram, TSI	457 features; 5 sec audio; multidomain; TSI converts MFCCs to images
(Hammoud et al., 2024)	Donate a Cry	ZCR, RMS, Mel-spectrogram, MFCC, TSI	5-sec audio segments; time (ZCR, RMS), frequency (Mel-spectrogram), time-frequency (MFCC); MFCC transformed with multiple TSI algorithms (GADF, GASF, MTF, RP, RGB-GAF)
(Zayed et al., 2023)	Cry audio recordings from newborns with neonatal RDS, sepsis, and healthy cries; collected in hospitals, segmented into expiratory segments, balanced to 3396 samples	GFCC (cepstral), Harmonic Ratio (prosodic), Spectrogram (image-based)	GFCCs and HR extracted via MATLAB; spectrograms generated via Python; VGG16 CNN used for feature extraction; fusion applied both by simple concatenation and through the deep learning process
(Lee et al., 2020)	Audio from video recordings of 39 infants (ASD & TD); collected at Seoul National University Bundang Hospital; ages 6–24 months	Extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS, 88 features); AutoEncoder bottleneck features	eGeMAPS features extracted using OpenSMILE (frame size: 25 ms, overlap: 10 ms); normalized; AutoEncoder compresses to latent dimension (54); joint optimization with BLSTM; feature map evaluated with t-SNE

consistency by analyzing the correlation between a signal and its delayed version. Their block-wise detection approach enabled the precise localization of sound events within longer cry or community audio segments.

4.6 Time series imaging (TSI)-based techniques

To benefit from the strengths of Deep Learning (DL) models designed for visual input, several studies have converted sequential acoustic features into 2D images. This includes techniques such as the following:

(i) Gramian angular summation field (GASF) and Gramian angular difference field (GADF) (Alsalemi et al., 2023) transform a time series into a polar coordinate system and compute the Gramian matrix on the basis of the angular cosine (summation) or sine (difference), encoding temporal correlations as textures.

(ii) The Markov transition field (MTF) (Zhao et al., 2022) encodes transition probabilities between discrete quantile bins of the time series, capturing dynamics as spatial maps.

(iii) Recurrence plot (RP) (Marwan et al., 2007) plots when states of a system recur in phase space, visualizing similarity patterns.

(iv) RGB-GAF (Chai et al., 2025) stacks multiple GAF matrices into RGB channels, enriching feature diversity.

Kumar Nukala et al. (2024) and Hammoud et al. (2024) used these TSI methods to convert MFCC sequences into image-like representations, enabling convolutional neural networks to learn spatial correlations in temporal data.

4.7 Autoencoder-based dimensionality reduction

To address the high dimensionality and redundancy of large acoustic feature sets, Lee et al. (2020) implemented an autoencoder (AE), a deep neural network, trained to compress input features into a compact latent representation and then reconstruct the original data. They used the extended Geneva minimalistic acoustic parameter set (eGeMAPS), comprising 88 prosodic and spectral features, extracted with OpenSMILE, a robust open-source toolkit for speech analysis. The AE reduced these vectors to a 54-dimensional latent vector, which was then fed to a bidirectional long short-term memory (BLSTM) network for sequential modeling.

5 Machine learning approaches in neonatal vocal analysis

Machine Learning has emerged as a transformative tool in the analysis of neonatal vocalizations, empowering researchers to uncover subtle, clinically meaningful patterns hidden within complex, high-dimensional acoustic signals. As summarized in **Table 3**, various supervised ML techniques, including support vector machines (SVMs), random forests (RFs), logistic regression (LR), decision trees (DTs), k-nearest neighbors (KNNs), and gradient boosting methods such as XGBoost, are commonly deployed to classify infant vocal samples into diagnostically relevant categories, such as typical versus atypical neurodevelopment, respiratory distress, sepsis, or signs of birth asphyxia. Supervised learning relies on datasets annotated with clear ground-truth labels, enabling these algorithms to learn optimal boundaries and rules for differentiating pathological cries from healthy cries with remarkable precision.

For example, Xu et al. (2025) coupled advanced feature selection (INCA) with a Bayesian-optimized SVM, achieving an impressive accuracy of 98.81% for cry classification. Similarly, Dey et al. (2025) systematically compared traditional ML classifiers (such as LR, RF, SVM, KNN, and NB) and reported that logistic regression achieved a near-perfect accuracy of 99.16% for asphyxia detection, outperforming several deep neural variants in the same study (**Table 3**). Compared with these conventional models, ensemble techniques have demonstrated significant gains in prediction robustness. Hammoud et al. (2024) and Kumar Nukala et al. (2024) reported how combining multiple learners, such as RF, XGBoost, and bagging, can exploit the strengths of individual algorithms while mitigating their weaknesses, increasing the classification accuracy above 98% in some cases. Narrayanan et al. (2024) applied a diverse suite of models, including decision trees, naive Bayes, multilayer perceptron (MLP), light gradient boosting machine (LGBM), and KNN, and achieved standout results for rapid and lightweight prediction pipelines suitable for real-time applications in resource-constrained settings.

In addition to classic ML, Deep Learning architectures have become indispensable in neonatal vocal research

Table 3 *Machine learning (ML) and deep learning (DL) algorithms with performance metrics for cry and sound analysis.*

Study	ML/DL Type	Algorithms Used	Performance Measures
(Xu et al., 2025)	ML	INCA (feature selection) + Bayesian-optimized SVM	Accuracy: 98.81%; high Recall, Precision, F1-score (~98.8%)
(Dey et al., 2025)	Combined ML & DL	ML: Logistic Regression (LR), SVM, RF, KNN, DT, NB; DL: custom ANN, ANN1, CNN, CNN1, CNN2 with hidden layers	ML: Best Logistic Regression: 99.16% accuracy, 0.008% error; DL: Best ANN1: 98.20% accuracy, 0.018% error; evaluated using Precision, Recall, F1-score, Confusion Matrix, ROC
(Ozcan & Gungor, 2025)	DL	Structure-Tuned Artificial Neural Network (ANN) with GridSearch + Data Augmentation	Accuracy: 90%, F1-score: 90%
(Narayanan et al., 2024)	ML	DT, RF, NB, MLP, LGBM, KNN	Best SE: NB 99.29% (fast, smallest model); Best overall accuracy: RF 93.77% (higher cost); DT: fast (1.07 ms); KNN: balanced SE 92.82%, SP 94.08%
(Kumar Nukala et al., 2024)	ML (Ensemble focus)	Random Forest, XGBoost, SVM, DT, KNN, LR	Best: RF & XGBoost with 98.03% accuracy (10-fold CV); strong feature importance on MFCCs, ZCR, RMS; visualized via confusion matrices
(Hammoud et al., 2024)	ML Ensemble	RF, SVM, DT, KNN, Bagging	MFCC-RF: Accuracy 96.39% (outperforms prior SOTA 95.17%); multiple TSI variants show robust F1 & precision
(Zayed et al., 2023)	Hybrid ML and DL with feature fusion	Support Vector Machine (SVM), Random Forest (RF), Deep Neural Network (DNN) with VGG16 for spectrogram; GridSearchCV and Keras Tuner for hyperparameter optimization	Achieved highest accuracy of 97.50% (spectrogram + GFCC + HR fused through learning); evaluated using accuracy, precision, recall, F1-score, confusion matrix, ROC curves
(Lee et al., 2020)	Hybrid DL with AutoEncoder feature compression and BLSTM	SVM with linear kernel; vanilla BLSTM; BLSTM jointly optimized with AutoEncoder bottleneck features	Joint optimized BLSTM showed improved ASD detection over vanilla BLSTM; results scored with Unweighted Average Recall (UAR) & Weighted Average Recall (WAR); t-SNE shows clearer feature separability

because of their ability to automatically discover and hierarchically encode complex acoustic patterns without the need for extensive manual feature crafting. Convolutional neural networks (CNNs) excel at extracting localized spectral features from spectrogram images of cry signals, capturing fine-grained details such as pitch modulations, formant transitions, and transient noise bursts that can distinguish healthy from pathological cries. For example, Ozcan & Gungor (2025) leveraged a structure-tuned ANN optimized via GridSearch to enhance the performance of cry classification, whereas Zayed et al. (2023) fused spectrogram-derived CNN features with handcrafted GFCC and prosodic measures, yielding a peak accuracy of 97.50% when deep neural networks with joint feature learning were used (**Table 3**).

Recurrent neural networks (RNNs), especially bidirectional long short-term memory (BLSTM) networks, are equally pivotal for modeling the temporal dependencies inherent in cry sequences. Lee et al. (2020) innovatively combined an AE for feature compression with a BLSTM classifier, refining the latent representation of high-dimensional speech parameters such as eGeMAPS and improving the discrimination of autism spectrum disorder vocal sign-

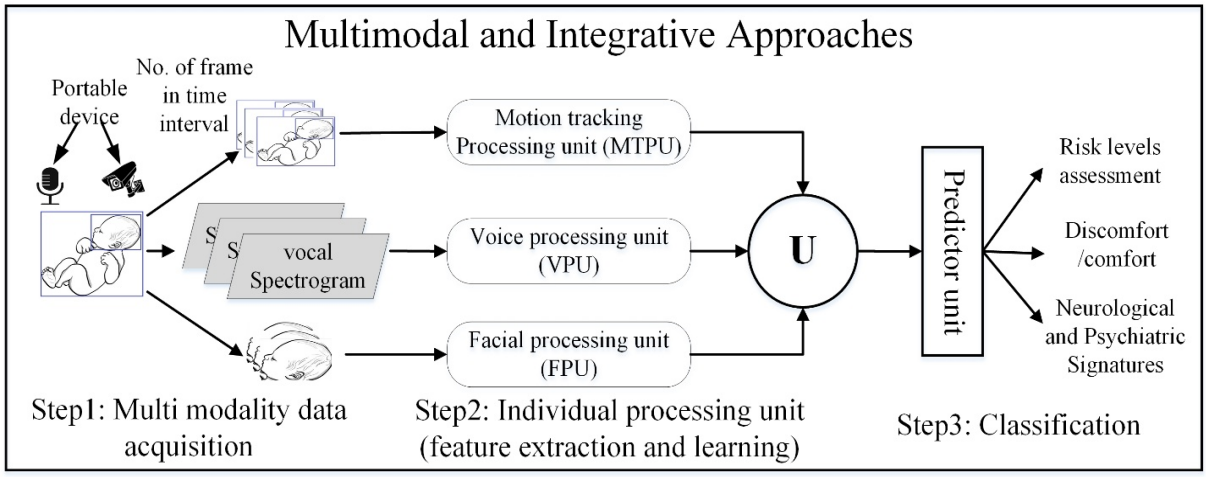

Figure 2 *A conceptual multimodal workflow integrating motion, voice, and facial data for comprehensive neonatal state assessment.*

atures. Their approach demonstrated that compressing features into a meaningful bottleneck and jointly optimizing it with a sequence model can enhance both detection accuracy and interpretability when dealing with sparse and noisy infant vocal data.

A critical consideration across studies is the balance between manual feature engineering and automated feature learning. Traditional pipelines often depend on carefully selected prosodic (pitch, intonation), cepstral (MFCC, GFCC), and spectral (harmonics-to-noise ratio, formant frequencies) features, which are grounded in phonetic and neurological knowledge. These handcrafted features are transparent and interpretable but may overlook nuanced, nonlinear relationships. In contrast, DL's automated feature learning, which is evident in CNNs and autoencoders, allows models to discover latent signal representations directly from raw waveforms or time-frequency images. While powerful, this strategy demands larger datasets and can reduce interpretability, a challenge partially mitigated by hybrid approaches such as those of Zayed et al. (2023) and Lee et al. (2020), who combine handcrafted and learned features to exploit the best of both worlds (**Table 3**).

As models grow more sophisticated and are increasingly entrusted with supporting early diagnosis and clinical decision-making, explainability becomes paramount. Explainable AI (XAI) techniques such as SHAP, LIME, and neural attention mechanisms are now being integrated into neonatal vocal analysis pipelines. These tools clarify which features or time segments drive a model's predictions, bolster clinician confidence and facilitate trust in automated systems. This transparency not only helps validate the biological plausibility of discovered vocal biomarkers but also aids in refining models by exposing biases or misclassifications. Ultimately, embedding XAI transforms black-box predictors into intelligible, clinically actionable tools, closing the gap between advanced computation and practical neonatal healthcare.

6 Multimodal and integrative approaches

In neonatology and early developmental neuroscience, the interpretation of infant vocalizations is increasingly recognized as a valuable window into the developing brain and nervous system (Narayanan et al., 2022). However, the diagnostic and predictive utility of infant cry and vocal sound analysis is substantially amplified when these acoustic features are interpreted within a multimodal framework that includes concurrent physiological and behavioral signals (Pigueiras-del-Real et al., 2024b). Modern research and clinical practice emphasize that no single data stream, whether audio, visual, or physiological, can fully capture the complexity of an infant's internal state or developmental trajectory.

As depicted in **Figure 2**, a multimodal and integrative pipeline typically begins with synchronized data acquisition from portable sensors, such as microphones and cameras, capturing voice, facial expressions, and body movements within defined time intervals. Each modality is then processed through specialized units such as motion tracking, vocal spectrogram analysis, and facial feature extraction before being unified in a

central fusion module. This combined information feeds into a predictive unit that classifies the infant's state, enabling risk assessment, discomfort detection, and the identification of early neurological or psychiatric signatures. Such architectures highlight how leveraging diverse and complementary data streams can significantly improve the sensitivity and reliability of neonatal monitoring systems.

Combining vocal features with physiological measures such as heart rate variability, respiration, or cortical signals (EEGs) can reveal how the autonomic and central nervous systems coordinate during stress, pain, or social engagement (Shah et al., 2025), as described in **Table 4**. For example, synchronizing cry acoustics with heart rate or oxygen saturation can clarify whether a seemingly normal cry actually masks distress or autonomic dysregulation. Similarly, coupling vocalizations with EEG signals helps identify the neural circuits involved in phonation control, which can be disrupted in certain neurological conditions. Video-based monitoring adds further depth by capturing facial expressions and body movements that contextualize vocal sounds, distinguishing a pain cry from a hunger cry or a cry accompanied by unusual posturing that may suggest a neuromotor problem.

To achieve this integration, researchers deploy multisensor platforms that combine microphones, high-definition cameras for pose or facial keypoint tracking, wearable biosensors for heart rate and oxygen saturation, and sometimes bedside devices for cortical or hemodynamic measurements (e.g., near-infrared spectroscopy, NIRS) (Pigueiras-del-Real et al., 2024a). These sensors collect time-synchronized data, creating a detailed temporal map of behaviors and physiological states. This is crucial because temporal cooccurrences such as a cry immediately following a noxious stimulus and accompanied by a spike in heart rate can provide robust evidence of pain perception and the effectiveness of analgesic interventions.

Recent studies have demonstrated the ability of ML and DL to process rich, multimodal data. For example, Natraj et al. (2024) combined OpenPose (Cao et al., 2021) for extracting body keypoints from video with a deep feature extractor (VGG16) and a temporal sequence model (LSTM) to interpret body movement patterns over time. In parallel, audio streams are processed using signal features such as Mel-frequency cepstral coefficients (MFCCs) and spectrograms and then classified via 1D-CNN architectures designed for sequential audio data. These modality-specific predictions are then fused via ensemble decision rules (logical OR/AND) or trained decision tree classifiers, increasing accuracy by leveraging complementary strengths: movement cues can disambiguate ambiguous cries, and vice versa.

Similarly, Salekin (2022) used a bilinear VGG16 network, a variant that models fine-grained interactions between features extracted from different modalities (e.g., face and body), followed by an LSTM to capture the temporal progression of pain-related facial expressions and body movements in neonates in the NICU. For the audio channel (crying), spectrogram images were analyzed with another VGG16 network. The outputs of these unimodal branches were fused via decision fusion, where predictions from each channel were combined to produce a final estimate of pain or behavior. The results showed that this integrated approach outperformed single-channel models, with the multimodal system achieving an area under the curve (AUC) of 0.90 compared with 0.78--0.87 for individual modalities.

Additionally, Shah (2025) presented a unified, explainable framework for neonatal health monitoring that integrates four complementary data streams: facial features, vocal expressions, electrocardiograms (ECGs), and motion keypoints. For facial analysis, high-quality facial regions are automatically cropped via MediaPipe and then processed via a hybrid CNN architecture (FRAI) that combines GoogleNet and AlexNet, which achieves precision, recall, and F1 scores of approximately 92-95% on benchmark datasets. Vocal expressions are converted to spectrograms and analyzed with modified XceptionNet, a computationally efficient variant of the Xception network, which delivers robust emotion and pain detection with precision and recall ranging from 90% to 98% in both the adult and infant datasets. ECG signals from out-of-hospital cardiac arrest cases are transformed into visual descriptors via SIFT and bag-of-visual-words, fused with patient metadata, and classified with an ensemble machine learning model, attaining a balanced accuracy of 0.82 and an AUC-ROC of 0.90. For motion tracking, advanced angular and inertia-based interpolation combined with LSTM networks accurately imputes missing trajectory data, significantly improving error metrics such as the MSE, MAE, RMSE, cosine, and Huber losses. This multisensor, hybrid-fusion approach demonstrates strong real-time performance and interpretability through SHAP, LIME, and Grad-CAM, offering clinicians a transparent and holistic tool for the early detection of neurodevelopmental issues and distress in neonates.

Table 4 *Overview of recent studies on multimodal automatic pain and behavior monitoring using diverse sensor data and AI methods.*

Study	Multimodal Data	Features Used	Machine/Deep Learning Techniques	Performance Measures
(Natraj et al., 2024)	Video (pose estimation) Audio (ADOS sessions)	Video: OpenPose keypoints Audio: MFCC, Mel spectrogram, tonal & spectral features	Video: VGG16-LSTM Audio: 1D-CNN Fusion: Ensemble (OR, AND, Decision tree)	Video: 80% acc Audio: 78.8% acc Ensemble: 82.5% acc, F1: 0.816 OR: Sensitivity 90% AND: Specificity 92.5%
(Salekin, 2022)	Neonates in NICU (USF-MNPAD-I & II)	Video (face/body), Audio (crying), Vital signs, NIRS	Facial/Body: Bilinear VGG16 + LSTM Sound: Spectrogram VGG16 Multimodal: Decision fusion	Unimodal AUC: Face (0.82), Body (0.78), Sound (0.87) Multimodal AUC: 0.90)
(Shah, 2025)	Facial features, Vocal expressions, ECG, Motion keypoints	Facial: Cropped face regions via MediaPipe; Vocal: Spectrograms processed with lightweight Xception (VocalXpressNet); ECG: Visual SIFT + BoVW; Motion: Angular & inertia interpolation + LSTM for missing data reconstruction	Facial: GoogleNet + AlexNet; Vocal: Modified Xception; ECG: Ensemble ML with visual features + demographic metadata; Motion: LSTM; fusion via hybrid early-late fusion	Facial: M3B Diseases: Acc 0.89, Bal. Acc 0.94; Vocal (Adult/Infant): Precision 90–97%, Recall 90–98%, F1 >90%; ECG (OHCA): Balanced Accuracy 0.82, AUC-ROC 0.90; Motion: Improved MSE, MAE, RMSE, Cosine & Huber Loss; reliable real-time imputation

A crucial advantage of these frameworks is that they provide contextualization: an isolated abnormal vocal pattern may have low specificity but becomes clinically meaningful when corroborated by co-occurring abnormal motor or physiological signs. This mitigates false positives and adapts more flexibly to interindividual variability, which is a major challenge in neonatal care, where rapid developmental changes are the norm. Multimodal systems can thus support personalized baselines and help clinicians distinguish pathology-driven anomalies from benign situational variations, enhancing both sensitivity and specificity.

7 Addressing methodological challenges

As neonatal vocal analysis becomes increasingly reliant on data-driven models, several methodological challenges must be overcome to ensure the reliability, robustness, and clinical applicability of these technologies (Gómez-Vilda et al., 2022). One of the foremost obstacles is the variability introduced by differences in recording environments, devices, and population characteristics across datasets. Models trained on a single dataset often fail to generalize effectively to new clinical settings or demographic groups, thereby limiting their scalability and practical utility. To mitigate this, domain adaptation and transfer learning have emerged as powerful strategies (Gichoya et al., 2023). These techniques enable models to utilize knowledge learned from one domain, such as a well-annotated dataset from a specific hospital, and apply it to new, unseen domains with minimal retraining (S. T. H. Shah et al., 2024). Through mechanisms such as fine-tuning, adversarial learning, and feature

alignment, models can adapt to new acoustic conditions or patient demographics, thereby enhancing cross-context reliability without requiring extensive new annotations.

Another critical tool for addressing methodological constraints is data augmentation. Given the scarcity and heterogeneity of high-quality neonatal vocal datasets, augmentation techniques play a vital role in expanding the effective size and diversity of training data (Fayaz et al., 2024). Traditional methods such as pitch shifting, time stretching, background noise injection, and waveform perturbation help simulate real-world variability and improve model robustness (Wen et al., 2025). More advanced techniques, including generative adversarial networks (GANs) and synthetic speech generation, are now being explored to produce realistic infant vocalizations that preserve meaningful acoustic features while introducing novel examples. These synthetic datasets not only increase model performance but also help reduce overfitting, especially for small or imbalanced datasets (Ma et al., 2025).

Finally, rigorous cross-dataset and cross-population validation is essential to establish the generalizability of predictive models. Too often, algorithms are evaluated only on the dataset used for training or within a narrow clinical context, which can result in overestimated performance and limited real-world applicability. By testing models across multiple datasets collected from diverse institutions, geographic regions, and patient subgroups, researchers can detect performance inconsistencies, potential biases, and applicability constraints. This level of validation is particularly critical in neonatal care, where interindividual variability is high and the clinical stakes are significant. Ensuring that models maintain accuracy across diverse populations not only builds clinical confidence but also advances the field toward the deployment of reliable, inclusive tools for early neurodevelopmental assessment.

8 Clinical translation and validation

The successful deployment of neonatal vocal analysis technologies in clinical practice hinges on the seamless translation of computational models into practical, reliable tools that meet the rigorous demands of real-world healthcare settings (Ganti, 2025). While laboratory-based models often demonstrate promising accuracy in detecting vocal biomarkers, transitioning these models to bedside and remote applications requires addressing critical factors such as clinical workflow compatibility, regulatory approval, interpretability, and ease of use (Arya et al., 2023). This means that models must not only perform well under controlled conditions but also maintain their diagnostic accuracy and consistency amid the noisy, unpredictable nature of clinical environments. To gain acceptance among health care professionals, these tools must produce outputs that are understandable, actionable, and clinically meaningful, which often necessitates the integration of explainable AI techniques and intuitive, user-friendly interfaces.

Clinical translation is inherently interdisciplinary. It demands active, sustained collaboration between data scientists, engineers, clinicians, speech-language pathologists, neonatologists, and healthcare administrators (Dalwai, 2021; Murphy et al., 2025). This collaboration ensures that the development of neonatal vocal analysis tools is firmly grounded in real-world needs, constraints, and practical opportunities. Clinicians contribute crucial domain knowledge about symptom presentation, patient variability, and the clinical relevance of vocal and physiological features, whereas engineers and data scientists design robust algorithms and ensure seamless system integration. These partnerships are also vital for addressing the ethical, privacy, and logistical complexities of data collection, especially when working with vulnerable populations such as neonates.

In addition to hospital environments, interest in the use of neonatal vocal analysis for remote and continuous monitoring is increasing. Real-world applications include home-based tracking of high-risk infants post-discharge, early screening in rural or underserved communities, and integration with telehealth platforms (Chiang et al., 2021). Wearable or ambient sensors, combined with cloud-based processing and mobile interfaces, can empower caregivers and clinicians to monitor developmental signals in real time and intervene promptly when anomalies arise. These applications represent a paradigm shift in pediatric healthcare, moving from reactive treatment to proactive, data-driven early intervention (Shah et al., 2022, 2025).

However, for such systems to be clinically viable, they must undergo rigorous validation in diverse, uncontrolled environments and be designed to safeguard privacy, minimize caregiver burden, and function with minimal calibration or technical oversight (Al-Worafi, 2024). When effectively translated, these innovations have the potential to transform neonatal care, enabling earlier diagnosis,

personalized treatment plans, and improved neurodevelopmental outcomes for at-risk infants.

9 Future directions

Future directions must prioritize advancing and refining multimodal integration approaches to fully utilize the diverse streams of information available from neonates. While significant progress has been made in analyzing vocalizations (Jeong & Ha, 2025), physiological signals (Gentile et al., 2023), motion patterns (Bruschetta et al., 2025), and facial features (Shah et al., 2023) independently, the greatest clinical benefit will come from combining these data sources in an intelligent, synchronized manner. The development of flexible and robust fusion strategies, including hybrid architectures that blend early, intermediate, and late fusion methods, will enable systems to capture subtle interactions across modalities, increasing accuracy and reliability in real-world settings (Guarrasi et al., 2025). Advanced techniques such as attention mechanisms, transformer-based fusion, and graph-based relational models hold great promise for modeling complex cross-modal relationships.

Equally crucial is the establishment of standardized, high-quality multimodal datasets that represent diverse clinical contexts, populations, and recording conditions. Current datasets often vary widely in quality and completeness, hindering generalizability (Krones et al., 2025). Collaborative efforts to build large-scale, open-access repositories can support the training and validation of models that are resilient to variations and bias and can better reflect the full spectrum of neonatal health states. Additionally, research should investigate innovative data augmentation and synthetic data generation methods to address data imbalance and scarcity, particularly for rare conditions.

To ensure that these technologies can be deployed effectively at the bedside and beyond, efficiency and scalability must be prioritized. Future work should focus on compressing complex multimodal models without sacrificing interpretability or diagnostic performance. Techniques (Violos et al., 2025) such as model pruning, quantization, and edge computing adaptations can enable real-time monitoring of portable devices or smart incubators, extending continuous care into both hospital and home environments.

Explainability and user interaction will remain central to the clinical adoption of these systems. As models become increasingly complex with multimodal inputs, new explainable AI frameworks must be designed to clearly communicate how each data type contributes to predictions and enable clinicians to understand the rationale behind alerts or risk scores (S. A. H. Shah et al., 2024). User-friendly interfaces and visualization tools support this goal, fostering trust among caregivers and healthcare professionals.

Finally, longitudinal and adaptive modeling should be a key research direction. Rather than relying solely on single-timepoint predictions, future systems should track developmental changes over time, detect subtle deviations from expected growth patterns, and adapt recommendations dynamically as more data becomes available (Kraus et al., 2023). This shift toward continuous, personalized monitoring can empower clinicians and families to intervene earlier, tailor care plans more precisely, and ultimately improve neurodevelopmental outcomes for infants at risk.

10 Conclusion

The analysis of neonatal vocal expression stands at the intersection of neuroscience, signal processing, and artificial intelligence, offering a promising avenue for the early detection of neurological and developmental conditions. This review outlines key methodologies, ranging from traditional spectro-temporal analysis and acoustic feature extraction to advanced machine learning models capable of capturing subtle vocal biomarkers. Foundational preprocessing steps, such as noise reduction and feature extraction, transform raw infant cries into informative representations for robust computational analysis. Both classical algorithms and modern deep learning architectures, including convolutional and recurrent neural networks, have shown substantial potential for detecting early signs of neurocognitive risk. Importantly, the integration of explainable AI techniques and the growing shift toward multimodal data fusion, which combines vocal cues with complementary physiological signals, motion patterns, and facial information, has enhanced the robustness and interpretability of these systems, thereby increasing their clinical utility. The implications for early diagnosis and timely intervention are profound. By capturing and contextualizing vocal signals within the broader framework of an infant's physiological and behavioral state, clinicians can detect risk factors for conditions such as autism spectrum disorder, cerebral palsy, or language impairments well before overt symptoms emerge. Early identification paves the way for targeted interventions during critical windows of neuroplasticity, improving long-term developmental outcomes. Moreover, noninvasive and

cost-effective voice analysis methods are ideally suited for widespread screening, including in low-resource settings where access to advanced diagnostic tools may be limited.

As this field evolves, the integration of wearable and ambient sensors, mobile platforms, and cloud-based analytics promises to enable continuous, personalized monitoring for high-risk infants, both in hospital and home settings. In the future, the field must converge on a unified framework for developing and validating neonatal cognitive and behavioral biomarkers. This should prioritize standardized data collection protocols, inclusive and diverse datasets, rigorous cross-population validation, and close interdisciplinary collaboration to ensure both technical excellence and clinical relevance. Equally, ethical safeguards, including data privacy, informed consent, and equitable access, must remain central as these tools transition into practice. By advancing toward this vision, neonatal vocal analysis and multimodal monitoring can transition from promising research prototypes to practical, trustworthy tools for neurodevelopmental health, laying the groundwork for a new era of precision medicine for newborns and providing every child with the best possible start in life.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Funding

"The present research was carried out as part of the PARENT and GALATEA projects, funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Innovative Training Network 2020, Grant Agreements No. 956394 (https://parenth2020.com/) and No. 101183057. (https://cordis.europa.eu/project/id/101183057).

Additional support was provided by the research program of the Extended Partnership "Future Artificial Intelligence Research - FAIR," Spoke 1 "Human-Centered AI," coordinated by the Università di Pisa. The study was developed within the TECHNET (Technology Assisted Narrative Exposure Therapy) project, funded under CUP: 83D24000260004, which aims to advance human-centered AI applications in therapeutic and socially impactful domains through interdisciplinary research and innovation."

Acknowledgement

The authors would like to express their sincere gratitude to Politecnico di Torino, particularly the PolitoBioMed Lab under the Department of Mechanical and Aerospace Engineering, for their valuable support and collaboration. Special thanks also go to GPI SpA, Department of Research and Development, for their technical insights and contributions, and to 7HC SRL for their partnership and continued encouragement throughout this work.

References

Alexlinander (2022) 2022DSPLab: Detecting baby sounds. https://kaggle.com/competitions/2022dsplab-detecting-baby-sounds, 2022. Kaggle.

Ali, S., Tanweer, S., Khalid, S.& Rao, N. (2021) Mel frequency cepstral coefficient: a review. In, *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development*, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India.

Alsalemi, A., Amira, A., Malekmohamadi, H. & Diao, K. (2023) Novel domestic building energy consumption dataset: 1D timeseries and 2D Gramian angular fields representation. *Data Brief* **47**, 108985.

Al-Worafi, Y.M. (2024) Patient care related issues in the developing countries: monitoring parameters. In, *Handbook of Medical and Health Sciences in Developing Countries*. Springer, Cham, pp. 1–23.

Andonotopo, W., Bachnas, M.A., Dewantiningrum, J., Pramono, M.B.A., Stanojevic, M. & Kurjak, A. (2025) AI and early diagnostics: mapping fetal facial expressions through development, evolution, and 4D ultrasound. *Journal of Perinatal Medicine* **53**, 263–285.

Arya, S.S., Dias, S.B., Jelinek, H.F., Hadjileontiadis, L.J. & Pappa, A.-M. (2023) The convergence of traditional and digital biomarkers through AI-assisted biosensing: a new era in translational diagnostics? *Biosensors and Bioelectronics* **235**, 115387.

Bartl-Pokorny, K.D., Pokorny, F.B., Garrido, D., Schuller, B.W., Zhang, D. & Marschik, P.B. (2022) Vocalisation repertoire at the end of the first year of life: an exploratory comparison of Rett syndrome and typical development. *Journal of Developmental and Physical Disabilities* **34**, 1053–1069.

Bellanca, J.L., Lowry, K.A., VanSwearingen, J.M., Brach, J.S. & Redfern, M.S. (2013) Harmonic ratios: a quantification of step to step symmetry. *Journal of Biomechanics* **46**, 828–831.

Bruschetta, R., Caruso, A., Micai, M., Campisi, S., Tartarisco, G., Pioggia, G. & Scattoni, M.L. (2025) Marker-less video analysis of infant movements for early identification of neurodevelopmental disorders. *Diagnostics* **15**, 136.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. & Sheikh, Y. (2021) OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 172-186.

Cappelli, G. & Noccetti, S. (2022) A Linguistic Approach to the Study of Dyslexia. Multilingual Matters, Bristol, UK.

Chai, Y., Deng, L., Shao, R., Zhang, J., Xing, L., Zhang, H. & Liu, Y. (2025) GAF: Gaussian action field as a dynamic world model for robotic manipulation. arXiv.org. *Preprint.*

Chiang, M.F., Starren, J.B. & Demiris, G. (2021) Telemedicine and telehealth. In, Shortliffe, E.H., Cimino, J.J. (Eds.), *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Springer, Cham, pp. 667–692.

Dalwai, S.H. (2021) *IAP Handbook of Developmental and Behavioral Pediatrics*. Jaypee Brothers Medical Publishers.

Dey, S.K., Mohi Uddin, K.M., Howlader, A., Mahbubur Rahman, Md., Babu, H.Md.H., Biswas, N., Siddiqi, U.R. & Mazumder, B., (2025) Analyzing infant cry to detect birth asphyxia. *Neuroscience Informatics* **5**, 100193.

Ding, H. & Zhang, Y. (2023) Speech prosody in mental disorders. *Annual Review of Linguistics* **9**, 335–355.

Fayaz, S., Shah, S.Z.A., Din, N.M. ud, Gul, N. & Assad, A. (2024) Advancements in data augmentation and transfer learning: a comprehensive survey to address data scarcity challenges. *Recent Advances in Computer Science and Communications* **17**, 14–35.

Filippa, M., Della Casa, E., D'amico, R., Picciolini, O., Lunardi, C., Sansavini, A. & Ferrari, F. (2021) Effects of early vocal contact in the neonatal intensive care unit: study protocol for a multi-centre, randomised clinical trial. *International Journal of Environmental Research and Public Health* **18**, 3915.

Filippa, M. & Kuhn, P. (2024) Early parental vocal contact in neonatal units: rationale and clinical guidelines for implementation. *Frontiers in Neurology* **15**, 1441576

Frank, M.C. (2020) Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science* **3**, 24–52.

Fu, M., Li, D., Gadhiya, A., Lambright, B., Alowais, M., Bahnassy, M., Elletter, S.E.D., Toyin, H.O., Jiang, H., Zhang, K., Aldarmaki, H. (2025) Infant cry detection using causal temporal representation. In, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Ganti, V.K.A.T. (2025) Beyond the stethoscope: how artificial intelligence is redefining diagnosis, treatment, and patient care in the 21st century. Deep Science Publishing.

Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M. & Ritter, M. (2017) Audio set: an ontology and human-labeled dataset for audio events. In, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), pp. 776–780.

Gentile, F.R., Shah, S.T.H., Sperti, M., Panagiotopoulos, K., Primi, R., Bendotti, S., Currao, A., Compagnoni, S., Baldi, E., Lopiano, C., Vicini Scajola, L., Marconi, G., Deriu, M.A. & Savastano, S. (2023) An innovative medical decision support tool for neurological outcome prediction from post-resuscitation electrocardiograms (MILESTONE). *European Heart Journal* **44**, ehad655.650.

Gichoya, J.W., Thomas, K., Celi, L.A., Safdar, N., Banerjee, I., Banja, J.D., Seyyed-Kalantari, L., Trivedi, H. & Purkayastha, S. (2023) AI pitfalls and what not to do: mitigating bias in AI. *British Journal of Radiology* **96**, 20230023.

Gómez-Vilda, P., Gómez-Rodellar, A., Palacios-Alonso, D., Rodellar-Biarge, V. & Álvarez-Marquina, A. (2022) The role of data analytics in the assessment of pathological speech—a critical appraisal. *Applied Sciences* **12**, 11095.

Guarrasi, V., Aksu, F., Caruso, C.M., Di Feola, F., Rofena, A., Ruffini, F. & Soda, P. (2025) A systematic review of intermediate fusion in multimodal deep learning for biomedical applications. *Image and Vision Computing* **158**, 105509.

Hammoud, M., Getahun, M.N., Baldycheva, A. & Somov, A. (2024) Machine learning-based infant crying interpretation. *Frontiers in Artificial Intelligence* **7**,1337356.

Hassani, H., Royer-Carenzi, M., Mashhad, L.M., Yarmohammadi, M. & Yeganegi, M.R. (2024) Exploring the depths of the autocorrelation function: its departure from normality. *Information* **15**, 449.

Hou, X., Zhang, P., Mo, L., Peng, C. & Zhang, D. (2024) Neonatal sensitivity to vocal emotions: a milestone at 37 weeks of gestational age. *eLife* **13**, RP95393.

Husain, A., Knake, L., Sullivan, B., Barry, J., Beam, K., Holmes, E., Hooven, T., McAdams, R., Moreira, A., Shalish, W. & Vesoulis, Z. (2025) AI models in clinical neonatology: a review of modeling approaches and a consensus proposal for standardized reporting of model performance. *Pediatric Research* https://doi.org/10.1038/s41390-025-04207-6

Jeong, Y. & Ha, S. (2025) Early developmental changes in infants' vocal responses in interactions with caregivers. *Infant Behavior and Development* **78**, 102022.

Joo, S., Choi, J., Kim, N. & Lee, M.C. (2021) Zero-crossing rate method as an efficient tool for combustion instability diagnosis. *Experimental Thermal and Fluid Science* **123**, 110340.

Kamiloğlu, R.G. & Sauter, D.A. (2021) Voice production and perception. In: *Oxford Research Encyclopedia of Psychology.* Oxford University Press, Oxford.

Kao, C. & Zhang, Y. (2025) Age and sex differences in infants' neural sensitivity to emotional prosodies in spoken words: a multifeature oddball study. *Journal of Speech, Language, and Hearing Research* **68**, 332–348.

Keles, E. & Bagci, U. (2023) The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *NPJ Digital Medicine* **6**, 220.

Kraus, B., Zinbarg, R., Braga, R.M., Nusslock, R., Mittal, V.A. & Gratton, C. (2023) Insights from personalized models of brain and behavior for identifying biomarkers in psychiatry. Neuroscience and Biobehavioral Reviews 152, 105259.

Kristian, Y., Simogiarto, N., Sampurna, M.T.A., Hanindito, E. & Visuddho, V. (2023) Ensemble of multimodal deep learning autoencoder for infant cry and pain detection. *F1000Research* **11**, 359.

Krones, F., Marikkar, U., Parsons, G., Szmul, A. & Mahdi, A. (2025) Review of multimodal machine learning approaches in healthcare. *Information Fusion* **114**, 102690.

Kumar Nukala, V., Reddy Motheline, S., Wesley Kolasanakoti, J., Vankayalapati, S., Velupula, V. & Reddy Dodda, V. (2024) Advanced machine learning approaches for infant cry classification using audio feature extraction. In, *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS),* pp.1–7.

La Rosa, V.L., Geraci, A., Iacono, A. & Commodari, E. (2024) Affective touch in preterm infant development: neurobiological mechanisms and implications for child–caregiver attachment and neonatal care. *Children* **11**, 1407.

Lee, J.H., Lee, G.W., Bong, G., Yoo, H.J. & Kim, H.K. (2020) Deep-learning-based detection of infants with autism spectrum disorder using auto-encoder feature representation. *Sensors* **20**, 6762.

Liu, Q., Song, L., Xu, D. & Long, Y. (2025) ICSD: an open-source dataset for infant cry and snoring detection.arXiv.org. *Preprint.*

Long, B., Xiang, V., Stojanov, S., Sparks, R.Z., Yin, Z., Keene, G.E., Tan, A.W.M., Feng, S.Y., Zhuang, C., Marchman, V.A., Yamins, D.L.K. & Frank, M.C.(2024) The BabyView dataset: high-resolution egocentric videos of infants' and young children's everyday experiences.arXiv.org. *Preprint.*

Long, H.L., Eichorn, N. & Oller, D.K. (2023) A probe study on vocal development in two infants at risk for cerebral palsy. *Developmental Neurorehabilitation* **26**, 44–51.

Ma, F., Li, Y., Xie, Y., He, Y., Zhang, Y., Ren, H., Liu, Z., Yao, W., Ren, F., Yu, F.R. & Ni, S. (2024) A review of human emotion synthesis based on generative technology.*IEEE Transactions on Affective Computing (Preprint)*.

Mallegni, N., Molinari, G., Ricci, C., Lazzeri, A., La Rosa, D., Crivello, A. & Milazzo, M. (2022) Sensing devices for detecting and processing acoustic signals in healthcare. *Biosensors* **12**, 835.

Marschik, P.B., Widmann, C.A.A., Lang, S., Kulvicius, T., Boterberg, S., Nielsen-Saines, K., Bölte, S., Esposito, G., Nordahl-Hansen, A., Roeyers, H., Wörgötter, F., Einspieler, C., Poustka, L. & Zhang, D. (2022a) Emerging verbal functions in early infancy: lessons from observational and computational approaches on typical development and neurodevelopmental disorders. *Advances in Neurodevelopmental Disorders* **6**, 369–388.

Marschik, P.B., Widmann, C.A.A., Lang, S., Kulvicius, T., Boterberg, S., Nielsen-Saines, K., Bölte, S., Esposito, G., Nordahl-Hansen, A., Roeyers, H., Wörgötter, F., Einspieler, C., Poustka, L. & Zhang, D. (2022b) Emerging verbal functions in early infancy: lessons from observational and computational approaches on typical development and neurodevelopmental disorders. *Advances in Neurodevelopmental Disorders* **6**, 369–388.

Marwan, N., Romano, M.C., Thiel, M. & Kurths, J. (2007) Recurrence plots for the analysis of complex systems. *Physics Reports* **438**, 237-329

Meissen, F., Breuer, S., Knolle, M., Buyx, A., Müller, R., Kaissis, G., Wiestler, B. & Rückert, D. (2024) (Predictable) performance bias in unsupervised anomaly detection. *eBioMedicine* **101**,105002.

Murphy, M.M., Colquitt, G.T., Ryals, P.S., Shin, K., Kjeldsen, W.C., McIntyre, A., Whitten, S.V.W., Modlesky, C.M. & Maitre, N.L. (2025) Synergies, discrepancies, and action priorities: a statewide engagement study to strengthen clinical research in cerebral palsy. *Health Expectations* **28**, e70257.

Narayanan, D.Z., Takahashi, D.Y., Kelly, L.M., Hlavaty, S.I., Huang, J. & Ghazanfar, A.A. (2022) Prenatal development of neonatal vocalizations. *eLife* **11**, e78485.

Narayanan, S.P., Manikandan, M.S. & Cenkeramaddi, L.R. (2024) Fast autocorrelation feature-based infant cry detector for resource-efficient affordable edge cry sound analysis systems. In, *2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA)*, pp.1–6.

Natraj, S., Kojovic, N., Maillart, T. & Schaer, M. (2024) Video-audio neural network ensemble for comprehensive screening of autism spectrum disorder in young children. *PLOS ONE* **19**, e0308388.

Nussbaum, C., Frühholz, S. & Schweinberger, S.R. (2025) Understanding voice naturalness. *Trends in Cognitive Sciences* **29**, 467–480.

Onciul, R., Tataru, C.-I., Dumitru, A.V., Crivoi, C., Serban, M., Covache-Busuioc, R.-A., Radoi, M.P. & Toader, C. (2025) Artificial intelligence and neuroscience: transformative synergies in brain research and clinical applications. *Journal of Clinical Medicine* **14**, 550.

Ozcan, T. & Gungor, H. (2025) Baby cry classification using structure-tuned artificial neural networks with data augmentation and MFCC features. *Applied Sciences* **15**, 2648.

Piczak, K.J. (2015) ESC: dataset for environmental sound classification. In, *Proceedings of the 23rd ACM International Conference on Multimedia, MM '15. Association for Computing Machinery*, New York, NY, USA, pp.1015–1018.

Pigueiras-del-Real, J., Gontard, L.C., Benavente-Fernández, I., Lubián-López, S.P., Gallero-Rebollo, E. & Ruiz-Zafra, A. (2024a) NRP: a multi-source, heterogeneous, automatic data collection system for infants in neonatal intensive care units. *IEEE Journal of Biomedical and Health Informatics* **28**, 678–689.

Pigueiras-del-Real, J., Ruiz-Zafra, A., Benavente-Fernández, I., Lubián-López, S.P., Shah, S.A.H., Shah, S.T.H. & Gontard, L.C. (2024b) NeoVault: empowering neonatal research through a neonate data hub. *BMC Pediatrics* **24**, 787.

Reyes-Galaviz, O.F., Cano-Ortiz, S.D. & Reyes-García, C.A. (2008) Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies. In, *2008 Seventh Mexican International Conference on Artificial Intelligence*, pp.330–335.

Ribolsi, M., Fiori Nastro, F., Pelle, M., Medici, C., Sacchetto, S., Lisi, G., Riccioni, A., Siracusano, M., Mazzone, L. & Di Lorenzo, G. (2022) Recognizing psychosis in autism spectrum disorder. *Frontiers in Psychiatry* **13**, 768586.

Romo, N., Robb, M.P., Lee, J. & Wermke, K. (2024) Noise phenomena in distress cries of term and very preterm infants at term-equivalent age. *Logopedics Phoniatrics Vocology* **50**, 48-54.

Rudenko, Y. (2023) Neurophysiological and neuropsychological mechanisms of vocalization contrasting with music perception. *SSRN Blog.*

Salekin, M.S. (2022) Generative spatio-temporal and multimodal analysis of neonatal pain. *Ph.D.thesis* University of South Florida, United States – Florida.

Sanna, M. (2025) Proprioceptive resonance and multimodal semiotics: readiness to act, embodied cognition, and the dynamics of meaning. *NeuroSci* **6**, 42.

Shah, S.A.H., di Terlizzi, A. & Deriu, M.A. (2022) Intelligent system development to monitor neonatal behaviour: a review. In, *Conference: International Workshop in Neurodevelopmental Impairments in Preterm Children - Computational Advancements* (DETERMINED 2022). Ljubljana, Slovenia.

Shah, S.A.H., Shah, S.T.H., Khaled, R., Buccoliero, A., Shah, S.B.H., Di Terlizzi, A., Di Benedetto, G. & Deriu, M.A. (2024) Explainable AI-based skin cancer detection using CNN, particle swarm optimization and machine learning. *Journal of Imaging* **10**, 332.

Shah, S.T.H., 2025. Multimodal AI tools for predicting neurological and neurodevelopmental trajectories. *PhD thesis*. Politecnico di Torino, Italy – Torino.

Shah, S.T.H., Shah, S.A.H., Khan, I.I., Imran, A., Shah, S.B.H., Mehmood, A., Qureshi, S.A., Raza, M., Di Terlizzi, A., Cavaglià, M. and Deriu, M.A., 2024. Data-driven classification and explainable-AI in the field of lung imaging. *Frontiers in Big Data,* **7**, 1393758.

Shah, S.T.H., Shah, S.A.H., Panagiotopoulos, K., Pigueiras-del-Real, J., Qayyum, K., Shah, S.B.H., Qureshi, S.A., Di Terlizzi, A., Di Benedetto, G. & Deriu, M.A. (2025) Artificial intelligence coupled with the Internet of Things targeting neurodevelopmental challenges in preterm neonates. *Journal of Multiscale Neuroscience* **4**, 32–56.

Shah, S.T.H., Shah, S.A.H., Qureshi, S.A., Di Terlizzi, A. & Deriu, M.A. (2023) Automated facial characterization and image retrieval by convolutional neural networks. *Frontiers in Artificial Intelligence* **6**, 1230383.

Sheikh, S.A., Sahidullah, M. & Kodrasi, I. (2025) Deep learning for pathological speech: A survey. *arXiv preprint.*

Sullivan, J., Mei, M., Perfors, A., Wojcik, E. and Frank, M.C., 2021. SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind* **5**, 20–29.

Veres, G. (2025) gveres/donateacry-corpus. [online] GitHub, Inc.

Violos, J., Diamanti, K.-C., Kompatsiaris, I. & Papadopoulos, S. (2025) Frugal machine learning for energy-efficient, and resource-aware artificial intelligence. *arXiv preprint.*

Wagner, L., Banchik, M., Tsang, T., Okada, N.J., Altshuler, R., McDonald, N., Bookheimer, S.Y., Jeste, S.S., Green, S. & Dapretto, M. (2025) Atypical early neural responses to native and non-native language in infants at high likelihood for developing autism. *Molecular Autism* **16**, 6.

Wang, T.V. & Song, P.C. (2022) Neurological voice disorders: A review. *International Journal of Head and Neck Surgery* **13**, 32–40.

Wen, Y., Innuganti, A., Ramos, A.B., Guo, H. & Yan, Q. (2025) SoK: How robust is audio watermarking in generative AI models? *arXiv preprint.*

Xu, L., Yildiz, A.M., Tuncer, I., Ozyurt, F., Dogan, S. & Tuncer, T. (2025) Detection of community emotions through sound: An investigation using the FF-Orbital chaos-based feature extraction model. *Ain Shams Engineering Journal* **16**, 103248.

Zayed, Y., Hasasneh, A. & Tadj, C. (2023) Infant cry signal diagnostic system using deep learning and fused features. *Diagnostics* **13**, 2107.

Zhang, E.Q. (2025) The influence of prenatal auditory input on newborn vocalizations. *SSRN Blog.*

Zhao, X., Sun, H., Lin, B., Zhao, H., Niu, Y., Zhong, X., Wang, Y., Zhao, Y., Meng, F., Ding, J., Zhang, X., Dong, L. & Liang, S. (2022) Markov transition fields and deep learning-based event-classification and vibration-frequency measurement for φ-OTDR. *IEEE Sensors Journal* **22**, 3348–3357.