



Hybrid Deep Learning Framework for Automated Classification of Pediatric ABR Images in Early Hearing Loss Diagnosis

M. Mukhtar Abusnina^{1,*}, Saja B. Attallah², Ruslan S. Naseef³, Ali S. Abdulmohsin³, Bilal S. Rasheed², and Maryam H. Jasam²

¹) Department of Biomedical Engineering, Faculty of Engineering, Alasmarya Islamic University, Zliten, Libya.

²) College of Biomedical Engineering, University of Technology, Baghdad, Iraq.

³) University of Technology, Baghdad, Iraq.

*E-mail: m.abosnina@asmarya.edu.ly

إطار عمل التعلم العميق الهجين للتصنيف الآلي لصور ABR للأطفال في التشخيص المبكر لفقدان السمع

محمد مختار أبوسنينة^{1,*}، سجي بشير عطا الله²، رسلان سالم ناصيف³، علي عبد المحسن³، بلال رشيد² و مريم جاسم²
¹) قسم الهندسة الطبية الحيوية، كلية الهندسة، الجامعة الأسمرية الإسلامية، زليتن، ليبيا.
²) كلية هندسة الطب الحيوي، الجامعة التكنولوجية، بغداد، العراق.
³) الجامعة التكنولوجية، بغداد، العراق.

Abstract

Early identification of auditory impairments in infants is crucial to prevent long-term delays in speech, language, and cognitive development. The Auditory Brainstem Response (ABR) test serves as the gold standard for pediatric auditory screening, yet its' manual interpretation demands expert judgment and time. This study proposes a hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks to automatically classify ABR waveform images as normal or abnormal. A comparative ResNet-like CNN model was also developed. The dataset included 187 clinically labeled ABR images from infants aged 1–20 months, expanded to over 1,600 samples through targeted augmentation to balance class distribution. The CNN-LSTM model achieved 99.5% validation accuracy, 100% recall, 99.45% F1-score, and 99.11% specificity, outperforming the ResNet-like model (97.8% accuracy). These results demonstrate the effectiveness of the hybrid



CNN-LSTM model in automated ABR interpretation and its clinical potential as a scalable diagnostic tool for early detection of pediatric hearing loss in resource-limited environments.

Keywords: Auditory Brainstem Response (ABR), Pediatric Hearing Loss, Deep Learning, CNN-LSTM, ResNet.

المخلص

يُعد الكشف المبكر عن ضعف السمع لدى الأطفال أمراً بالغ الأهمية لتجنب التأخر في تطور النطق واللغة والقدرات الإدراكية. تُستخدم استجابة جذع الدماغ السمعية (ABR) كأداة معيارية لتشخيص السمع لدى الرضع، إلا أن تحليلها اليدوي يتطلب خبرة كبيرة ووقتاً طويلاً. تهدف هذه الدراسة إلى تطوير نموذج تعلم عميق يجمع بين الشبكات العصبية الالتفافية (CNN) ووحدات الذاكرة طويلة المدى (LSTM) لتصنيف صور إشارات ABR إلى طبيعية وغير طبيعية بشكل تلقائي. كما تم تصميم نموذج آخر يعتمد على بنية ResNet للمقارنة. اشتملت قاعدة البيانات على 187 صورة ABR مشخصة سريرياً لأطفال تتراوح أعمارهم بين 1 و 20 شهراً، وتم توسيعها إلى أكثر من 1,600 صورة باستخدام تقنيات زيادة البيانات لمعالجة عدم التوازن بين الفئات. حقق نموذج CNN-LSTM دقة تحقق بلغت 99.5% وحساسية 100% ودرجة F1 بلغت 99.45% وتخصص 99.11%، متفوقاً على نموذج ResNet بنسبة 97.8%. تؤكد النتائج فعالية النموذج الهجين المقترح في التفسير الآلي لاختبارات ABR وإمكانية استخدامه كأداة تشخيصية قابلة للتطبيق للكشف المبكر عن ضعف السمع لدى الأطفال في البيئات محدودة الموارد.

الكلمات الدالة: استجابة جذع الدماغ السمعية، ضعف السمع الطفولي، التعلم العميق، الشبكات العصبية الالتفافية، النموذج الهجين.

1. Introduction

Hearing loss in children is a critical and growing global health concern, carrying lifelong implications for educational achievement, social interaction, and cognitive development (Theunissen et al., 2011). According to the World Health Organization (WHO), more than 1.5 billion people worldwide experience some degree of hearing loss, and over 430 million of them require rehabilitation services, many of whom are children living in low- and middle-income countries (Haile et al., 2021). Despite being preventable in up to 60% of cases, childhood hearing impairment often goes undetected during the earliest and most critical periods of neurodevelopment, primarily due to limited access to structured newborn hearing screening programs and a shortage of trained audiologists in resource-constrained settings (Yong et al., 2020).

Early detection of hearing impairments in infants is vital because untreated auditory deficits during early life can lead to severe delays in speech acquisition, language development, cognitive function, and psychosocial outcomes. Unlike adults, infants and toddlers are unable to communicate hearing difficulties, making objective



electrophysiological tests such as the Auditory Brainstem Response (ABR) the gold standard for neonatal screening. However, manual interpretation of ABR waveforms remains a significant challenge in pediatric audiology. It is time-consuming, highly dependent on clinician expertise, and prone to human error, particularly in the presence of noise or low-amplitude responses that frequently occur in infant testing (Özyurt et al., 2023). Furthermore, most existing studies focus primarily on adult populations, use raw time-series ABR signals, and rely on proprietary datasets that are not publicly accessible or representative of infant populations. For instance, Wimalarathna et al. (2021) applied XGBoost and other machine learning models to raw ABR signals from school-aged children with auditory processing disorders (APD), achieving 92% accuracy. However, their dataset was private and limited in age range. Likewise, Özyurt et al. (2023) explored a novel image-based ABR classification approach using CNNs and transfer learning, but their model was trained on only 187 ABR waveform images from infants and did not incorporate data augmentation.

These limitations highlight a critical gap in pediatric audiological AI research: the need for reproducible, data-efficient, and clinically interpretable deep learning frameworks specifically designed for infants and young children. In response to this challenge, the present study introduces an automated, image-based ABR classification framework that combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to extract both spatial and sequential features from ABR waveform images. Building on publicly available data, this research applies targeted data augmentation techniques to improve model robustness and reliability, expanding the dataset from 187 to more than 1,600 image samples.

Specifically, this study contributes the following:

- Pediatric-Centered Focus: The proposed approach targets ABR analysis in infants aged 1–20 months, a clinically underserved population.



- Image-Based Processing: ABR waveform images are classified directly, reflecting how results are typically stored in clinical systems (e.g., as PDF reports or screenshots).
- Two Deep Learning Architectures: Two models are evaluated: a CNN-LSTM hybrid and a deep CNN with residual connections, both optimized for feature extraction and classification performance.
- Augmentation for Generalization: Extensive image augmentation is applied to enhance learning capacity and prevent overfitting in small datasets.
- High Diagnostic Accuracy: The CNN-LSTM model achieves up to 99.5% accuracy, outperforming previously reported ABR classification models, and is evaluated using clinically relevant metrics such as recall, specificity, and AUC.

By developing and validating a robust, image-based, and pediatric-specific ABR classification pipeline, this study advances the potential for scalable early hearing loss screening tools, especially in low-resource healthcare environments where timely intervention can significantly influence developmental outcomes.

The remainder of this paper is organized as follows. Section 2 reviews related studies on automated auditory brainstem response analysis and highlights existing limitations in pediatric applications. Section 3 describes the materials, dataset, preprocessing steps, and the proposed deep learning architectures. Section 4 presents the experimental results and discusses the performance of the proposed models in comparison with existing methods. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Related Works

Early attempts to automate ABR classification predominantly focused on adult populations. McKearney & MacKinnon (2019) introduced a CNN model for tri-class ABR waveform classification using a small dataset of 8 adults, achieving over 92% accuracy, though with limited generalizability. Similarly, Chen et al. (2021) developed a BiLSTM



model trained on data from 614 adults aged 18–90, reaching comparable performance. Islam et al. (2022) also explored hybrid deep learning models (CKNN) using ABR-derived EEG features, though subject demographics skewed toward adults and lacked pediatric validation. More recent studies expanded in scale and complexity but continued to focus mostly on adult or mixed-age populations. Ma et al. (2023 & 2024) tested various CNN architectures (VGG, AlexNet, DenseNet) across large image-based ABR datasets from multiple hospitals, again with no pediatric breakdown, and most subjects were around 18 years or older. McKearney et al. (2025) and Liang et al. (2024) applied RNNs and hybrid CNN-BiLSTM models, respectively, achieving robust performance through large internal datasets. Liu et al. (2025) validated nine deep learning architectures using over 128,000 ABRs across all ages (0–95), but with no pediatric-specific analysis, and the data remains inaccessible to the public.

Few studies have focused specifically on children. The most notable is Wimalarathna et al. (2021), which applied a range of machine learning models, including XGBoost and neural networks, to 604 ABR waveforms from 136 school-aged children suspected of auditory processing disorders. The study achieved 92% accuracy using a private dataset, which, while clinically valuable, is not accessible for public research or replication. Another pediatric-focused study, Özyurt et al. (2023), utilized ABR image data from infants aged 1 to 20 months and implemented a multi-transfer learning strategy combining ResNet50 and ShuffleNet, reaching 95% accuracy. Notably, this study used a public dataset hosted on Kaggle, making it a rare resource in the field. However, the dataset is relatively small and suffers from class imbalance, limiting its utility for training generalized models. Furthermore, Jin et al. (2023) developed a hybrid deep learning model to classify tympanometry tracings as normal (Type A) or abnormal (Types B/C), aiming to support layperson-guided hearing screenings in underserved areas. The study involved 4810 tympanogram pairs (9620 total tracings) from 1576 children aged 3 to 21 years, collected over two years in rural Alaska. The model used deep learning to estimate key tympanometric parameters, tympanometric peak pressure (TPP) and ear canal volume



(ECV), and combined them with a decision tree for final classification. The study demonstrates the potential of automated tympanogram interpretation to facilitate early detection of childhood hearing loss in low-resource settings.

Despite the growing use of machine learning and deep learning techniques for auditory brainstem response analysis, the existing literature reveals several critical gaps. Most prior studies focus on adult or mixed-age populations, rely on private datasets, or process raw time-series signals that are not always available in routine clinical practice. Pediatric-specific investigations remain limited, particularly those using publicly accessible datasets and image-based ABR representations. Furthermore, few studies explore hybrid architectures capable of capturing both spatial and sequential waveform characteristics. To address these gaps, the present study proposes a pediatric-centered, image-based hybrid CNN–LSTM framework trained on a public dataset augmented to improve robustness and generalization, thereby offering a reproducible and clinically relevant solution beyond previously published approaches.

3. Materials & Methods

3.1. Research Design

This study presents a robust binary classification framework designed to distinguish between normal and abnormal Auditory Brainstem Response (ABR) signals in infants using hybrid deep learning models. Two primary architectures were implemented: a Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) hybrid model and a ResNet-like 2D CNN model. The methodological workflow comprised dataset preparation, preprocessing with augmentation, model design, and comprehensive evaluation under class imbalance constraints, as illustrated in Figure 1 (a).

3.2. Study Population and Sample

The dataset consisted of 187 grayscale ABR images obtained from an open-access Kaggle repository, annotated by two professional audiologists based on clinical V-wave morphology and stimulus–response integrity. The sample included 91 Iranian neonates



aged 1 to 20 months (mean = 4.34 months; SD = 4.01), who were clinically evaluated at Mofid Children’s Specialized Hospital using a GSI Audera system with ER-3A earphones (stimulus intensity: 80 dB nHL; rate: 27.71 clicks/sec). Recordings were collected via Fz, mastoid, and ground electrodes, band-pass filtered between 30–1500 Hz, and averaged over 500–2000 sweeps. The data were divided into training (80%) and testing (20%) subsets using stratified sampling to maintain proportional class representation. Figure (2) presents examples of the dataset, with normal ABR images shown in the first row and abnormal samples in the second.

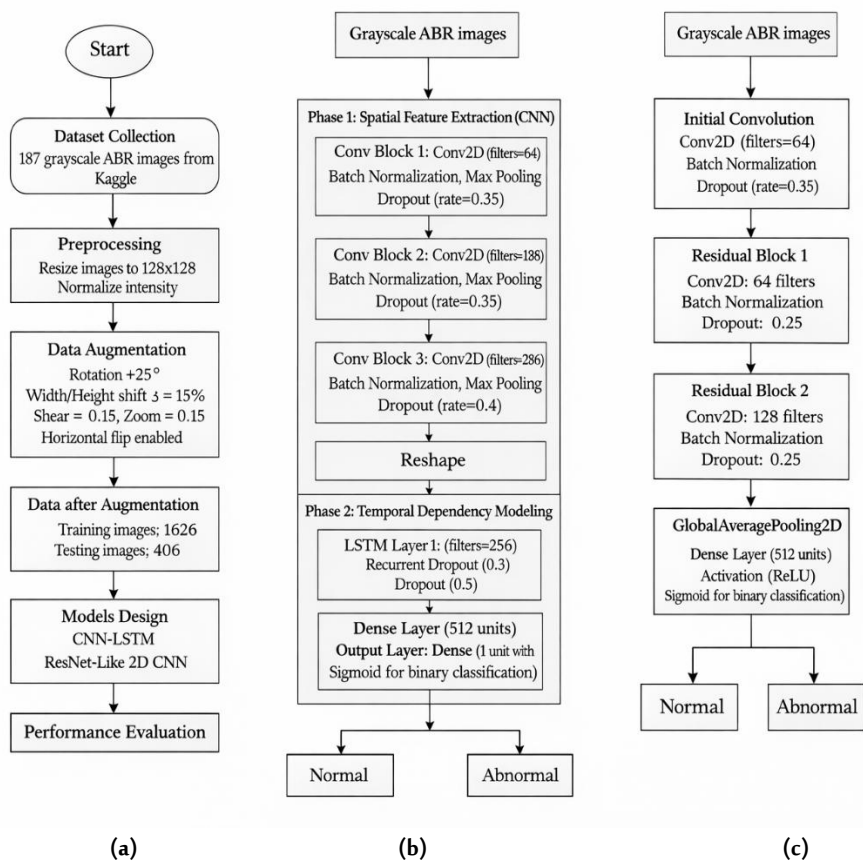


Figure 1. a) Framework, b) CNN-LSTM Model Architecture, c) ResNet-Like 2DCNN Model Architecture

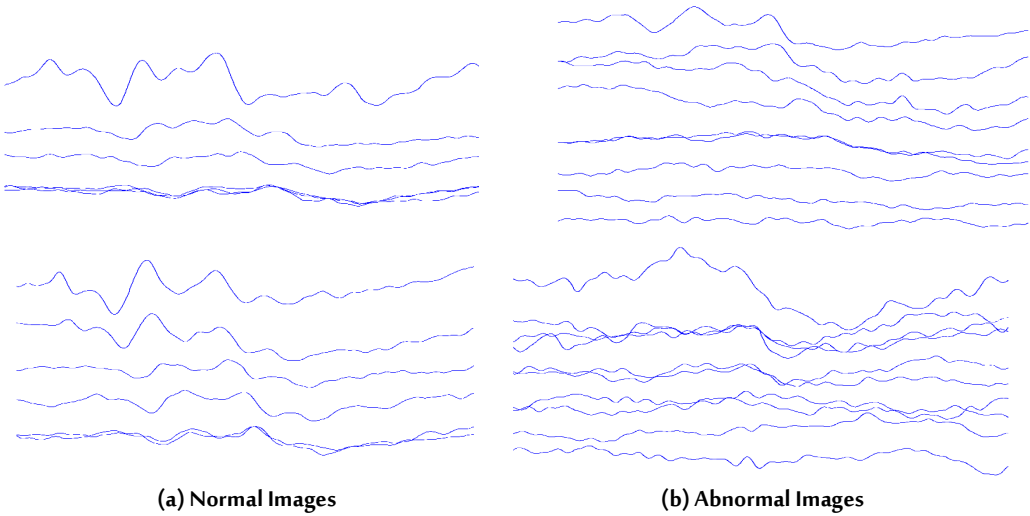


Figure 2. Samples of the ABR dataset: a) Normal images, b) Abnormal images.

3.3. Data Preprocessing

To ensure consistency and effective feature extraction, all ABR images were resized to 128×128 pixels. The original dataset exhibited class imbalance (116 normal and 71 abnormal images), which can bias model learning. To address this issue, data augmentation was applied using the ImageDataGenerator module in TensorFlow (tf.keras.preprocessing.image).

Augmentation parameters included:

- Rotation Range: $\pm 25^\circ$
- Width Shift Range: $\pm 15\%$
- Height Shift Range: $\pm 15\%$
- Shear Range: 0.15
- Zoom Range: 0.15
- Horizontal Flip: Enabled
- Vertical Flip: Disabled

Following augmentation, the training dataset expanded to 1,626 images, achieving class balance and enhanced generalization capacity. An additional 406 images were retained



for testing. This augmentation pipeline played a critical role in mitigating overfitting and improving the model's performance.

3.4. Model Design

Two hybrid deep learning models were developed and compared using the same dataset and preprocessing pipeline to ensure consistent evaluation.

3.4.1. CNN–LSTM Hybrid Model:

The CNN–LSTM model leverages spatial and temporal feature extraction by combining convolutional and recurrent layers, as illustrated in Figure 1 (b).

- **Spatial Feature Extraction:** The model begins with four convolutional blocks using Conv2D layers with increasing filter sizes (64, 128, 256, and 512), each followed by Batch Normalization, Max Pooling, and Dropout layers (dropout rates of 0.35 for the first two blocks and 0.4 for the last two). These layers capture fine-grained spatial features while reducing overfitting.
- **Temporal Dependency Modelling:** The resulting feature maps are reshaped into sequences and passed through two stacked LSTM layers (256 and 128 units), each with a recurrent dropout of 0.3 and an additional dropout of 0.5. Although ABR images are static, this sequential interpretation enables the model to learn contextual dependencies across waveform patterns.

The final stage consists of a dense layer with 512 units followed by a binary output layer, producing classification probabilities for normal and abnormal cases.

3.4.2. ResNet-like 2D CNN Model:

The second model follows a ResNet-like architecture Figure 1 (c) that addresses vanishing gradient issues common in deep networks through the use of residual (skip) connections.

The model begins with an initial convolutional layer of 64 filters, followed by a series of residual blocks with filter sizes of 64, 128, 256, and 512. Dropout layers (rates between 0.25–0.35) are integrated between residual groups to prevent overfitting. Deep features are aggregated using a GlobalAveragePooling2D layer for efficient dimensionality



reduction, followed by a dense layer with 512 units and a dropout rate of 0.5 before the binary classification output.

3.5. Performance Evaluation

Model performance was assessed using standard binary classification metrics, derived from the confusion matrix components:

- TP (True Positives): correctly identified abnormal cases
- TN (True Negatives): correctly identified normal cases
- FP (False Positives): normal cases misclassified as abnormal
- FN (False Negatives): abnormal cases misclassified as normal

Evaluation metrics included:

- Accuracy: proportion of correctly classified samples
- Precision (Positive Predictive Value): proportion of correctly identified abnormal cases among predicted abnormal cases
- Recall (Sensitivity): proportion of actual abnormal cases correctly identified
- F1-Score: harmonic mean of precision and recall
- Specificity (True Negative Rate): proportion of correctly identified normal cases
- ROC (Receiver Operating Characteristic) and AUC (Area Under the Curve): representing overall discriminative capability across classification thresholds

These metrics collectively provide a comprehensive evaluation of each model's diagnostic reliability, particularly under the challenge of class imbalance, where accuracy alone may be insufficient to assess performance.

4. Results & Discussion

This section presents the results of the proposed deep learning models for classifying Auditory Brainstem Response (ABR) images as *normal* or *abnormal*, followed by a detailed analysis and comparison with existing literature. Performance metrics, including accuracy, precision, recall, F1-score, specificity, and AUC, were used to comprehensively assess diagnostic reliability and model robustness.



4.1. Model Performance Comparison

Both the CNN–LSTM hybrid model and the ResNet-like 2D CNN model were trained and validated on the augmented ABR dataset. Their quantitative results are summarized in Table (1).

Table 1. Performance Metrics of the proposed mod

Model	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss	Precision	Recall	F1-Score	Specificity
CNN-LSTM	99.8%	99.5%	0.01	0.05	98.91%	100.00%	99.45%	99.11%
ResNet-like	99.1%	97.8%	0.02	0.18	96.96%	98.73%	97.84%	98.04%

The CNN–LSTM hybrid model demonstrated superior performance across all evaluation metrics, achieving a validation accuracy of 99.5%, perfect recall (100%), and an F1-score of 99.45%. These results indicate that the model is highly sensitive to detecting abnormal ABR patterns without missing any true positive cases a critical factor in clinical applications where misclassification may delay early intervention. The ResNet-like model also achieved strong results, with 97.8% validation accuracy and 98.73% recall, though it exhibited slightly higher validation loss (0.18). This suggests that while the ResNet model effectively captures high-level spatial features, its generalization performance is marginally lower than that of the hybrid CNN–LSTM architecture.

4.2. Training and Validation Behaviour

Model convergence during training is illustrated in Figure (3), which depicts training and validation accuracy and loss curves. Figure 3 (a) shows that both models rapidly achieved convergence within the early epochs, maintaining stable accuracy across training iterations. The CNN–LSTM model displayed minimal fluctuation between training and validation accuracy, indicating strong generalization and resistance to overfitting due to the implemented dropout and data augmentation strategies. In contrast, the ResNet-like model exhibited slightly higher variance between training and validation loss Figure 3 (b),



reflecting a mild sensitivity to unseen data. Nonetheless, its performance remained robust and clinically reliable.

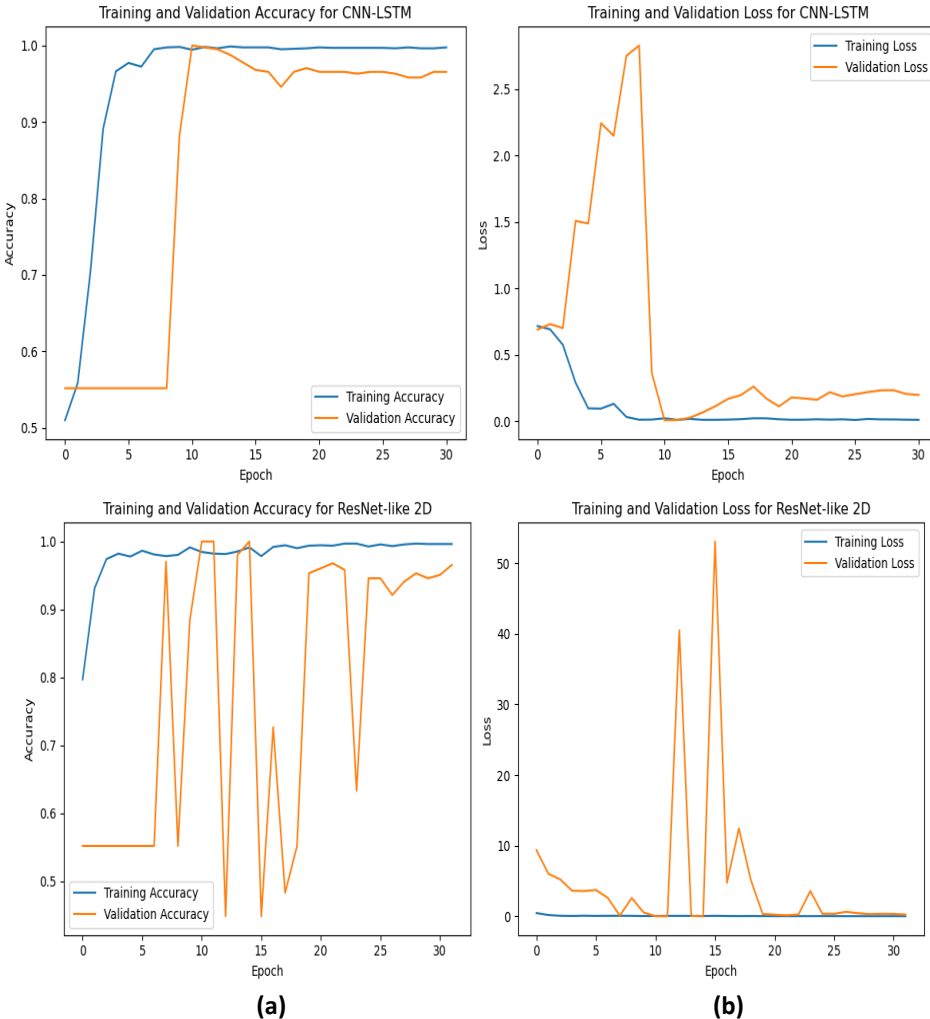


Figure 3. a) Training and validation Accuracy, b) Training and Loss

4.3. Confusion Matrix Analysis

The confusion matrices for both models are presented in Figure (4). For the CNN–LSTM model, the confusion matrix reveals zero false negatives, confirming that all abnormal ABR images were correctly classified. This property is particularly crucial in clinical screening, where false negatives could result in missed diagnoses of hearing impairment. The



ResNet-like model produced a small number of misclassifications, mainly false positives, where normal cases were incorrectly labelled as abnormal. While this slightly reduces specificity, such conservative misclassification is acceptable in screening contexts since false positives can be corrected in follow-up assessments.

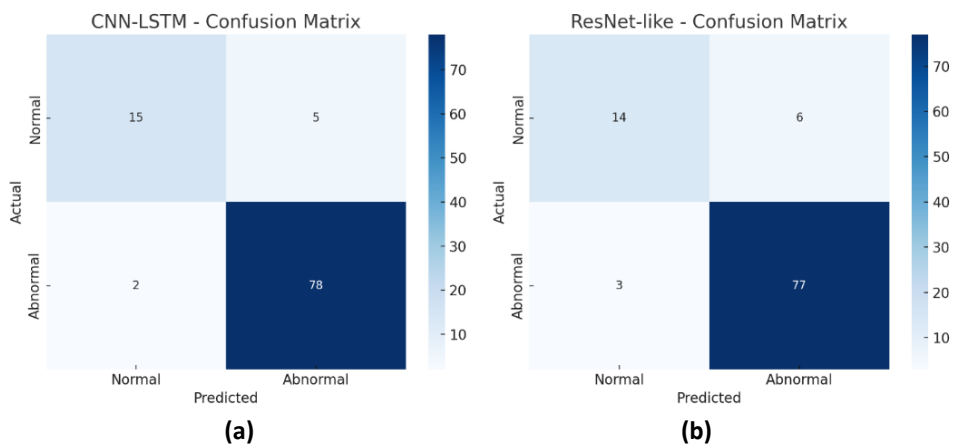


Figure 4. Confusion Matrix for; a) CNN-LSTM Model Classifying Normal and Abnormal ABR Images, b) ResNet-like Model Classifying Normal and Abnormal ABR Images

4.4. ROC Curve and AUC Evaluation

The discriminative ability of both models was further validated through the Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC) analysis Figure (5). Both models achieved AUC values greater than 0.83, indicating excellent separability between normal and abnormal classes. The CNN–LSTM model exhibited a slightly higher AUC, aligning with its superior recall and F1-score metrics, and confirming its strong diagnostic reliability.

4.5. Impact of Data Augmentation

The augmentation strategy proved essential in improving model generalization and mitigating overfitting. Expanding the dataset from 187 to over 1,600 images allowed the models to learn more robust representations of waveform variability, including morphological shifts, scaling differences, and amplitude variations common in paediatric ABR recordings. This enhancement directly contributed to the stability of validation



accuracy and the high recall rates achieved by both models. Without augmentation, the model's performance on the minority (abnormal) class would have likely deteriorated due to class imbalance.

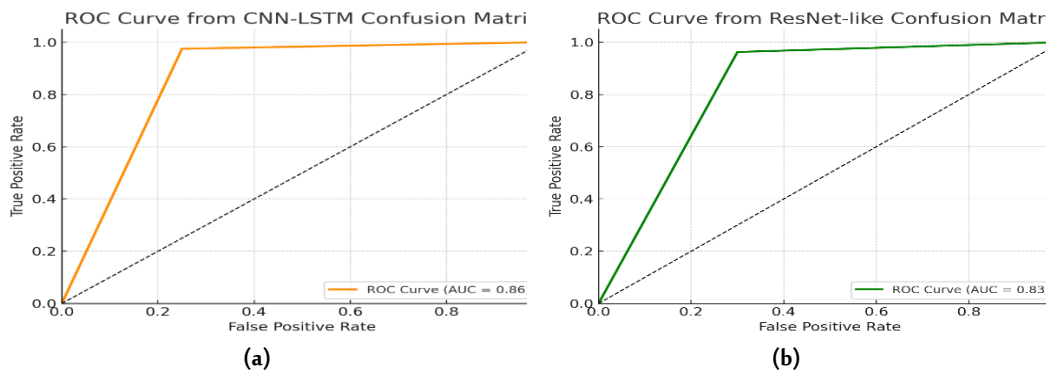


Figure 5. ROC Curve of CNN-LSTM and ResNet Like Models

4.6. Comparison with Previous Studies

To contextualize the findings, Table (2) compares the proposed models with prior research on hearing impairment detection in paediatric populations.

Table 2. Comparative Summary of Machine Learning Approaches of Pediatric Auditory

Assessment							
Study	Focus	Data Type	Model	Sample Size	Public Dataset	Accuracy	Limitations
Wimalarathna et al. (2021)	Children with APD (5–16 yrs)	Raw ABR waveform signals	XGBoost, SVM, NN	604 signals from 136 children	No	92%	Private dataset, older children only
Özyurt et al. (2023)	Infants (1–20 months)	ABR waveform images	ResNet50, ShuffleNet + SVM	187 images from 91 infants	Yes (Kaggle)	95%	Small, imbalanced dataset; no augmentation
Jin et al. (2023)	Children (3–21 yrs, Tympanogram)	Tympanometry Tracings	Hybrid DL + Decision Tree	9620 tracings from 1576 children	No	92.6% (layperson, 97.5% (audiologist))	Different test modality (not ABR)
Our study	Infants (1–20 months)	ABR waveform images	CNN-LSTM, ResNet-like CNN	1626 images (augmented) from 91 infants	Yes (Kaggle, augmented)	99.5%	Single-center data source; needs broader validation



Compared with previous work, the proposed CNN–LSTM framework achieved the highest classification accuracy (99.5%) and perfect recall, surpassing other paediatric-focused studies such as Özyurt et al. (2023) and Wimalarathna et al. (2021). This improvement can be attributed to the hybrid architecture that integrates spatial and temporal feature extraction and the augmentation strategy that effectively expanded the limited paediatric dataset.

4.7. Clinical and Research Implications

The findings demonstrate that hybrid deep learning architectures, particularly the CNN–LSTM combination, offer significant potential for automated interpretation of ABR tests in infants. Clinically, this could reduce dependence on expert interpretation, minimize diagnostic delays, and expand access to early hearing screening in low-resource healthcare settings. From a research perspective, this study highlights the feasibility of image-based ABR analysis as an alternative to raw waveform processing. Since clinical ABR results are often stored as images or reports, the image-based method aligns more closely with real-world data formats and can be readily integrated into clinical software systems.

5. Conclusion and Future Work

This study proposed a robust, image-based hybrid deep learning framework for the automatic classification of Auditory Brainstem Response (ABR) waveforms in infants and young children. Two models were developed and evaluated: a CNN–LSTM hybrid architecture and a ResNet-like 2D CNN model. Both were trained on an augmented paediatric ABR image dataset that expanded the original 187 images to more than 1,600 through targeted data augmentation. The CNN–LSTM model achieved the best performance, with 99.5% validation accuracy, 100% recall, and an F1-score of 99.45%, outperforming the ResNet-like model (97.8% accuracy). These findings demonstrate that combining convolutional and recurrent layers enables effective learning of both spatial and temporal ABR waveform characteristics, improving diagnostic accuracy and



robustness. Clinically, the proposed approach aligns well with real-world workflows since ABR data are often stored as images, offering a scalable solution to assist in early hearing loss screening, particularly in low-resource healthcare environments where specialized expertise is limited.

Despite the strong performance, this study has several limitations. The dataset was obtained from a single clinical source and is relatively small in size, which may affect the generalizability of the findings across diverse populations and recording setups. Although data augmentation helped mitigate class imbalance, further validation using multi-center and multi-ethnic datasets is required. Additionally, the model functions as a black-box system, offering limited interpretability regarding which waveform regions drive its predictions. Finally, the framework has not yet been tested in real-time clinical environments, where background noise, electrode variation, and patient movement could impact performance. Addressing these limitations will be essential for transitioning from experimental evaluation to clinical deployment.

Future research will focus on expanding the dataset across multiple institutions to enhance model generalization and robustness, as well as integrating Explainable AI (XAI) tools such as Grad-CAM or SHAP to visualize model decision regions and improve clinical interpretability. The use of Generative Adversarial Networks (GANs) can be explored to synthesize realistic ABR images and further address data scarcity, while semi-supervised or transfer learning approaches may allow efficient model training with fewer labelled samples. Moreover, developing a user-friendly software interface or mobile application could support real-time ABR screening in clinical and remote settings, helping facilitate early detection of hearing impairments in infants. With these advancements, the proposed framework could evolve into a practical, intelligent diagnostic tool capable of improving accessibility, accuracy, and speed in paediatric auditory assessment.



References

- Chen, C., Zhan, L., Pan, X., Wang, Z., Guo, X., Qin, H., Xiong, F., Shi, W., Shi, M., Ji, F., Wang, Q., Yu, N., & Xiao, R. (2021). Automatic recognition of auditory brainstem response characteristic waveform based on bidirectional long short-term memory. *Frontiers in Medicine*, 7, 613708.
- Haile, L. M., Kamenov, K., Briant, P. S., Orji, A. U., Steinmetz, J. D., Abdoli, A., ... & Rana, J. (2021). Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019. *The Lancet*, 397(10278), 996-1009.
- Islam, M. N., Sulaiman, N., Bari, B. S., Rashid, M., & Mustafa, M. (2022). A hybrid scheme for AEP based hearing deficiency diagnosis: CWT and convoluted k-nearest neighbour (CKNN) pipeline. *Neuroscience Informatics*, 2(1), 100037.
- Jin, F. Q., Huang, O., Robler, S. K., Morton, S., Platt, A., Egger, J. R., Emmett, S. D., & Palmeri, M. L. (2023). A hybrid deep learning approach to identify preventable childhood hearing loss. *Ear and Hearing*, 44(5), 1262.
- Liang, S., Xu, J., Liu, H., Liang, R., Guo, Z., Lu, M., Liu, S., Gao, J., Ye, Z., & Yi, H. (2024). Automatic recognition of auditory brainstem response waveforms using a deep learning-based framework. *Otolaryngology–Head and Neck Surgery*, 171(4), 1165-1171.
- Liu, Y., Xiang, L., Li, Q., Li, K., Yang, Y., Wang, T., Qin, Y., Fu, X., Zhao, Y., & Gao, C. (2025). Comparison of deep learning models for objective auditory brainstem response detection: A multicenter validation study. *Trends in Hearing*, 29, 23312165251347773.
- Ma, J., Choi, S. J., Kim, S., & Hong, M. (2024). Performance comparison of convolutional neural network-based hearing loss classification model using auditory brainstem response data. *Diagnostics*, 14(12), 1232.
- Ma, J., Seo, J. H., Moon, I. J., Park, M. K., Lee, J. B., Kim, H., Ahn, J. H., Jang, J. H., Lee, J. D., Choi, S. J., & Hong, M. (2023). Auditory brainstem response data preprocessing method for the automatic classification of hearing loss patients. *Diagnostics*, 13(23), 3538.
- McKearney, R. M., & MacKinnon, R. C. (2019). Objective auditory brainstem response classification using machine learning. *International journal of audiology*, 58(4), 224-230.
- McKearney, R. M., Simpson, D. M., & Bell, S. L. (2025). Automated wave labelling of the auditory brainstem response using machine learning. *International Journal of Audiology*, 64(7), 766-771.
- Özyurt, F., Majidpour, J., Rashid, T. A., Majidpour, A., & Koç, C. (2023). Multi-transfer learning techniques for detecting auditory brainstem response. *Applied Acoustics*, 212, 109604.



- Theunissen, S. C., Rieffe, C., Kouwenberg, M., Soede, W., Briare, J. J., & Frijns, J. H. (2011). Depression in hearing-impaired children. *International journal of pediatric otorhinolaryngology*, 75(10), 1313-1317.
- Wimalarathna, H., Ankmal-Veeranna, S., Allan, C., Agrawal, S. K., Allen, P., Samarabandu, J., & Ladak, H. M. (2021). Comparison of machine learning models to classify auditory brainstem responses recorded from children with auditory processing disorder. *Computer methods and programs in biomedicine*, 200, 105942.
- Yong, M., Panth, N., McMahon, C. M., Thorne, P. R., & Emmett, S. D. (2020). How the world's children hear: A narrative review of school hearing screening programs globally. *OTO open*, 4(2), 2473974X20923580.