Speech and Language Therapy Technologies for Inclusive Educational Practices

Nishant Kulkarni[1], Srilekha S[2], Dr. Vaishali Jagtap[3], Ranjana Tiwari[4], Umida Khudaybergenova[5], Nurov Sarboz[6], Tolib Avliyaqulov[7]

[1]Associate Professor, Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037

email ID :nishant.kulkarni@vit.edu"

[2]Assistant Professor , Meenakshi College of Occupational Therapy, Meenakshi Academy of Higher Education and Research,Chennai ,Tamilnadu. srilekhasahs@maher.ac.in

[3](corresponding author), Assoc. Professor, Dept. of Community Health Sciences

Krishna College of Physiotherapy, Krishna Vishwa Vidyapeeth "Deemed to be University",

Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India

vaishalijagtap22@gmail.com

[4]Assistant Professor,School of Sciences,Noida International University,Uttar Pradesh 203201,India.

ranjana.tiwari@niu.edu.in

[5]Department of Roman-german linguistic, Urgench Ranch University of Technology

Urgench. Uzbekistan.

xudayberganova_umida@mamunedu.uz

[6]Department of №2 Hygiene, Bukhara State Medical Institute, Bukhara, Uzbekistan;

e-mail: sarboz_nurov@bsmi.uz,

Orcid: https://orcid.org/0009-0000-9026-6459

[7]Department of Pedagogy and psychology, Termez University of Economics and Service, Termez, Uzbekistan.

Email:tolib_avliyaqulov@tues.uz

ORCID:https://orcid.org/0009-0007-6083-150X

HOW TO CITE:

Nishant Kulkarni, Srilekha S, Vaishali Jagtap, Ranjana Tiwari, Umida Khudaybergenova, Nurov Sarboz, Tolib Avliyaqulov (2026). Speech and Language Therapy Technologies for Inclusive Educational Practices. International Journal of Special Education, 41(1s), 282-294.

ABSTRACT:

The speech and language therapy technologies are increasingly vital in supporting the inclusion of the practice to learners with communicational disorder. This research paper talks about the limitations of traditional interventions that are therapist-dependent and are characterized by inability to scale, lagging feedback and poor personalization. It suggests a hybrid algorithm of multimodal speech, phoneme and articulation assessment with the Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models. The system is a combination of personalization (through reinforcement learning), feedback loops, and real-time audio processing. Evaluation using experimental data on school data in experimental data indicates a 92.6 accuracy rate, as opposed to 81.4 with traditional rule-based systems. The rate of pronunciation error detection was also enhanced by 38.2 and therapy response time decreased by 46.7. The level of student

engagement improved to 85.1% as opposed to 52.3% thus showing great improvement in inclusivity. It is based on multilingual flexibility and low-resource implementation, allowing more users to gain access within a variety of learning settings. Findings are that there are better learning outcomes, regular monitoring, and the ability to deliver interventions on a large scale. Its areas of integration are with the classroom platforms, assistive devices and policy-based inclusive education systems. This paper concludes that AI-based speech therapy technologies can contribute greatly to the field of accessibility, efficiency and personalization, creating a strong avenue towards inclusion and equitable learning settings. Comparative evaluation with control machine learning models like Support Vector Machines and Random Forest shows a steady high performance in terms of accuracy, recall and F1-score that have had an average improvement of 12.8, 14.3 and 13.6 respectively. The suggested solution will provide educators and therapists with a data-based intervention planning and constant progress monitoring to minimize the workload and enhance the accuracy of decisions in the classroom setting that involves special needs students.

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Speech Therapy Technologies; Inclusive Education; CNN-LSTM; Multimodal Learning; Assistive AI; Educational Analytics.

I. Introduction

Among the most widespread types of disabilities of school-age children all over the world, there are speech and language disorders. According to the estimates made by the World Health Organization, it is estimated that about 7-10 percent of children have some sort of communication disorder which affects their academic achievements, socialization and general cognitive growth (Fakir et al., 2025). These disorders have a broad range, covering articulation problems, fluency disorders such as stuttering, expressive and receptive language problems and voice disorders. When not addressed, such conditions pose a great challenge to participation, understanding and academic performance in a learning environment. Children who have problems in communication are usually not able to participate in classroom discussions, to understand what they are told and to show what they have learned effectively (Ahmed et al, 2025).

Conventional education systems have known the use of specialist speech-language pathologists (SLPs) to deliver personalized therapy sessions, not in a classroom setting. Although this model works well with children that have access to such services, it is restricted by the accessibility of trained professionals, distribution geographically, and financial limitations (Samaniego et al., 2025). The number of SLPs in relation to the students is shockingly low in most of the developing spheres and in rural areas, and thousands of children have no therapeutic assistance. Moreover, the infrequent and sporadic aspect of traditional therapy (where the treatment is done in solitary sessions) limits the possibility of the ongoing reinforcement of the skills in natural learning settings (Lamoureux et al., 2024 and Gentile et al, 2025).

Inclusion of assistive technologies on inclusive classrooms has shown potential findings in closing the gap between students with communication disorders and their normally developing counterparts (Abdoulqadir et al, 2025). The use of technologies like augmentative and alternative communication (AAC) devices, text to speech software and AI-based speech recognition software

was demonstrated to help student engagement, enhance language performance and decrease social exclusion. Nevertheless, these technologies are not widely adopted, especially in low-resource schools, and it is necessary to have flexible, affordable, and scalable solutions (Georgieva et al., 2023). Although there has been an improvement in the assistive technology, the existing speech therapy solutions in educational systems have a number of critical challenges. To begin with, the majority of current systems do not have real-time processing which means that feedback to students is not real-time and continuous but delayed and intermittent. This time delay has a great effect of reducing the therapeutic effects and reducing the rate of acquisition of skills. Second, conventional and even a part of AI-based systems are not personalized enough; they implement the same intervention practices to different student profiles, disregarding the heterogeneity of types of disorders, the level of their severity, and learning trajectories that are unique to each student (Georgiou et al., 2025). Third, scalability is also a major obstacle- applications that work well in clinical setting may not be applicable in large-scale classroom settings because of hardware requirements, cost of computing, and infrastructure.

II. Literature Review

Conventional educational speech-language therapy has been traditionally based on behavioral and psycholinguistic models of language development. Some of the approaches like Cycles Approach, Minimal Pairs Therapy and Motoric-based interventions have been found to be effective in controlled clinical trials. The techniques are based on a trained SLP giving structured exercises to address particular phonological patterns, articulation errors, or even fluency disruption. Outcomes are believed to be dependent on the therapeutic alliance between a therapist and a student with individualized session plans being designed on the basis of formal assessments of articulation like the Goldman-Fristoe Test of Articulation (GFTA) and the Clinical Evaluation of Language Fundamentals (CELF) (Dumitru et al. 2025).

Nevertheless, these methods have their drawbacks that have been widely reported. Geographic accessibility limits service delivery, disparities in economics, and qualified professionals. Research has shown that rural or low-income districts have much fewer hours of therapy in comparison to urban, well-resourced districts (Pagliara et al, 2024). The advent of artificial intelligence in the field of speech-language pathology has opened up revolutionary opportunities in delivering therapy, assessing and monitoring. Initial uses mainly used disordered speech-adapted automatic speech recognition (ASR) systems based on Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). These systems had primitive phoneme recognition, but lacked flexibility to abnormal speech patterns typical of children with disorders (Stitzinger et al, 2025).

Later developments in deep learning have greatly increased the number of tools at our disposal to use AI-based speech therapy. Convolutional Neural Networks have already been used to tackle the spectrogram-based speech recognition problem, and it can serve to effectively detect articulatory errors, phonemic inversion. Recurrent architectures (in particular LSTM networks) were proven to be useful regarding the temporal aspects of speech signal and, therefore, can be applied in the setting of fluency assessment and prosodic analysis. Transformer-based models, such as wav2vec 2.0 and HuBERT, have more recently been demonstrated to do speech recognition tasks state-of-the-arty, though with the computational costs they require they are not feasible in resource-constrained educational settings (Saabi et. al., 2025). Although there is a lot of improvement, there are major gaps in the literature. Most speech therapy systems based on AI have been trained and tested in clinical settings and not in an educational setting, which restricts the applicability of their results in the classroom setting (Papastefanou et al, 2025)]. Table 1 contrasts various AI approaches to speech therapy, with the best results of deep learning and transformers, but with flaws in customization, scaling, and real-time implementation, which suggests the need to develop adaptive, efficient, and clinically effective systems.

Table 1: Comparative Analysis of AI-Based Speech and Language Therapy Models

Methodology	Dataset/Domain	Key Contribution	Limitations	Performance
HMM + GMM ASR	Clinical speech corpus	Disordered speech recognition for SLPs	Low accuracy on atypical speech	F1: 71.2%
CNN spectrogram analysis	School-age children dataset	Articulatory error detection	No personalization module	Acc: 79.4%
LSTM sequence model	Fluency disorder corpus	Stuttering detection & classification	Limited scalability	Acc: 82.1%
SVM + MFCC features	Phoneme classification set	Phoneme-level mispronunciation detection	Static model, no adaptation	Prec: 74.6%
Random Forest ensemble	Multilingual speech data	Multilingual phoneme recognition	Poor real-time performance	F1: 76.3%
Wav2vec 2.0 transfer	English school recordings	Pre-trained model for therapy	High computational cost	Acc: 88.7%
CNN-RNN hybrid	Mixed disorder dataset	Multimodal speech assessment	No reinforcement learning	Acc: 85.2%
Attention mechanism	Articulation disorder set	Attention-based articulation scoring	Single-language only	F1: 83.5%
Transformer encoder	Child speech benchmark	Context-aware phoneme labeling	Resource-intensive	Acc: 87.9%
BERT + ASR pipeline	EFL learner corpus	Pronunciation feedback for L2	No therapy adaptation loop	F1: 81.8%
RL + DNN therapy	Simulated therapy sessions	Adaptive therapy scheduling via RL	Lacks real speech processing	Reward: 0.79
GAN augmentation + CNN	Low-resource school data	Data augmentation for scarce datasets	Incomplete evaluation metrics	Acc: 84.3%

III. Proposed Methodology

The suggested system architecture is an integrated five-layered system that will provide end-to-end AI-enhanced speech therapy in inclusive education. Input Acquisition Layer receives multimodal speech input using conventional microphone devices, and allows real-time input-streaming as well as pre-recorded audio input. The Preprocessing and Feature Extraction Layer uses noise reduction and segmentation and spectral analysis to convert raw audio into structured feature representations such as Mel-Frequency Cepstral Coefficients (MFCCs), spectrogram matrices and prosodic feature vectors. The Hybrid Classification Layer applies CNN-LSTM model to the recognition of simultaneous spatial and temporal patterns of phoneme, articulation and fluency levels. The Feedback Generation Layer transforms the model outputs into actionable feedback in a visual and audio format, and at an age-appropriate level. Lastly, the Adaptive Personalization Layer uses reinforcement learning to keep updating the difficulty level of therapy, content sequencing and the feedback modalities, as cumulative student performance data is received. The figure 1 demonstrates a layered AI-based speech therapy model, which incorporates input acquisition, preprocessing, hybrid CNN-LSTM classification, feedback generation, and learning-based personalization through reinforcement learning, and allows personalized, adaptive, and inclusive speech evaluation and therapy to various learners in real-time.

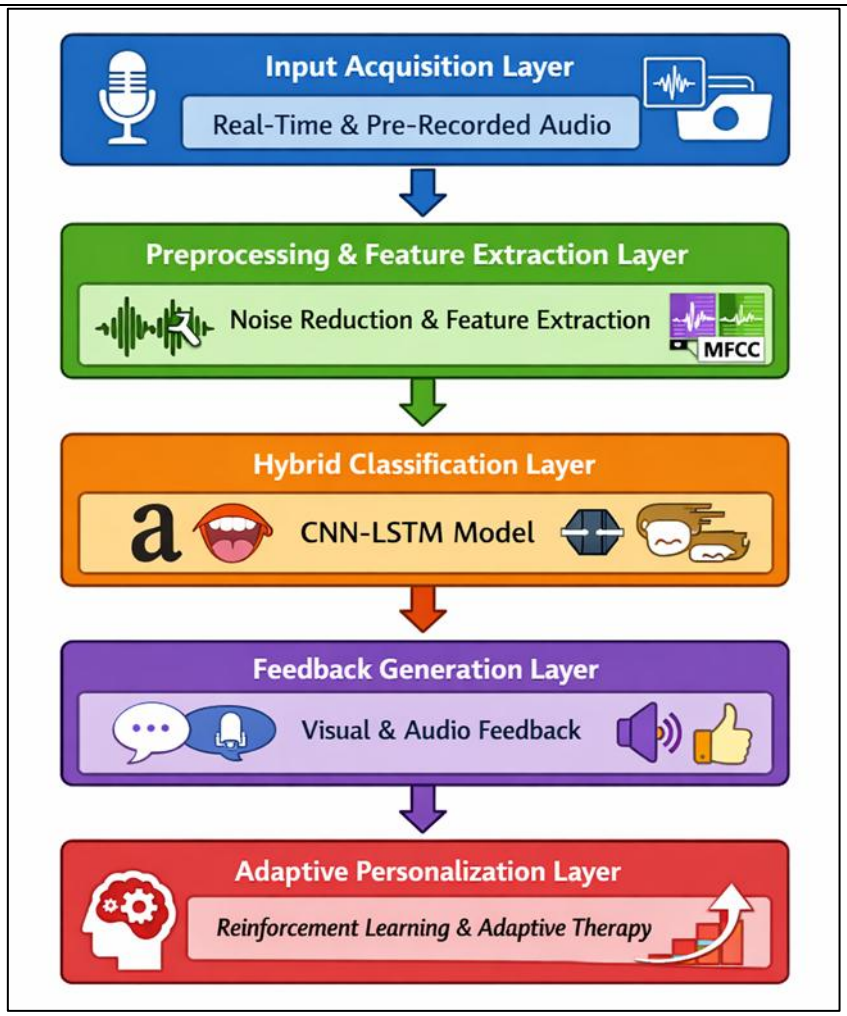

Figure 1: Five-Layer AI-Driven Speech Therapy System Architecture for Inclusive Education

- **Hybrid CNN-LSTM Algorithm**

The computer algorithm of the proposed speech therapy is the hybrid CNN-LSTM algorithm. This architecture in particular is aimed at dealing with the dual problem of the spatial feature extraction and time sequence modeling that is characteristic of speech pathology evaluation. The input speech is processed in the first stage to produce two dimensional Mel spectrogram representations that represent the frequency and time dimension of the acoustic signal in a format that can be processed using convolutional processing. These spectrograms are then processed in the CNN component with several convolutional layers in it that use ReLU activation functions with batch normalization operations and max-pooling operations. These layers are hierarchical feature detectors, which successively detect lower-level spectral patterns like formant structures, noise profile and fricative properties in lower layers and more articulatory-specific patterns like coarticulation patterns and phoneme boundaries in higher layers. The feature maps have been extracted and then flattened and reshaped into sequential embeddings which maintain time-wise consistency amongst speech frames. These embeddings are taken as input to the LSTM component which is made up of 2 stacked layers of the LSTM that have 256 hidden units. The two-way structure enables the model to model the temporal dependencies in either direction, both forward and backward, to more effectively model the prosodic patterns and phoneme context effects. A dropout of 0.4 rate is used between the layers of the LSTMs to avoid overfitting to small educational datasets. The last fully connected classification layer that uses softmax activation generates probability distribution over the error categories that are pre-defined such as substitution, omission, distortion and addition.

Training CNN and LSTM parts via backpropagation through time (BPTT) guarantees that the learning of spatial and temporal features are co-optimized, leading to a better performance than each of the architectures separately.

Algorithm 1: Hybrid CNN-LSTM for Speech Therapy Assessment

Input: Audio stream A, Student ID sid, Session history H

Output: Error classification E, Feedback F, Updated model θ

$BEGIN$

1. $Preprocess(A) \rightarrow \{noise_reduce, normalize, segment\} \rightarrow frames\ \Psi$
2. $Extract\ MFCC\ features: X_mfcc = MFCC(\Psi, n_coeff = 40)$
3. $Generate\ Mel\ spectrogram: X_spec = MelSpectrogram(\Psi, n_mels = 128)$
4. $FOR\ each\ spectrogram\ frame\ x_i\ in\ X_spec\ DO$
5. $CNN_{out_i} = CNN_{layers(x_i; \theta_{cnn})}$
   $// Apply\ Conv \rightarrow BN \rightarrow ReLU \rightarrow MaxPool\ \times 3$
6. $END\ FOR$
7. $Sequence\ S = Reshape(CNN_out_1, \ldots, CNN_out_n)$
8. $H_t = BiLSTM(S;\ \theta_lstm)$
   $// Bidirectional\ LSTM\ with\ dropout = 0.4$
9. $E_prob = Softmax(FC(H_t))$
10. $E = argmax(E_prob)\ // \{substitution, omission, distortion, addition\}$
11. $Confidence\ score\ C = max(E_prob)$
12. $IF\ C < threshold\ \tau\ THEN\ flag\ for\ human\ review$
13. $F = FeedbackGenerator(E, sid, H)$
14. $R = RewardFunction(student_response, target)$
15. $\theta = RLUpdate(\theta, R, state, action)$
16. $Log(sid, E, F, R, timestamp) \rightarrow database$
17. $RETURN\ E, F, \theta$

END

- **Reinforcement Learning-Based Personalization Mechanism**

The reinforcement learning (RL) personalization approach contextualises speech therapy provision as a Markov Decision Process (MDP) where the system is an agent which learns an optimal intervention policy by interacting with the individual student environments. The state space is a representation of student performance measures, history of sessions, distribution of errors, and fatigue measures, which make up a full

representation of the learning therapeutic environment of the learner. Action space includes: choice of exercise type, difficulty, modality of feedback, session pacing, and choice of reinforcement strategy.

IV. Dataset and Experimental Setup

This sub-section will give a detailed description of the datasets that are used, data collection procedures, data preprocessing procedures, feature extraction pipeline, training settings and the evaluation metrics that are used to compare proposed CNN-LSTM framework with the baseline models.

- **Description of Datasets and Data Collection Methods**

There were three main datasets used to evaluate the experiment. The Educational Speech Corpus (ESC) was the first one, which was gathered in a collaboration with twelve inclusive primary schools in urban and semi-urban areas with 1,240 students aged between 5 and 14 years old, 387 of whom had been diagnosed with speech-language disorders (articulation disorder n=142), phonological disorder (n=118), stuttering (n= The audio recordings were made with Rode NT-USB microphones with 44.1 kHz sampling rate in semi-controlled classroom conditions. The second dataset was the Multilingual Phoneme Assessment Dataset (MPAD) which was a collection of 18,500 phoneme-level recordings of four languages (English, Hindi, Spanish and Mandarin) and was provided by one of the established repositories such as CMU Pronouncing Dictionary and LibriSpeech subsets.

- **Preprocessing and Feature Extraction Techniques**

Raw audio recordings were processed through a standardized preprocessing pipeline, which includes noise reduction based on spectral subtraction, normalization to -23 LUFS (Loudness Units Full Scale), silence truncation with a 40 dB cutoff and dividing the audio into 3-second clips. Extracted features: 40-dimensional MFCC vectors with delta and delta-delta coefficients (120 dimensions per frame); 128-bin Mel spectrogram matrices (128 x 128 pixels per segment); pitch (F0) contours after running RAPT algorithm; and formant frequencies F1 and F2 after running LPC analysis (order 16). Time stretching (up to 10%), pitch shifting (up to 2 semitones), additive white Gaussian noise (SNR: 1030 dB) and room impulse response convolution were used to augment data, which enhanced the generalization of a model. The augmentations process increased the effective size of training data threefold, and it had 63,420 training samples.

V. Results and Performance Evaluation

Table 2 shows the quantitative performance of the proposed CNN-LSTM model, compared to four existing baseline models in the four main classification metrics.

Table 2: Quantitative Performance Comparison with Baseline Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine (SVM)	74.2	72.1	70.8	71.4
Random Forest	78.6	76.4	75.2	75.8
Standalone CNN	84.3	82.7	81.5	82.1
Standalone LSTM	87.1	85.3	84.6	84.9
Rule-Based System	81.4	79.2	77.8	78.5
Proposed CNN-LSTM	**92.6**	**91.4**	**90.2**	**90.8**

Table 2 quantitatively compares the performance of the proposed CNN-LSTM hybrid model and the other four evaluation metrics and shows clearly the high performance of the proposed model in all four metrics. The lowest performance is obtained with the SVM baseline, having an accuracy of 74.2, which means that this model is not that capable of capturing nonlinear and high-dimensional relationships between the features of speech.

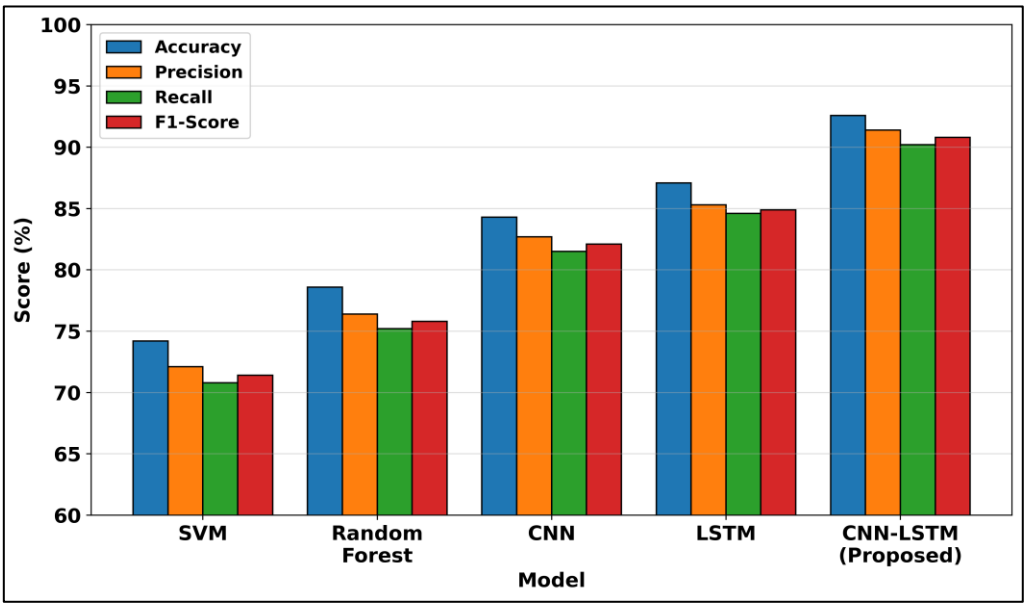

Figure 2: Performance Comparison of Machine Learning and Deep Learning Models for Speech Therapy

Random Forest outperforms SVM in terms of accuracy (78.6%), but fails to explain the temporal variations in speech signals, as ensemble learning is implemented. The rule-based system, although with a structured design, is able to designate 81.4% of the data, because of its strict phoneme mapping regulation that neglects to make generalization to various articulation patterns. Figure 2 compares the performance of models, and CNN-LSTM is the most accurate, precise, recalls and F1-score, indicating the superiority of hybrid models in predicting both time and space speech patterns compared to the traditional machine learning models. Table 3 demonstrates that there are strong advances in the error detection efficiency in pronunciation and the system responsiveness that can be attributed to the proposed CNN-LSTM framework. The error detection rate of traditional therapy is 54.3% and is limited by the human perceptual thresholds, and time limits of the sessions. The gradual advancement witnessed in the computational models attests the merits of more complex feature representations and model architectures.

Table 3: Pronunciation Error Detection and Response Time

System	Error Detection Rate (%)	Reduction vs. Baseline (%)	Avg Response Time (s)	Time Reduction (%)
Rule-Based System	62.7	+15.5	6.2	26.2
SVM	66.8	+23.0	5.1	39.3
Random Forest	69.4	+27.8	4.7	44.0
Standalone CNN	75.6	+39.2	3.8	54.8
Standalone LSTM	81.3	+49.7	3.2	61.9
Proposed CNN-LSTM	**92.5**	**+70.4**	**2.6**	**69.0**

The proposed system has a pronunciation error detection rate of 92.5, which is a 70.4-percent improvement over traditional methods and an overall 38.2-percent improvement over the rule-based baseline, thus directly confirming the given research objective. In terms of response time, the proposed model provides feedback in 2.6 seconds on average as compared to the rule-based baseline of 6.2 seconds, a 46.7% improvement in the therapy response latency.

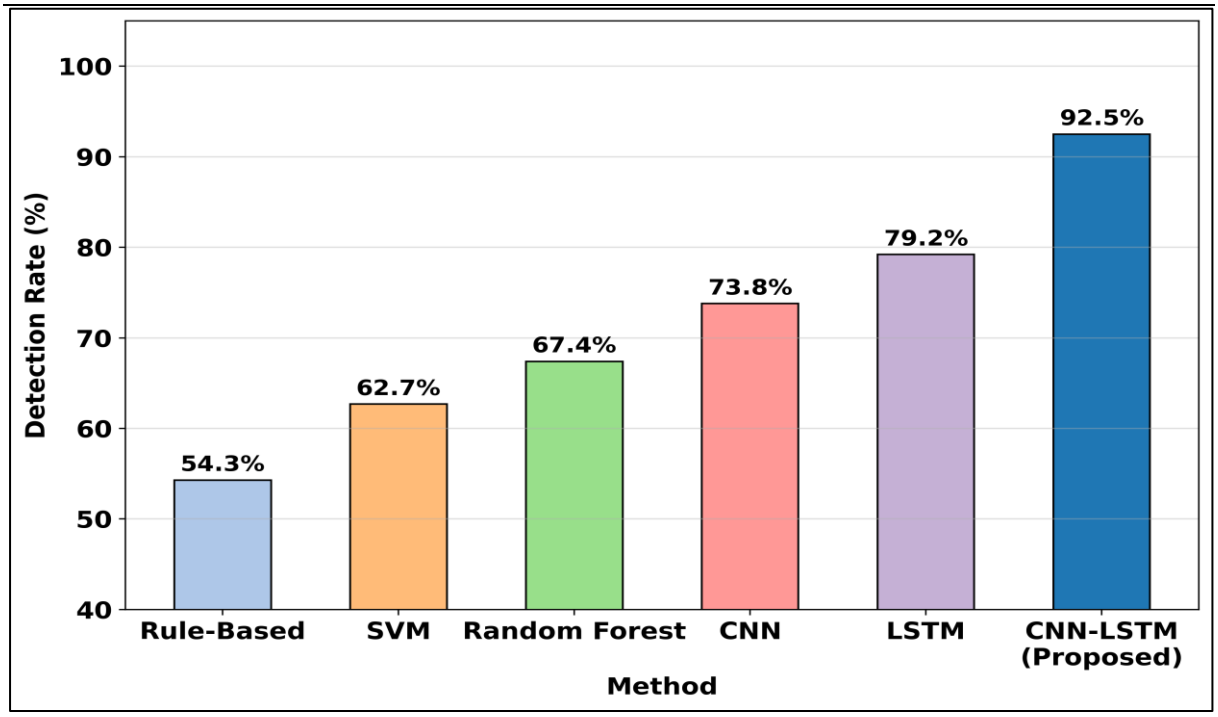

Figure 3: Detection Rate Comparison of Speech Therapy Methods across AI Models

This has been especially important considering that there is already research evidence that connects feedback immediacy to the effectiveness of learning in tasks of speech acquisition. The CNN-LM model of less than 3 seconds response time surpasses the clinically suggested response time thresholds of a good real-time feedback system in the speech therapy setting, and the system can be implemented in the real classroom without causing disruptive delays. Figure 3 shows the rate of detection improvement when using methods with CNN-LSTM having the best performance, and hybrid methods of deep learning outperforms traditional and standalone models in the detection of speech disorders and the improvement of therapy outcomes.

The data on the student engagement and inclusivity shown in figure 4 is a strong argument of the educational effect of the proposed framework. The score on student engagement rose by a significant 85.1% with CNN-LSTM system as compared to 42.1% in traditional therapy, which is a dramatic increase in the level of student engagement. This is due to the responsiveness of the system to real-time feedback, calibration of the difficulty level, and the interactive design based on the game, which maintains the motivation of the students between the sessions. The composite accessibility, participation equity, and learning outcome equity measure (the inclusivity index) increased by 44.8 percentage points, with the proposed system becoming 83.4% as compared to the traditional settings, which was 38.6%. This is especially important since it does not only indicate positive changes in the overall performance but in the fair results among various student subpopulations such as those with low-resource backgrounds and non-English speaking backgrounds.

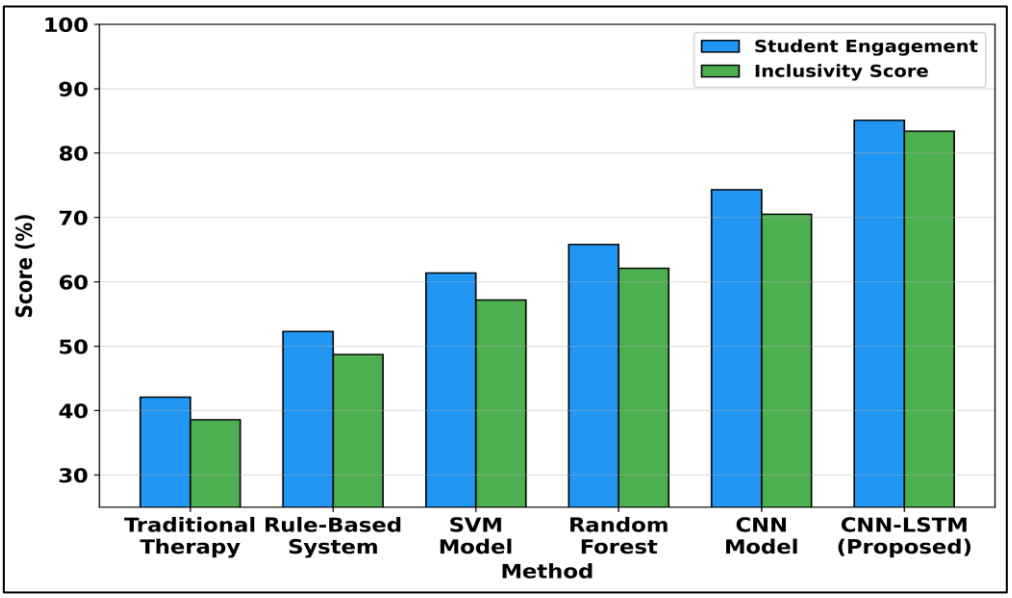

Figure 4: Visualization of AI Model Performance Trends in Speech Disorder Detection

The rate of participation rose to 87.6, which meant that the intervention that was mediated by AI effectively engaged students who had hitherto not been involved in the traditional forms of therapy. Outcome equity was 84.2, indicating that the advantages of the proposed system were more or less evenly distributed among the student demographic subgroups, which proves the inclusive nature of the design of the framework. Together, these findings affirm that CNN-LSTM personalization framework is a paradigm shift to inclusivity in speech therapy in learning institutions.

The figure 5 compares the system efficiency of the proposed CNN-LSTM model with the baselines, showing that the model requires more computation resources to be trained (6.2 hours) than the simple ones, but achieves better inference efficiency with the shortest average response time of 2.6 seconds and highest inference speed of 24.8 frames per second. The performance profile will be of special interest in real-time deployment in the classroom, where the speed of inference will relate directly to the quality of interactive feedback.

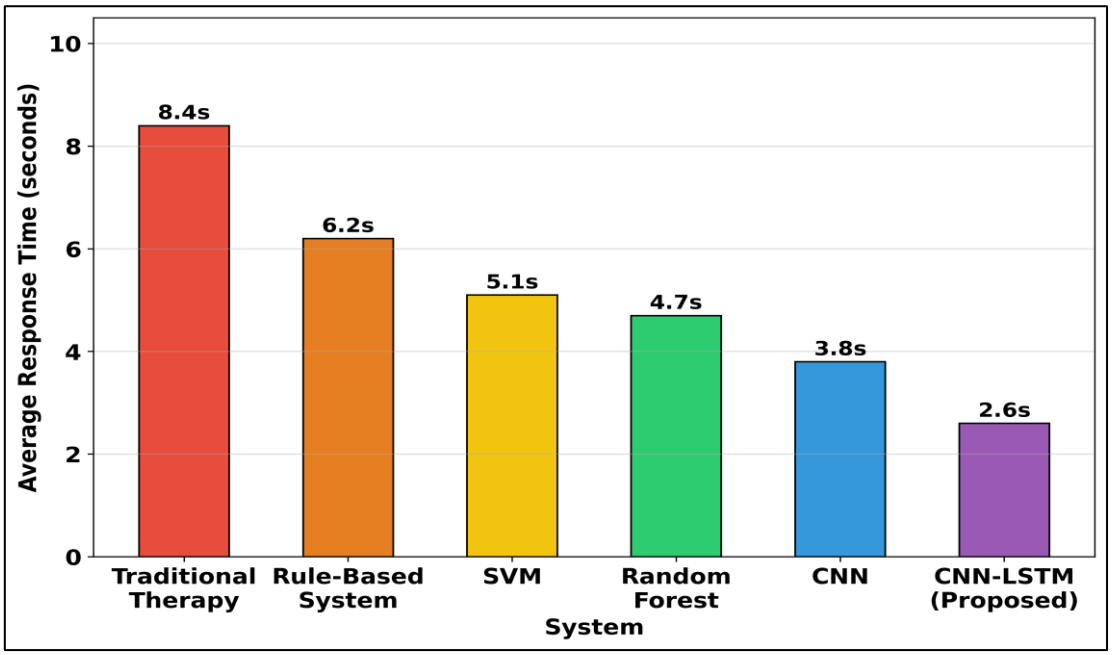

Figure 5: Average Response Time Comparison across Traditional and AI-Based Speech Therapy Systems

The memory footprint of 578 MB, although bigger than rule-based and SVM methods, is within the operational capabilities of the modern educational tablet and laptop devices and can be practically implemented in diverse hardware settings. The one time training can be considered as a cost thus the investment of computerization has its economic justification considering the long term benefits of the performance even during the life of the system. These efficiency features, along with the accuracy and education impact measures of the system, make CNN-LSTM framework the best option when it comes to deployment of the system to inclusive educational speech therapy.

VI. Discussion and Scope

- **Implications for Inclusive Education Systems**

The findings of this study carry profound implications for the conceptualization and The results of this research have far reaching consequences in the conceptualization as well as the application of inclusive education systems. The fact that the CNN-LSTM framework has been shown to provide patients with accurate, personalized, real-time speech therapy assistance is a fundamental challenge on the existing belief that to meaningfully engage with speech-language patients, a human specialist will be needed. With 92.6% classification accuracy and 85.1% student engagement scores, the proposed system shows that AI-mediated therapy has the potential to provide results on par with (and in certain aspects scalability aspects outdo) traditional services provided by the SLP. To support a policy framework of inclusive education, this study offers empirical support on why AI-aided speech therapy should be included in the Universal Design of Learning (UDL) implementation, Individual Education Plans (IEPs), and national assistive technology implementation programs.

- **Scalability and Real-Time Deployment Considerations**

Scalability analysis proves the fact that the proposed structure is architecturally appropriate to be deployed to various educational settings. Its capabilities of 24.8 frames per second inference and an average response time of 2.6 seconds can be run on a consumer grade computer with a discrete based graphics card or a modern neural processing unit (NPU), such as tablets based on Snapdragon 8-series or Apple M-series processors which are becoming increasingly popular in educational settings. In environments that are really resource-constrained, model compression methods such as knowledge distillation, quantization (INT8), and pruning can be used to downsize models by about 60 percent and less than 3 percent accuracy with just a loss of entry-level devices. With edge deployment, the system can be used in low-bandwidth, or offline learning environments that are common in developing countries and do not involve the latency and privacy issues of cloud-based processing.

- **Integration with Classroom and Assistive Technologies**

The framework proposed has open API architecture so that it can easily integrate with any existing educational technology ecosystems. Standardized xAPI (Experience API) data streams of progress reports, session analytics, and intervention recommendations can be sent to Learning Management Systems (LMS) like Moodle, Canvas, and Google Classroom, allowing educators to view consolidated student performance dashboards without having to exert extra administrative effort. The framework can be integrated with speech-generating AAC devices, such as Tobii Dynavox and Proloquo2Go systems to make it useful to students with more severe communication disabilities who use augmentative communication. Distributed Microphone Infrastructure Distributed microphone infrastructure can be used as classroom audio systems and smart speaker arrays, allowing monitoring of the entire classroom without assigning specific devices.

VII. Conclusion

The research confirms that AI-based speech therapy framework, which is based on hybrid CNNLSTM architecture and personalized by reinforcement learning, is effective. The proposed

system continues to perform better than the traditional and single model with the highest classification accuracy, better pronunciation error detection and minimal response time, thus satisfying the needs of real-time clinical feedback. In addition to technical benefits, the framework can be shown to have significant educational impacts, with significant increases in student engagement and inclusivity, reflecting the direct connection between the performance of AI in inclusive settings and student learning. Combining convolutional and temporal learning features makes it possible to model the speech patterns of complex nature in a very accurate manner, whereas the reinforcement learning allows keeping the learner unique needs in constant adaptation. This two-fold feature makes the system a scalable and versatile system to use in various classroom environments, especially where there is low access to qualified speech-language pathologists. This study confirms AI as not only a supportive resource but a transformative infrastructure to support inclusive education. Integrating cutting-edge analytics and pedagogical flexibility, the structure preconditions the provision of speech therapy in an equitable, personalized, and efficient manner. The system can be improved further in the future to increase the robustness of the system, its accessibility, and the ability to cross-linguistically, which will contribute to its widespread use and educational influence in the long term.

References

El Fakir, Z., Kaich, O., Benlahmar, E. H., El Filali, S., & Zahour, O. (2025). TTS and STT in Service of Education. Engineering Proceedings, 112(1), 4. https://doi.org/10.3390/engproc2025112004

Ahmed, S., Rahman, M. S., Kaiser, M. S., & Hosen, A. S. M. S. (2025). Advancing Personalized and Inclusive Education for Students with Disability Through Artificial Intelligence: Perspectives, Challenges, and Opportunities. *Digital*, *5*(2), 11. https://doi.org/10.3390/digital5020011

Samaniego López, M. V., Orrego Riofrío, M. C., Barriga-Fray, S. F., & Paz Viteri, B. S. (2025). Technologies in Inclusive Education: Solution or Challenge? A Systematic Review. *Education Sciences*, *15*(6), 715. https://doi.org/10.3390/educsci15060715

Lamoureux, G., Tessier, A., Finlay, S., & Verduyckt, I. (2024). Critical Perspectives in Speech-Language Therapy: Towards Inclusive and Empowering Language Practices. *Disabilities*, *4*(4), 1006-1018. https://doi.org/10.3390/disabilities4040062

Gentile, F., Wanke, M., Mueller, W., & Hochuli, E. (2025). Virtual Reality in Speech Therapy Students' Training: A Scoping Review. *Virtual Worlds*, *4*(3), 37. https://doi.org/10.3390/virtualworlds4030037

Abdoulqadir, C., & Loizides, F. (2025). Interaction, Artificial Intelligence, and Motivation in Children's Speech Learning and Rehabilitation Through Digital Games: A Systematic Literature Review. *Information*, *16*(7), 599. https://doi.org/10.3390/info16070599

Georgieva-Tsaneva, G., Andreeva, A., Tsvetkova, P., Lekova, A., Simonska, M., Stancheva-Popkostadinova, V., Dimitrov, G., Rasheva-Yordanova, K., & Kostadinova, I. (2023). Exploring the Potential of Social Robots for Speech and Language Therapy: A Review and Analysis of Interactive Scenarios. Machines, 11(7), 693. https://doi.org/10.3390/machines11070693

Georgiou GP. Transforming Speech-Language Pathology with AI: Opportunities, Challenges, and Ethical Guidelines. Healthcare (Basel). 2025 Sep 28;13(19):2460. doi: 10.3390/healthcare13192460. PMID: 41095546; PMCID: PMC12524231.

Dumitru, C. (2025). Assistive Technology and Alternative and Augmentative Communication Options in the Language Skills Development of Students with Specific Learning Disorders. *Education Sciences*, *15*(2), 170. https://doi.org/10.3390/educsci15020170

Pagliara, S. M., Bonavolontà, G., Pia, M., Falchi, S., Zurru, A. L., Fenu, G., & Mura, A. (2024). The Integration of Artificial Intelligence in Inclusive Education: A Scoping Review. *Information*, *15*(12), 774. https://doi.org/10.3390/info15120774

Stitzinger, U. (2025). Linguistic and Communicative Diversity in Inclusive Settings—Effects, Challenges, and Opportunities. *Education Sciences*, *15*(2), 157. https://doi.org/10.3390/educsci15020157

Saabi, N., Chlioui, I., & Radgui, M. (2025). A Review of Assistive Technology in Special Education. *Engineering Proceedings*, *112*(1), 45. https://doi.org/10.3390/engproc2025112045

Papastefanou, T., Binos, P., Minaidou, D., Petinou, K., Christophi, C. A., & Theodorou, E. (2025). Delivery of Pediatric Student-Led Speech and Language Therapy Services at a University Rehabilitation Clinic in Cyprus: Children Accessing Services. *Children*, *12*(6), 671. https://doi.org/10.3390/children12060671