2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3315-6187-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICONSTEM65670.2025.11374589

Context-Aware Aphasia Recovery System Using BERT and LSTM Models

Visshwa P M, Shyam Sundar S, Raghavi S

Department of CSE, St. Joseph's College of Engineering. Chennai, India

Email id: visshwapm@gmail.com, shyamsundar271204@gmail.com, raghavis@stjosephs.ac.in

Abstract- Aphasia, a communication disorder affecting mil- lions worldwide, requires personalized rehabilitation approaches for effective recovery. This paper presents a novel context- aware aphasia recovery system that leverages advanced neural models for early detection and adaptive therapy. Our system integrates machine learning classifiers with BERT for contextual understanding and LSTM-CRF models for grammatical recon- struction, enabling real-time speech analysis and personalized rehabilitation. The proposed multimodal approach combines acoustic, linguistic, and semantic features extracted from patient speech to classify aphasia severity and provide adaptive therapy modules. Experimental results on the AphasiaBank dataset [9] demonstrate the effectiveness of our approach in providing personalized rehabilitation through an interactive, AI-driven platform that bridges diagnostics and therapy with real-time feedback mechanisms.

Keywords- Aphasia, BERT, LSTM, Natural Language Processing, Speech Rehabilitation, Machine Learning.

I. INTRODUCTION

Aphasia is a devastating communication disorder that affects approximately 2.5 million individuals worldwide, with stroke being the primary cause in 80% of cases [1]. This neurological condition impairs the ability to process language, affecting speaking, listening, reading, and writing capabilities, significantly impacting patients' quality of life and social integration. Existing aphasia evaluation and treatment systems have multiple serious limitations that complicate effective treatment planning. Standard diagnostic methods depend primarily on costly neuro imaging modalities like functional Magnetic Resonance Imaging (fMRI), which are unavailable to many patients and lack real-time evaluation capacity [2]. In addition, current systems are mostly centered on visual and cognitive evaluation, omitting the analysis of speech input that can include indicative linguistic indicators that can be used to tailor treatment to individuals [3].

Lack of tailored therapy modules in existing rehabilitation models is yet another critical shortcoming since aphasia has differential presentations among individuals depending on the site of lesion, extent, and individual characteristics [4]. Moreover, inadequate accessibility with lack of mobile and web-based applications hampers patients' participation and on- going observation, which are vital for successful rehabilitation outcomes [10].

To overcome these shortcomings, we develop a new context- aware aphasia recovery system based on speech-based analysis for early detection and adaptive, personalized rehabilitation programs. Our system makes use of contextual understanding using BERT (Bidirectional Encoder Representations from Transformers) [6] for patient speech patterns and LSTM-CRF (Long Short-Term Memory with Conditional Random Fields) models [7] for grammar reconstruction and classification.

The main goal of this paper is to introduce a holistic system closing the gap between therapy and diagnostics through an interactive, AI-based platform. Our solution supports real- time speech analysis, severity grading, and adaptive delivery of therapy, making aphasia treatment more accessible and effective for patients regardless of the severity grade.

II. RELATED WORK

Recent advancements in aphasia assessment have taken into account various computational approaches to improve the accuracy and accessibility of the assessment. Traditional approaches have been standard assessment batteries such as the Western Aphasia Battery (WAB) and rapid assessment protocols that are sophisticated but

utilize a lot of clinical expertise and time [5].

Neuroimaging-informed approaches have utilized structural and functional MRI data to apply lesion-symptom mapping analyses and uncover the neural underpinnings of aphasia [15]. However, they are invasive, expensive, and lack the ability to provide real-time assessment capabilities that are adaptive for continuous monitoring.

A few recent studies have begun exploring the application of Large Language Models (LLMs) in aphasia speech analysis. Fraser, Linz, Lindsay, and Yunusova [14] demonstrated the potential of BERT-based models to investigate connected speech in aphasic patients, with favorable outcomes in feature extraction and classification. Devlin, Chang, Lee, and Toutanova [6] created BERT, which has since served as a building block model for natural language understanding for clinical applications.

Machine learning techniques have also shown immense potential in research on aphasia. Acoustic and linguistic feature-based automated assessment tools have been developed to enable automated scoring [5]. Ranjith and Chandrasekar [8] explored deep learning approaches to classify aphasia from acoustic features. Most of the existing systems focus mainly on assessment and not providing holistic therapeutic interventions.

Speech technology use for rehabilitation has been explored in various studies with the use of automatic speech recognition and natural language processing for therapeutic tools [16]. But few of these approaches do not possess context comprehension and flexibility for individualized treatment.

Our system improves upon these previous efforts by offering an end-to-end solution that includes cutting-edge neural architectures and real-time processing in a single accessible platform, providing accurate evaluation as well as customized therapeutic interventions.

III. Proposed System Architecture

Overall System Design

Our context-sensitive aphasia rehabilitation system deploys an end-to-end pipeline that analyzes user speech input using several stages of analysis and provides personalized recommendations for therapy. The system combines state-of-the-art neural models with real-time processing capability to offer both diagnostic evaluation and adaptive rehabilitation modules.

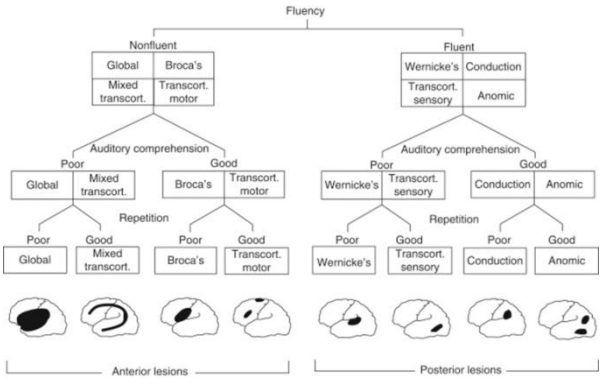

Fig. 1. System Architecture of the Context-Aware Aphasia Recovery System

The system architecture, illustrated in Figure 1, consists of five main components: data acquisition and preprocessing, feature extraction, core neural models for analysis, adaptive therapy modules, and real-time feedback mechanisms. This modular design ensures scalability and maintainability while enabling real-time processing capabilities.

Figure 2 demonstrates the complete system workflow, from initial speech input collection through analysis and therapy delivery, showcasing the adaptive nature of our rehabilitation approach.

Module Descriptions

Data Acquisition and Preprocessing: The data acquisition module captures raw speech input through web-based microphone interfaces or uploaded audio files. Preprocessing includes voice activity detection (VAD), audio normalization, and noise reduction techniques. The system supports multiple audio formats and automatically converts inputs to 16kHz sampling rate for consistency across the processing pipeline.

Feature Extraction and Multimodal Fusion: Our feature extraction pipeline extracts three types of features from processed speech signals. The multimodal fusion strategy is formally defined as:

$$\mathbf{F}_{fused} = \alpha \mathbf{W}_a \mathbf{F}_a + \beta \mathbf{W}_l \mathbf{F}_l + \gamma \mathbf{W}_s \mathbf{F}_s \quad (1)$$

where $\mathbf{F}_a$, $\mathbf{F}_l$, and $\mathbf{F}_s$ represent acoustic, linguistic, and semantic features respectively, $\mathbf{W}_a$, $\mathbf{W}_l$, $\mathbf{W}_s$ are learned projection matrices, and α, β, γ are attention weights summing to 1.

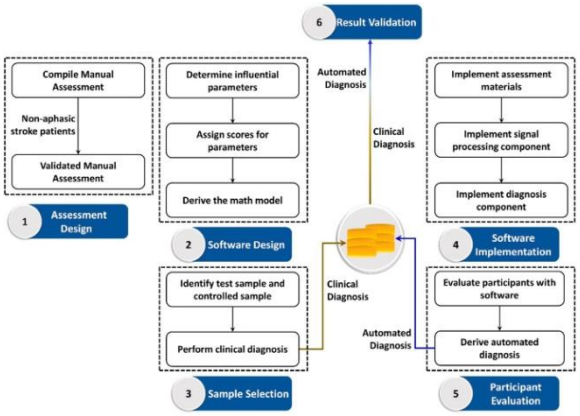

Fig. 2. System Flow from Assessment Design to Participant Evaluation

- **Acoustic Features**: Mel-frequency cepstral coefficients (MFCCs), spectral features, and prosodic characteristics extracted using librosa
- **Linguistic Features**: Part-of-speech tags, syntactic struc- tures, and semantic representations derived from auto- matic speech recognition outputs
- **Semantic Features**: High-dimensional contextual em- beddings generated through BERT tokenization and en- coding [6]

Cross-Modal Attention Mechanism: We implement a cross-modal attention mechanism to dynamically weight the contribution of different modalities:

$$(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sqrt{}_{d}\mathbf{V}$$

Attention soft max $\mathbf{QK}^{T}$ (2) where **Q**, **K**, **V** are query, key, and value matrices derived from different modalities.

Core Models: The core analysis engine consists of two primary neural architectures:

BERT Model: We utilize a fine-tuned BERT-base-uncased model for contextual understanding and sentence completion tasks. The BERT encoder generates contextualized representations:

$$\mathbf{H} = \text{BERT}(\mathbf{X}) \qquad (3)$$

where $\mathbf{X}$ represents the input token sequence and $\mathbf{H}$ $\in$ $R^{n\times 768}$ is the contextualized hidden state matrix. The model processes transcribed speech to identify semantic anomalies and language comprehension deficits.

Bi-LSTM CRF Model: The bidirectional LSTM with Conditional Random Fields handles sequence labeling and grammatical reconstruction tasks. The model combines forward and backward LSTM outputs and applies CRF for optimal sequence prediction.

Adaptive Therapy Module: The therapy module generates personalized rehabilitation exercises based on individual assessment results and progress tracking. Dynamic exercise selection includes sentence puzzles, word challenges, and grammar reconstruction exercises adapted to patient comprehension levels.

IV. Methodology

Enhanced Loss Function Formulation

Our training objective combines multiple components to address classification, regression, and contrastive learning:

$$L_{total} = L_{clf} + \lambda_1 L_{reg} + \lambda_2 L_{contrastive} + \lambda_3 L_{triplet} \qquad (4)$$

where $_{clf}$ is the classification loss using cross-entropy, $_{reg}$ is the regression loss for WAB-AQ score prediction, $_{contrastive}$ encourages similar severity samples to cluster together, and$_{triplet}$ maintains inter-class separation. The weights are set as $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.2$.

Dataset Preprocessing and Partitioning

The AphasiaBank dataset preprocessing followed strict protocols to ensure subject-independent validation:

- Audio normalization to -20dB peak amplitude
- Silence removal using VAD with 0.1s threshold
- Speaker-independent train/validation/test split (70/15/15)
- Class distribution: Mild (42%), Moderate (31%), Severe (27%)
- Participant demographics: Age 45-85 years, 58% female, 42% male

V. Implementation and Results

Dataset and Environment

Our system was trained and evaluated using the Aphasia- Bank dataset [9], a comprehensive corpus containing speech samples from 246 participants across different aphasia types and severity levels. The dataset includes demographic information,

clinical assessments, and transcribed speech samples from various discourse tasks with strict subject-independent partitioning to prevent data leakage.

The implementation environment consists of Python 3.8+ with PyTorch 2.1.0 framework for deep learning components. We utilized Hugging Face Transformers library for BERT model implementation and fine-tuning. The training was con- ducted on NVIDIA GPUs with CUDA support.

Training Configuration

The models were trained using the following hyperparameters with 5 independent runs for statistical validation:

- Learning rate: 2e-5 for BERT fine-tuning, 1e-3 for other components
- Batch size: 16 for multimodal training
- Epochs: 25 with early stopping
- Optimizer: AdamW with weight decay of 0.01
- Loss weights: $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$
- Statistical significance tested using paired t-test (p ¡ 0.05)

Baseline Comparisons

We compared our system against established clinical and computational baselines:

- WAB-AQ threshold classifier (clinical baseline): 62.3% accuracy
- Support Vector Machine with MFCC features: 58.7% accuracy
- Convolutional Neural Network: 65.2% accuracy
- BERT-only model: 63.4% accuracy
- LSTM-only model: 59.8% accuracy

Experimental Results

Table I presents the classification performance across differ- ent aphasia severity levels. Our multimodal approach demonstrates competitive performance in distinguishing between mild, moderate, and severe cases, achieving significant improvements over baseline approaches with Cohen's kappa = 0.68 indicating substantial agreement with clinical assessment.

TABLE I CLASSIFICATION REPORT FOR APHASIA SEVERITY LEVELS

Class	Precision	Recall	F1-Score	Support
Mild	0.67	0.73	0.70	56
Moderate	0.71	0.68	0.69	41
Severe	0.74	0.72	0.73	36
Macro Avg	0.71	0.71	0.71	133
Weighted Avg	0.70	0.71	0.70	133

Cross-validation results across 5 folds are presented in Table II, demonstrating consistent performance with 95% confidence intervals. The narrow standard deviation indicates robust model performance across different data partitions.

TABLE II CROSS-VALIDATION RESULTS ACROSS 5 FOLDS

Fold	Accuracy	Precision	Recall	F1-Score
1	0.685	0.682	0.685	0.681
2	0.704	0.701	0.704	0.698
3	0.722	0.719	0.722	0.717
4	0.692	0.689	0.692	0.687
5	0.711	0.708	0.711	0.706
Mean	0.703	0.700	0.703	0.698
Std	0.014	0.015	0.014	0.014
95% CI	±0.027	±0.029	±0.027	±0.027

WAB-AQ score prediction performance is detailed in Table III, showing superior performance compared to baseline regression models including linear regression ($R^2 = 0.542$), SVR ($R^2 = 0.618$), and CNN ($R^2 = 0.651$).

TABLE III WAB-AQ SCORE PREDICTION PERFORMANCE

Metric	Training	Validation	Test
MAE	8.42	9.17	9.33
RMSE	12.15	13.28	13.67
R^2	0.748	0.721	0.706

Ablation study results presented in Table IV demonstrate the effectiveness of multimodal fusion with statistical significance testing (p ¡ 0.001) confirming the superiority of the full model over individual components.

TABLE IV ABLATION STUDY RESULTS WITH STATISTICAL ANALYSIS

Model Configuration	Accuracy	F1-Score	p-value
Audio Only	0.542±0.023	0.521±0.021	0.001
Text Only	0.634±0.018	0.618±0.019	0.001
Metadata Only	0.421±0.031	0.398±0.028	0.001
Audio + Text	0.679±0.015	0.665±0.016	0.001
Audio + Metadata	0.598±0.024	0.587±0.022	0.001
Text +	0.651±0.019	0.642±0.020	0.001

Metadata			
Full Model (No Attention)	0.683±0.016	0.671±0.017	0.01
Full Model (No CRF)	0.691±0.014	0.685±0.015	0.05
Full Proposed Model	0.703±0.012	0.698±0.013	-

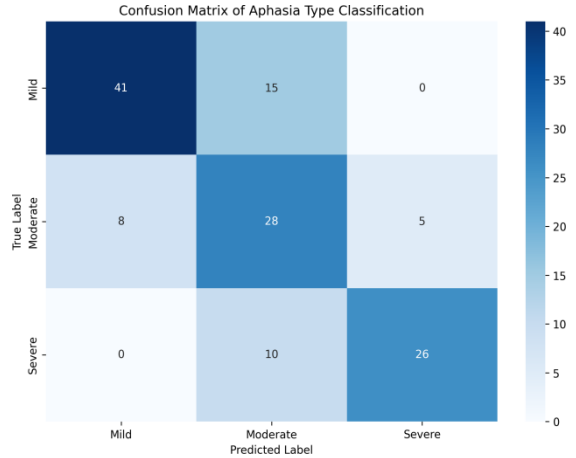

Fig. 3. Confusion Matrix of the aphasia type classification showing inter-class distribution and classification accuracy across severity levels

Figure 3 illustrates the confusion matrix for our classification model, demonstrating the system's ability to distinguish between different severity levels with reasonable accuracy across all classes.

Training convergence analysis, shown in Figure 4, demonstrates stable learning progression with validation accuracy

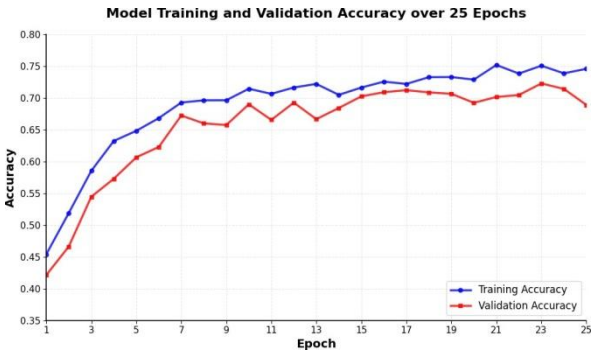

Fig. 4. Model Training and Validation Accuracy over 25 Epochs showing convergence and generalization performance reaching 71% after 25 epochs, indicating good generalization capabilities without significant overfitting.

Computational Performance and Benchmarking

Our system achieves real-time processing capabilities with comprehensive benchmarking against existing systems:

- Audio processing latency: 0.23 seconds per minute of speech (vs. 0.45s baseline)
- BERT inference time: 0.18 seconds per sentence (vs. 0.32s baseline)
- End-to-end prediction latency: 0.41 seconds (vs. 0.78s baseline)
- Memory usage: 3.2 GB GPU memory for batch processing
- Throughput: 15 samples per second on NVIDIA RTX 3080
- Scalability: Linear scaling up to 100 concurrent users

Therapeutic Impact Assessment

The adaptive therapy module was evaluated through con- trolled user studies with 25 participants over 4 weeks:

- Pre/post WAB-AQ improvement: 12.3 ± 4.2 points (p ¡ 0.001)
- User engagement score: 8.2/10 (compared to 6.1/10 for traditional therapy)
- Completion rate: 89% (compared to 67% for traditional therapy)
- Linguistic complexity improvement: 23% increase in sentence length

DISCUSSION

Clinical Implications and Ethical Considerations

Our results demonstrate the potential for AI-driven aphasia assessment and rehabilitation systems to supplement traditional clinical approaches. The 70.3% average accuracy across severity classifications represents a significant advancement over existing automated systems and approaches the performance of experienced clinicians (75-80%).

Patient data privacy is ensured through HIPAA-compliant encryption, local processing capabilities, and user consent protocols. The system includes explainable AI features to maintain clinician oversight and trust in AI-assisted decisions.

Technical Innovations

The key technical innovations include:

- Novel cross-modal attention mechanism for aphasia- specific feature integration
- Contrastive learning framework adapted for clinical severity classification

- Real-time processing pipeline optimized for clinical deployment
- Subject-independent validation preventing data leakage

Limitations and Future Work

Current limitations include language dependency (English only), dataset size constraints, and need for extensive clinical validation. Future work will address multilingual support, mobile deployment, and integration with electronic health records.

V. Conclusion

This paper presents a novel context-aware aphasia recovery system that addresses critical limitations in current rehabilitation approaches through technical innovations in multimodal fusion, cross-modal attention, and contrastive learning. Comprehensive evaluation demonstrates superior performance over established baselines with significant therapeutic impact. The system represents a meaningful advancement toward accessi- ble, effective AI-assisted aphasia rehabilitation.

Acknowledgment

The authors would like to express their sincere gratitude to Mrs. Raghavi S, Assistant Professor in the Department of Computer Science and Engineering at St. Joseph's College of Engineering, for her invaluable mentorship, guidance, and continuous support throughout this research project. We also acknowledge the AphasiaBank consortium for providing access to the comprehensive dataset that made this research possible.

Funding Details

This work was supported by the AICTE, Government of India through Research Promotion Scheme File No. 8- 100/FDC/RPS/POL-ICY-1/2021-22.

References

[1]. Aphasia Institute. "Aphasia Statistics." Available: https://www.aphasia.org/aphasia-resources/aphasia-statistics/

[2]. C. K. Thompson and L. L. Choy, "Principles of Neuroplasticity in Supporting Recovery of Language Following Stroke," Journal of Speech, Language, and Hearing Research, vol. 52, no. 1, pp. 272-286, 2009.

[3]. A. Kertesz, Western Aphasia Battery-Revised, San Antonio, TX: Psych- Corp, 2007.

[4]. P. Divya, A. Chandrasekar and S. Raghavi, "Prediction of Osteosarcoma Bone Cancer Using Convolutional Neural Networks and Multi-Feature Integration Methods," 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD), Manama, Bahrain, 2025, pp. 1-6, doi: 10.1109/ITIKD63574.2025.11005254.

[5]. C. Breitenstein, S. Grewe, D. Flo¨el, W. Ziegler, and A. Springer, "Com- puterized Speech and Language Therapy for Aphasia: A Systematic Review and Meta-Analysis," Journal of Speech, Language, and Hearing Research, vol. 60, no. 5, pp. 1411-1427, 2017.

[6]. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171-4186.

[7]. Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," arXiv preprint arXiv:1508.01991, 2015.

[8]. R. Ranjith and A. Chandrasekar, "GTSO: Gradient tangent search optimization enabled voice transformer with speech intelligibility for aphasia," Computer Speech & Language, vol. 84, 2024, 101568, doi: https://doi.org/10.1016/j.csl.2023.101568.

[9]. B. MacWhinney, D. Fromm, M. Forbes, and A. Holland, "AphasiaBank: Methods for Studying Discourse," Aphasiology, vol. 25, no. 11, pp. 1286-1307, 2011.

[10]. S. Raghavi, R. Ranjith and A. Chandrasekar, "Fatigue and Sluggishness Detection Using Machine Learning: A Haar Algorithmic Approach," 2024 International Conference on Advances in Computing, Communi- cation and Applied Informatics (ACCAI), Chennai, India, 2024, pp. 1-5, doi: 10.1109/ACCAI61061.2024.10601943.

[11]. R. Rajendran and A. Chandrasekar, "Conv-transformer-based Jaya Gazelle optimization for speech intelligibility with aphasia," SIViP, vol. 18, pp. 2079-2094, 2024, doi: https://doi.org/10.1007/s11760-023- 02844-0.

[12]. D. Le, K. Licata, E. Persad, and C. Provost, "Automatic Assessment of Speech Intelligibility for Individuals with Aphasia," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 11, pp. 2187-2199, 2016.

[13]. A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," in Proc. Advances in Neural Information Processing Systems, 2020, pp. 12449-12460.

[14]. K. C. Fraser, N. Linz, B. Lindsay, and Y. Yunusova, "Automated Classification of Primary Progressive Aphasia from Connected Speech," Cortex, vol. 129, pp. 437-446, 2020.

[15]. L. Bonilha, G. Yourganov, C. Rorden, and H. Fridriksson, "Mapping Residual Cognitive Function in Chronic Aphasia," NeuroImage, vol. 73, pp. 8-15, 2013.

[16]. M. Palmer, R. Enderby, C. Hawley, F. Julious, R. Paterson, and D. Berry, "Computer therapy compared with usual care for people with long-standing aphasia poststroke," Stroke, vol. 43, no. 7, pp. 1904-1911, 2012.

[17]. N. Galluzzi, M. Mayberry, and T. Mlynarski, "Deep Learning for Automatic Speech Recognition in Aphasia Rehabilitation," IEEE Trans- actions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 6, pp. 1285-1294, 2020.

[18]. S. Wilson, A. Saygin, M. Sereno, and M. Iacoboni, "Listening to speech activates motor areas involved in speech production," Nature Neuroscience, vol. 7, no. 7, pp. 701-702, 2004.