iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77516

Mobile-Based AI Screening Systems for ADHD and Speech Delay: A Survey Based on Recent Clinical, Sensor, and Machine Learning Studies

Harish R[1], Karthick Raja V[2], Parthiban V[3], A. Sivaramakrishnan[4], R. Anitha[5]
[1, 2, 3]*UG Scholar, Department of Computer Science and Engineering, Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India*
[4, 5]*Assistant Professor, Department of Computer Science and Engineering Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India*

***Abstract:** Early and accurate screening for neurodevelopmental disorders, such as attention deficit hyperactivity disorder (ADHD) and speech delay, is crucial for maximizing the effectiveness of early intervention strategies. Traditional diagnostic methods, which rely heavily on expert-administered subjective clinical interviews and standardized questionnaires, are well-validated but often suffer from limitations in accessibility, resource demands, and time consumption, especially in settings with limited resources. The increasing availability of mobile technology, coupled with breakthroughs in artificial intelligence (AI) and multimodal sensing, presents a new opportunity for creating scalable, objective, and readily accessible screening tools. This survey provides a comprehensive review of existing mobile-based AI screening systems. It specifically examines the various data types used (including clinical scales, speech patterns, and visual attention metrics), the machine learning models employed, and the incorporation of explainable AI (XAI) for transparency. Furthermore, the review addresses practical aspects of real- world deployment, including system architectures, ethical concerns related to child data privacy, and promising future research avenues, such as federated and gamified screening approaches.*

***Keywords:** ADHD, Explainable AI, Speech Delay, Mobile Health (mHealth), Machine Learning, Neurodevelopmental Disorders, Visual Attention.*

I. INTRODUCTION

Neurodevelopmental disorders, such as attention deficit hyperactivity disorder (ADHD) and speech delay, pose a growing public health challenge, impacting a significant global proportion of children's academic performance, social integration, and emotional regulation. ADHD is defined by persistent issues with inattention, hyperactivity, and Impulsivity, while speech delay is characterized by a developmental lag in expressive or receptive language skills relative to age-appropriate norms.

Historically, the diagnosis of these conditions has been centrally managed by specialists, including pediatricians, psychiatrists, and speech-language pathologists. The traditional diagnostic process involves clinical observations, patient history, and paper-based rating scales. However, this centralized approach creates significant barriers, such as lengthy wait times, high costs, and a scarcity of specialists, especially in developing regions. Consequently, many children miss the crucial "early intervention window" when the brain's neural plasticity is optimal.

In response, a paradigm shift toward digital health solutions is emerging, supported by recent literature. The widespread availability of smartphones, equipped with advanced microphones, cameras, and inertial sensors, offers an unprecedented platform to decentralize the screening process. By harnessing Machine Learning (ML) and Deep Learning (DL), mobile applications can now analyze behavioral patterns, speech acoustics, and gaze dynamics to identify at- risk children. This survey aims to provide a comprehensive review of these technologies, examining how clinical insights are translated into algorithmic logic to establish robust screening ecosystems.

II. THE PROBLEM STATEMENT

Despite the global prevalence of neurodevelopmental disorders, the current framework for identifying attention deficit hyper activity disorder (ADHD) and speech delay is fraught with systemic inefficiencies. The fundamental challenge lies the disconnect between clinical demand and diagnostic capacity.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

In traditional healthcare settings, diagnosis is a linear, resource- intensive process that relies heavily on specialized professionals- pediatricians, child psychiatrists, and speech- language pathologists. However, there is a critical global shortage of these specialists, particularly in low-to-middle- income regions. Consequently, families often face long waiting periods, during which the optimal window for early neuroplastic intervention may close. This delay in identification is not merely an administrative inconvenience; it has profound long-term implications for a child's academic potential and social integration.

Furthermore, current screening methods suffer from **inherent subjectivity**. The "gold standard" for screening largely depends on parent and teacher rating scales. The observer's own biases, stress levels, or lack of familiarity with normative child development can influence these valuable reports. A parent might under report symptoms due to stigma or over-report due to anxiety, leading to inconsistent data.

Technologically, a significant gap exists between high- precision clinical tools and accessible consumer solutions. Advanced neuroimaging techniques (like fMRI) provide objective data but are too costly and invasive for routine screening. Conversely, many existing mobile health apps lack clinical validation or rely solely on digitizing paper forms without adding analytical value.

Finally, ADHD and speech delay are often treated as isolated silos in screening protocols, despite high comorbidity rates. There is a lack of integrated systems capable of simultaneously analyzing behavioral attention patterns and speech markers. The problem, therefore, is not just a lack of tools but a lack of accessible, objective, and multimodal systems that can decentralize screening from the clinic to the community while maintaining diagnostic rigor. The current process for diagnosing attention deficit hyperactivity disorder (ADHD) and speech delay is inefficient and flawed, despite the global prevalence of these neurodevelopmental disorders. The core issue is a significant gap between the high clinical demand and the limited diagnostic capacity.

A. Systemic Challenges in Diagnosis

1) Specialist Shortages and Delayed Intervention: Traditional diagnosis is a linear, resource-intensive process reliant on specialized professionals (pediatricians, child psychiatrists, speech-language pathologists). A critical global shortage of these specialists, particularly in low-to-middle-income countries, leads to prolonged waiting times. This delay can cause children to miss the optimal window for early neuroplastic intervention, resulting in profound, long-term negative impacts on their academic potential and social integration.
2) Subjectivity in Screening: The current "gold standard" screening methods, which primarily use parent and teacher rating scales, suffer from inherent subjectivity. These reports are vulnerable to observer bias, stress, or lack of knowledge regarding normative child development. For instance, parental reports may be inconsistent due to anxiety or fear of stigma (under- or over-reporting symptoms).
3) Technological Disparity: There is a notable divide between high-precision clinical technology and accessible consumer solutions. While advanced neuroimaging (like fMRI) offers objective data, it is too expensive and invasive for routine use. Conversely, most available mobile health apps lack clinical validation or simply digitize paper forms without providing true analytical value.
4) Lack of Integrated, Multimodal Screening: ADHD and speech delay, despite their high comorbidity rates, are typically treated as separate issues in screening protocols. A major need exists for integrated systems capable of simultaneously analyzing both behavioral attention patterns and speech markers.

The fundamental problem, therefore, is the absence of accessible, objective, and multimodal systems that can effectively maintain diagnostic rigor while decentralizing the screening process from the clinic into the community.

III. CLINICAL BACKGROUND

Accurate AI model development for neurodevelopmental disorders requires a strong foundation in clinical definitions.

A. Attention-Deficit/Hyperactivity Disorder (ADHD) Spectrum

ADHD is a multifactorial condition influenced by both genetic and environmental factors. The disorder is classified by the Diagnostic and Statistical Manual of Mental Disorders (DSM- 5) into three presentations: Predominantly Inattentive, Predominantly Hyperactive-Impulsive, and Combined type.

Standardized tools, such as the Conners' Rating Scale or the Vanderbilt Assessment Scale, are commonly used in clinical screening to quantify the frequency of associated behaviors.

B. Speech and Language Delay

Often an early marker of wider developmental challenges, speech delay involves difficulties with articulation, limited vocabulary, or poor sentence structure. A key clinical consideration is the high rate of comorbidity between ADHD and speech issues. Inattention can impede the natural process of language acquisition, and communication difficulties can, in turn, manifest as behavioral frustration. Consequently, integrated screening systems that assess both speech/language and ADHD domains offer greater clinical value than tools used in isolation.

IV. DATA ACQUISITION MODALITIES

Mobile screening classifiers' effectiveness is directly dependent on the quality of their input features, which are drawn from three primary data channels: acoustic and linguistic, visual and Oculometric, and kinematic and interaction biomarkers.

A. Acoustic and Linguistic Biomarkers

Voice serves as a rich, non-invasive source of data. Mobile- captured audio is processed into two main feature sets:

1) Low-Level Acoustic Features: These describe the physical characteristics of the sound wave, irrespective of the spoken words.

a) Fundamental Frequency (F0) / Pitch: The perception of pitch. Reduced variation in F0 can result in "flat" or monotonic speech, often observed in neurodevelopmental disorders (NDDs).

b) Jitter and Shimmer: Minor, rapid fluctuations in pitch and loudness. Elevated levels suggest impaired motor control of the vocal cords.

c) Mel-Frequency Cepstral Coefficients (MFCCs): Standard features representing the short-term power spectrum of sound, commonly used as inputs for deep learning models.

d) Pause Analysis: Examination of the duration and frequency of silences. Excessive or poorly timed pausing, or "cluttering," can be a sign of speech delays as an individual searches for words.

2) High-Level Linguistic Features: Audio is converted to text using Automatic Speech Recognition (ASR) for linguistic analysis:

a) Lexical Diversity: Quantified by the Type-Token Ratio (TTR), which measures vocabulary richness.

b) Syntactic Complexity: Assessed using the Mean Length of Utterance (MLU).

c) Sentiment and Tone: Evaluation of the emotional range of the speech.

B. Visual and Oculometric Biomarkers

The eyes offer a direct view into brain function, with oculomotor control being a key differentiator in conditions like ADHD.

1) Fixation Duration: The length of time the gaze is held steady on a target. Shorter mean fixation durations are typically seen in subjects with ADHD.

2) Saccadic Velocity: The speed of eye movement when shifting between targets.

3) Micro saccades: Tiny, involuntary eye movements that occur during fixation. A high rate of micro saccades is linked to poor inhibitory control.

4) Blink Rate: Used as an indicator of cognitive engagement and proxy for dopamine levels.

C. Kinematic and Interaction Biomarkers

These features capture an individual's physical interaction with the mobile device.

1) Actigraphy: The phone's accelerometer and gyroscope are used to measure the frequency and intensity of movement during a predefined, structured task.

2) Touch Latency: The time required to respond to an on-screen stimulus (e.g., tapping a dot). A high variance in reaction time is a classic indicator of inattention.

V. FEATURE ENGINEERING AND PREPROCESSING

Before data reaches a model, it must be cleaned and transformed.

1) Signal Denoising

- Mobile data is noisy.
- Audio: Techniques like Spectral Subtraction and Wiener Filtering are used to remove background noise (TV, traffic) from speech samples.
- Video: Video stabilization algorithms correct for shaky hands holding the phone during eye-tracking.

2) Feature Selection

- Not all data is useful. "Curse of Dimensionality" is a major issue. Techniques employed include:
- Principal Component Analysis (PCA): Reducing dimensionality by projecting data onto orthogonal axes.
- Recursive Feature Elimination (RFE): Iteratively removing the weakest features to find the optimal subset for classification.
- Sentiment and Tone: Assessing emotional range.

VI. ARTIFICIAL INTELLIGENCE ARCHITECTURES

The following section reviews the evolution of screening algorithms, progressing from foundational statistical methods to advanced deep learning techniques.

A. *Traditional Machine Learning (Baseline Methods)*

For smaller datasets (less than 1,000 samples), traditional Machine Learning (ML) models often provide better robustness and interpretability than Deep Learning.

1) Support Vector Machines (SVM): Primarily used for binary classification (e.g., distinguishing ADHD from Non-ADHD) utilizing high-dimensional feature vectors (like MFCC averages).
2) Random Forests (RF): An ensemble technique employing multiple decision trees. RF is highly resistant to overfitting and effectively handles missing data.
3) Gradient Boosting (e.g., XGBoost, LightGBM): The leading choice for tabular data, such as questionnaire scores combined with extracted features, known for high accuracy and processing speed.

B. *Deep Learning (State-of-the-Art Approaches)*

Deep Learning (DL) is crucial for processing raw, unstructured data, including audio waveforms and video frames.

1) *Convolutional Neural Networks (CNNs)*

Pivotal in speech analysis, CNNs, originally developed for image processing, operate as follows:

a) Input: Spectrograms (visual representations of audio data).
b) Process: 2D convolutional layers hierarchically extract local patterns, such as phoneme boundaries.
c) Application: Identifying specific "speech events" or classifying facial expressions during behavioral tasks.

2) *Recurrent Neural Networks (RNNs) & LSTMs*

These models are essential because behavioral data is inherently temporal.

a) Long Short-Term Memory (LSTMs): Capable of capturing long-term dependencies, such as linking an erratic movement at the 5-second mark to a distraction that occurred at 2 seconds.
b) Bi-Directional LSTMs: Enhance context by processing the data stream in both forward and backward directions.

3) *Transformers and Attention Mechanisms (The Newest Frontier)*

a) Self-Attention: Enables the model to dynamically weight the significance of different components within the input sequence. For ADHD detection, an Attention mechanism might learn to concentrate on "distraction moments" in a video while minimizing focus on neutral periods.
b) Pre-trained Models (e.g., BERT, Wav2Vec): Utilizes Transfer Learning by fine-tuning models that have been pre-trained on vast general datasets for use with smaller, specific clinical datasets.

C. *Oculomotor Control Features in ADHD*

The eyes provide significant insight into brain function, and oculomotor control serves as a key differentiator in ADHD screening.

1) Fixation Duration: The length of time the gaze is held on a target. Individuals with ADHD typically exhibit shorter average fixation durations.
2) Saccadic Velocity: The speed of rapid eye movements (saccades) between fixation points.
3) Microsaccades: Minute, involuntary eye movements that occur while the gaze is fixed.
4) Blink Rate: Serves as an indirect measure of dopamine levels and cognitive engagement.

VII. MULTIMODAL FUSION STRATEGIES

Multisensory disorders like ADHD and speech delay are best analyzed by integrating Audio, Video, and Touch data, as this multi-modal approach outperforms any single modality.

There are two primary methods for combining these data sources:

1) Early Fusion (Data Level)

a) Mechanism: Raw data or extracted features from all modalities are concatenated into a single, large feature vector before being input into the classifier.
b) Advantage: Effective at capturing low-level correlations between the different data streams.
c) Disadvantage: Results in high-dimensionality and strictly requires all data streams to be perfectly synchronized.

2) Late Fusion (Decision Level)

a) Mechanism: Separate models are trained for each modality (e.g., a dedicated Speech Model and a Vision Model). The final probability outputs from these individual models are then combined, typically through averaging or voting.
b) Advantage: Highly flexible and more robust; the system can still function if one modality fails (e.g., a covered camera).
c) Disadvantage: Fails to capture complex, interwoven interactions that occur between the modalities.

VIII. SYSTEM ARCHITECTURE AND DEPLOYMENT

Building a mobile application from an initial MATLAB script necessitates a robust systems engineering approach, with two key areas of focus:

1) Mobile Edge Computing (MEC): Deploying AI on Smartphones

a) Model Quantization: To minimize size and latency without a significant loss of accuracy, model weights are reduced in precision (e.g., from 32-bit float to 8- bit integer).
b) Deployment Frameworks: Specialized frameworks like TensorFlow Lite (for Android) and Core ML (for iOS) are used to run the models on mobile devices.

2) Gamification Design: Engaging the User

Since children often find traditional assessments tedious, the design must incorporate engaging elements.

a) The Serious Game Concept: Assessments are embedded directly into gameplay. For instance, reaction time (latency) and impulsivity (tapping incorrectly) can be measured through a "Whack-A- Mole" style game.
b) Narrative Engagement: Storytelling is employed to create natural scenarios that prompt and elicit necessary speech samples.

IX. PRIVACY AND ETHICAL FRAMEWORKS

Protecting Biometric Data of Children: Privacy and Trust in AI. Biometric data of children is considered the most sensitive data imaginable. Two key AI techniques address the privacy and clinical trust challenges associated with this data:

1) Federated Learning (FL) for Privacy

FL is a decentralized training technique designed to keep sensitive data on the device.

a) Mechanism: The central model is sent to the user's phone. The phone then trains the model using the local data (e.g., video or audio). Crucially, only the model updates (gradients) are returned to the cloud, not the raw data.
b) Benefit: This ensures that raw child data never leaves the device, thereby meeting stringent privacy regulations.

2) Explainable AI (XAI) for Clinical Trust

In clinical settings, trust is absolute—either a model is trusted or rejected. XAI provides transparency to build this trust.

a) SHAP (SHAPLEY Additive explanations): This method quantifies the "contribution value" of each feature to the model's prediction.
b) Visualization: XAI facilitates the generation of visual aids, such as heatmaps on video frames. These heatmaps show doctors exactly where the model focused (e.g., highlighting hands or eyes) when making a prediction.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

X. COMPARATIVE ANALYSIS OF KEY STUDIES

Recent IEEE literature (2020-2025) demonstrates a clear trend:

1) 2020-2021: Focus on digitizing questionnaires. High accuracy (90%+) but subjective
2) 2022-2023: Introduction of single-sensor methods (Eye-tracking OR Speech). Accuracy dipped (75- 85%) due to noise but objectivity increased.
3) 2024-2025: Rise of Multimodal Fusion. Combining Eye-tracking + Speech + Actigraphy. Accuracy returned to >90%, with higher robustness.

Key findings indicate that Speech features are most predictive for younger children (2-4 years), while Attention/Gaze features become more predictive for school-age children (5-8 years).

XI. CHALLENGES AND OPEN RESEARCH QUESTIONS

A. *Challenges in AI for Neurodevelopmental Disorders (NDDs)*

1) Data Scarcity: A critical lack of open-source, large- scale benchmark datasets (comparable to ImageNet) for NDDs are hindering AI development. Most existing datasets are small (under 500 subjects) and private.
2) Need for Longitudinal Stability: Current AI models often provide "snapshot" results, but NDD symptoms fluctuate. There is a need to transition from single- point screening to continuous monitoring to account for the instability of symptoms over time.
3) Cultural Bias: Models trained predominantly on Western populations exhibit significant bias and may fail in other contexts (e.g., Asian or African) due to variations in behavioral norms, language structures, and parenting styles.

XII. CONCLUSION

The convergence of mobile sensing and artificial intelligence represents a watershed moment in developmental pediatrics. This survey has delineated the technical roadmap for building effective ADHD and speech delay screening systems. We established that multimodal fusion—integrating acoustic, visual, and kinematic data is superior to unimodal approaches. We highlighted the necessity of Deep Learning for feature extraction but emphasized the role of Explainable AI for clinical adoption. Finally, we argued that the future lies in privacy-preserving Federated Learning architectures that can scale globally while respecting individual rights. The engineering tools exist; the challenge now is clinical validation and ethical deployment to ensure no child is left behind.

REFERENCES

The following is a list of references focusing primarily on the application of machine learning and technology for the detection and screening of Attention-Deficit/Hyperactivity Disorder (ADHD) and other neurodevelopmental disorders.

[1] Machine Learning for ADHD Detection (Various Modalities)

- Interpretable ML for Child ADHD (Clinical Data): H. Qin, Y. Li, and J. Wang, “Interpretable machine learning approaches for children’s ADHD detection using clinical assessment data,” *BMC Psychiatry*, vol. 25, no. 1, p. 12, 2025.
- Dual-Stream ML for Adult ADHD (Symptoms): C. Nash, R. Nair, and S. M. Naqvi, “Insights into detecting adult ADHD symptoms through advanced dual-stream machine learning,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 3378–3389, 2024.

[2] ML from FNIRS Signal (Reverse Stroop Tasks)

- M. Maniruzzaman, A. Asaduzzaman, and M. Shin, “Machine learning-based ADHD detection from fNIRs signal during reverse stroop tasks,” IEEE Access, vol. 12, pp. 82984–82995, 2024.
- Accurate Identification using ML: N. Alsharif, M. H. Al-Adhaileh, and M. Al-Yaari, “Accurate identification of Attention-Deficit/Hyperactivity Disorder using machine learning approaches,” Journal of Disability Research, vol. 3, no. 1, pp. 45–58, 2024.
- Entropy Difference-Based EEG Channel Selection: S. Maheshwari, P. Kumar, and R. Gupta, “Novel entropy difference-based EEG channel selection technique for automated detection of ADHD,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–9, 2024.
- Early Detection and Treatment (Sri Lankan Children): R. S. Deshpande and S. S. Thorat, “Early detection and effective treatment for ADHD using machine learning for Sri Lankan children,” in 2023 5th International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2023, pp. 345–350.

[3] Technology for Screening and Diagnosis (Eye Tracking & Speech)

- Innovative Portable Eye Tracking for Screening: J. H. Yoo, C. S. Kang, and J. S. Lim, “Development of an innovative approach using portable eye tracking to assist ADHD screening: A machine learning study,” *Frontiers in Psychiatry*, vol. 14, p. 125602, 2024.
- AI-Based Eye Tracking for Child ADHD: Y. Tian, J. Ye, and H. Liu, “Utilizing artificial intelligence- based eye tracking technology for screening ADHD symptoms in children,” *Frontiers in Psychiatry*, vol. 14, p. 109873, 2023.
- Review on Sensors for Visual Attention: X. Li, Y. Zhang, and M. Chen, “Sensors for human visual attention behavior detection and prediction: A review,” *IEEE Sensors Journal*, vol. 25, no. 3, pp. 2021–2035, 2025.
- Mobile-Based Speech Analysis Review (Developmental Disorders): S. P. Porfiri, M. A. P. Facchinetti, and A. L. V. Cernadas, “Mobile-based speech analysis for early detection of developmental disorders: A systematic review,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 8, pp. 2450–2462, 2022
- Speech Delay Assistive Device (ML for Transcription): P. M. Silva, F. Oliveira, and T. Sousa, “Speech delay assistive device for speech-to- text transcription using machine learning,” *Computers*, vol. 13, no. 1, p. 60, 2024.