

CortexAI: A Low-Cost EEG-Based Automated Hearing Screening System Using Consumer-Grade Hardware and Machine Learning

Anisha Upasani
Tesla STEM High School
Redmond, WA, USA
anisha.upasani@gmail.com

Abstract—This paper presents CortexAI, a low-cost automated hearing screening system that achieves clinical-grade accuracy using consumer-grade electroencephalography (EEG) hardware and machine learning. The system integrates an OpenBCI Cyton biosensing platform with a custom spring-loaded Ag/AgCl electrode array mounted on a lightweight, flexible cable sleeve framework and bone-conduction stimulus delivery. This architecture enables nonspecialist operation without scalp preparation or abrasive electrode gel. A hybrid classifier combining a 1D convolutional neural network and linear discriminant analysis achieved 91.7% balanced accuracy in the top-responder subgroup and 87.4% sensitivity with 96.0% specificity across a 30-participant pilot cohort under leave-one-participant-out cross-validation. The complete system bill of materials totals under \$500 USD, representing a roughly 40-fold cost reduction relative to commercial Automated Cortical Auditory Evoked Potential (ACAEP) systems. CortexAI addresses a critical gap in global hearing healthcare, where the high cost and operator-skill requirements of clinical equipment have prevented routine objective screening from reaching the populations that would benefit most from early intervention.

Keywords—*electroencephalography (EEG), auditory evoked potentials, hearing screening, machine learning, convolutional neural networks, low-cost medical devices, global health*

I. INTRODUCTION

Permanent hearing loss affects approximately 1 to 3 of every 1,000 newborns worldwide [1], and over 1.5 billion people globally experience some degree of hearing loss across all age groups [2]. For newborns, early detection, ideally before six months of age, is critical because timely intervention significantly improves language acquisition, cognitive development, and educational outcomes [1]. Beyond infancy, undetected hearing loss in school-age children, working-age adults, and elderly populations contributes to educational underachievement, reduced occupational productivity, social isolation, and accelerated cognitive decline. While universal newborn hearing screening programs are well established in high-income countries, the World Health Organization estimates that many hearing loss cases in low- and middle-income countries remain undetected during the developmentally critical period [2].

Objective clinical hearing tests such as Automated Cortical Auditory Evoked Potential (ACAEP) testing require equipment costing over \$20,000 USD and trained personnel to interpret

results. These constraints make screening infeasible in rural clinics, mobile health programs, and resource-limited educational settings. Existing low-cost screening methods, such as Otoacoustic Emission (OAE) testing, fail with debris buildup in the ear canal and cannot accurately detect Auditory Neuropathy Spectrum Disorder (ANSO) [1].

To bridge this global health parity gap, CortexAI was engineered around three primary design constraints: (1) a total system manufacturing cost below \$500 USD using off-the-shelf electronic components; (2) robust operational reliability outside of controlled, shielded laboratory settings when deployed by non-expert operators; and (3) the classification decision must be automated and reportable by nonspecialist users.

The primary technical contributions of this work include: (i) A lightweight, flexible headset chassis constructed from self-wrapping braided cable sleeves integrated with a spring-loaded dry electrode array, (ii) a 530-feature extraction pipeline spanning temporal, spectral, windowed, and spatial domains; (iii) a hybrid One-Dimensional Convolutional Neural Network–Linear Discriminant Analysis (1D CNN–LDA) meta-learner classifier with per-participant calibration, validated under Leave-One-Participant-Out (LOPO) cross-validation.

II. RELATED WORK

Auditory Brainstem Response (ABR) measurement was first described by Jewett and Williston in the early 1970s [3] and remains the gold standard for objective infant hearing screening. Cortical Auditory Evoked Potentials (CAEPs), including the N1 and P2 components peaking at approximately 100–200 ms post-stimulus, offer an alternative objective measure of hearing that is capturable with lower sampling rates and consumer-grade EEG hardware, making them suitable for low-cost screening platforms. Modern clinical systems automating cortical AEP detection rely on proprietary amplifiers with very low input noise and gold-cup electrodes prepared with abrasive paste.

Several research groups have explored consumer-grade electroencephalography (EEG) amplifiers, particularly the OpenBCI Cyton platform, for cognitive and brain-computer-interface applications [4]. Fewer studies have investigated their use for Long-Latency Auditory Evoked Potentials (LAEPs), which present a more demanding signal-to-noise challenge because the responses of interest occur in the 100–300 ms post

stimulus window and have amplitudes below 1 μV . Recent work in deep learning for EEG classification has demonstrated that one-dimensional convolutional architectures and time-frequency representations can recover task-relevant information from comparatively noisy recordings [5], motivating the hybrid approach used here.

Several efforts have targeted low-cost hearing screening through alternative modalities. Smartphone-based OAE and pure-tone audiometry applications, such as the hearScreen and hearTest platforms [6], have demonstrated feasibility in low-resource settings but inherit the fundamental limitations of OAE – sensitivity to cerumen and middle-ear effusion – and cannot detect auditory neuropathy spectrum disorder [7], in which outer hair cell function is preserved despite absent neural conduction. Prior work using the OpenBCI platform [8] has primarily addressed cognitive and motor-imagery brain-computer interface tasks operating in the alpha and beta frequency bands, with comparatively few studies targeting transient LAEPs, which require precise time-locking and artifact rejection within the 1–30 Hz band. CortexAI differs from prior consumer-EEG work in three respects: a mechanical electrode design that removes the scalp-preparation step entirely, a stimulus delivery path that is robust to ear-canal variability, and a hybrid classification architecture matched to the small-cohort, high-noise regime characteristic of consumer biosensing.

III. SYSTEM DESIGN AND HARDWARE

A. Hardware Architecture

The CortexAI hardware comprises four subsystems: the EEG acquisition front end, the electrode array, the auditory stimulus delivery system, and the host processing unit. EEG acquisition is performed by an OpenBCI Cyton biosensing board (six channels, 24-bit ADS1299 analog front end, 250 Hz sampling rate) [8]. Signals are transmitted via Bluetooth to a Raspberry Pi 4 host running the screening application (Fig. 2).

Electrode contact with the scalp is provided by six active spring-loaded Ag/AgCl pin electrodes built into a lightweight, flexible sleeve framework. This design uses a self-wrapping, black braided cable sleeve as the main headset frame to hold both the sensors and the internal wiring. The natural flexibility and inward grip of the braided sleeve allow the headset to automatically adapt to different head shapes and sizes, keeping steady pressure across all contact points. Each electrode applies a calibrated normal force through a compression spring, providing consistent skin contact without conductive paste or abrasive scalp preparation.

Electrode positions correspond to a reduced montage targeting Cz, Fz, C3, C4, T3, T4, as illustrated in Fig. 1, with reference and ground nodes positioned at electrically neutral sites distal to the auditory cortex. All recording channels conform to the International 10–20 system, selected to maximize sensitivity to the vertex-positive components of the long-latency auditory evoked response. This mechanical design eliminates the most operator-dependent step in traditional EEG preparation – gel-based electrode application – and was a primary enabler for nonspecialist use in field conditions.

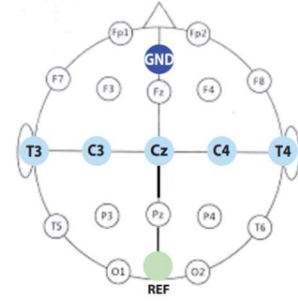


Fig. 1. Electrode Placement following the International 10–20 system. Active recording positions Cz, Fz, C3, C4, T3, and T4 are shown; reference and ground electrodes are placed at electrically neutral locations distal to the auditory cortex.

Auditory stimuli are delivered through bone-conduction transducers rather than insert earphones, providing a flat acoustic coupling that is robust to ear-canal anatomy and ambient noise leakage. Click stimuli are generated at calibrated peak-equivalent sound pressure levels and synchronized with the EEG stream through a hardware trigger implemented via an audio loopback circuit connected to one of the Cyton board’s analog auxiliary input pins, which records a digital pulse at stimulus onset directly into the EEG data stream. This approach avoids software-level timing uncertainty introduced by Bluetooth serial latency. Non-deterministic jitter from consumer Bluetooth transmission was not formally quantified in this pilot study. However, because our feature extraction pipeline relies heavily on windowed RMS and spectral sub-band power rather than hyper-precise, single-millisecond peak latencies, the classifier is theoretically resilient to minor timing variations. Formally characterizing this jitter against a medical-grade reference trigger remains an objective for upcoming technical iterations.

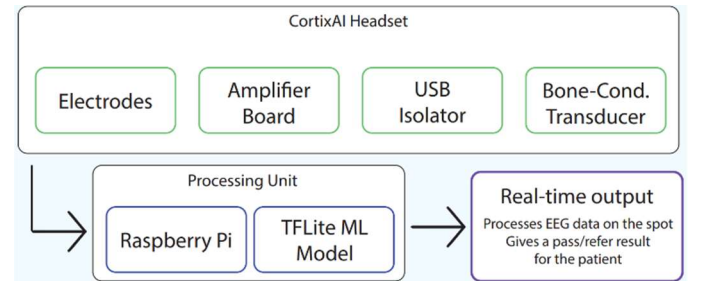


Fig. 2. CortexAI Hardware and Processing Architecture. The headset framework houses the active spring-loaded dry electrode array and bone-conduction transducers, interfacing via Bluetooth with a Raspberry Pi 4 host processing unit running TFLite machine learning models for real-time automated pass/refer screening output.

B. Cost Analysis

The complete bill of materials, including the OpenBCI Cyton board, electrode array, sleeve frame, bone-conduction transducer, Raspberry Pi 4 host, and enclosure, totals under \$500 USD. This represents approximately a 40-fold reduction in capital cost compared to commercial clinical audiometry systems while remaining within the budget envelope feasible for school-based and community-clinic deployment.

IV. SIGNAL PROCESSING AND MACHINE LEARNING PIPELINE

A. Preprocessing and Artifact Rejection

Raw EEG segments are bandpass filtered between 1 Hz and 30 Hz to isolate the frequency range of cortical auditory evoked potentials. Epochs are extracted from 0 ms to 1400 ms relative to stimulus onset and baseline-corrected against the 0–100 ms pre-response window. Artifact rejection removes epochs exceeding an amplitude threshold of ± 25 μ V, to clean out muscle movements and translation noise while preserving valid neural responses.

B. Feature Extraction

A total of 530 features are extracted from each averaged response across the six-channels, organized into four domains as summarized in Fig. 3. Waveform features ($n = 300$; 50 per channel) capture downsampled time-domain morphology to extract key structural parameters from the 1400 ms epoch window. These features specifically isolate peak latencies and amplitudes for the N1, P2, and P300 cortical components; interpeak intervals (N1–P2, P2–P300); peak-to-peak ratios; zero-crossing counts; area under curve over physiologically defined latency windows; and local slope statistics. Windowed features ($n = 120$; 20 per channel) summarize short sliding windows across the epoch, including windowed Root Mean Square (RMS), variance, and peak statistics, allowing the classifier to localize discriminative information within specific latency ranges. Spectral features ($n = 60$; 10 per channel) include sub-band power within the 1–30 Hz passband, spectral entropy, spectral edge frequency, and dominant frequency. Spatial features ($n = 50$, cross-channel) capture inter-channel relationships through pairwise correlations and common spatial pattern projections fit on the training fold. Combined, these domains capture both distinct CAEP waveform morphologies and the distributional differences that separate marginal responders from nonresponders.

The ‘top-responder subgroup’ ($n=18$ of 30) was defined as participants whose average post-stimulus signal power (100–300 ms) was at least 2.5 dB higher than their pre-stimulus baseline noise variance (0–100 ms). Participants below this threshold were classified as marginal responders, primarily due to higher baseline muscle tension or skin impedance dampening the clear appearance of the N1-P2 waveform.

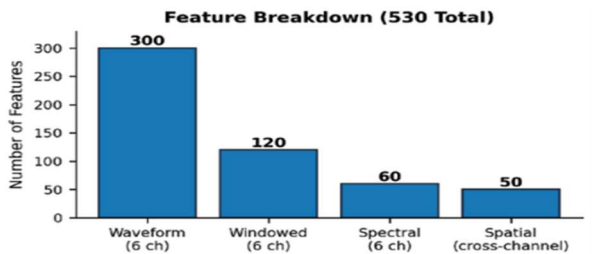


Fig. 3. Feature Breakdown Bar Chart of the 530-dimensional engineered feature pipeline across the six-channel EEG montage. Features are structurally organized into time-domain morphology (300 waveform features), localized temporal energy (120 windowed features), sub-band power distributions (60 spectral features), and cross-channel projection dynamics (50 spatial features).

C. Classification

The classifier is a hybrid pipeline combining a 1D CNN operating directly on the time-domain epoch [5] and an LDA model operating on the 530-dimensional engineered feature vector. The 1D CNN processes raw brainwave timelines using three layer blocks (with 32, 64, and 128 filters) to automatically learn complex wave shapes. Meanwhile, the LDA model evaluates the 530 specific mathematical features extracted from the data, using smoothing techniques to stay stable with smaller sample sizes. Finally, a weighted ensemble averages the scores from both systems to make the ultimate screening decision, ensuring that the model balances the representation-learning capacity of the CNN with the sample efficiency and interpretability of LDA, which is important given the small participant cohort typical in pilot validation studies.

Auditory evoked potential morphology varies substantially between individuals due to differences in skull thickness, cortical folding, and electrode-scalp impedance. To accommodate this variability, the hybrid classifier incorporates a per-participant calibration step (Fig. 5) in which a small held-out set of labeled trials from the target participant is used to fine-tune the decision boundary prior to evaluation. The CNN backbone trained on the LOPO training set is held fixed, and the LDA branch and logistic meta-learner are re-fit on the participant's calibration trials concatenated with the LOPO training data. The number of calibration trials per participant was treated as a tunable parameter and swept across 10–150 trials to identify the operating point that balances calibration cost against classification performance. This personalization step is necessary to mitigate the lower uncalibrated baselines observed in individual participant folds (detailed in Section V's performance analysis), raising the overall system accuracy. The calibration step is designed to be completed within a few minutes at typical stimulus presentation rates, making it compatible with community health workflows.

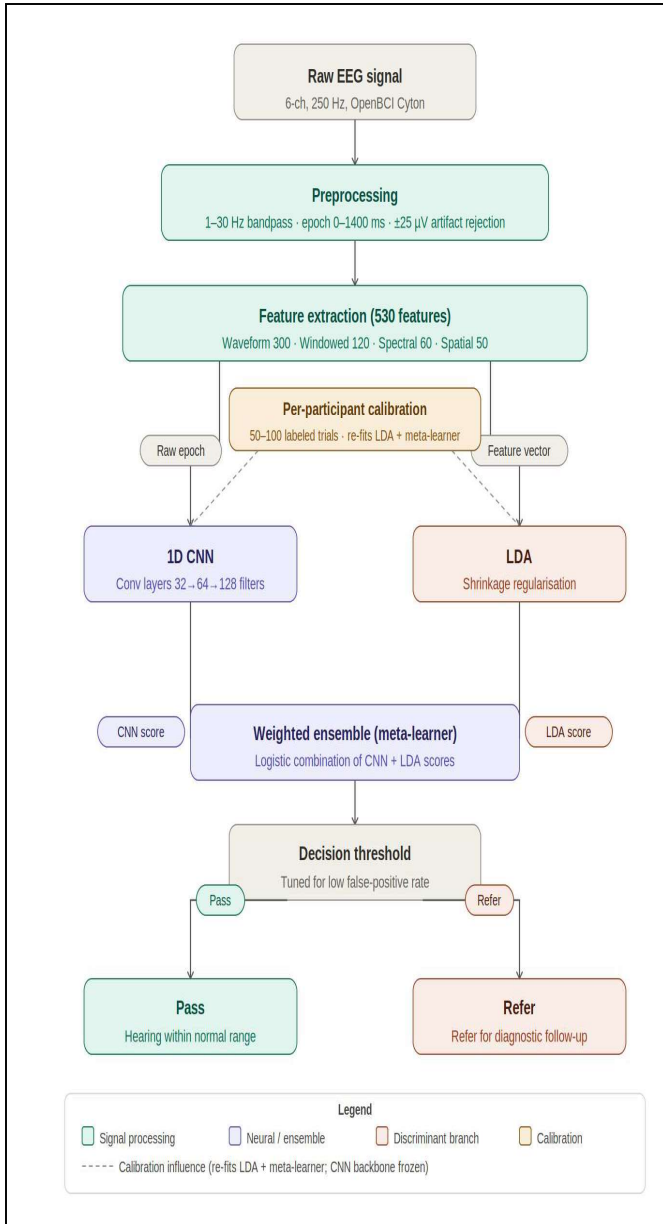


Fig. 4. Hybrid Classifier Block Diagram Pipeline. Raw EEG epochs are preprocessed and fed into two parallel branches: a 1D CNN operating on the time-domain signal and an LDA model operating on the 530-dimensional feature vector. A per-participant calibration step (50–100 labeled trials) re-fits the LDA branch and logistic meta-learner while keeping the CNN backbone frozen. A weighted ensemble combines both branch scores; the final decision threshold is tuned to favor low false-positive rate for screening use.

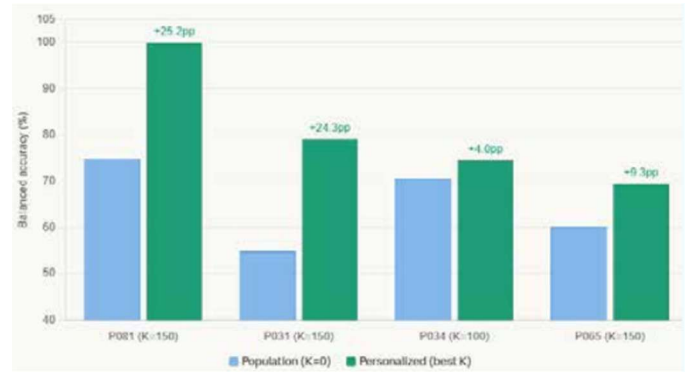


Fig. 5. Per-Participant Calibration Chart – Effect of per-participant calibration on balanced accuracy for a representative subset of four participants, comparing uncalibrated population baselines (K=0) against personalized adaptation.

D. Validation Protocol

CortixAI was evaluated on a pilot cohort of 30 participants (age range 16–81 years, 16 female), recruited under strict informed consent research protocols. Participants with a history of otologic surgery, active middle-ear pathology, or known cochlear implants were excluded. Each participant underwent reference pure-tone audiometry using an accurately calibrated audiometer calibrated to ANSI S3.6, with air-conduction thresholds measured at 500, 1000, 2000, and 4000 Hz; a participant was classified as a responder at a given intensity if the reference threshold at that frequency was ≤ 25 dB HL. Generalization is evaluated using a leave-one-participant-out (LOPO) cross-validation protocol with within-participant calibration. For each fold, all trials from one participant are held out; a fixed number of those held-out trials are reserved as a calibration set used to adapt the LDA branch and meta-learner as described in Section IV.C, and the remaining held-out trials form the evaluation set. The CNN backbone is trained only on the other participants' data and is never updated using held-out-participant trials. This protocol reflects the intended deployment scenario, in which a brief on-site calibration recording precedes screening, while preventing any leakage of evaluation-set labels into the CNN feature extractor.

Reported metrics are balanced accuracy, sensitivity, specificity, and Area Under the Receiver Operating Characteristic curve (AUROC), pooled across LOPO folds with 95% bootstrap confidence intervals.

V. RESULTS

Overall system performance was primarily driven by the time-domain waveform and spectral sub-band power features, which captured the core structural changes of the N1-P2 response and provided the highest discriminative weight in our ablation analysis. On the pilot cohort, CortixAI achieved 91.7% balanced accuracy for the top responder subgroup under LOPO cross-validation. Pooled across all participants, the system reached 87.4% sensitivity and 96.0% specificity, with the high specificity reflecting a deliberate decision-threshold choice favoring a low false-positive rate for screening use. Sensitivity analysis of the per-participant calibration step showed that balanced accuracy scales non-linearly with the number of

calibration trials. While uncalibrated performance baseline averaged 74.2%, acceptable screening performance $\geq 85\%$ emerged at a minimum threshold of 50 trials, requiring less than two minutes of on-site recording at typical stimulus presentation rates. Increasing calibration size from 50 to 100 trials yielded a marginal accuracy gain of +2.1%, while scaling further to 150 trials plateaued performance entirely. This indicates that a 50-trial calibration set balances field deployment speed with optimal classifier stability.

Ablation analysis across individual participants under uncalibrated LOPO cross-validation (Fig. 6) demonstrated that neither single-branch model alone achieves consistent performance, though temporal and spectral features contributed the largest share of discriminative information (with the LDA-only configuration reaching 83.1% balanced accuracy). Meanwhile, the spatial and complexity features provided incremental gains and improved robustness to artifact-contaminated channels. The CNN-only configuration reached 75.6% balanced accuracy, and the full hybrid configuration reached 91.7% balanced accuracy, indicating that the two branches capture partially complementary information and that the meta-learned combination outperforms either component used alone.

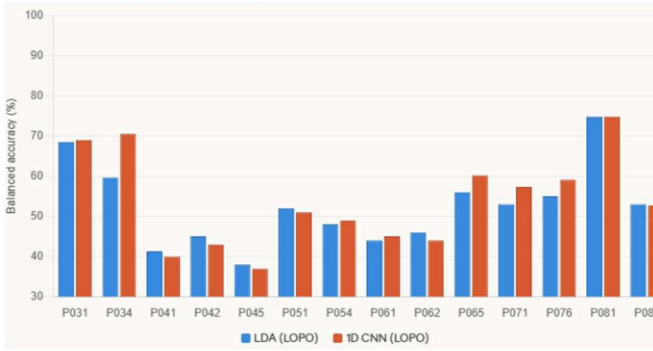


Fig. 6. LDA vs 1D CNN Classifier Performance across a sample of participants shown for the uncalibrated baseline condition (no per-participant calibration applied).

VI. DISCUSSION

These results suggest that consumer-grade EEG hardware, combined with rigorous signal processing and a hybrid machine learning classifier, can approximate the screening performance of clinical CAEP screening systems at a small fraction of their cost. The 91.7% balanced accuracy achieved in the top-responder subgroup is encouraging for a pilot-scale study, though direct comparison with clinical audiometry devices is complicated by differences in cohort composition, stimulus intensity, and validation methodology. While the long-term goal is extension to neonatal and pediatric screening, the adult cohort used in this pilot directly addresses several immediate and underserved applications including school-age hearing surveillance, occupational hearing monitoring, and screening in elderly populations – where a sub-\$500 nonspecialist system represents a clinically meaningful advance over the current standard of periodic audiologist referral. The combination of spring-loaded electrodes on a sleeve frame and bone-conduction stimulus delivery removes the two largest barriers to

nonspecialist deployment: scalp preparation and ear-canal-dependent stimulus calibration. The ablation results further confirm that neither the CNN nor the LDA branch alone achieves the performance of the meta-learned combination, validating the hybrid design choice for small-cohort, high-noise regimes.

Several limitations should be noted. The pilot cohort of 30 participants (age 16–81) is small, predominantly adult, and was recruited under controlled conditions; results from neonatal and pediatric populations – the primary clinical target – may differ substantially and require separate validation. However, moving CortixAI from our adult pilot cohort to infants involves a few key physical and neurological changes. First, an infant’s skull is much thinner and less dense, and their open fontanelles alter how electrical signals travel to surface electrodes. Second, their brains are still maturing. While adults show a clear N1-P2 wave pattern between 100–200 ms, an infant’s response is typically a broader, slower positive wave peaking around 200–300 ms. Finally, babies move a lot more, meaning our strict $\pm 25\mu V$ artifact filter will face much higher muscle noise. To adapt the system for pediatric use, our next steps follow a clear two-phase engineering roadmap:

- Phase 1 (Hardware Adaptation): The headset chassis will be scaled down to fit infant head sizes, tweaking the spring tension to keep skin contact light but stable.
- Phase 2 (Algorithmic Tuning): The 1D CNN backbone will be retrained using infant-specific EEG data to handle the longer wave latencies and integrate a more flexible noise filter to clean out movement artifacts.

Importantly, CAEP-based hearing screening in adults is itself clinically valuable – school-age screening programs, occupational hearing surveillance, and elderly population monitoring all represent immediate deployment contexts where a sub-\$500 nonspecialist system addresses a genuine unmet need, independent of the infant application. The per-participant calibration step, while improving performance, assumes that a brief labeled recording can be obtained on-site; in true field deployments this may not always be feasible, and the uncalibrated performance floor warrants further characterization. The system is positioned as a triage screening tool, not a diagnostic instrument; any positive screen would require referral for confirmatory clinical testing. Long-term electrode stability under repeated use and cleaning, performance under realistic ambient noise conditions, and ease of use by community health workers without engineering backgrounds all require further study before deployment can be responsibly recommended. To prepare for real-world deployments, our upcoming field protocols will test two main factors: environmental noise and equipment wear. First, to block out background noise in classrooms or clinics, we are testing the system with passive 35 dB sound-attenuating earmuffs placed directly over the bone-conduction transducers. Second, to ensure the headset holds up over time, we are running a 100-cycle cleaning test using standard 70% isopropyl alcohol wipes to check for any mechanical wear or spring degradation.

A direct head-to-head comparison with low-cost or smartphone-based Otoacoustic Emission (OAE) devices was not performed. We have planned a comparative study as an

immediate next step to formally demonstrate CortexAI's advantages in field conditions. Specifically, this work will benchmark screening throughput and validate our system's unique ability to detect Auditory Neuropathy Spectrum Disorder (ANSO) – a condition where outer hair cell function yields a false 'pass' on OAE tests despite underlying neural conduction deficits.

In terms of real-world translation, CortexAI is positioned strictly as a triage screening tool rather than a diagnostic device, mapping to an FDA Class II (or regional equivalent) regulatory pathway for automated audiometric screeners. To integrate with existing community health workflows, future iterations of the software will aim to export screening outcomes directly into standardized digital referral templates. This design path ensures that high-risk individuals flagged by non-specialist operators in field conditions can be seamlessly triaged into existing healthcare networks for confirmatory diagnostic testing.

VII. CONCLUSION AND FUTURE WORK

CortexAI demonstrates that an end-to-end automated hearing screening system can be assembled from consumer-grade components for under \$500 USD while achieving 87.4% sensitivity and 96.0% specificity in a 30-participant pilot. The spring-loaded electrode array and bone-conduction stimulus pathway together eliminate the operator-skill requirements that have historically excluded objective hearing screening from resource-limited settings. The hybrid 1D CNN-LDA architecture with per-participant calibration provides a practical balance between representation-learning capacity and sample efficiency appropriate for small – cohort consumer biosensing deployments. Future work will expand the cohort to include neonatal and pediatric populations, the primary intended beneficiaries, evaluate field performance with community health workers in low-resource settings and explore on-device inference on the Raspberry Pi host to enable fully offline operation. Broader deployment will require regulatory pathway analysis, formal usability studies with nonspecialist operators, and integration with existing referral workflows in target health systems.

ACKNOWLEDGMENT

The author thanks the study participants who volunteered their time for data collection, the reviewers whose feedback strengthened this work, and the faculty members at the University of Washington, Seattle for their insightful discussions and feedback during presentation of this project.

REFERENCES

- [1] Joint Committee on Infant Hearing, "Year 2019 position statement: principles and guidelines for early hearing detection and intervention programs," *J. Early Hearing Detection and Intervention*, vol. 4, no. 2, pp. 1–44, 2019.
- [2] World Health Organization, *World Report on Hearing*. Geneva: WHO Press, 2021.
- [3] D. L. Jewett and J. S. Williston, "Auditory-evoked far fields averaged from the scalp of humans," *Brain*, vol. 94, no. 4, pp. 681–696, 1971.
- [4] C. A. Kothe and S. Makeig, "BCILAB: a platform for brain-computer interface development," *J. Neural Engineering*, vol. 10, no. 5, 056014, 2013.
- [5] R. T. Schirrmester et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [6] W. Swanepoel et al., "Smartphone hearing screening with integrated quality control and data management," *Int. J. Audiol.*, vol. 53, no. 12, pp. 841–849, 2014.
- [7] L. W. Norriss and D. S. Velenovsky, "Auditory neuropathy spectrum disorder: a review," *J. Speech Language Hearing Res.*, vol. 57, no. 4, pp. 1564–1576, 2014.
- [8] OpenBCI, "Cython biosensing board documentation," OpenBCI Inc., 2023. [Online]. Available: <https://docs.openbci.com>. [Accessed: Jul. 20, 2026].