scientific reports

OPEN

Behavior recognition technology based on deep learning used in pediatric behavioral audiometry

Wen Xie[1], Chunhua Li[1], Haisen Peng[1], Yuehui Liu[1], Zhilin Zhang[1], Xiaogang Cheng[2]✉ & Jiali Liu[1]✉

This study aims to explore the feasibility and accuracy of deep learning-based pediatric behavioral audiometry. The research provides a dedicated pediatric posture detection dataset, which contains a large number of video clips from children's behavioral hearing tests, encompassing various typical hearing test actions. A detection platform based on this dataset is also constructed, named intelligent diagnostic model of pediatric hearing based on *o*ptimized transformer (DoT); further, an estimation model of patient skeletal keypoints based on *o*ptimized transformer (POTR) was proposed to estimate human skeleton points. Based on this, the DoT approach was handled to perform posture recognition on videos of children undergoing behavioral hearing tests, thus enabling an automated hearing testing process. Through this platform, children's movements can be monitored and analyzed in real-time, allowing for the assessment of their hearing levels. Moreover, the study establishes decision rules based on specific actions, combining professional knowledge and experience in audiology to evaluate children's hearing levels based on their movement status. Firstly, we gathered image and video data related to posture in the process of conditioned play audiometry to test the hearing of 120 children aged 2.5 to 6 years old. Next, we built and optimized a deep learning model suitable for pediatric posture recognition. Finally, in the deployment and application phase, we deployed the trained pediatric posture recognition model into real-world application environments. We found that for children aged 2.5 - 4 years, the sensitivity of artificial behavior audiometry (0.900) was not as high as that of AI behavior audiometry (0.929), but the specificity of artificial behavior audiometry (0.824) and Area Under Curve (AUC) (0.901) was higher than that of AI behavior audiometry. For children aged 4–6 years, the sensitivity (0.943), specificity (0.947), and AUC (0.924) of artificial behavioral audiometry were higher than those of AI behavioral audiometry. The application of these rules facilitates objective assessment and diagnosis of children's hearing, providing essential foundations for early screening and treatment of children with hearing disorders.

Trial Registration: **Chinese Clinical Trial Registry: Registration number ChiCTR2100050416.**

Keywords Pediatric hearing loss, Pediatric behavioral audiometry, Transformer, Skeleton keypoints, Image processing, AI

Pediatric hearing loss is a common sensory deficit, with a incidence of 0.44% among children worldwide[1]. In China, the number of children under the age of 7 who suffer from hearing loss and speech disabilities has reached 800,000, with an increase of 30,000 annually[2]. The harm of hearing loss in children is more serious comparing to adults. For young children, hearing impairment not only affects their interpersonal communication, academic performance, and mental health, but also can cause language development delay[3]. Therefore, early detection and appropriate interventions are crucial for young children with hearing loss. In view of this, many countries, including China, have launched early hearing detection and intervention programs[4,5].

The pediatric hearing tests strategies include objective and subjective auditory tests. The objective tests are used to estimate the subjects' cochlear function and hearing threshold by testing their cochlear electrical activity and brainwave response during acoustic stimulation. However, objective hearing tests are not the real hearing tests no matter how reliable they are, but rather a physiologic process that occurs in a part of the auditory pathway after specific acoustic stimulations[6]. Compared with objective hearing tests, subjective hearing tests can

[1]Department of Otolaryngology, Head and Neck Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China. [2]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China. ✉email: chengxg@njupt.edu.cn; 475302391@qq.com

truly reflect the children's hearing condition and examine the sense of hearing as a whole, because it can not only evaluate the children's hearing threshold, but also evaluate the entire auditory pathway function, as well as the cognitive and social development by observing their response to sound in the process of testing[5,7]. Therefore, subjective hearing test method remains important for quantitative detection of hearing in infants and children[8]. However, many times, it is difficult to measure the children's hearing threshold solely through subjective hearing tests. Therefore, in order to improve the accuracy of pediatric hearing test results the cross-checking principle (combined with objective and subjective hearing tests) is often used in pediatric hearing tests.

The main subjective hearing test method used in infant and children is behavioral audiometry. It includes behavioral observative audiometry (BOA), conditioned orientation reflex (COR), visual reinforcement audiometry (VRA) and conditioned play audiometry (CPA). BOA is commonly used in infants within 6 months; VRA and COR are suitable for children aged 6 months to 2.5 years old. CPA is mainly applicable to children aged 2.5 to 6 years old[9]. The principle of BOA is testing the hearing of infants and young children by observing their natural responses to sound stimulus. For older children, COR and VRA are used by establishing a controlled conditioned reflex of stimuli through combining auditory sound stimulus with shiny moving toy signals to obtain their hearing threshold. The method of CPA is as follows: testers have children participate in a simple and fun game, teaching them to respond clearly and reliably to given sound stimulus, and testing their hearing based on their reactions in the game.

Although subject hearing test is the essential hearing assessment for infants and young children, it is often challenging. Firstly, it requires the testers to have extensive experience in testing, as training young infants to respond to sounds and judging whether they have response to sound stimuli is a challenging task. Secondly, it requires the children to have a certain level of cognitive and cooperative ability. Infants and young children, or children with autism or attention deficit hyperactivity disorder, are difficult to cooperate in the test. Thirdly, the subjective hearing tests are time consuming. In our hospital, the estimated time to complete each set of pediatric behavioral audiometry is at least 40 to 60 min.

To compensate for these shortcomings of pediatric behavioral audiometry, we plan to introduce artificial intelligence (AI) to assist in the testing[10]. Machine Learning algorithms can provide efficient and effective data analysis models to reveal the patterns of children' response in behavioral audiometry. Deep Neural Networks (DNN) is a multi-step feature learning technique where data is filtered through multiple cascades[11–13].

As far as we know, there is currently limited studies on the direct application of pose recognition technology in the field of pediatric behavioral audiology. For the first time, we have conducted interdisciplinary cooperation to apply human posture recognition technology to pediatric behavioral audiometry[14]. The true value data was captured from professional audiologist. The specific method of behavioral audiometry is that the expert audiologist to judge which behaviors of the subject responded to sound and which behaviors did not. Artificial intelligence techniques, such as transformer, diffusion, etc., have grown rapidly in recent years, which can be used to identify human behaviors[15–17]. Simultaneously, using high-definition cameras to capture participants' reactions during the test. Computers trained (deep learning) based on these classifications and collected video data combined with test results to generate algorithm models, which can then be applied in real-time in specific detection[18–20].

The aim of this study is to estimate the specificity and sensitivity of AI CPA in children between 2.5 and 6 years old. As the first part of a series of studies and due to the higher degree of cooperation in older children, we first attempted to include older children aged 2.5 to 6 years who underwent CPA. In addition, due to the uncertainty of the accuracy of AI application in pediatric hearing detection, we attempted to s to investigate the application value of AI in CPA, as subjects have clearer reactions during the CPA process, and are more easily defined and trained by AI technology. We realize that compared with older children, AI is more valuable for hearing test in infants and younger children, as their response to sound is more ambiguous and cannot be clearly defined. Our next research plan is to investigate the accuracy of AI behavioral audiometry in children under 2.5 years old.

The main contributions of this paper are shown as follows.

(1) Pediatric Posture Dataset: A dedicated dataset for pediatric pose detection in hearing diagnosis which contains 20,000 video clips with annotations from behavioral hearing tests in children.

(2) Intelligent diagnostic model of pediatric hearing based on optimized transformer (DoT): A detection platform that utilizes deep learning algorithms to perform pose recognition on videos of children undergoing behavioral hearing tests.

(3) Estimation model of patient skeletal keypoints based on optimized transformer (POTR): A part of DoT model that mainly accomplishes the extraction of skeletal joint points.

The structure of this paper includes the following sections: Sect. 1 presents a research framework for a deep learning-based algorithm for child pose recognition. Section 2 displays how we included the subjects, conducted hearing tests, and collected data. Section 3 primarily focuses on analyzing detection data and comparing experimental results. Section 4 mainly discusses the experimental results and provides prospects for future research directions. Section 5 provides a summary and statement.

Methods

Overview of approach application

This paper presents a research framework for a deep learning-based pediatric posture recognition algorithm focused on auditory diagnosis. The framework consists of several key steps. The overall diagram is shown in the Figs. 1 and 2.

Firstly, in the data collection phase, we gathered image or video data related to pediatric posture. To ensure dataset diversity, we included pediatric samples of different ages, genders, and poses. Secondly, in the data annotation phase, we annotated the collected dataset by labeling the pediatric posture information in each

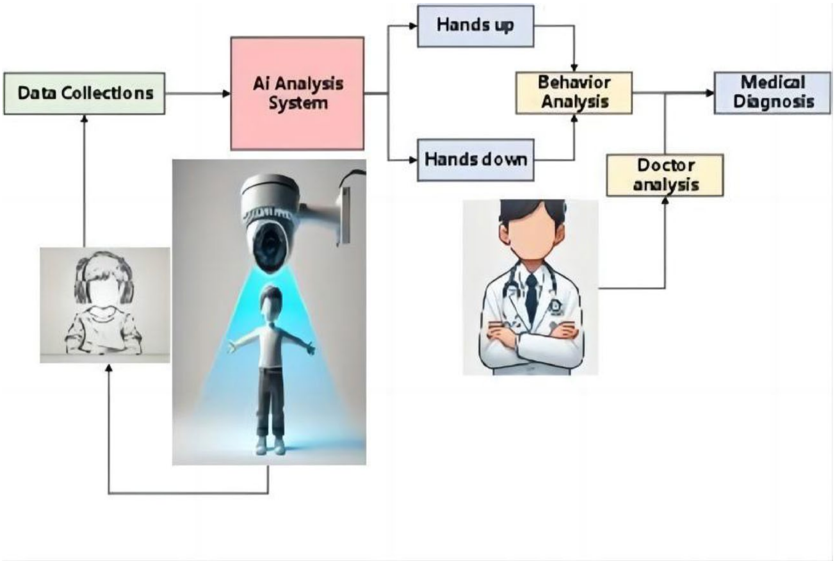

Fig. 1. Schematic of DoT approach application (DoT: Intelligent diagnostic model of pediatric hearing based on optimized transformer). This dataset is based on behavioral audiometry video data of 120 children included in this study.

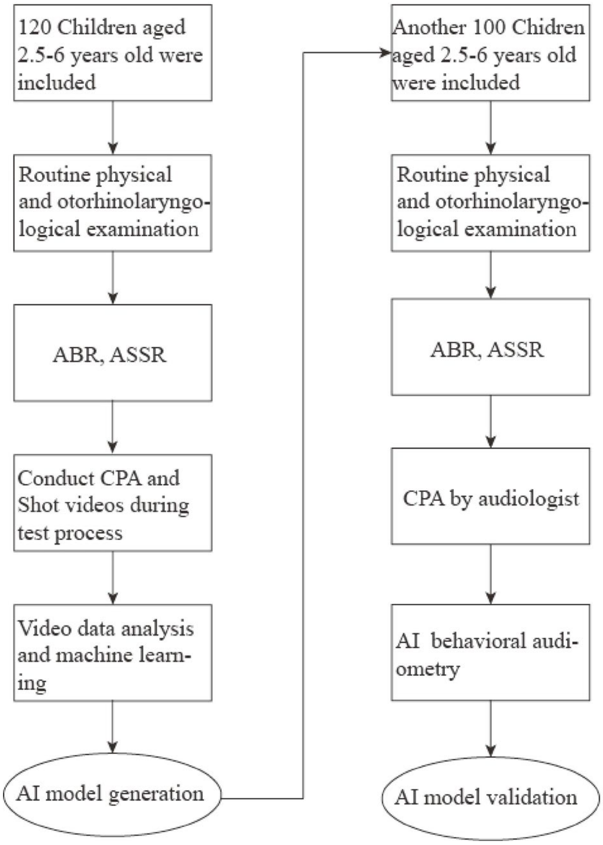
120 Children aged 2.5-6 years old were included
Routine physical and otorhinolaryngological examination
ABR, ASSR
Conduct CPA and Shot videos during test process
Video data analysis and machine learning
AI model generation
Another 100 Chidren aged 2.5-6 years old were included
Routine physical and otorhinolaryngological examination
ABR, ASSR
CPA by audiologist
AI behavioral audiometry
AI model validation

Fig. 2. Flow gram of the study.

sample. Annotations included information such as joint positions and posture angles, which were necessary for training and inference of the network model. We used professional annotation tools or manual annotation methods for this task.

Next, in the network model training phase, we designed and built a deep learning model suitable for pediatric posture recognition. We prepared a training set and a validation set by dividing the dataset into two parts for model training and validation. Using the annotated dataset, we trained the network model and optimized its parameters by using the backpropagation algorithm.

Subsequently, in the inference phase, we used the trained model to predict and infer new input data. For pediatric posture images or video data to be identified, we employed the trained model for inference. By applying the forward propagation algorithm, we input the data into the network model and obtained posture prediction results.

During the evaluation and optimization phase, we evaluated the model using a test set and calculated metrics such as recognition accuracy and recall rate. Based on the evaluation results, we adjusted and optimized the model parameters to improve its performance and generalization ability. Cross-validation and other methods were used to verify the stability and reliability of the model.

Finally, in the deployment and application phase, we deployed the trained pediatric posture recognition model into real-world application environments. We used the hearing test results tested by the audiologist as the 'gold standard", and compared the results of AI behavioral audiometry and artificial behavioral audiometry (behavioral audiometry conducted by audiologist). Depending on specific requirements, we integrated the model into an auditory diagnosis system to monitor pediatric posture in real-time.

Medical preparation

Patients

In this study we enrolled children aged 2.5 to 6 years who undergone CPA from January 2020 to June 2022. The study was performed in accordance with the ethical principles and approved by the Second Affiliated Hospital of Nanchang University Institutional Review Board. (reference number IIT-O-2021 -002). Written, informed consent was obtained from all children's guardians to participate in this study and to publish their processed images online.

Inclusion and exclusion criteria

Inclusion criteria Inclusion criteria were as follows: (1) Infants and young children aged 2.5 to 6 years. (2) Children who could successfully cooperate to complete CPA. (3) Children whose objective and subjective hearing test results were highly consistent (The hearing thresholds gap of objective and subjective hearing tests at each frequency was less than 10 dB).

Exclusion criteria Exclusion criteria comprised of: (1) Infants and young children who could not cooperate with subjective and objective hearing tests. (2) Subjects had unreliable hearing test results, such as inconsistent result between objective and subjective hearing tests, excessive noise during auditory steady-state response (ASSR) testing (the noise of reaction wave is greater than 30 nV), or poor ABR waveform repeatability.

After screening, 120 children (79 males and 41 females) were ultimately included for the AI model generation.

Test procedure

All children were assessed by the neuropsychological development scale for children aged 2.5 to 6 years old. All children underwent a detailed clinical interview. Clinical data, demographic information, past medical history and personal history were obtained. Routine physical examination and otorhinolaryngological examination were conducted in all subjects. All children underwent auditory brainstem responses (ABR), ASSR and CPA.

CPA procedure

Measurements of CPA were performed by an expert audiologist in a sound treated room. The subject was seated in a child chair facing the audiologist, with one of the parents sit behind their child. Then, the audiologist taught the child to respond to sound by stacking rings on a stick. After four or five demonstrations, conditioning could be established. Reinforcing conditioning by praising the child. By using this technique, frequency specific and ear specific air and bone hearing threshold could be obtained. The hearing test was performed using a diagnostic audiometer (Madsen Astera 2, Type 1066 Otometrics©, Taastrup, Denmark). The air conduction hearing threshold was tested using standard circumaural earphone, while the bone conduction hearing threshold was measured using bone vibrator. The given hearing stimulus were pure tone. Air conduction hearing thresholds were measured at 1k Hz, 2k Hz, 4k Hz, 8k Hz, 500 Hz, and 250 Hz; bone conduction hearing thresholds were assessed at 250 Hz to 8k Hz in sequence. Pure tone audiogram (PTA) was calculated by averaging air conduction thresholds at 0.5, 1, 2, and 4 kHz. The hearing loss levels were categorized into four grades: mild (26–40 dB HL), moderate (41–55 dB HL), moderate to severe (56–70 dB HL), severe (71–90 dB HL), and profound (> 90 dB HL)[21]. Audiogram patterns were classified into 5 types: ascending (the average threshold of 0.25–0.50 kHz was 20 dB higher than that of 4–8 kHz), descending (the average threshold of 4–8 kHz was 20 dB higher than that of 0.25–0.50 kHz), flat (all frequencies presented similar thresholds and hearing threshold was below 80 dB HL), profound (all frequencies showed similar threshold and hearing threshold was over 80 dB HL), and concave or convex type (average hearing degree of the mid-tone frequency was 20 dB higher than low and high frequencies)[22].

The air conduction test started at a frequency of 1k Hz, and the initial test sound level was lowered by 10 dB compared to the hearing level that the child just had response to. If the child responded to the sound, it was further lowered by 10 dB until the child had no response, then it was increased by 5 dB, and the test was repeated 5 times. Among them, the same hearing level which child had responses to was considered as the hearing threshold of pure tones at that frequency.

One small lamp, emitting yellow light, was placed on the wall behind the child, and the light was not too bright. During the test, kept the child focused and avoided turning their head to notice the lights on or off. Each time a sound was made, the audiologist observed the child's response. If the audiologist considered the child had no response to the sound, the tester would press the light button once to make the yellow light on once, and if the audiologist considered the child was responsive to sound, the audiologist would press the light button twice to make the yellow light on twice. s.

AI method of DoT

Pediatric posture dataset

During the detection process, a high-definition camera was placed behind the audiologist, capturing the facial expressions, head and neck, and body movements of the subjects. As shown in Fig. 3, it is the key points of human skeleton detected by POTR, showing a total of 18 points, arranged in numerical order as nose, neck, right shoulder, right elbow, right wrist, left shoulder, left elbow, left wrist, right hip, right knee, right ankle, left hip, left knee, left ankle, right eye, left eye, right ear, and left ear. Each skeleton key point of the subjects is represented by x, y, and z, where x and y represent the coordinates of key points of the bone, z represents the confidence value. So, when the subject appeared in the perspective of the visual sensor, the corresponding points were captured. That is to say, through the POTR method, the corresponding points of the main parts of the subject's face and body were be captured. The algorithm of this project set images with a set confidence value greater than 0.5 to be saved and those with a value less than 0.5 to be discarded. The retained images could be identified for corresponding behaviors through the algorithm designed in this project. Because according to the AI visual recognition technology, at least 64 samples with positive reactions were required to achieve statistically significant. Therefore, if too many children could not cooperate with behavioral audiometry, resulting in the positive samples size less than 64, additional cases would be added until the sample size needs were met.

DoT

This project collected a large number of children's behavior videos and classified them through audiologist's subjective evaluation to complete label calibration. In the process of the project, at least 120 samples were collected, with a cumulative video volume of 120 people * 20 min/person * 60 * 30 frames/second = 4,320,000 frames of images, and we classified the video frame images using raise and release labels.

We developed an intelligent diagnostic model for pediatric hearing based on DoT. The workflow of the Dot construction of this study is shown in Fig. 4. We used a video clip as input and applied the DoT model to detect whether the child's hearing was normal or not. First, we built an estimation model of patient skeletal keypoints based on POTR, which detected the skeletal keypoints of the child in each frame of the video.

Next, we calculated the relative distances and angles between the skeletal keypoints in each frame to determine whether the child is lifting or putting down their hands. If the child was classified as being in the same state for five consecutive frames, we recorded it as either "hands up" or "hands down." The alternating process of hands up and hands down was considered one response to sound. Finally, by comparing the number of responses to a set threshold, we determined whether the child's hearing was normal or not.

The structure of our POTR model is shown in Fig. 5. POTR took each frame of the video as input and first extracted multi-scale image features using a backbone network, which adapted to objects of different sizes. Then, position codes were concatenated with the multi-scale features. Through the interaction of features in the Transformer blocks, we obtained more robust representations. Finally, two prediction heads were used to predict the skeletal keypoints and their corresponding confidence scores.

To determine the child's movement status, we calculated the Euclidean distances between skeletal keypoints, including the distance between shoulders and elbows (2–3, 5–6), wrists and elbows (3–4, 6–7), and wrists and neck (1–4, 1–7). Additionally, we computed the angles between keypoints, specifically the elbow angle (2–3-

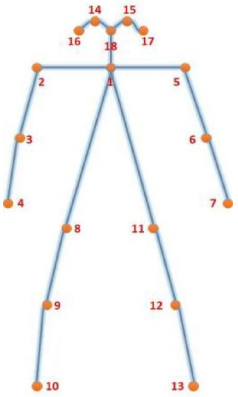

Fig. 3. Pediatric skeleton keypoints. (This image is generated using software PowerPoint version 2013, URL link: https://zenodo.org/records/15016815.).

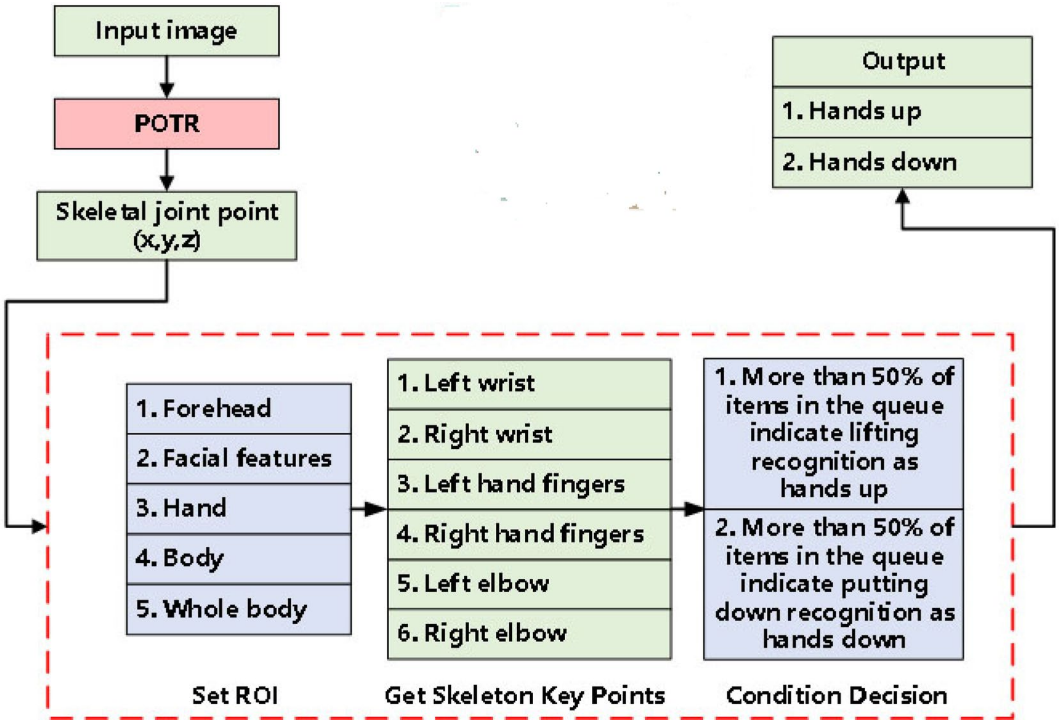

Fig. 4. Dot model for pediatric behavioral audiometry (DoT: Intelligent diagnostic model of pediatric hearing based on optimized transformer. POTR is a part of DoT model that mainly accomplishes the extraction of skeletal joint points and the remaining part accomplishes the estimation of pediatric posture). This dataset is based on behavioral audiometry image data from 100 children who included in this study.

4, 5–6- 7) and shoulder angle (1–2- 3, 1–5- 6). To facilitate subsequent state assessment, these angles were converted to cosine values. Since both distances and angles between skeletal keypoints were calculated, any set of keypoints with a confidence score below 0.5 was discarded.

We evaluated the child's state based on 10 sets of data. If more than 50% of the data indicated a lifting motion, we classified it as "hands up"; if more than 50% indicated a lowering motion, we classified it as "hands down." We defined a continuous sequence of five frames which were classified as the same action state (either hands up or hands down) as one complete hands-up or hands-down event. Alternating hands-up and hands-down events were counted as one response to sound.

After processing all video clips, we tallied the number of responses the child made to the sound. If this number exceeded the set threshold M, the DoT model concluded that the child had a normal response to sound; otherwise, there might be a risk of hearing issues, and further expert evaluation was recommended. The threshold M could be personalized based on the child's age and the duration of the video. In this study, M was set to 20.

Code availability

The codes are available at the website: https://zenodo.org/uploads/14913784?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjBmMjhmMDdjLTBjYjEtNDBjMC1iNjBlLTgwYmUzZjcxNDkwMSIsImRhdGEiOnt9LCJyYW5kb20iOiJlMzQ3ZDQzMjQxNzVjOWVkOGQ4MTY5ZWQxYmYyNjFkNiJ9.PSGZvIYQ7oyL2wR_hmiuo4F9vSb72Bk6FNIZAcJVQgEuMpCIWPrBpt50scbJ-McJR3lpY2X3i_8BD3fhhGpcvA.

Validation and performance evaluation strategy

A total of 100 subjects were included in the objective hearing test, with ages ranging from 2.5 to 6 years old. The inclusion and exclusion criteria were the same as in the previous screening stage of the subjects. After undergoing routine physical examination, otorhinolaryngological examination and intellectual and psychological assessment, they underwent artificial behavior audiometry and AI behavior audiometry. Compared each response result of subjects to sound stimulation in AI audiometry and artificial audiometry. The sensitivity and specificity of AI audiometry were calculated and compared with the sensitivity and specificity of artificial audiometry in subjects. The model output was a classification of the videos as one possible label: (1) responsive to sound, (2) unresponsive to sound.

The area under the receiver operating characteristic curves was also plotted to assess model performance. All analyses were conducted using SPSS version 25 for Windows. All the experiments are implemented under an image workstation, Windows NT OS, CPU Intel(R) Xeon(R) E5 - 2680 v4 @ 2.40 GHz, GPU NVIDIA 3080 ti. TensorFlow library was used for building and training models.

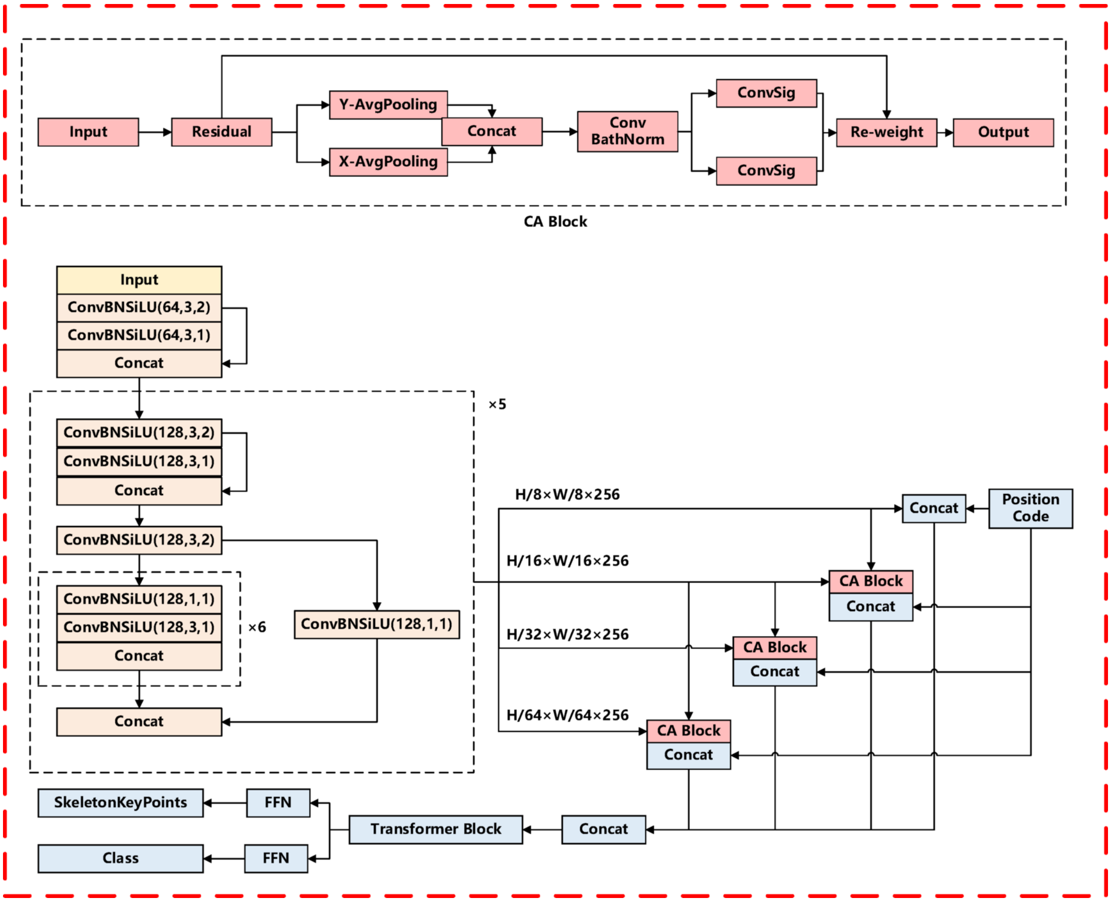

Fig. 5. POTR model (POTR: Estimation model of patient skeletal keypoints based on optimized transformer).

Diagnostic method	Sensitivity	Specificity	AUC
AI behavioral audiometry	0.929	0.778	0.862
Artificial behavioral audiometry	0.900	0.824	0.901

Table 1. Comparison between AI behavioral audiometry and artificial behavioral audiometry (Group A).

Results

Clinical characteristics

A total of 100 subjects (58 males and 42 females) were invited in this study for validation, including 36 children with normal hearing and 64 children with hearing loss. Among the 64 children with hearing loss, 12 were classified as mild, 13 as moderate, 9 as moderate to severe, 20 as severe, and 10 as profound hearing loss. In terms of the audiogram shape, the most common one was flat (51 children), followed by profound (10 children), ascending (2 children) and descending type (1 child). None of them suffered from concave and convex hearing loss. There were 46 children aged 2.5 to 4 years old, including 16 children with normal hearing and 30 children with hearing loss. There were 54 children aged 4 to 6 years old, including 20 children with normal hearing and 34 children with hearing loss.

Comparison of evaluation indicators between AI behavioral audiometry and artificial behavioral audiometry

Due to cognitive difference among children of different ages, we divided all subjects into group A (2.5 to 4-year-olds group) and group B (4 to 6-year-olds group) based on their ages and) and compared the sensitivity, specificity, and Area Under Curve (AUC) of AI behavioral audiometry and artificial behavioral audiometry (behavioral audiometry done by audiologist) between group A and B. For Group A, the sensitivity of artificial behavior audiometry (0.900) was not as high as that of AI behavior audiometry (0.929), but the specificity of artificial behavior audiometry (0.824) and AUC (0.901) was higher than that of AI behavior audiometry. For Group B, the sensitivity (0.943), specificity (0.947), and AUC (0.924) of artificial behavioral audiometry were higher than those of AI behavioral audiometry. The results are shown in Tables 1 and 2, the detection results of this experiment are shown in the Fig. 6.

Diagnostic method	Sensitivity	Specificity	AUC
AI behavioral audiometry	0.914	0.895	0.873
Artificial behavioral audiometry	0.943	0.947	0.924

Table 2. Comparison between AI behavioral audiometry and artificial behavioral audiometry (Group B).

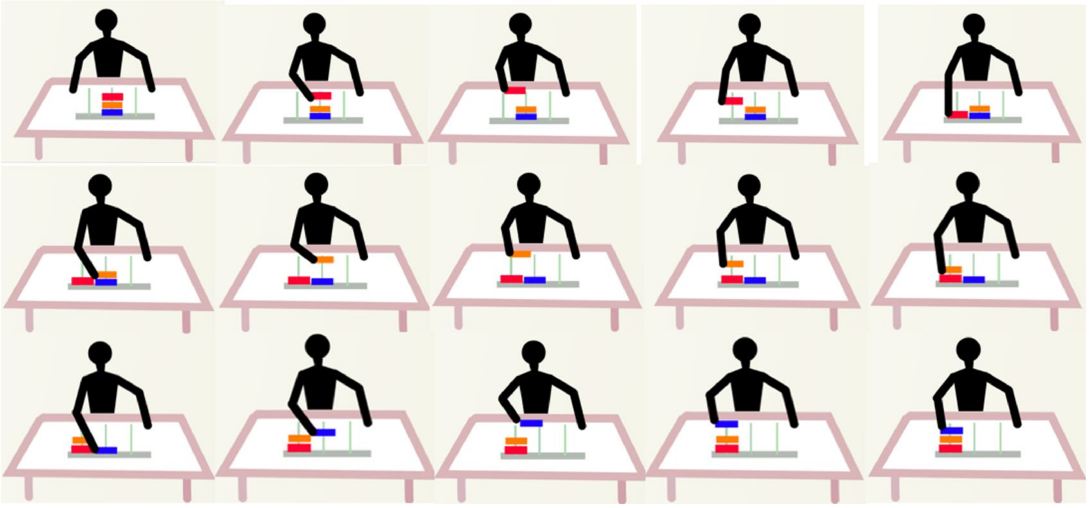

Fig. 6. Detection results of PORT approach proposed in this paper.

Discussion

To our knowledge, few studies have reported the application of deep learning-based approaches and technologies in pediatric behavior audiometry. Deep learning is an emerging algorithm in the biometrics field, enabling us to tackle the covariates and produce highly accurate results[23]. Previous studies have reported that a few validated automated procedures had been used for audiology researches. However, these studies typically only focused on using DNN models to judge audiogram results, including ABR, air and bone conduction audiometry results[24–28]. Currently, no specific and direct machine learning–based audiometry approaches have been developed for children or hard-to-test populations. In this study, we conducted interdisciplinary collaboration and applied the existing deep learning-based gait recognition approach in the process of pediatric behavior audiometry, in order to evaluate the clinical application value of this technology. It is gratifying that the accuracy of AI applied in CPA in our study was significantly high, with high sensitivity and specificity. This suggests that our models performed comparably with a human expert, or even superior, in conducting CPA. In the separate real-world testing cohort, robust performances were also observed in predicting subjects' hearing level with small variability, suggesting that the models had good generalizability. On the other hand, our study paves the way for AI applied in more complex hearing test task among younger children in future. In the next step of our work, we will try to apply this technology in children under 2.5 years old to explore the application value of this technology in other types of behavioral audiometry.

As we mentioned previously, pediatric behavioral hearing test is a big challenge. It requires professionally trained and experienced audiologists, who are very scarce in grassroots hospitals, particularly in developing countries[29]. In addition, due to the time-consuming nature of these examinations, many hospitals are unwilling to conduct this test, even in tertiary hospitals. Our study could facilitate to solve these problems through the following aspects. Firstly, this technology contributes to improve the efficiency and accuracy of pediatric behavioral hearing test, making more hospitals more willing to carry out this examination, thereby enabling more children with hearing impairment to be diagnosed in a timely manner. In addition, our ultimate goal is to truly apply AI to assist audiologist in conducting this test, and ultimately achieve remote and self-administered hearing test by using mobile healthcare apps and online websites. This automated approach contributes to children undergo hearing tests by their parents, and has the potential to increase accessibility and scalability without the direct involvement of professionals, avoiding multiple long-distance trips to hospitals for testing. After testing, hearing test result dada can be analyzed preliminarily by apps or computer software, or sent to the doctors, who can make an accurate hear assessment. The onset of the COVID- 19 pandemic has further emphasized the importance of self-testing approaches[30,31].

DNN models become more and more accurate when they process large scale data, which enables them to outperform many classical machine learning models. Therefore, in the model generation stage, we recruited many children for auditory tests. However, the accuracy of AI hearing test results is based on the accuracy of artificial hearing test results. Therefore, in order to improve the accuracy of artificial auditory test results, we conducted both subjective and objective hearing tests, due to the cross-check principle, which means that the results of a single test are cross checked by an independent test, and a definite diagnosis can only be made after mutual verification of multiple test results[32]. Based on the large amount of image data, we developed a novel algorithm designed for pediatric behavioral audiometry. We built the corresponding algorithm program through POTR, extracted the corresponding data and made the corresponding preprocessing rules. Finally, we verified the test results repeatedly until the goal of this project was achieved, thus reaching the purpose of reducing the cost and standardizing the detection process at this stage. Due to the novelty of this research field, deep learning model engineering is still a sandbox with no unified optimal path. Continuous studies on model constructs and hyperparameters will contribute to improve the accuracy of model performance, ultimately achieving or even exceeding human standards.

Our research is a successful application of deep learning approaches in audiology. However, there are still many challenges and urgent issues to be addressed in this field. For example, in practical applications, this program may experience short-term misjudgment affected by the range of coordinate axis values. One reason for this phenomenon is that the amplitude of the subjects' movements is too high, making data capture difficult. Another reason is that the values such as threshold in the code is unable to meet the current sample tester's needs, and need further improvements, with a focus on the data. In addition, only the lifting and lowering movements can be accurately detected, but the movements of subjects using their left or right hands for operations cannot be effectively identified by AI technology. Therefore, there still exists room for improvement in this field, making the results of children's hearing measurements more specific and easier to be analyzed accordingly.

In summary, although this AI technology can be able to well used in CPA, there still exist some issues and a risk of instability. It is worth further exploration in our future research. At the same time, we can also consider not only identifying the two states of lifting and lowering to make A → B → A judgment. It is also possible to improve recognition accuracy from identifying lifting movements to left-handed or right-handed actions, enabling testers to better assess the degree of hearing impairment among child subjects.

Conclusion

Hearing loss is and will continue to be a significant public health issue. Reengineering the process of hearing test with a machine learning innovation may make the audiologist services available to a large number of children with hearing loss. Our results suggest that deep learning may be a transformative technology that enables automatic and accurate pediatric behavioral hearing test.

Data availability

Data such as the method of obtaining code is provided within the manuscript. Due to privacy or ethical restrictions, data including personal information and videos of subjects are not publicly available.

Received: 23 December 2024; Accepted: 4 April 2025
Published online: 05 May 2025

References

1. Organization, W. H. Childhood hearing loss: strategies for prevention and care. 1st ed. *Geneva (Switzerland): WHO.* (2016).
2. Leading Group of the Second China National Sample Survey on Disability, N. Main Data bulletin of the second national sampling survey on persons with disabilities. *Chin. J. Rehabil. Theory Pract.* **12**, 1013 (2006).
3. Osei, A. O., Larnyo, P. A., Azaglo, A., Sedzro, T. M. & Torgbenu, E. L. Screening for hearing loss among school going children. *Int. J. Pediatr. Otorhinolaryngol.* **111**, 7–12. https://doi.org/10.1016/j.ijporl.2018.05.018 (2018).
4. American Academy of Pediatrics & o. I., J. C. Year 2007 position statement: principles and guidelines for early hearing detection and intervention programs. *Pediatrics* **120**, 898–921. https://doi.org/10.1542/peds.2007-2333 (2007).
5. Commission, E. G. & o. U. N. H. S. O. T. N. H. A. F. P. Guideline for T.e early diagnostic evaluation A.d intervention O. hearing loss in infants. *Chin. J. Otorhinolaryngol. Head Neck Surg.* **53**, 181–188. https://doi.org/10.3760/cma.j.issn.1673-0860.2018.03.004 (2018).
6. Psarommatis, I., Valsamakis, T., Raptaki, M., Kontrogiani, A. & Douniadakis, D. Audiologic evaluation of infants and preschoolers: a practical approach. *Am. J. Otolaryngol.* **28**, 392–396. https://doi.org/10.1016/j.amjoto.2006.11.011 (2007).
7. Prekopp, P. et al. Recognition and complex diagnostics of functional hearing loss. *Orv. Hetil.* **164**, 283–292. https://doi.org/10.1556/650.2023.32712 (2023).
8. Ricalde, R. R., Chiong, C. M. & Labra, P. J. P. Current assessment of newborn hearing screening protocols. *Curr. Opin. Otolaryngol. Head Neck Surg.* **25**, 370–377 (2017).
9. Sabo, D. L. The audiologic assessment of the young pediatric patient: the clinic. *Trends Amplif.* **4**, 51–60. https://doi.org/10.1177/108471389900400205 (1999).
10. Sheng, B. et al. A markless 3D human motion data acquisition method based on the binocular stereo vision and lightweight open pose algorithm. *Measurement* **225**, 113908. https://doi.org/10.1016/j.measurement.2023.113908 (2024).
11. Wang, J. P. et al. Open pose mask R-CNN network for individual cattle recognition. *IEEE Access.* **11**, 113752–113768. https://doi.org/10.1109/access.2023.3321152 (2023).
12. Saiki, Y. et al. Reliability and validity of openpose for measuring hip-knee-ankle angle in patients with knee osteoarthritis. *Sci. Rep.* https://doi.org/10.1038/s41598-023-30352-1 (2023).
13. Liu, Z. S. et al. Recent progress in transformer-based medical image analysis. *Comput. Biol. Med.* https://doi.org/10.1016/j.compbiomed.2023.107268 (2023).
14. Han, W. Y. et al. Design and application of the transformer base editor in mammalian cells and mice. *Nat. Protoc.* https://doi.org/10.1038/s41596-023-00877-w (2023).
15. He, K. et al. Transformers in medical image analysis. *Intell. Med.* **3**, 59–78. https://doi.org/10.1016/j.imed.2022.07.002 (2023).
16. Su, J. L. et al. RoFormer: enhanced transformer with rotary position embedding. *Neurocomputing* https://doi.org/10.1016/j.neucom.2023.127063 (2023).
17. Qiang et al. Group DETR: Fast DETR training with group-wise one-to-many assignment. *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 6633–6642, (2023).
18. Adrian Bulat, R., Guerrero, B. & Tzimiropoulos, G. Martinez Fs-DETR: Few-shot detection transformer with prompting and without re-training. *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 11793–11802, (2023).

19. Feng Li et al. Lite DETR: An interleaved multi-scale encoder for efficient detr. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18558–18567, (2023).
20. Nakano, N. et al. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Front. sports Act. living* **2**, 50. https://doi.org/10.3389/fspor.2020.00050 (2020).
21. Organization, W. H. World report on hearing. *Available from*: (2021). https://www.who.int/publications/i/item/world-report-on-hearing.
22. Editorial Board of Chinese Journal of Otorhinolaryngology Head and Neck Surgery. Society of otorhinolaryngology head and neck surgery, C. M. A. Guideline of diagnosis and treatment of sudden deafness (2015). *Chin. J. Otorhinolaryngol. Head Neck Surg.* **50**, 443–447 (2015).
23. Wasmann, J. W., Pragt, L., Eikelboom, R. & Swanepoel, W. Digital approaches to automated and machine learning assessments of hearing: scoping review. *J. Med. Internet. Res.* **24**, e32581. https://doi.org/10.2196/32581 (2022).
24. Mahomed, F., Swanepoel de, W., Eikelboom, R. H. & Soer, M. Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear.* **34**, 745–752. https://doi.org/10.1097/01.aud.0000436255.53747.a4 (2013).
25. Song, X. D. et al. Fast, continuous audiogram Estimation using machine learning. *Ear Hear.* **36**, e326–335 https://doi.org/10.1097/aud.0000000000000186 (2015).
26. Barbour, D. L. et al. Online machine learning audiometry. *Ear Hear.* **40**, 918–926. https://doi.org/10.1097/aud.0000000000000669 (2019).
27. McKearney, R. M. & MacKinnon, R. C. Objective auditory brainstem response classification using machine learning. *Int. J. Audiol.* **58**, 224–230. https://doi.org/10.1080/14992027.2018.1551633 (2019).
28. Crowson, M. G. et al. AutoAudio: deep learning for automatic audiogram interpretation. *J. Med. Syst.* **44**, 163. https://doi.org/10.1007/s10916-020-01627-1 (2020).
29. Bhutta, M. F., Bu, X. K., de Muñoz, P. C., Garg, S. & Kong, K. Training for hearing care providers. *Bull. World Health Organ.* **97**, 691–698. https://doi.org/10.2471/blt.18.224659 (2019).
30. Manchaiah, V., Eikelboom, R. H., Bennett, R. J. & Swanepoel, W. International survey of audiologists during the COVID-19 pandemic: effects on the workplace. *Int. J. Audiol.* **61**, 265–272. https://doi.org/10.1080/14992027.2021.1937348 (2022).
31. Saunders, G. H. & Roughley, A. Audiology in the time of COVID-19: practices and opinions of audiologists in the UK. *Int. J. Audiol.* **60**, 255–262. https://doi.org/10.1080/14992027.2020.1814432 (2021).
32. Lu, T. M., Wu, F. W., Chang, H. & Lin, H. C. Using click-evoked auditory brainstem response thresholds in infants to estimate the corresponding pure-tone audiometry thresholds in children referred from UNHS. *Int. J. Pediatr. Otorhinolaryngol.* **95**, 57–62. https://doi.org/10.1016/j.ijporl.2017.02.004 (2017).

Author contributions

Xiaogang Cheng and Jiali Liu conceived and designed the study. Wen Xie, Xiaogang Cheng, and Chunhua Li planned and conducted human tests. Wen Xie, Xiaogang Cheng, Haisen Peng and Yuehui Liu wrote the initial manuscript draft. All authors involved in proofreading and editing of the final version of the manuscript.

Funding

The National Natural Science Foundation of China (No.62371253, No. 52278119). The Science and Technology Department of Jiangxi Province (No. 20202BBGL73017).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.C. or J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.