Abnormal brain activation during speech perception and production in children and adults with reading difficulty

Check for updates

Yang Fu[1,2], Xiaohui Yan [1,2], Jiaqi Mao[3], Haibin Su[4] & Fan Cao[1,2] ✉

Reading difficulty (RD) is associated with phonological deficits; however, it remains unknown whether the phonological deficits are different in children and adults with RD as reflected in foreign speech perception and production. In the current study, using functional Near-infrared spectroscopy (fNIRS), we found less difference between Chinese adults and Chinese children in the RD groups than the control groups in the activation of the right inferior frontal gyrus (IFG) and the dorsolateral prefrontal cortex (DLPFC) during Spanish speech perception, suggesting slowed development in these regions associated with RD. Furthermore, using multivariate pattern analysis (MVPA), we found that activation patterns in the left middle temporal gyrus (MTG), premotor, supplementary motor area (SMA), and IFG could serve as reliable markers of RD. We provide both behavioral and neurological evidence for impaired speech perception and production in RD readers which can serve as markers of RD.

Reading difficulty (RD) is a common neurodevelopmental disorder with a prevalence of 5–17% across language systems[1–3], which is characterized by difficulties in accurate and fluent reading, despite appropriate cognitive ability and instruction[4]. Previous studies have shown that challenges in acquiring proficient reading and spelling skills persist into adulthood[5–7]. Deficits in phonological processing have been well-documented as a signature of RD, which is reflected in speech processing in early life[8]. Abnormal brain activities during speech processing might serve as markers of RD and facilitate early diagnosis because brain responses to speech are automatic which can even be collected in newborns. A recent longitudinal study[9] has shown that RD can be successfully predicted through lower functional connectivity between the left primary auditory cortex and the left planum temporale at the age of 5. Several studies have documented the predictive relationship between early childhood speech perception and later literacy skills[10–14].

Phonological deficits could be reflected in both speech perception and production in early life[15,16]; however, relatively fewer studies have concerned speech production in RD, presumably because it is challenging to collect evidence in infants/young children for speech production. An alternative is to examine foreign speech perception and production in older children and adults with RD and to test whether brain activity patterns during these tasks can be reliable markers of RD, which would also facilitate the diagnosis of RD.

A substantial amount of research has documented deficits in speech perception in individuals with RD, for example, a significantly lower discrimination score between minimum contrast syllables (e.g., /ba/-/da/) than

age controls and reading controls[17–19]. Individuals with RD were found to have less sharply defined phoneme boundaries in the categorical perception of the synthetic syllable continuum[20].

Neuroimaging studies have also provided evidence for speech perception deficits in individuals with RD. For example, reduced brain activation in the left prefrontal cortex was found in children with RD compared to control children during auditory rhyming judgment in English[21], and reduced activation was also found in the superior temporal cortex during auditory speech processing[22–25]. An fMRI study with multivoxel pattern analysis showed less distinct activation patterns in the bilateral superior temporal regions for /bA/ and /dA/ in beginning readers with a high familial risk of RD[26], suggesting a low quality of phonemic representations.

Evidence of speech perception deficits in RD also comes from EEG research. For example, smaller amplitude and longer latency of mismatch negativity/response (MMN/MMR) and/or late discriminative negativity (LDN) elicited by deviant stimuli have been observed over frontocentral sites in children and adults with RD[27], and in 20-month-old infants who have high risk of RD[28], suggesting reduced phoneme discrimination. Gu and Bi[29], in their review and meta-analysis, reported persistent speech perception deficits (reduced MMN amplitude) in individuals with RD in alphabetic languages with an even larger effect size in adults than in children with RD. It suggests that speech perception deficits do not disappear with age in RD, which is consistent with the finding that adults with RD continue to show phonological deficits even though they may have improved word decoding accuracy[30].

[1]Department of Psychology, The University of Hong Kong, Hong Kong, China. [2]State Key Lab of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China. [3]Basque Center on Cognition, Brain and Language, BCBL, San Sebastian, Spain. [4]Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. ✉e-mail: fancao@hku.hk

In Chinese RD, speech perception deficits have also been documented, such as reduced brain activation in the dorsal left IFG during an auditory rhyming judgment task[31], and reduced MMN elicited by deviant stimuli in the frontal sites[32] in children with RD compared to control children. In a Mandarin Chinese tone categorical perception study, it was also found that children without RD but not children with RD showed greater MMN peak amplitude over frontal sites for cross-category deviants than within-category deviants[33], suggesting reduced categorical perception in Chinese children with RD compared to children without RD. However, no studies have concerned differences between adults and children in speech perception deficits in Chinese RD.

Speech perception and production are tightly connected because the same phonological representations are involved in the two processes[34]. The poorly specified phonological representation in individuals with RD makes it challenging to formulate a speech-motor plan during production[35]; therefore, resulting in a slower speaking rate and increasing pauses in individuals with RD[36]. In a longitudinal study, toddlers who were later identified as having RD at school age showed slower speaking rates, longer pauses and reduced production of syllables per speaking turn at ages 2 and 3 compared to not-at-risk peers[35], suggesting early difficulties in articulation planning for multi-syllabic utterances in children with RD. As children with RD reach school age, their difficulties with multi-syllabic words persist. They continue to struggle with repeating multi-syllabic words and nonwords[37], or producing the names of multi-syllabic items[38] compared to both age-matched control children and reading-matched control children.

Remarkably, the deficits in multi-syllabic production extend into adulthood. Adults with RD demonstrate slower rates and more errors than controls when repeating multi-syllabic phrases rapidly[39]. Other than deficits in multi-syllabic production, it has also been found that information about articulatory movements for specific phonemes is less accessible in adults with RD[40], because they showed difficulty in matching phonemes with drawings of the articulators positions when making a specific sound. However, one study found no deficits in articulatory speed in adults with RD, despite deficits in phonemic awareness[41], whereas another more recent study found slower articulatory rate in adults with RD, regardless of their comorbidity with motor coordination disorder[42].

In summary, previous studies have focused on speed, pauses, and errors in speech production and more sophisticated measures of speech quality such as the voice onset time (VOT) of consonants and the frequency formants of vowels have not been used. VOTs and vowel frequency formants are objective acoustic analyses that quantitatively describe the quality of speech sounds. VOT refers to the time between the stop burst and the onset of the voicing for a consonant. A vowel frequency formant is a concentration of acoustic energy around a particular frequency in the speech wave. According to Marchetti et al.[42], each vowel has three formants. The first formant (F1) is inversely related to vowel height. The second formant (F2) is related to the degree of backness of a vowel. We distinguish one vowel from another by the differences in these formants. These acoustic analyses of speech sound would provide a more accurate and detailed measure of phonological deficits associated with RD. Furthermore, no studies have examined brain activity patterns during speech production in individuals with RD.

Deficits have also been found in individuals with RD during foreign speech perception and production. Ylinen et al.[43] found weaker MMN amplitude over the right temporal sites in RD children than in children without RD for second-language words but not for native words, suggesting specific deficits in perceiving second-language speech. Soroli & Ramus[44] found that French adults with RD showed deficient foreign lexical stress discrimination but normal foreign stress and plosive production. However, in Bouhon et al.'s[45] study, French adults with RD showed difficulties in producing English vowel contrasts. Specifically, the duration difference between /iː/-/ɪ/ was smaller than in adults without RD. Taken together, there might be specific deficits in foreign speech perception and production that are not evident in native languages.

One important factor that needs to be taken into account in foreign speech perception and production is the language distance to the native language. The phonological similarities between native and foreign languages have been reported to influence foreign speech production[46,47]. When the foreign speech sounds are very contrastive to the native speech sounds, it may present a greater challenge for perception and production. Therefore, in the current study, we chose to use Spanish which is contrastive to Chinese in speech sounds, so that we may have a greater chance to identify deficits in RD.

One research gap in the literature is whether and how phonological deficits in RD are reflected in foreign speech perception and production, and how it differs in children and adults with RD. In the current study, we examined foreign speech production quality and brain activities during foreign speech perception and production in Chinese children and adults with RD using fNIRS. fNIRS is especially suitable for studying speech production, because of its relative tolerance with motion artifacts. For the measures of VOT and vowel formants, we expected children to be more similar to a native Spanish speaker than adults, and readers without RD to be more similar than readers with RD. For the fNIRS data, we expected that individuals with RD show different brain activation patterns compared to individuals without RD during foreign speech perception and production, especially in the fronto-tempo-parietal regions. Moreover, we expected different age effects in individuals with RD compared to those without RD. For example, in some regions, there may be greater differences between adults and children in readers without RD than those with RD and vice versa. Finally, using a machine learning approach, we examined whether brain activity patterns during speech perception and production could serve as reliable markers of RD. We expected activation patterns in the fronto-tempo-parietal regions to be strong predictors of RD classification.

Results

Behavioral assessments

We compared chronological age and Raven score between RD readers and readers without RD separately for children and adults. No significant difference was found in chronological age between children with RD and children without RD ($t(41) = -2.01$, $p = 0.072$), neither between adults with RD and adults without RD ($t(40) = 0.05$, $p = 0.958$). There was no significant difference in Raven scores between children with RD and children without RD ($t(38) = 1.81$, $p = 0.078$), while adults with RD scored lower on Raven than adults without RD ($t(27.48) = 4.91$, $p < 0.001$). We did sub-group analyses with matched Raven in RD adults ($N = 14$) and adults without RD ($N = 12$) for the behavioral tests and fNIRS data, and we found similar results as for the whole sample. Therefore, we report results from the whole sample. Results from the sub-group analyses are reported in the supplementary results and Supplementary Table 1. Furthermore, we added Raven as a covariate in all statistics. Table 1 presents the results of the behavioral tests and more detailed reports on the behavioral tests are presented in the supplementary results and Supplementary Figs. 3 & 4.

Phonetic analyses for frequency formants

Figure 1 shows normalized vowel charts for the 5 vowels in each group. The vowel chart of children without RD showed relatively clear boundaries and less overlap across the vowels. The other groups of participants showed a considerable degree of overlap across the vowels. A Group by Age ANCOVA with Raven as a covariate revealed a significant main effect of Age ($F(1,84) = 12.10$, $p_{FDR} = 0.003$, $\eta_p^2 = 0.13$) for the averaged distance to the native speaker across all 5 vowels. Children were more similar to the native speakers than adults. The main effect of Group ($F(1,84) = 2.20$, $p_{FDR} = 0.430$, $\eta_p^2 = 0.03$) and the Age by Group interaction were not significant ($F(1,84) = 2.22$, $p_{FDR} = 0.140$, $\eta_p^2 = 0.03$). In order to understand whether the main effect of age was driven by individuals without RD, individuals with RD, or both, we conducted a simple effect analysis. We found that children without RD had a greater similarity to the model speaker than adults without RD ($t(25.25) = -3.66$, $p = 0.001$), but there was no difference between RD children and RD adults ($t(38) = -1.52$, $p = 0.14$), suggesting that the main effect of age was mainly driven by readers without

Table 1 | Demographic information and results on behavioral tests

Group	Control children	Children with RD	Control adults	Adults with RD
N (Male)	24 (9)	20 (12)	23 (10)	20 (9)
Age in years	10.6 ± 0.6	11.0 ± 0.7	19.7 ± 1.5	19.6 ± 1.0
Raven	107.3 ± 13.2	100.7 ± 8.6	129.9 ± 9.2	109.6 ± 16.0[†††]
Chinese Character Naming (150)	126.4 ± 8.1	91.1 ± 13.2[***]	143.1 ± 2.8	133.2 ± 8.9[†††]
Sentence Reading Fluency (3505)	1100.9 ± 238.1	621.9 ± 171.2[***]	1714.8 ± 419.0	741.0 ± 204.1[†††]
One-minute Character Naming (regular) (150)	70.2 ± 16.2	42.3 ± 17.9[***]	103.1 ± 14.7	69.1 ± 9.0[†††]
One-minute Character Naming (irregular) (150)	46.0 ± 14.2	22.8 ± 14.8[***]	83.6 ± 18.0	47.1 ± 10.8[†††]
Initial Sound Deletion (30)	18.6 ± 7.3	9.8 ± 7.8[**]	26.3 ± 2.9	12.4 ± 9.6[†††]
Pseudowords Rhyming (40)	32.9 ± 5.1	27.4 ± 4.9[**]	36.1 ± 2.4	31.1 ± 3.2[†††]
Homophonic Morpheme (30)	27.2 ± 1.8	20.6 ± 5.2[***]	29.6 ± 0.6	28.5 ± 1.7[††]
Homographic Morpheme (30)	22.7 ± 3.2	17.4 ± 3.4[***]	26.9 ± 2.1	26.1 ± 2.3
Character Correction (60)	38.0 ± 7.4	31.0 ± 6.3[**]	49.8 ± 5.0	42.3 ± 6.5[††]
Delayed Copy (30)	22.2 ± 3.5	19.8 ± 2.6[*]	26.9 ± 1.7	24.6 ± 2.5[†††]
Forward Digit Span	8.2 ± 1.1	6.9 ± 1.1[**]	9.5 ± 1.1	9.3 ± 1.6
Backward Digit Span	5.1 ± 1.1	3.5 ± 0.8[***]	7.6 ± 1.4	5.8 ± 2.0[††]
RAN (digits) (reaction time in seconds)	31.6 ± 6.2	40.7 ± 8.0[***]	24.2 ± 2.8	28.2 ± 3.0[†††]
RAN (pictures) (reaction time in seconds)	51.7 ± 11.0	62.4 ± 13.7[**]	40.6 ± 4.9	47.9 ± 6.9[†††]

[*]$p < 0.05$, [**]$p < 0.01$, [***]$p < 0.001$, significance between the two child groups; [††]$p < 0.01$, [†††]$p < 0.001$, significance between the two adult groups. Numbers in the parenthesis following the name of each test is the full score for each test.
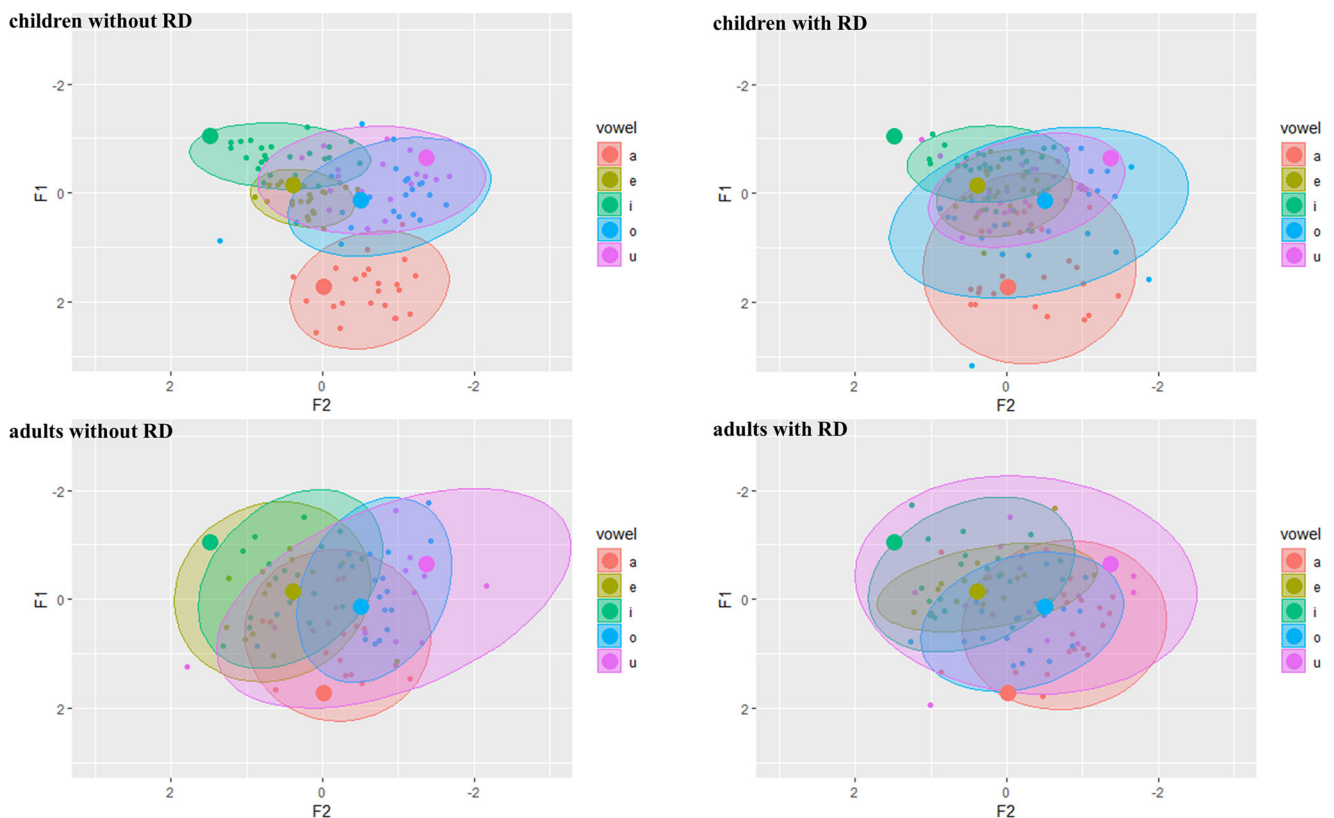

Fig. 1 | Normalized vowel charts for children without RD (upper left), children with RD (upper right), adults without RD (lower left), and adults with RD (lower right). The bigger dots in each graph are normalized vowel formant frequencies for the native model speaker while the small dots represent individual participants.

RD. Furthermore, simple effect analysis also revealed that children without RD had a greater similarity to the native speaker than children with RD in the vowels' frequency formants across all five vowels ($t(40) = -4.69$, $p < 0.001$, Fig. 2a), and no significant difference was found between adults with RD and adults without RD ($t(40) = -0.29$, $p = 0.77$).

Phonetic analyses for VOTs

We analyzed the voice onset time for /b/ and /d/ in the 5 Spanish words (i.e., difícil, dado, brazo, bueno, bebe). We found that every group showed a significant difference from the model speaker for both /b/ and /d/ (Fig. 2b); however, neither the main effect of Age ($F(1,82) = 1.10$, $p_{FDR} = 0.30$,
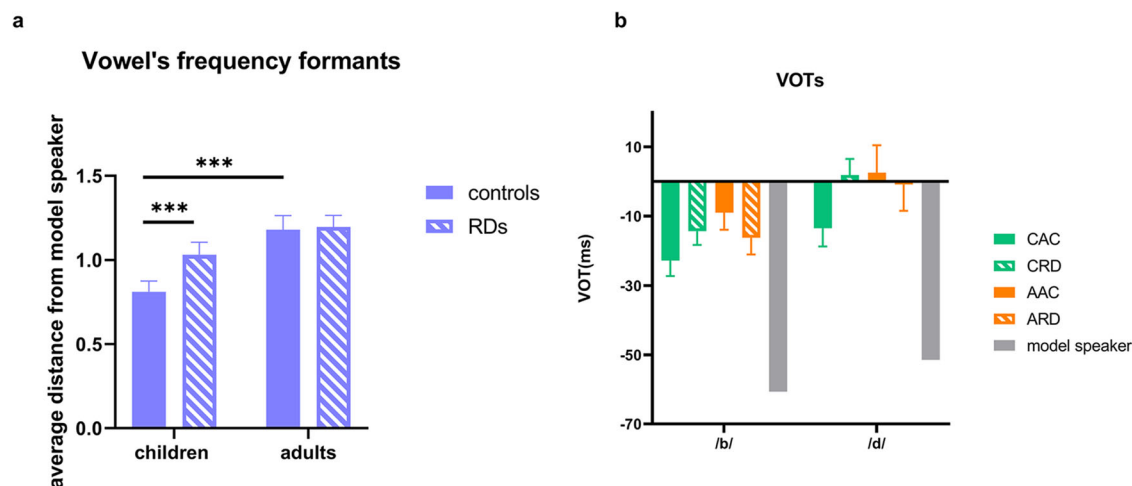
Fig. 2 | Results of vowels' frequency formants and consonants' VOT in each group. Vowels' frequency formants are presented in (**a**) and consonants' VOT results are presented in (**b**). No group differences were found for VOTs. CAC age-control children, CRD children with RD, AAC age-control adults, ARD adults with RD. Error bars depict SEM. *$p < 0.05$, **$p < 0.01$.

$\eta_P^2 = 0.01$ for /b/, $F(1,65) = 3.24$, $p_{FDR} = 0.12$, $\eta_P^2 = 0.05$ for /d/) nor the main effect of Group ($F(1,82) = 0.03$, $p_{FDR} = 0.97$, $\eta_P^2 < 0.001$ for /b/, $F(1,65) = 0.002$, $p_{FDR} = 0.97$, $\eta_P^2 < 0.001$ for /d/) reached statistical significance. Furthermore, the interaction effect between Age and Group was not significant either ($F(1,82) = 4.23$, $p_{FDR} = 0.093$, $\eta_P^2 = 0.05$ for /b/, $F(1,65) = 3.65$, $p_{FDR} = 0.093$, $\eta_P^2 = 0.06$ for /d/).

fNIRS general linear model (GLM) results for the speech perception task

We conducted a repeated-measure ANCOVA of group (RD, AC) by age (children, adults) by similarity to Chinese (high similarity, low similarity) by syllable consistency (identical, different) with Raven score as a covariate for each channel. Results showed no main effect of age, group, similarity, or syllable consistency. A significant three-way (age × group × similarity) interaction was found in CH31 (right IFG, $F(1,80) = 10.07$, $p_{FDR} = 0.048$, $\eta_P^2 = 0.11$) and CH37 (right DLPFC, $F(1,80) = 16.16$, $p_{FDR} = 0.006$, $\eta_P^2 = 0.17$). Simple effect analysis showed that adults with RD had a greater hemodynamic response than adults without RD in perception of Spanish syllables with high similarity to Chinese ($F(1,80) = 6.92$, $p = 0.010$ for CH31, $F(1,80) = 6.92$, $p = 0.010$ for CH37), while no significant differences were found between children without RD and children with RD. No group differences were found for Spanish syllables with low similarity to Chinese (Fig. 3a, b). Another way to explain the interaction is a greater decrease in activation from children without RD to adults without RD than that from children with RD to adults with RD for the perception of Spanish syllables with high similarity to Chinese but not for Spanish syllables with low similarity to Chinese ($F(1,80) = 6.05$, $p = 0.016$ for CH31, $F(1,80) = 6.66$, $p = 0.012$ for CH37).

fNIRS general linear model (GLM) results for the speech production task

An ANCOVA of group (RD, AC) by age (children, adults) by language (Chinese, Spanish) was conducted with Raven as a covariate for each channel's data in the speech production task. We found no significant main effects of Age, Group, or Language, as well as no interactions between Age and Group. However, a significant Language by Group interaction was observed in channel 22 (left MTG, $F(1,74) = 11.53$, $p_{FDR} = 0.048$, $\eta_P^2 = 0.14$) (Fig. 3d). Simple effect analysis indicated that individuals without RD exhibited greater deactivation for Spanish than for Chinese ($F(1,74) = 8.12$, $p = 0.006$), whereas individuals with RD showed greater deactivation for Chinese than for Spanish ($F(1,74) = 5.64$, $p = 0.020$). The interaction could also be explained by the fact that individuals with RD had reduced deactivation compared to individuals without RD for Spanish production

($F(1,74) = 7.10.25$, $p = 0.002$), while no significant group difference was observed for Chinese production ($F(1,74) = 0.17$, $p = 0.68$).

Brain-behavioral correlations

In order to understand how abnormal brain activation is correlated with phonological deficits, we conducted Pearson's correlations between the activation of channels exhibiting a significant interaction in each task (beta values of CH31 and CH37 for the high similarity to Chinese condition in the speech perception task and beta values of CH22 in the speech production task) and phonological awareness for children with RD and adults with RD separately. A negative correlation was found between the activation of CH37 during perception of Spanish syllables with high similarity to Chinese and pseudoword rhyming judgment ($r = -0.629$, $p_{FDR} = 0.084$) for children with RD (Fig. 3c). No significant correlation was found between activation and phonological awareness in adults with RD. Steiger's Z test revealed a significant difference between children with RD and adults with RD in the correlation between activation of CH37 during Spanish perception and pseudoword rhyming judgment ($z = 2.90$, $p = 0.004$).

Classification performance

Using LOOCV, the SVM classifier yielded an accuracy ranging from 60% to 90% when classifying children, adults, or all participants across the two fNIRS tasks (Table 2). For classifications with higher accuracy than the permutation tests, we listed the p-value. The null hypothesis distribution obtained from permutation testing is displayed in Supplementary Fig. 5.

Brain regions with high discriminative power

For the five classifications that showed a significantly higher accuracy than the permutation tests (Table 2), we calculated the frequency of channels appearing in the optimum feature set during cross-validation. As Fig. 4a shows, several regions exhibited relatively large weights (appearing in at least 80% of the optimum feature sets of cross-validation folds) for classifying RD in children using the Spanish production task, including the left MTG (CH22) and the right supramarginal gyrus (CH29).

For classifying RD in adults using the perception of Spanish syllables with low similarity to Chinese, Fig. 4b shows that the left DLPFC (CH2, CH13), left premotor and SMA (CH8), left MTG (CH22), right postcentral (CH34) and right IFG (CH31) exhibited relatively large weights.

For classifying RD in adults using the Chinese production task, Fig. 4c shows that the bilateral IFG (CH9, CH31), bilateral premotor and SMA (CH11, CH47), left MTG (CH24), right primary somatosensory cortex (CH33), right PMC (CH36, CH48) exhibited large weights.
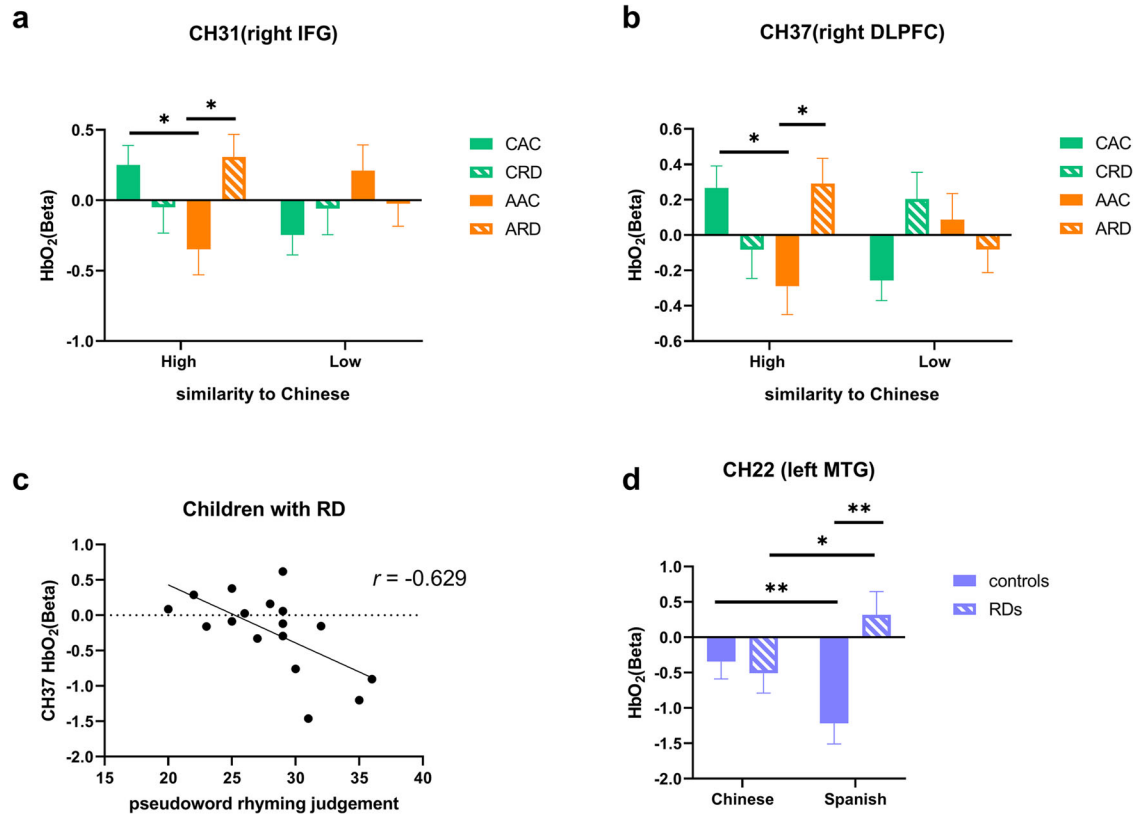
Fig. 3 | Results from the GLM analysis of the fNIRS tasks. a, b Significant three-way interaction for CH31 and CH37 during the speech perception task. CAC age-control children, CRD children with RD, AAC age-control adults, and ARD adults with RD. **c** Children with RD showed a significant negative correlation between pseudoword rhyming and activation in CH37 (right DLPFC) during perception of Spanish syllables with high similarity to Chinese. **d** Significant interaction between language and group for CH22 during the speech production task. $*p_{FDR} < 0.05$, $**p_{FDR} < 0.01$.

Table 2 | Accuracies of SVM across languages and tasks in the classification model

	Spanish perception		Production		Production+Perception	
	High similarity to Chinese	Low similarity to Chinese	Chinese	Spanish	Chinese/high similarity to Chinese	Spanish/low similarity to Chinese
Children ($N = 44$)	68%	72%	75%	81% ($p = 0.031$)	66%	72%
Adults ($N = 43$)	66%	82% ($p = 0.002$)	82% ($p = 0.004$)	65%	87% ($p = 0.003$)	77%
Children +adults ($N = 87$)	68%	65%	73% ($p = 0.012$)	63%	71%	68%

In Fig. 4d for classifying RD in adults using the Chinese production task and the perception of Spanish syllables with high similarity to Chinese, activations that had large weights include the left IFG (CH9), left frontal eye fields (CH1), left PMC (CH10), right postcentral (CH27) left postcentral (CH11), left MTG (CH24), right DLPFC (CH37), right premotor and SMA (CH46 CH47 CH48).

When classifying RD in both children and adults using the Chinese production task, regions that showed large weights include the bilateral premotor and SMA (CH11, CH16, CH39, CH47), bilateral postcentral (CH19, CH27, CH36), bilateral MTG (CH24, CH25, CH28), left IFG (CH14), left supramarginal gyrus (CH18), and right IPL (CH41) (Fig. 4e).

Taken together, channels located in the left MTG were the most consistent region with high weights, because they appeared in all five classifications, and the left premotor cortex, left SMA, and the left IFG had high weights in three classifications.

Discussion

In the current study, we compared brain activation during speech perception and production in children and adults with or without RD. We

found reduced Spanish pronunciation accuracy in children with RD compared to children without RD in the vowel frequency formants analysis. In the brain, we found reduced differences between adults and children in RD readers compared to readers without RD in the perception of Spanish syllables in the right inferior frontal gyrus (IFG) and right DLPFC, suggesting slowed development in these regions in individuals with RD. We also found reduced language differentiation between Chinese and Spanish in the left MTG in individuals with RD compared to individuals without RD in the production task. Moreover, using a machine learning approach, we found that brain activity patterns in the left MTG, left premotor, SMA, and left IFG during the speech tasks were the most reliable features for classifying individuals with RD. Our findings provide evidence for brain abnormalities associated with phonological deficits during foreign speech processing in RD from a developmental perspective. Discussion on the findings of behavioral assessments is presented in the supplementary discussion.

In the phonetic analysis, we found that children had better Spanish vowel pronunciation than adults, which is consistent with previous findings that children in general have advantages compared to adults in foreign
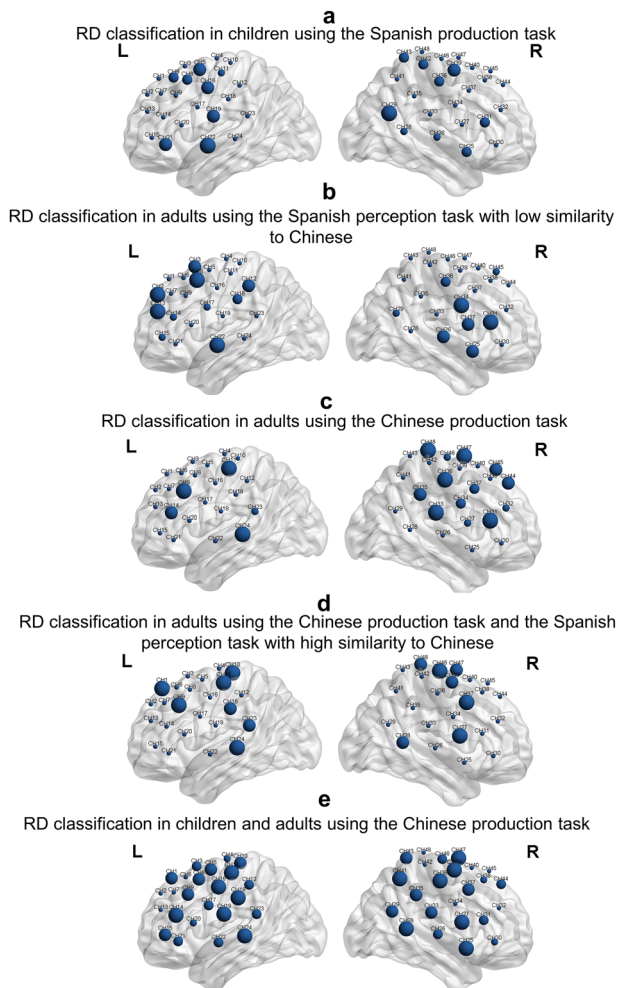
Fig. 4 | Feature weight maps for RD classification. a In children using the Spanish production task, **b** in adults using the Spanish perception task with low similarity to Chinese, **c** in adults using the Chinese production task, **d** in adults using the Chinese production task and the Spanish perception task with high similarity to Chinese, and **e** in children and adults using the Chinese production task. The Figure shows the weight assigned to each feature (fNIRS channel) in all folds of the LOOCV. The diameter of the sphere at each feature reflects its relative weight, with larger diameters indicating greater weights.

speech imitation[48,49]. According to Yeni-Komshian et al.[50], 11-year-old children achieved a better score in second-language pronunciation than adults, even though their scores were not as high as children younger than 6. Our study had exactly the same finding, suggesting that 11-year-old children still have advantages compared to adults in foreign speech learning.

We also found poorer vowel pronunciation in children with RD than children without RD but adults with RD did not differ significantly from adults without RD. This is because children without RD had a better performance than adults without RD but such an advantage was not found in children with RD compared to adults with RD. Our finding suggests that children with RD have less accurate foreign speech production than control children, presumably due to their phonological deficits. We speculate that phonological deficits in children with RD affect their learning to produce foreign speech sounds so that they do not show an age advantage compared to adults on the speech production task.

In the brain, for the speech perception task, we found that adults showed decreased activation compared to children in readers without RD but not in readers with RD in the right IFG and right DLPFC for Spanish syllables with high similarity to Chinese. Less involvement of these regions in adults than in children is generally interpreted as less effort in adults than

in children[51], especially when these syllables are similar to the native language. The role of the right IFG in phoneme discrimination has been repeatedly documented. For example, a study by Myers et al.[52] found that the right IFG exhibited greater activation in discriminating both between-category and within-category trials than identical trials, suggesting its involvement in phoneme discrimination. Furthermore, Kovelman, Yip, and Beck[53] found that the right IFG showed greater activation to deviant stimuli than standard stimuli in native and non-native phoneme discrimination, while the left IFG was specifically sensitive to native phonemes.

The dorsolateral prefrontal cortex (DLPFC) is primarily associated with working memory and other executive functions[54]. The less involvement of this region in adults than in children suggests less effort in working memory[55], especially when the Spanish syllables were similar to Chinese. Furthermore, we found a negative correlation between brain activation in the right DLPFC and pseudoword rhyming judgment in RD children in the current study, further suggesting that RD children with better phonological awareness need less effort of working memory in this region for speech perception. However, there was a lack of difference between adults with RD and children with RD, suggesting reduced development in the speech network in RD readers compared to typical readers, probably because of the influence of their phonological deficits.

For the speech production task, RD readers showed reduced deactivation compared to readers without RD in the anterior part of the left MTG (i.e. channel 22) during Spanish but not Chinese production. Another way to interpret the interaction is that readers without RD showed greater deactivation in this region in Spanish than in Chinese production, but RD readers showed greater deactivation in Chinese than in Spanish production. The MTG is believed to be part of the default mode network[56,57] and greater deactivation for Spanish than Chinese in readers without RD suggests greater challenge in doing the Spanish production task than the Chinese production task. Nonetheless, RD readers could not efficiently deactivate the default mode network for the more challenging foreign speech imitation to a greater degree. The default mode network has been found to play an important role in learning[58,59], and its abnormality has been reported in developmental disorders, such as RD[60–62], ASD[63–65], and ADHD[66–68]. In a recent study, the default mode network was found to show the largest developmental changes in brain signal complexity for participants at 6–13 years of age, compared to five other networks, namely, the vision, motor, dorsal attention, ventral attention, and frontal-parietal network[69]. Compared to the early-developing vision and motor networks, late-developing networks such as the DMN have a longer developmental window and therefore might be more influenced by learning experiences and environment. Therefore, the abnormality in the default mode network might be due to the atypical learning experiences in individuals with developmental disorders.

Using MVPA, individuals with RD were distinguished from age-matched non-RD counterparts with relatively high accuracy based on brain activation patterns during the speech tasks, suggesting a reliable classifier of RD. We found that the fNIRS channels located in the left MTG left premotor cortex, left SMA, and left IFG consistently showed high discriminative power.

The left MTG includes an anterior channel 22, and a posterior channel 24 (Fig. 4). channel 22 is part of the DMN network as discussed above, whereas channel 24 is involved in speech perception and phonological representation. In the model by Hickok and Poeppel[70], the posterior MTG supports the sound-based representation of speech. It has been found that the posterior MTG exhibits greater activation for speech perception in noise than in normal condition[71], suggesting its importance in speech perception. Previous studies have also reported decreased activation of the left MTG in individuals with RD compared to individuals without RD across both speech and reading tasks[72–76], suggesting deficient phonological representation and speech processing in individuals with RD. Consistent with previous studies, our results from a machine learning approach, further suggest that brain activation pattern in the left MTG during speech tasks is a reliable marker of RD.

The premotor cortex has been found to play an important role in speech-motor planning, phonological short-term memory, and

sensorimotor integration[77–79]. The SMA has also been found to play a crucial role in planning complex motor sequences, which is essential for handling the "complex speech demands" in tasks with unfamiliar words and complex nonwords production[80–82]. The premotor cortex and SMA have been found to be involved in not only speech production but also speech perception[83,84], because perception and action share a representation system[85,86]. Abnormalities in these two regions during speech tasks make them reliable features of RD. Consistent with our finding, previous studies have also shown reduced activations in the SMA and premotor regions in individuals with RD during phonological rhyming tasks and phonological short-term memory tasks[87,88], suggesting that abnormal function of these regions might be associated with the phonological deficits in RD. Taken together, with a machine learning approach, we suggest a functional abnormality of these regions associated with phonological deficits during speech processing to be key features of Chinese RD.

The left IFG, where Broca's area is housed, is also involved in speech production planning[70,89]. In addition to its role in speech production, the left IFG has been found to be involved in other phonological processing such as phonemic discrimination[90], phonological working memory[91,92], phonological competition and selection[93,94]. Previous research has consistently found abnormal structure and function of the left IFG associated with RD, especially in Chinese RD[95]. Our study using a machine learning method, also confirms that the abnormal activation patterns of the left IFG during speech tasks can be a reliable marker of RD in Chinese.

The relatively small sample size and cross-sectional design might have limited our capability of revealing developmental changes in the RD readers and typical readers in foreign speech perception and production. Moreover, future research is also needed to examine whether these findings are replicable in other languages.

To conclude, in the current study, we revealed neurological differences that are associated with phonological deficits reflected in low quality of foreign speech perception and production in individuals with RD. Moreover, we found that brain activation patterns in the left MTG, left premotor, SMA, and left IFG can serve as reliable classifiers of RD regardless of age and speech tasks. Our findings provide important evidence for abnormal foreign speech processing in RD from a developmental perspective.

Methods
Participants
We recruited fifth-grade children from public elementary schools, and students from associate degree colleges in the local city. Participants with RD met the following criteria: (1) the standard score on Raven was above 80; (2) the z-score was below −1.5 on at least one of three reading tests, namely, a Chinese character naming test, a Chinese sentence reading fluency test, and a one-minute Chinese character naming test. The inclusion criteria for participants without RD were: (1) the standard score on Raven was above 80; (2) the z-score was above −1 on all of the three reading tests. We had 20 children with RD (mean age = 11.00 years, range 10–12, 12 males), 24 age-matched children without RD (mean age = 10.58 years, range 10–12, 9 males), 20 adults with RD (mean age = 19.63 years, range 18–22, 9 males), and 23 age-matched adults without RD (mean age = 19.65 years, range 18–24, 10 males). All participants were native Chinese speakers, right-handed, without neurological or psychiatric diseases, have not learned Spanish. All adults with RD and parents of children in the RD group reported a history of reading difficulties including poor reading accuracy and fluency. The IRB at Sun Yat-Sen University approved the study and consent procedures. All participants/parents of child participants signed written consent before we conducted any testing. Children also gave assent.

Behavioral assessments
The Chinese character naming test is a measure of word decoding accuracy, in which the participant is asked to read aloud 150 Chinese characters without a time limit. The total number of characters read correctly is the raw score. The Chinese sentence reading fluency test is a measure of reading fluency and reading comprehension, in which the participant is asked to

silently read 100 sentences of varying length and make a judgment whether each sentence makes sense in meaning, and the time limit is 3 min. The total number of characters in sentences that are correctly judged is the raw score. Norms for fifth-grade children on these two tests are available from a previous study[96]. We tested 215 adults without RD from the same colleges where we found adults with RD to develop a norm for adults on the character naming test (mean ± standard deviation: 140.02 ± 6.41 characters) and the sentence reading fluency test (mean ± standard deviation: 1379.88 ± 377.21 characters).

The one-minute Chinese character naming test is a character reading fluency test that is composed of two parts: 150 regular characters and 150 irregular characters. Regular characters are those that share the same pronunciation with the phonetic radical, while irregular characters are those that have a different pronunciation from the phonetic radical. The test requires participants to read the characters as quickly and accurately as possible within one minute. The one-minute character naming test was administered to 201 college students (100.45 ± 17.70 for regular characters, 80.92 ± 19.10 for irregular characters) and 217 fifth-grade children (71.18 ± 18.11 for regular characters, 48.72 ± 18.29 for irregular characters) to develop norms for adults and children, respectively.

In addition to the three reading tests used for screening, all participants also completed meta-linguistic awareness tests for phonological awareness, morphological awareness, and orthographic awareness, as well as cognitive ability tests for working memory and rapid automatized naming (RAN). Phonological awareness was tested with English words and pseudowords while morphological awareness and orthographic awareness were tested with Chinese materials. All of the participants were Chinese-English bilinguals and English but not Chinese materials were used in the phonological awareness tests because all Chinese characters are monosyllabic and there may be a ceiling effect in adults if we use Chinese materials in the phonological awareness tests. Using English pseudowords was also helpful in equalizing the material familiarity among participants.

Phonological awareness was measured using a 30-item initial sound deletion test and a 40-item pseudoword rhyming test. In the initial sound deletion test, participants were orally presented with a word and asked to delete the first consonant sound and then pronounce the rest part of the word (e.g., the word "sock" /sɑk/ should be pronounced as "ock" [ɑk] after the initial sound is deleted). In the pseudoword rhyming task, participants were orally presented with a pair of English pseudowords and asked to determine if the two pseudowords rhyme.

Morphological Awareness was tested in a 30-item homophonic morpheme test and a 30-item homographic morpheme test. In the homophonic morpheme test, participants were asked to choose one character from four homophones to form a meaningful word with a given character. For example, __段, (线/xian4/"line," 献/xian4/ "dedicate," 羡/xian4/ "envy," 县/xian4/ "town"). In the homographic morpheme test, participants were presented with a pair of two-character words containing the same morpheme (e.g., 道理/dao4-li3/ "reason" and 理会/li3-hui4/ "pay attention to") and they were asked to judge whether the morpheme had the same meaning in the two words.

Orthographic awareness was measured with a 60-item character correction test and a 30-item delayed copy test. In the character correction test, participants were asked to identify and correct wrongly written characters. In the delayed copy test, participants were presented with infrequently-used characters for 500 ms and asked to write down the character they had just seen. Raw scores for the phonological awareness, morphological awareness, and orthographic awareness tests were the number of correct items.

Working Memory was tested using forward and backward digit spans. *Rapid Automatized Naming (RAN)* was tested using digit RAN and picture RAN. In each RAN test, there are 50 items, and the time taken to name all of the 50 items was recorded in seconds and used as the raw score.

fNIRS procedures and stimuli
A passive speech perception task was used to examine Spanish perception with a rapid event-related design. A total of 120 pairs of Spanish consonant-

vowel (CV) syllables were used in this task with a consonant and a vowel in each syllable. There were four types of CV syllable pairs: (1) the two Spanish syllables were identical and the sounds had high similarity to those in Chinese (e.g., /pi/-/pi/), (2) the syllables were different but the sounds had high similarity to those in Chinese (e.g., /pi/-/bi/), (3) the syllables were identical but the sounds had low similarity to those in Chinese (e.g., /je/-/je/), and (4) the syllables were different and the sounds had low similarity to those in Chinese (e.g., /je/-/ge/). Participants were asked to listen carefully to the stimuli and to keep their heads as still as possible during the task in order to reduce motion artifacts. There were also 60 baseline trials, for which two black crosses were presented on the screen sequentially. All trials were randomly presented and divided into four runs with around 4 min for each run.

The experimental procedure is displayed in Supplementary Fig. 1. At the very beginning of each run, there was a 5000 ms fixation cross. Each single trial began with a brief black fixation cross (200 ms) warning the onset of a new trial and then two Spanish CV syllables were presented sequentially in the auditory modality with a duration of 800 ms for each and a 200 ms blank between the two syllables. The SOA was jittered between 3.5–4 s.

In the speech production task, participants were asked to imitate 26 multi-syllabic Chinese pseudowords and 26 multi-syllabic Spanish words. Each word/pseudoword was repeated 3 times sequentially, resulting in 156 trials in total divided into two runs with 78 trials per run. Chinese pseudowords and Spanish words were randomized in the presentation. Each run began with a silent period (5000 ms) prior to the onset of the first word. For each trial, the audio stimulus was displayed for 1500 ms with a black fixation cross shown on the screen, followed by a red cross to cue the start of the imitation phase which lasted for 1500 ms. The ITI was jittered at 250, 500, 750 or 1000 ms. A baseline trial was arranged after the third imitation of each word, during which, a cross was presented on the screen for 3000 ms and the participant did not need to do anything. An additional baseline trial was inserted randomly after the first or second imitation phase for each word. Responses were recorded through a microphone connected to the monitor through the E-prime SRBOX. The whole process took approximately 20 min to complete. The experimental procedure is displayed in Supplementary Fig. 1. The mean number of syllables per word/pseudoword was matched in Chinese and Spanish. The speech perception and speech production tasks were counterbalanced across participants.

Phonetic analyses in speech production
To examine whether participants with RD performed worse than individuals without RD in the foreign speech imitation task, we measured the VOT of two initial stops (i.e. /b/, /d/) and vowel's frequency formants for 5 vowels (i.e. a, o, i, e, u) in 5 Spanish words in the speech production task (i.e. difícil, dado, brazo, bueno, bebe). VOT and formant extraction were performed in Praat[97].

Since formant frequencies are influenced by anatomical/physiological differences (e.g., vocal tract shape, and gender)[98], a vowel normalization procedure was employed to eliminate the impact of these variables among participants. The vowel frequency formant normalization was performed using the Vowels R package[99]. We followed the approach in Lobanov[100], which is speaker-intrinsic, vowel-extrinsic, and formant-extrinsic, and it performs best on mitigating the effects of speakers' gender and age-related variations while preserving valuable sociolinguistic information[101]. In order to quantify the foreign speech imitation performance, we calculated the Euclidean distance between each participant's vowels and a native Spanish model speaker's vowels in the F1-F2 vowels space.

fNIRS data acquisition
Changes in the oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) concentrations were measured with a continuous-wave (CW) NIRSport2 system (NIRx, Medical Technologies LLC, Berlin, Germany) sampled at 4.4 Hz. Two wavelengths of near-infrared light (760 and 850 nm) were used, with a distance between pairs of source and detector probes set at 3.0 cm. Two 4 × 4 probe sets were placed on the bilateral frontal, parietal, and temporal areas, with each comprising 8 emitter and 8 detector probes,

forming 48 channels in total. The international 10–20 system was used to guide and standardize the optode placement, with the D8 and D10 detectors aligned with T7 and T8, respectively, (Supplementary Fig. 2).

To determine the anatomical localization of each optode, we collected T1-weighted images from a typical adult participant using a 3.0 Tesla Prisma Siemens scanner with the following parameters: time repetition = 2300 ms; time echo = 3.39 ms; flip angle = 7°; slice thickness = 1 mm; voxel size = 1 × 1 × 1 mm. The images were normalized to MNI coordinate space using SPM12 and brain regions under each optode were determined using the AAL template. The position of a channel was defined as the center of the two adjacent emitting and receiving optodes.

fNIRS data pre-processing
Data pre-processing began with a manual visual check on signal quality following Liang et al.[102]. The spectrograms of all channels were plotted and the NIRS channels without a clear, visible cardiac component (a spike at ~1–1.5 Hz in the spectrograms) or only with random noise were regarded as low quality. Visual inspections were conducted by two experienced researchers and for inconsistent inspections, we invited a third researcher. If more than 20% of a participant's channels were low quality, then that participant's data would be excluded from further analysis. If the number of channels with low quality did not exceed 20%, these channels were excluded from further analysis and the rest channels from this participant were included for further analysis. On average, 3 channels were excluded in each group (Children with RD: 2.52; children without RD: 3.39; adults with RD: 3.8; adults without RD: 3.26).

Then we used Homer 2[103] for further pre-processing. First, the raw fNIRS intensity signals were converted into optical density using the Homer2 *hmrIntensity2OD* function. Next, wavelet filtering was conducted for motion correction using the *hmrMotionCorrectWavelet* function (iqr = 0.8). According to Di Lorenzo et al.[104], the use of 0.8 was recommended for analyzing short event-related data. The data were then bandpass filtered between 0.02 Hz and 0.5 Hz to attenuate low-frequency drift and cardiac oscillations. Optical density signals were converted to concentration changes (μmol/L) of HbO and HbR using the modified Beer–Lambert law with a default partial pathlength factor of 6.0 for each wavelength.

General linear model (GLM) analysis
The preprocessed fNIRS data were imported to the NIRS-KIT toolbox[105] based on the MATLAB environment for individual-level analysis. A general linear model (GLM) was used to evaluate channel-wise task-evoked neural activation for each individual participant. Because of the lower signal-to-noise ratio of HbR compared to HbO[106], only concentration changes of HbO were investigated in the GLM.

For the speech perception task, five conditions were included in the model (identical Spanish syllables that are similar to Chinese, different Spanish syllables that are similar to Chinese, identical Spanish syllables that are dissimilar to Chinese, different Spanish syllables that are dissimilar to Chinese, and baseline trials). For the speech production task, three conditions were included (Chinese pseudowords, Spanish words, and baseline trials). The model was convolved with the canonical hemodynamic response function, and then model estimation was conducted to calculate how well the model fits with the real brain signal at each channel. The contrast of each lexical condition minus the baseline condition was then defined to estimate signal magnitudes specifically related to each type of stimuli. Finally, beta values from the model estimation were entered for subsequent group-level statistical analysis using ANCOVAs for each channel. Multiple comparison correction was conducted using FDR correction[107], since we had 48 channels.

Classification based on fNIRS data
The beta values for the contrast of lexical minus baseline for each channel from the general linear model estimation were extracted, resulting in a feature vector of 1 × 48 for each participant in each task.

We used supporting vector machine (SVM) for classification of RD readers from readers without RD, due to its higher accuracy than other methods for small datasets[108,109]. We used an open-source machine learning library in Python, scikit-learn[110] for the SVM implementation.

First, the feature vector was normalized across participants. In SVM, C is a regularization parameter that determines the trade-off between maximizing the margin and minimizing the classification error while γ is a parameter that influences the shape of the decision boundary. We optimized these two parameters in the radial basis function kernels (RBF-SVM) using a cross-validation grid search among the values of 2^N (N from -5 to 11 for C and from -9 to 13 for γ) in the training dataset, and then optimal parameters were used to test the classifier. We performed a leave-one-out cross-validation (LOOCV) to assess the classifier's performance until all participants were tested. Last, the number of correct predictions was divided by the total number of participants to calculate the accuracy of cross-validation.

In order to speed up computation and improve performance, we employed recursive feature elimination (RFE) to reduce the impact of irrelevant features in this study. The RFE approach involves a nested LOOCV strategy, wherein the inner LOOCV is performed on the training set of each outer SVM LOOCV fold. The primary objective is to identify an optimal subset of features that contribute most significantly to the classification task. Since we used LOOCV to estimate the generalization ability of the classifier, the optimum feature set was different in the training dataset for each fold of LOOCV. Therefore, when analyzing the contributions of different brain regions, the weights of the features were defined as the frequency of appearing in the optimum feature set across all cross-validation folds. By employing RFE within the LOOCV framework, we were able to optimize the feature selection process and obtain a robust and reliable set of features that consistently contributed to accurate classification across different folds.

We performed a permutation test to evaluate whether the predictive validity of the model was higher than chance. Participants' feature vectors were shuffled across participants to generate a randomized matrix, and the model was trained and cross-validated as previously described. The data randomization procedure was repeated 1000 times to obtain a null distribution of accuracies. The p-value is the proportion of permutation tests with an accuracy higher than the actual classification accuracy. A significance threshold of 5% ($p < 0.05$) was employed.

In order to find the task with the highest classification accuracy for distinguishing RD readers and readers without RD in children, adults or children and adults combined, we compared the model performance on the speech perception task, the speech production task and a combination of the two tasks.

Data availability
Data analyzed in this study are available upon request.

Code availability
The code used for data analysis is available upon request.

References
1. Chan, D. W., Ho, C. S. H., Tsang, S. M., Lee, S. H. & Chung, K. K. H. Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. *Educ. Stud.* **33**, 249–265 (2007).
2. Korhonen, J., Linnanmäki, K. & Aunio, P. Learning difficulties, academic well-being and educational dropout: a person-centred approach. *Learn. Individ. Differ.* **31**, 1–10 (2014).
3. Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M. & Escobar, M. D. Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *Jama* **264**, 998–1002 (1990).
4. Lyon, G. R., Shaywitz, S. E. & Shaywitz, B. A. A definition of dyslexia. *Ann. Dyslexia* **53**, 1–14 (2003).
5. Callens, M., Tops, W. & Brysbaert, M. Cognitive profile of students who enter higher education with an indication of dyslexia. *PLoS One* **7**, e38081 (2012).
6. Lefly, D. L. & Pennington, B. F. Spelling errors and reading fluency in compensated adult dyslexics. *Ann. Dyslexia* **41**, 141–162 (1991).
7. Swanson, H. L. & Hsieh, C.-J. Reading disabilities in adults: a selective meta-analysis of the literature. *Rev. Educ. Res.* **79**, 1362–1390 (2009).
8. Goswami, U. In *The Cambridge Handbook of Dyslexia and Dyscalculia* (ed. Michael A. Skeide) 5–24 (Cambridge University Press, 2022).
9. Kuhl, U. et al. The emergence of dyslexia in the developing brain. *NeuroImage* **211**, 116633 (2020).
10. Molfese, D. L. Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain Lang.* **72**, 238–245 (2000).
11. Eberhard-Moscicka, A. K., Jost, L. B., Daum, M. M. & Maurer, U. Predicting reading from behavioral and neural measures–a longitudinal event-related potential study. *Front. Psychol.* **12**, 733494 (2021).
12. Lohvansuu, K. et al. Unveiling the mysteries of dyslexia—lessons learned from the prospective Jyväskylä longitudinal study of dyslexia. *Brain Sci.* **11**, 427 (2021).
13. Hong, T. et al. Cortical responses to Chinese phonemes in preschoolers predict their literacy skills at school age. *Dev. Neuropsychol.* **43**, 356–369 (2018).
14. Snowling, M. J., Lervag, A., Nash, H. M. & Hulme, C. Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Dev. Sci.* **22**, e12723 (2019).
15. Keshavarzi, M., et al. Atypical speech production of multisyllabic words by children with developmental dyslexia. *Dev. Sci.* **27**, e13428 (2024).
16. Gerrits, E. & de Bree, E. Early language development of children at familial risk of dyslexia: speech perception and production. *J. Commun. Disord.* **42**, 180–194 (2009).
17. Calcus, A., Lorenzi, C., Collet, G., Colin, C. & Kolinsky, R. Is there a relationship between speech identification in noise and categorical perception in children with dyslexia? *J. Speech Lang. Hear Res*. **59**, 835–852 (2016).
18. Hakvoort, B. et al. The role of categorical speech perception and phonological processing in familial risk children with and without dyslexia. *J. Speech Lang. Hear. Res*. **59**, 1448–1460 (2016).
19. Serniclaes, W., Sprenger-Charolles, L., Carré, R. & Demonet, J. F. Perceptual discrimination of speech sounds in developmental dyslexia. *J. Speech Lang. Hear. Res*. **44**, 384–399 (2001).
20. Messaoud-Galusi, S., Hazan, V. & Rosen, S. Investigating speech perception in children with dyslexia: is there evidence of a consistent deficit in individuals? *J. Speech Lang. Hear. Res*. **54**, 1682–1701 (2011).
21. Kovelman, I. et al. Brain basis of phonological awareness for spoken language in children and its disruption in dyslexia. *Cereb. Cortex* **22**, 754–764 (2012).
22. Jaffe-Dax, S., Kimel, E. & Ahissar, M. Shorter cortical adaptation in dyslexia is broadly distributed in the superior temporal lobe and includes the primary auditory cortex. *Elife* **7**, e30018 (2018).
23. Steinbrink, C., Groth, K., Lachmann, T. & Riecker, A. Neural correlates of temporal auditory processing in developmental dyslexia during German vowel length discrimination: an fMRI study. *Brain Lang.* **121**, 1–11 (2012).
24. Gaab, N., Gabrieli, J., Deutsch, G., Tallal, P. & Temple, E. Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: an fMRI study. *Restor. Neurol. Neurosci.* **25**, 295–310 (2007).
25. Temple, E. et al. Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc. Natl Acad. Sci. USA* **97**, 13907–13912 (2000).
26. Vandermosten, M. et al. Brain activity patterns of phonemic representations are atypical in beginning readers with family risk for dyslexia. *Dev. Sci.* **23**, e12857 (2020).

27. Hommet, C. et al. Topography of syllable change-detection electrophysiological indices in children and adults with reading disabilities. *Neuropsychologia* **47**, 761–770 (2009).

28. Chen, A. Later but not weaker: neural categorization of native vowels of children at familial risk of dyslexia. *Brain Sci.* **12**, 412 (2022).

29. Gu, C. & Bi, H.-Y. Auditory processing deficit in individuals with dyslexia: a meta-analysis of mismatch negativity. *Neurosci. Biobehav. Rev.* **116**, 396–405 (2020).

30. Del Tufo, S. N. & Earle, F. S. Skill profiles of college students with a history of developmental language disorder and developmental dyslexia. *J. Learn. Disabil.* **53**, 228–240 (2020).

31. Cao, F. et al. Neural signatures of phonological deficits in Chinese developmental dyslexia. *NeuroImage* **146**, 301–311 (2017).

32. Meng, X. et al. Auditory and speech processing and reading development in Chinese school children: behavioural and ERP evidence. *Dyslexia* **11**, 292–310 (2005).

33. Zhang, Y. et al. Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *J. Child Psychol. Psychiatry* **53**, 874–882 (2012).

34. Harwood, V., Preston, J., Baron, A., Kleinman, D. & Landi, N. Event-related potentials to speech relate to speech sound production and language in young children. *Dev. Neuropsychol.* **47**, 105–123 (2022).

35. Smith, A. B., Smith, S. L., Locke, J. L. & Bennett, J. A longitudinal study of speech timing in young children later found to have reading disability. *J. Speech Lang. Hear Res.* **51**, 1300–1314 (2008).

36. Fawcett, A. J. & Nicolson, R. I. Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia* **8**, 189–203 (2002).

37. Larrivee, L. S. & Catts, H. W. Early reading achievement in children with expressive phonological disorders. *Am. J. Speech Lang. Pathol.* **8**, 118–128 (1999).

38. Swan, D. & Goswami, U. Picture naming deficits in developmental dyslexia: the phonological representations hypothesis. *Brain Lang.* **56**, 334–353 (1997).

39. Catts, H. W. Speech production/phonological deficits in reading-disordered children. *J. Learn. Disabil.* **19**, 504–508 (1986).

40. Griffiths, S. & Frith, U. Evidence for an articulatory awareness deficit in adult dyslexics. *Dyslexia* **8**, 14–21 (2002).

41. Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A. & Haith, M. M. Phonological processing skills and deficits in adult dyslexics. *Child Dev.* **61**, 1753–1778 (1990).

42. Marchetti, R. et al. Phoneme representation and articulatory impairment: insights from adults with comorbid motor coordination disorder and dyslexia. *Brain Sci.* **13**, 210 (2023).

43. Ylinen, S. et al. Diminished brain responses to second-language words are linked with native-language literacy skills in dyslexia. *Neuropsychologia* **122**, 105–115 (2019).

44. Soroli, E., Szenkovits, G. & Ramus, F. Exploring dyslexics' phonological deficit III: foreign speech perception and production. *Dyslexia* **16**, 318–340 (2010).

45. Bouhon, M., Ferreira, C., Bahuon, S., Tillmann, B. & Bedoin, N. Improving non-native duration contrast with dichotic training in dyslexic and non-dyslexic individuals. *Dyslexia* **29**, 151–158 (2023).

46. Costa, A., Santesteban, M. & Caño, A. On the facilitatory effects of cognate words in bilingual speech production. *Brain Lang.* **94**, 94–103 (2005).

47. Ghazi-Saidi, L. & Ansaldo, A. N. A. I. The neural correlates of semantic and phonological transfer effects: language distance matters. *Bilingualism Lang. Cogn.* **20**, 1080–1094 (2017).

48. Munro, M. J., Flege, J. E. & MacKay, I. R. The effects of age of second language learning on the production of English vowels. *Appl. Psycholinguist.* **17**, 313–334 (1996).

49. Tahta, S., Wood, M. & Loewenthal, K. Foreign accents: factors relating to transfer of accent from the first language to a second language. *Lang. speech* **24**, 265–272 (1981).

50. Yeni-Komshian, G. H., Flege, J. E. & Liu, S. Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism Lang. Cogn.* **3**, 131–149 (2000).

51. Shaywitz, S. E. & Shaywitz, B. A. The neurobiology of reading and dyslexia. *ASHA Lead.* **12**, 20–21 (2007).

52. Myers, E. B., Blumstein, S. E., Walsh, E. & Eliassen, J. Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.* **20**, 895–903 (2009).

53. Kovelman, I., Yip, J. C. & Beck, E. L. Cortical systems that process language, as revealed by non-native speech sound perception. *Neuroreport* **22**, 947–950 (2011).

54. Hertrich, I., Dietrich, S., Blum, C. & Ackermann, H. The role of the dorsolateral prefrontal cortex for speech and language processing. *Front. Hum. Neurosci.* **15**, 645209 (2021).

55. Tamm, L., Menon, V. & Reiss, A. L. Maturation of brain function associated with response inhibition. *J. Am. Acad. Child Adolesc. Psychiatry* **41**, 1231–1238 (2002).

56. Raichle, M. E. The brain's default mode network. *Annu. Rev. Neurosci.* **38**, 433–447 (2015).

57. Buckner, R. L. & DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).

58. Howard-Jones, P. A., Jay, T., Mason, A. & Jones, H. Gamification of learning deactivates the default mode network. *Front. Psychol.* **6**, 1891 (2016).

59. Lin, P. et al. Dynamic default mode network across different brain states. *Sci. Rep.* **7**, 46088 (2017).

60. Mateu-Estivill, R. et al. Functional connectivity alterations associated with literacy difficulties in early readers. *Brain Imaging Behav.* **15**, 2109–2120 (2021).

61. Liu, Z., Li, J., Bi, H.-Y., Xu, M. & Yang, Y. Disruption of functional brain networks underlies the handwriting deficit in children with developmental dyslexia. *Front. Neurosci.* **16**, 919440 (2022).

62. Boros, M. et al. Orthographic processing deficits in developmental dyslexia: Beyond the ventral visual stream. *NeuroImage* **128**, 316–327 (2016).

63. Washington, S. D. et al. Dysmaturation of the default mode network in autism. *Hum. Brain Mapp.* **35**, 1284–1296 (2014).

64. Assaf, M. et al. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage* **53**, 247–256 (2010).

65. Murdaugh, D. L. et al. Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. *PLoS One* **7**, e50064 (2012).

66. Qiu, M.-G. et al. Changes of brain structure and function in ADHD children. *Brain Topogr.* **24**, 243–252 (2011).

67. Castellanos, F. X. & Proal, E. Large-scale brain systems in ADHD: beyond the prefrontal–striatal model. *Trends Cogn. Sci.* **16**, 17–26 (2012).

68. Uddin, L. Q. et al. Network homogeneity reveals decreased integrity of default-mode network in ADHD. *J. Neurosci. methods* **169**, 249–254 (2008).

69. Jia, G. et al. Intrinsic brain activity is increasingly complex and develops asymmetrically during childhood and early adolescence. *NeuroImage* **227**, 120225 (2023).

70. Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131–138 (2000).

71. Zheng, Z. Z., Munhall, K. G. & Johnsrude, I. S. Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J. Cogn. Neurosci.* **22**, 1770–1781 (2010).

72. Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R. & Münte, T. F. Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain Imaging Behav.* **12**, 357–368 (2018).

73. Hancock, R., Richlan, F. & Hoeft, F. Possible roles for fronto-striatal circuits in reading disorder. *Neurosci. Biobehav. Rev.* **72**, 243–260 (2017).

74. Dębska, A. et al. Neural patterns of word processing differ in children with dyslexia and isolated spelling deficit. *Brain Struct. Funct.* **226**, 1467–1478 (2021).

75. Łuniewska, M. et al. Children with dyslexia and familial risk for dyslexia present atypical development of the neuronal phonological network. *Front. Neurosci.* **13**, 1287 (2019).

76. Martin, A., Kronbichler, M. & Richlan, F. Dyslexic brain activation abnormalities in deep and shallow orthographies: a meta-analysis of 28 functional neuroimaging studies. *Hum. Brain Mapp.* **37**, 2676–2699 (2016).

77. Hickok, G. & Poeppel, D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).

78. Hickok, G. The functional neuroanatomy of language. *Phys. Life Rev.* **6**, 121–143 (2009).

79. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).

80. Carreiras, M., Mechelli, A., Estévez, A. & Price, C. Brain activation for lexical decision and reading aloud: two sides of the same coin? *J. Cogn. Neurosci.* **19**, 433–444 (2007).

81. Carreiras, M., Mechelli, A. & Price, C. J. Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Hum. Brain Mapp.* **27**, 963–972 (2006).

82. Bohland, J. W. & Guenther, F. H. An fMRI investigation of syllable sequence production. *NeuroImage* **32**, 821–841 (2006).

83. Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* **7**, 701–702 (2004).

84. D'Ausilio, A. et al. The motor somatotopy of speech perception. *Curr. Biol.* **19**, 381–385 (2009).

85. Lotto, A. J., Hickok, G. S. & Holt, L. L. Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* **13**, 110–114 (2009).

86. McGettigan, C. & Tremblay, P. In *The Oxford Handbook of Psycholinguistics* (Oxford Academic, 2018).

87. Paulesu, E., Danelli, L. & Berlingeri, M. Reading the dyslexic brain: multiple dysfunctional routes revealed by a new meta-analysis of PET and fMRI activation studies. *Front. Hum. Neurosci.* **8**, 830 (2014).

88. Paulesu, E. et al. Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* **119**, 143–157 (1996).

89. Park, H., Iverson, G. K. & Park, H.-J. Neural correlates in the processing of phoneme-level complexity in vowel production. *Brain Lang.* **119**, 158–166 (2011).

90. Ruff, S., Marie, N., Celsis, P., Cardebat, D. & Démonet, J.-F. Neural substrates of impaired categorical perception of phonemes in adult dyslexics: an fMRI study. *Brain Cogn.* **53**, 331–334 (2003).

91. Chein, J. M., Fissell, K., Jacobs, S. & Fiez, J. A. Functional heterogeneity within Broca's area during verbal working memory. *Physiol. Behav.* **77**, 635–639 (2002).

92. Guenther, F. H. & Hickok, G. Role of the auditory system in speech production. *Handb. Clin. Neurol.* **129**, 161–175 (2015).

93. Xie, X. & Myers, E. Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: a comparison of clear and conversational speech. *J. Cogn. Neurosci.* **30**, 267–280 (2018).

94. Heim, S., Eickhoff, S. B., Friederici, A. D. & Amunts, K. Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Struct. Funct.* **213**, 441–456 (2009).

95. Yan, X. et al. Convergent and divergent brain structural and functional abnormalities associated with developmental dyslexia. *Elife* **10**, e69523 (2021).

96. Song, S. et al. Tracing children's vocabulary development from preschool through the school-age years: an 8-year longitudinal study. *Dev. Sci.* **18**, 119–131 (2015).

97. Boersma, P. & Weenink, D. *Praat: Doing Phonetics by Computer* [*Computer Program*]. http://www.praat.org/ (2021).

98. Adank, P., Smits, R. & van Hout, R. A comparison of vowel normalization procedures for language variation research. *J. Acoust. Soc. Am.* **116**, 3099–3107 (2004).

99. Kendall, T. & Thomas, E. R. *Vowels: Vowel Manipulation, Normalization, and Plotting. R Package Version 1.2-2.* https://CRAN.R-project.org/package=vowels (2018).

100. Lobanov, B. M. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* **49**, 606–608 (2005).

101. Adank, P. M. *Vowel Normalization: A Perceptual-acoustic Study of Dutch Vowels* (Catholic University of Nijmegen, 2003).

102. Liang, Z. et al. Tracking brain development from neonates to the elderly by hemoglobin phase measurement using functional near-infrared spectroscopy. *IEEE J. Biomed. Health Inform.* **25**, 2497–2509 (2021).

103. Huppert, T. J., Diamond, S. G., Franceschini, M. A. & Boas, D. A. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* **48**, D280–D298 (2009).

104. Di Lorenzo, R. et al. Recommendations for motion correction of infant fNIRS data applicable to multiple data sets and acquisition systems. *Neuroimage* **200**, 511–527 (2019).

105. Hou, X. et al. NIRS-KIT: a MATLAB toolbox for both resting-state and task fNIRS data analysis. *Neurophotonics* **8**, 010802 (2021).

106. Schaeffer, J. D. et al. An fNIRS investigation of associative recognition in the prefrontal cortex with a rapid event-related design. *J. Neurosci. Methods* **235**, 308–315 (2014).

107. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

108. Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H. & Stetter, M. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage* **28**, 980–995 (2005).

109. Kumari, R. SVM classification an approach on detecting abnormality in brain MRI images. *Int. J. Eng. Res. Appl.* **3**, 1686–1690 (2013).

110. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

Author contributions

Conception and design: F.C., Y.F. Study supervision: F.C. Funding acquisition: F.C. Sample collection and assessments: Y.F., X.Y., J.M. Analysis methodology: F.C., Y.F., X.Y., J.M., H.B. Data analysis: Y.F. Data interpretation: F.C., Y.F. Writing and editing of the manuscript: F.C. and Y.F. with input from all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41539-024-00266-2.

Correspondence and requests for materials should be addressed to Fan Cao.

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