

Article

Screening Children's Intellectual Disabilities with Phonetic Features, Facial Phenotype and Craniofacial Variability Index

Yuhe Chen ¹, Simeng Ma ², Xiaoyu Yang ^{3,4}, Dujuan Liu ^{2,*} and Jun Yang ^{5,6,*} ¹ School of Foreign Languages, Huazhong University of Science and Technology, Wuhan 430074, China² Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan 430060, China³ Department of Pharmacy, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China⁴ Hubei Province Clinical Research Center for Precision Medicine for Critical Illness, Wuhan 430030, China⁵ School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, China⁶ School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China

* Correspondence: fengdemimi@foxmail.com (D.L.); junyang_ie@whut.edu.cn (J.Y.)

Abstract: Background: Intellectual Disability (ID) is a kind of developmental deficiency syndrome caused by congenital diseases or postnatal events. This syndrome could be intervened as soon as possible if its early screening was efficient, which may improve the condition of patients and enhance their self-care ability. The early screening of ID is always achieved by clinical interview, which needs in-depth participation of medical professionals and related medical resources. Methods: A new method for screening ID has been proposed by analyzing the facial phenotype and phonetic characteristic of young subjects. First, the geometric features of subjects' faces and phonetic features of subjects' voice are extracted from interview videos, then craniofacial variability index (CVI) is calculated with the geometric features and the risk of ID is given with the measure of CVI. Furthermore, machine learning algorithms are utilized to establish a method for further screening ID based on facial features and phonetic features. Results: The proposed method using three feature sets, including geometric features, CVI features and phonetic features was evaluated. The best performance of accuracy was closer to 80%. Conclusions: The results using the three feature sets revealed that the proposed method may be applied in a clinical setting in the future after continuous improvement.

Keywords: facial features; phonetic features; machine learning; intellectual disability; craniofacial variability index



Citation: Chen, Y.; Ma, S.; Yang, X.; Liu, D.; Yang, J. Screening Children's Intellectual Disabilities with Phonetic Features, Facial Phenotype and Craniofacial Variability Index. *Brain Sci.* **2023**, *13*, 155. <https://doi.org/10.3390/brainsci13010155>

Academic Editors: Corrado Romano and Mark Burke

Received: 5 November 2022

Revised: 31 December 2022

Accepted: 9 January 2023

Published: 16 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intellectual disability (ID) is a kind of generalized neurodevelopmental disorder that mostly occurs before 18 years old [1]. The intelligence and adaptive function of patients are obviously limited, including many daily social and practical skills. The prevalence of ID is estimated to be between 1% and 3% [1,2]. The lifetime costs (direct and indirect) of patients with ID is estimated to be about \$1 million per person on average [3]. Common causes of ID include idiopathic genetic factors, infection or exposure to toxins during pregnancy, infant trauma at birth, malnutrition or metabolic disorders after birth and other unexplained causes [2,4–6].

Although there is no specific drug for the treatment of ID, there are a variety of rehabilitation programs. Early diagnosis is very important for patients' long-term rehabilitation and to learn social skills, and can predict the prognosis of patients and reduce unnecessary diagnostic experiments [7,8].

The American Association on Intellectual and Developmental Disability defines ID through measurements in three areas: intelligence (IQ), adaptive behavior and systems of supports afforded the individual [1]. The World Health Organization (WHO) defines

the severity of ID into 4 levels according to intelligence quotient (IQ) testing, i.e., mild ID (score range from 50 to 69), moderate ID (35 to 49), severe ID (20 to 34) as well as profound ID (below 20) [9].

IQ assessment needs to be conducted by highly trained doctors and is complex and time-consuming, requiring 60–90 min. In China, only the pediatrics or psychiatric departments of large hospitals have special outpatient services for intelligence assessment. However, these two departments have the most shortage of doctors, which results in the delayed diagnosis and subsequent intervention of many children. Limited by social, economic and medical restriction, insufficient attention has been paid to the diagnosis and rehabilitation of children with ID, which makes some children miss the best time for treatment. With the increase of age, there is a growing gap between these children and normal children, and there are many problems such as psychological and social adaptation, so it is very necessary to explore a simple, accurate and rapid method that can be used to screen ID.

With the advancement of artificial intelligence, related technologies have been utilized to achieve screening or preliminary diagnosis of some kinds of diseases [10]. Analyzing the subjects' phonetic and facial features to try to achieve disease diagnosis is becoming popular [11]. People with ID may have symptoms such as severe speech delay and facial deformities such as macrostomia and/or open mouth appearance [12]. Children with ID are at higher risk for speech and language disorders. Speech and language disorders are one of the key characteristics of people with ID and can have long term negative effects on a child's development if not treated early [13]. Some helpful clues for screening ID include delayed speech, dysmorphic features (minor anomalies), hypotonia of the extremities, general inability to do things for self, etc. [14].

Gurovich utilized computer vision and deep learning algorithms to develop a facial analysis framework, which extracted the facial features of hundreds of genetic syndromes by analyzing 2D face images. Then, the framework achieved 91% accuracy of top-10-possible-diseases in identifying 215 different genetic syndromes, which outperformed clinical experts in three different trials [15]. Abdul-Rahman analyzed 2D facial images using facial dysmorphology analysis technology, which evaluates the measurement ratio between different facial landmarks to determine whether there are deformity features. After comparing the performance of computer-based facial analysis technology against standard, manual examination in fetal alcohol spectrum disorders (FASD), the result showed that the facial dysmorphology analysis technology can potentially improve the diagnosis of alcohol-related neurodevelopmental disorder (ARND) by recognizing FASD-associated facial anomalies [16]. X-linked hypohidrotic ectodermal dysplasia is a kind of gene deficiency disease with a conspicuous facial phenotype. Hadj-Rabia designed an automated facial recognition system, which was non-invasive and achieved the diagnosis of ectodermal dysplasia for patients at all ages by analyzing their facial images [17]. 2D images could not represent patients' facial phenotype well, so more and more researchers utilize 3D facial features to represent patients' facial phenotype, which is a type of fundamental for disease diagnosis. Hallgrimsson explored whether syndromes can be diagnosed from 3D images of human faces, and the result showed that 3D facial imaging has considerable potential to facilitate syndrome diagnosis [18]. Gene defect diagnosis based on facial phenotype has become a new research hotspot.

Speech has been widely used in the auxiliary diagnosis of mental diseases [19]. For example, Karmele et al., proposed a non-linear multi-task method for Alzheimer's Disease detection based on automatic speech analysis [20]. Charalambos et al., aimed to analyze whether voice quality and speech fluency distinguished people with mild cognitive impairment from healthy individuals, and the results showed that there were significant differences between people with mild cognitive impairment and healthy individuals in parameters such as cepstral peak prominence, shimmer, articulation rate and averaged speaking time [21]. In order to detect depression and predict its severity with speech assistance, Emna Rejaibi et al., proposed a MFCC-based recurrent neural network framework

to perform the assessment of depression [22]. Lang He et al., combined handcrafted and deep-learned features to effectively measure the severity of depression from speech [23]. Ellen W. McGinnis et al., detected children with internalizing disorder from speech and analyzed the most discriminative speech features [24]. Meng et al., proposed a spontaneous speech-based framework, which merges mobile inverted bottleneck convolutional blocks and visual Transformer blocks, for screening mental retardation [25]. Liu et al., proposed a two stream Non-Local CNN-LSTM network to learn the features of upper body behavior and facial expression of patients to achieve preliminary screening of mental retardation [26]. In our study, the open-source tool Covara was used to extract some audio features, such as spectrum and formant, and the open-source tool OpenFace was used to extract information such as motion unit and eye gaze direction.

At present, the diagnosis of ID is based on clinical evaluation, which means clinicians need to evaluate the status of subjects through face-to-face communication. Providing an efficient and automatic screening method for ID without clinical evaluation is helpful for the diagnosis and early intervention of ID.

In this article, a benchmark data set has been established by collecting the child subjects' video data in a clinical setting. By extracting and analyzing the features of the children's voice and face, a new analysis system has been established for screening ID automatically. The contributions of this article are shown as follows:

- Benchmark Dataset: establishing a video data set for automatic screening ID between 6 years old and 17 years old;
- ID measurement based on CVI: By utilizing an open-source face analysis tools, high-quality 3D facial features are extracted, the subject's facial phenotype is measured with facial features and CVI and finally, an important reference for screening ID is produced;
- ID measurement based on Voice: By extracting multiple phonetic features from the subjects' audio, the correlation between acoustic features and ID is explored;
- Automatic Screening of ID: machine learning algorithms are utilized to analyze 3D facial features and phonetic features, and an analysis system is established to automatically screen children for ID.

However, we have to attack several challenges in order to achieve automatic ID screening for children between 7 years old and 16 years old. This article attempts to solve the related problems from the establishment of benchmark data set, extraction of high-quality 3D facial features and phonetic features, measurement of facial phenotype and the establishment of a decision-making mechanism for screening ID.

2. Materials and Methods

2.1. Dataset

The Wechsler intelligence scale for children—China revised edition (WISC-CR), which is adjusted based on WISC in order to match Chinese culture better, was used to evaluate the IQ of the subjects in clinical settings. WISC is an individually administered intelligence test for children between 7 years old and 16 years old, which can be completed without reading or writing [27]. The WISC consists of several subsets, such as Verbal Comprehension index (VCI), Perceptual Reasoning Index (PRI), Processing Speed Index (PSI) and Working Memory Index (WMI). The following four subsets were chosen to evaluate the cognitive ability of subjects, whose evaluation data constituted the benchmark dataset we used.

- Comprehension: questions about social situations or common concepts.
- Similarities: asking how two words are alike/similar.
- Picture Completion: children are shown artwork of common objects with a missing part and asked to identify the missing part by pointing and/or naming.
- Block Design: children put together red-and-white blocks in a pattern according to a displayed model. This is timed, and some of the more difficult puzzles award bonuses for speed.

Psychiatrists who have received professional training have been assigned to perform these evaluations in order to ensure the consistency of evaluation for different subjects. Meanwhile, the characteristics of WISC also guarantee the consistency of evaluation to some extent, even if the evaluation tasks have been performed by different psychiatrists [28].

During the evaluation of IQ, the subjects' audio-visual and social demographic data were collected simultaneously. A total of 147 children were evaluated and 128 children met the experimental requirements. Among those subjects who have met the experimental requirements, there were 92 male children (71.9%). The evaluation results showed that there were 9 severe patients, 24 moderate patients, 43 mild patients and 52 normal controls. An IP camera and voice recorder were used to collect audio and video data of each subject during the evaluation. The frame rate of the video was 30, and its resolution was 640×480 ; the audio consisted of dual channels, and its sampling rate was 48 kHz. The duration of evaluation for the subjects ranged from 10 to 34 min. The sociological data collected from the subjects mainly included gender, age, height, weight, place of residence, family medical history, whether the mother had specific diseases during pregnancy, whether there were abnormal situations at birth, and whether the subject had specific diseases as a newborn or infant. The ID label of each subject was determined after comprehensive consideration of the scale evaluation results and sociological data. Finally, the video data of 33 patients (i.e., patients with moderate and severe ID) and 37 normal controls were chosen to construct the benchmark data set. The video size of the data set is 1660 min, 520 min of which belong to 33 patients and the rest were recordings about the 37 normal controls. Table 1 shows the socio-demographic information of the subjects used to construct the dataset.

Table 1. Characteristics of the subjects.

Variables	Number (%) or Mean (SD)
Gender	
Male	49 (70.0)
Female	21 (30.0)
Age	9.63 (3.02)
Height	139.41 (19.21)
Weight	37.00 (14.43)
BMI	18.66 (4.80)

2.2. Analysis

- (1) Architecture: Figure 1 shows the architecture for screening ID. We processed WISC test videos of the subjects and extracted facial images and audio of the subjects. OpenFace and OpenSmile tools were used to extract facial features and phonetic features. Facial geometric features can be further transformed into CVI features. Furthermore, machine learning algorithms were utilized to establish a method for screening ID based on facial features and phonetic features.
- (2) 3D facial features: Openface2.0 is an open-source facial behavior analysis tool, which can implement facial landmark detection, head pose estimation, facial action unit recognition and eye-gaze estimation. It has been widely used in computer vision, affective computing and human–computer interaction [29]. In this study, Openface2.0 was used to extract 3D facial features from evaluation videos, including facial contour, eye-gaze [30] and head-pose [31]. The facial features of subjects were chosen, including 3D facial landmarks of the head, to analyze the facial phenotype of subjects. Figure 2a shows the 68 facial landmarks, which represent facial contour, eye shape, nose and mouth. Each landmark was represented by an $L(x, y, z)$ to indicate its position in 3D space. Subjects of different ages and genders possessed different scales of faces, which is not conducive to compare different faces directly. Therefore, the tool scaled the subject's face in 3D space, and the scaling ratio is represented by p_scale ; finally, the facial phenotypes of the subjects were compared at same scale. Figure 2b shows

the subjects' 2D faces images detected from videos, and Figure 2c shows the 3D facial landmarks extracted from the 2D face images. Combining the temporal information of neighbor frames, head-pose landmarks were extracted frame by frame from the videos. Neighbor frames contain subjects' action information, which helps to extract these landmarks more accurately. The head-pose features included three different features (p_{rx} , p_{ry} , p_{rz}), which measures the 3D rotation degree between the head and the IP camera. In order to boost the reliability of our algorithm, we filtered the data according to the condition of confidence ≥ 0.98 (how confident is the tracker in current landmark detection estimate), $|p_{rx}| \leq 0.5$, $|p_{ry}| \leq 0.25$ and $|p_{rz}| \leq 0.5$. Only those data meeting the above conditions, i.e., 22,602 data frames, were chosen for constructing the algorithm.

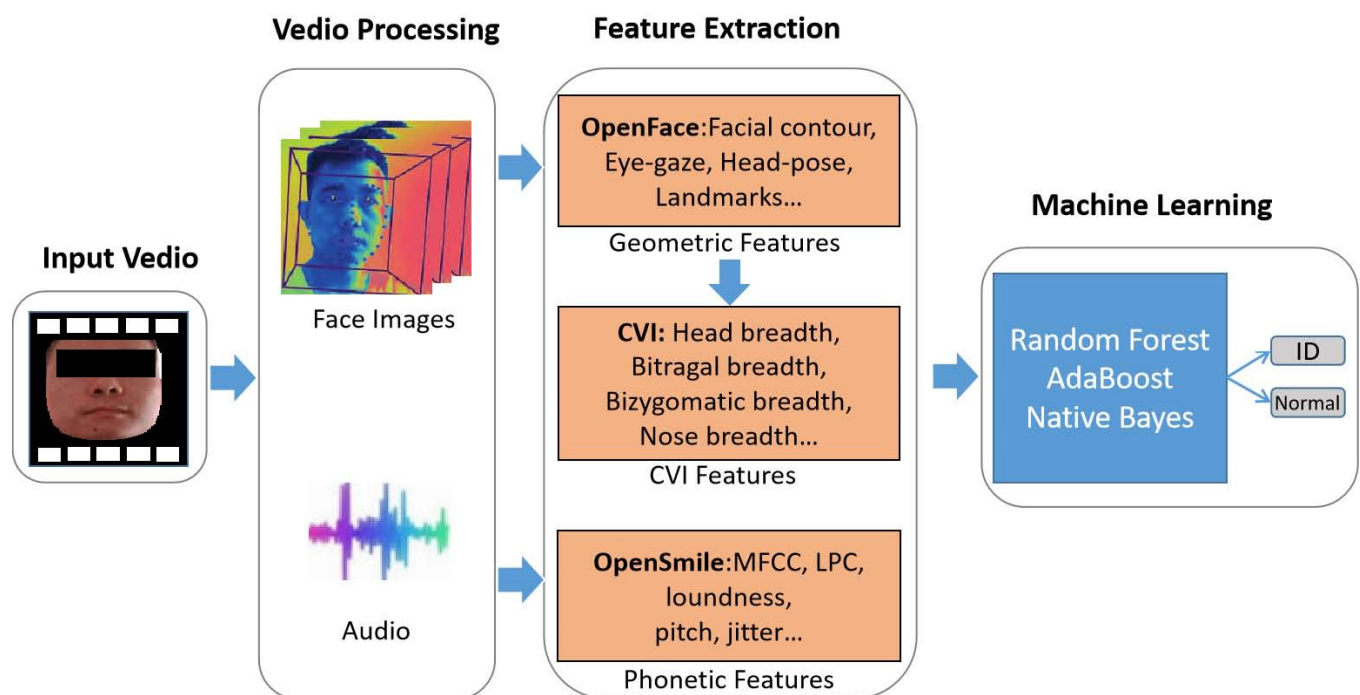


Figure 1. Architecture for screening ID.

- (3) **Phonetic Features:** Delays in speech development are common and may become more obvious when contrasted with the speech development of a sibling [14], which is a guidance for screening ID with phonetic features. Before extracting phonetic features, speech preprocessing is required, which mainly includes voice activity detection, speech enhancement and speaker-based speech segmentation. Voice activity detection can distinguish sound segments from silent segments in audio. The purpose of speech enhancement is to extract features as pure as possible from speech containing noise and improve the quality of speech. Speaker-based speech segmentation was mainly used to extract the audio of child subjects in order to improve the effectiveness of our algorithms. The speech segmentation based on Bayesian Information Criterion was adopted in this study [32].

Every audio was processed by the OpenSmile toolkit to extract the INTERSPEECH 2010 Paralinguistic Challenge feature set, which contains 1582 low-level features [33]. The INTERSPEECH 2010 Paralinguistic Challenge feature set consists of 34 low-level descriptors (LLDs) and 34 corresponding deltas as 68 LLDS contour values, and 1428 features can be obtained from 21 functions. In addition, for the four pitch-based LLDs and their four delta coefficient contours, 19 functions are used to obtain 152 features. Finally, the number of pitch onsets (pseudo syllables) and the duration of the total input (two features) are

appended [34]. The LLDS mainly include loudness, Mel Frequency Cepstral coefficients (MFCC), linear prediction coding (LPC) coefficients, jitter, shimmer and other phonetic features. Since the feature dimension is much larger than the sample size, kernel principal component analysis (KPCA) was used for feature dimension reduction, and radial basis function (RBF) was selected as the kernel function [35]. Finally, the phonetic dataset was reduced to 38 dimensions.

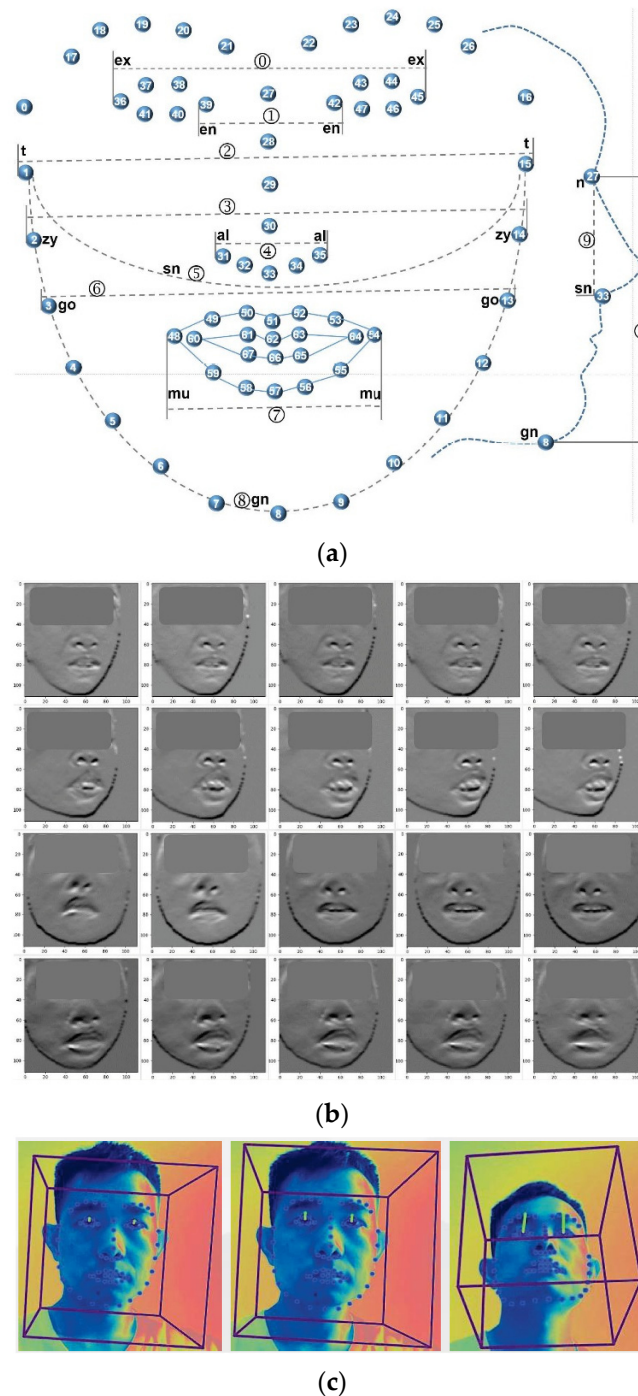


Figure 2. 3D facial feature extraction and facial phenotype definition. (a) The 11 Geometric Features referenced from CVI; (b) Face Detection; (c) 3D Feature Extraction.

- (4) Geometric Features and CVI: ID is often caused by gene deficiency syndrome, abnormal pregnancy, abnormal birth, brain injury, etc., which also often lead to abnormal

facial phenotypes. There is a certain extent of correlation between ID and facial phenotypes, so the severity of ID for subjects can be determined through the analysis of their facial phenotypes [36]. Facial landmarks can represent the facial phenotype of subjects to some extent, and the analysis results of facial landmarks may be utilized to judge the degree of ID. Using facial landmarks to define facial phenotypic abnormalities accurately is a key factor for the performance of the algorithm. Craniofacial variability index (CVI) has been utilized to describe, characterize, and evaluate craniofacial morphology, and has been widely used in evaluating dysmorphology, diagnosing auxiliary and assessing the effect of craniofacial surgery [37]. First, 16 characteristic measurements of the head and face are obtained, and each measurement is converted into a standardized z-score. The 16 z-scores are utilized to calculate standard deviation (SD, i.e., σ_z), which is the CVI score. Some studies have shown that the CVI of normal people has an approximate normal distribution, and the CVI of patients with craniofacial syndrome is significantly higher than a normal person; studies also showed that utilizing a subset of 16 characteristic measurements to calculate CVI can obtain similar conclusions [38,39]. Considering all videos collected by medical professionals capture the facial information of subjects, those characteristic measurements of the face were chosen to calculate the CVI, which included 11 geometric features, as shown in Figure 2a and Table 2. Table 2 defines the formulas for calculating the 30 geometric features, and the first 11 features of which were utilized to calculate CVI. The method of calculating σ_z , i.e., CVI, was given in [38].

Table 2. Geometric features of facial phenotype.

ID	Name	Formula	ID	Name	Formula	ID	Name	Formula
F01	ex-ex	$ P_{e08} - P_{e42} $	F11	n-gn	$ P_{f27} - P_{f8} $	F21	h-re	$ P_{e17} - P_{e11} $
F02	en-en	$ P_{e14} - P_{e36} $	F12	w-le	$ P_{e08} - P_{e14} $	F22	w-nb	$ P_{f31} - P_{f32} - P_{f33} - P_{f34} - P_{f35} $
F03	t-t	$ P_{f01} - P_{f15} $	F13	w-re	$ P_{e36} - P_{e42} $	F23	h-nb	$ P_{f27} - P_{f28} - P_{f29} - P_{f30} $
F04	zy-zy	$ P_{f02} - P_{f14} $	F14	la-le	$\angle(P_{e17} - P_{e08} - P_{e11})$	F24	a-nb	$\angle(P_{f31} - P_{f30} - P_{f35})$
F05	al-al	$ P_{f31} - P_{f35} $	F15	ra-le	$\angle(P_{e17} - P_{e14} - P_{e11})$	F25	ia-lm	$\angle(P_{f61} - P_{f60} - P_{f67})$
F06	t-sn-t	$ P_{f1} - P_{f33} - P_{f15} $	F16	la-re	$\angle(P_{e45} - P_{e36} - P_{e39})$	F26	oa-lm	$\angle(P_{f49} - P_{f48} - P_{f59})$
F07	go-go	$ P_{f03} - P_{f13} $	F17	ra-re	$\angle(P_{e45} - P_{e42} - P_{e39})$	F27	ia-rm	$\angle(P_{f63} - P_{f64} - P_{f65})$
F08	mu-mu	$ P_{f48} - P_{f54} $	F18	a-le-nb	$\angle(P_{e08} - P_{e14} - P_{e36})$	F28	oa-rm	$\angle(P_{f53} - P_{f54} - P_{f55})$
F09	t-gn-t	$ P_{f01} - P_{f08} - P_{f15} $	F19	a-re-nb	$\angle(P_{e42} - P_{e36} - P_{e14})$	F29	a-2e	$\angle(\text{vec}(P_{e14} - P_{e08}), \text{vec}(P_{e36} - P_{e42}))$
F10	n-sn	$ P_{f27} - P_{f33} $	F20	h-le	$ P_{e17} - P_{e11} $	F30	a-s	$\angle(\text{vec}(G_{lx}, G_{ly}, G_{lz}), \text{vec}(G_{rx}, G_{ry}, G_{rz}))$

Table 3 shows the distribution of CVI for the normal group and positive group. The CVI of the positive group was higher than that of normal group, which means that it may be feasible to screen ID using the CVI. However, there is no way to judge whether a subject with a certain CVI belongs to the positive group or not. To this end, we need to build a model to screen ID utilizing all the features we obtained, including CVI.

- (5) Machine Learning: The algorithms we chose to build the classification models for screening ID are Random Forest [40], AdaBoost [41] as well as Gaussian Naive Bayes [42]. All algorithms were implemented using the scikit-learn package [43] in Python.

Table 3. Percentile distribution of the CVI in the two groups.

Percentile	σ Nor	σ Pos	Percentile	σ Nor	σ Pos
5th	0.404	0.577	70th	0.951	1.334
10th	0.461	0.638	75th	1.016	1.427
15th	0.503	0.685	80th	1.085	1.521
20th	0.537	0.730	85th	1.173	1.631
25th	0.567	0.776	90th	1.297	1.807
50th	0.744	0.996	95th	1.502	2.479

3. Results

As shown in Table 4, non-CVI-related geometric features, all geometric features (including CVI) and phonetic features were utilized to establish a classification models for screening ID. Accuracy, precision, recall as well as the F1 score of the three algorithms for the three feature sets are given in Table 4. Based on geometric features, Native Bayes had the best performance compared to the other two algorithms, with an accuracy of 0.714 and F1 score of 0.715. Based on the fusion of CVI features and geometric features, Native Bayes also had the best performance, with an accuracy of 0.772 and F1 score of 0.749. Based on phonetic features, Native Bayes still had the best performance, with an accuracy of 0.796 and F1 score of 0.754. Among the three types of features, it was obvious that the best performance appeared when phonetic features were used for building the algorithms. The results also showed that phonetic features and all geometric features outperformed the non-CVI-related geometric features. In addition, the Gaussian Naive Bayes algorithm had the best performance among the three machine learning algorithms.

Table 4. Algorithm performance based on three different types of features.

	Accuracy	Precision	Recall	F1 Score
Classifier	Geometric Features (30)			
Random Forest	0.715	0.698	0.715	0.693
AdaBoost	0.653	0.644	0.653	0.648
Native Bayes	0.714	0.717	0.714	0.715
Classifier	CVI Features + Geometric Features (31)			
Random Forest	0.743	0.748	0.743	0.745
AdaBoost	0.715	0.714	0.715	0.715
Native Bayes	0.772	0.773	0.772	0.749
Classifier	Phonetic Features (38)			
Random Forest	0.759	0.677	0.677	0.677
AdaBoost	0.754	0.666	0.684	0.675
Native Bayes	0.796	0.690	0.832	0.754

The results with the best performance among 3 algorithms on three different feature sets are shown in bold font.

Next, we determined how to judge whether our proposed models are effective or not. An intuitive method is that the performance of these models should be at least better than random guesswork. In [44], a measurement was proposed to judge whether a classifier built on a particular data set performs better than random guesswork, and based on a custom data set, the classification accuracy of the proposed models must be larger than 0.646 if outperforms random guesswork. As shown in Table 4, the accuracy of the proposed models was almost 75%, which is 10 percentage points higher than 64.6%. The results showed that our work has the potential to be applied in clinical settings.

4. Discussion

Early diagnosis of ID is valuable because it allows the identification of children with risk, supportive counseling for parents, and potential stimulation programs for children. However, the diagnosis of ID in young children is frequently missed. Therefore, the automatic method for screening ID explored in this article provides a new perspective, which evaluates the risk of ID by analyzing subjects' phonetic features and facial phenotype. Developmental assessment should be a part of routine pediatric care for all preschool children [45], and the method we proposed is quite suitable in terms of massive throughput of assessment and feasible assessment for very young children. A developmental pediatrician or clinical psychologist should still perform a formal assessment once ID is suspected after assessment with proposed method.

Children should be examined closely for dysmorphic features or minor abnormalities, such as unusual eyebrow patterns, eyes that are widely or closely spaced, low-set ears or abnormal palmar crease patterns. Minor abnormalities are defined as defects that have unusual morphologic features [46]. Minor physical abnormalities involve the head, eyes, ears, hands, mouth or feet, and are readily recognized even on simple examination [47]. If children have head circumferences that falls below the 5th percentile (microcephaly) or above the 95th percentile (macrocephaly), ID is suspected [12]. The presence of three or more minor abnormalities in newborns is correlated with a 90% risk of coexistent major abnormalities [48].

The study of [49] showed that the presence of ID is closely related to the level of speech ability. ID is the most common factor in speech delay [50,51]. Hearing loss and speech dystonia are common in patients with ID. The speech of children with ID is influenced not only by their cognitive impairment, but also by certain specific factors [52]. In general, the more severe the ID, the slower the acquisition of communicative speech. In children with ID, the development of language is relatively more delayed than other aspects of development [50]. Incorrect pronunciation and slurred speech are one of the clinical manifestations of speech retardation. Our findings suggest that the phonetic features of children with ID are somewhat distinguishable from normal controls.

In the future, more and more video data of new subjects should be collected, so as to improve the performance of our proposed method. Furthermore, the authors will try to establish more efficient methods from the perspective of dynamic analysis of subjects' facial behavior patterns for screening ID. A valuable future study that we have planned is to mock the evaluation process done by a pediatrician into an applet of WeChat, which enables guardians help subjects to complete professional evaluation which usually done by pediatricians, so as to enable large-scale screening.

Author Contributions: Conceptualization, J.Y.; methodology, Y.C. and J.Y.; software, Y.C.; validation, S.M. and X.Y.; formal analysis, Y.C. and J.Y.; investigation, S.M. and D.L.; resources, J.Y. and D.L.; data curation, S.M. and D.L.; writing—original draft preparation, Y.C. and J.Y.; writing—review and editing, Y.C., X.Y. and J.Y.; visualization, J.Y.; supervision, J.Y. and D.L.; project administration, J.Y.; funding acquisition, J.Y. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Natural Science Foundation of China (grant number: 62002269) and the National Key R&D Program of China (grant number: 2018YFC1314600).

Institutional Review Board Statement: This study was approved by the ethics committee of Renmin Hospital of Wuhan University (WDRY2020-K191). All guardians have consented to this research on the behalf of all participants, including patients and normal controls. The individuals shown in Figure 2 have approved the publication of Figure 2.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author, Yang J. The data are not publicly available due to their containing information that could compromise the privacy of the research participants.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Schalock, R.; Luckasson, R.A.; Shogren, K.A. The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellect. Dev. Disabil.* **2007**, *45*, 116–124. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schaefer, G.; Bodensteiner, J. Evaluation of the child with idiopathic mental retardation. *Pediatr. Clin. N. Am.* **1992**, *39*, C929–C943. [\[CrossRef\]](#) [\[PubMed\]](#)
- Centers for Disease Control and Prevention (CDC). Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment—United States 2003. *MMWR Morb. Mortal. Wkly. Rep.* **2004**, *53*, 57–59.
- Amor, D. Investigating the child with intellectual disability. *J. Paediatr. Child Health* **2018**, *54*, 1154–1158. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chiurazzi, P.; Pirozzi, F. Advances in understanding genetic basis of intellectual disability. *F1000Research* **2016**, *5*. [\[CrossRef\]](#)
- Milani, D.; Ronzoni, L.; Esposito, S. Genetic Advances in Intellectual Disability. *J. Pediatr. Genet.* **2015**, *4*, 125–127.
- Schuit, M.; Segers, E.; Balkom, H.; van Balkom, H.; Verhoeven, L. Early language intervention for children with intellectual disabilities: A neurocognitive perspective. *Res. Dev. Disabil.* **2011**, *32*, 705–712. [\[CrossRef\]](#) [\[PubMed\]](#)
- Moeschler, J.; Shevell, M. Comprehensive evaluation of the child with intellectual disability or global developmental delays. *Pediatrics* **2014**, *134*, e903–e918. [\[CrossRef\]](#)
- WHO. *ICD-10 Guide for Mental Retardation*; World Health Organization: Geneva, Switzerland, 1996.
- Kulkarni, S.; Seneviratne, N.; Baig, M.S.; Khan, A.H.A. Artificial intelligence in medicine: Where are we now? *Acad. Radiol.* **2020**, *27*, 62–70. [\[CrossRef\]](#)
- Jin, B.; Cruz, L.; Goncalves, N. Deep Facial Diagnosis: Deep Transfer Learning from Face Recognition to Facial Diagnosis. *IEEE Access.* **2020**, *8*, 123649–123661. [\[CrossRef\]](#)
- Torring, P.M.; Larsen, M.J.; Brasch-Andersen, C.; Krogh, L.N.; Kibæk, M.; Laulund, L.; Illum, N.; Dunkhase-Heinl, U.; Wiesener, A.; Popp, B.; et al. Is MED13L-related intellectual disability a recognizable syndrome? *Eur. J. Med. Genet.* **2019**, *62*, 129–136. [\[CrossRef\]](#) [\[PubMed\]](#)
- Memisevic, H.; Hadzic, S. Speech and language disorders in children with intellectual disability in Bosnia and Herzegovina. *Disabil. CBR Incl. Dev.* **2013**, *24*, 92–99. [\[CrossRef\]](#)
- Daily, D.K.; Ardinger, H.H.; Holmes, G.E. Identification and evaluation of mental retardation. *Am. Fam. Physician* **2000**, *61*, 1059–1067. [\[PubMed\]](#)
- Gurovich, Y.; Hanani, Y.; Bar, O.; Nadav, G.; Fleischer, N.; Gelbman, D.; Basel-Salmon, L.; Krawitz, P.M.; Kamphausen, S.B.; Zenker, M.; et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat. Med.* **2019**, *25*, 60–64. [\[CrossRef\]](#) [\[PubMed\]](#)
- Valentine, M.; Bihm, D.C.; Wolf, L.; Hoyme, H.E.; May, P.A.; Buckley, D.; Kalberg, W.; Abdul-Rahman, O.A. Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders. *Pediatrics* **2017**, *140*, e20162028. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hadj-Rabia, S.; Schneider, H.; Navarro, E.; Klein, O.; Kirby, N.; Huttner, K.; Wolf, L.; Orin, M.; Wohlfart, S.; Bodemer, C.; et al. Automatic recognition of the XLHED phenotype from facial images. *Am. J. Med. Genet. Part A* **2018**, *173*, 2408–2414. [\[CrossRef\]](#)
- Hallgrímsson, B.; Aponte, J.D.; Katz, D.C.; Bannister, J.J.; Riccardi, S.L.; Mahasuwan, N.; McInnes, B.L.; Ferrara, T.M.; Lipman, D.M.; Neves, A.B.; et al. Automated syndrome diagnosis by three-dimensional facial imaging. *Genet. Med.* **2022**, *22*, 1682–1693. [\[CrossRef\]](#)
- Benevides, T.W.; Shore, S.M.; Palmer, K.; Duncan, P.; Plank, A.; Andresen, M.L.; Caplan, R.; Cook, B.; Gassner, D.; Hector, B.L.; et al. Listening to the autistic voice: Mental health priorities to guide research and practice in autism from a stakeholder-driven project. *Autism* **2020**, *24*, 822–833. [\[CrossRef\]](#)
- Karme, L.; Unai, M.; Calvo, P.M.; Jiri, M.; Blanca, B.; Nora, B.; Ainara, E.; Milkell, T.; Mirian, E.-T. Advances on Automatic Speech Analysis for Early Detection of Alzheimer Disease: A Non-linear Multi-task Approach. *Curr. Alzheimer Res.* **2018**, *15*, 139–148.
- Charalambos, T.; Eckerström, M.; Kokkinakis, D. Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. *PLoS ONE* **2020**, *15*, e0236009.
- Rejaibi, E.; Komaty, A.; Meriaudeau, F.; Agrebi, S.; Othmani, A. MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech. *Biomed. Signal Process. Control* **2022**, *71*, 103107. [\[CrossRef\]](#)
- He, L.; Cao, C. Automated depression analysis using convolutional neural networks from speech. *J. Biomed. Inform.* **2018**, *83*, 103–111. [\[CrossRef\]](#) [\[PubMed\]](#)
- McGinnis, E.W.; Anderau, S.P.; Hruschak, J.; Gurchiek, R.D.; Lopez-Duran, N.L.; Fitzgerald, K.; Rosenblum, K.L.; Muzik, M.; McGinnis, R.S. Giving voice to vulnerable children: Machine learning analysis of speech detects anxiety and depression in early childhood. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 2294–2301. [\[CrossRef\]](#) [\[PubMed\]](#)
- Meng, W.; Zhang, Q.; Ma, S.; Cai, M.; Liu, D.; Liu, Z.; Yang, J. A lightweight CNN and Transformer hybrid model for mental retardation screening among children from spontaneous speech. *Comput. Biol. Med.* **2022**, *151*, 106281. [\[CrossRef\]](#)

26. Liu, Q.; Cai, M.; Liu, D.; Ma, S.; Zhang, Q.; Liu, Z.; Yang, J. Two stream Non-Local CNN-LSTM network for the auxiliary assessment of mental retardation. *Comput. Biol. Med.* **2022**, *147*, 105803. [\[CrossRef\]](#)
27. Wechsler, D.; Kodama, H. *Wechsler Intelligence Scale for Children*; Springer: New York, NY, USA, 2011.
28. Ryan, J.; Glass, L.A.; Bartels, J.M. Internal Consistency Reliability of the WISC-IV among Primary School Students. *Psychol. Rep.* **2009**, *104*, 874–878. [\[CrossRef\]](#)
29. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial Behavior Analysis Toolkit. In Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, 15–19 May 2018; pp. 59–66.
30. Wood, E.; Baltrusaitis, T.; Zhang, X.; Sugano, Y.; Robinson, P.; Bulling, A. Rendering of Eyes for Eye-Shape Registration and Gaze Estimation Erroll Wood. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3756–3764.
31. Zadeh, A.; Chong Lim, Y.; Baltrusaitis, T.; Morency, L.P. Convolutional experts constrained local model for 3d facial landmark detection. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2519–2528.
32. Chen, S.; Gopalakrishnan, P. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In *Proc. DARPA Broadcast News Transcription and Understanding Workshop*; Morgan Kaufmann Publishers: Burlington, MA, USA, 1998; Volume 8, pp. 127–132.
33. Eyben, F.; Wöllmer, M.; Schuller, B. Opensmile: The munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 1459–1462.
34. Jassim, W.A.; Paramesran, R.; Harte, N. Speech emotion classification using combined neurogram and INTERSPEECH 2010 paralinguistic challenge features. *IET Signal Process.* **2017**, *11*, 587–595. [\[CrossRef\]](#)
35. Mika, S.; Schölkopf, B.; Smola, A.; Müller, K.R.; Scholz, M.; Rätsch, G. Kernel PCA and De-noising in feature spaces. *Adv. Neural Inf. Process. Syst.* **1998**, *11*, 536–542.
36. Giliberti, A.; Currò, A.; Papa, F.T.; Frullanti, E.; Ariani, F.; Coriolani, G.; Grosso, S.; Renieri, A.; Mari, F. MEIS2 Gene Is Responsible for Intellectual Disability, Cardiac Defects and a Distinct Facial Phenotype. *Eur. J. Med. Genet.* **2020**, *63*, 103627. [\[CrossRef\]](#)
37. Ward, R.; Jamison, P.; Farkas, L. Craniofacial variability index: A simple measure of normal and abnormal variation in the head and face. *Am. J. Med. Genet.* **1998**, *80*, 232–240. [\[CrossRef\]](#)
38. Ozdemir, M.B.; Ilgaz, A.; Dilek, A.; Ayten, H.; Esat, A. Describing Normal Variations of Head and Face by Using Standard Measurement and Craniofacial Variability Index (CVI) in Seven-Year-Old Normal Children. *J. Craniofacial Surg.* **2007**, *18*, 470–474. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Roelfsema, N.M.; Hop, W.C.J.; Van Adrichem, L.N.A.; Wladimiroff, J.W. Craniofacial Variability Index in Utero: A ThreeDimensional Ultrasound Study. *Ultrasound Obstet. Gynecol.* **2007**, *29*, 258–264. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Breiman, L. Random Forests. *Mach. Learn. Arch.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
41. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-Class AdaBoost*. *Stat. Its Interface* **2009**, *2*, 349–360. [\[CrossRef\]](#)
42. Rennie, J.D.; Shih, L.; Teevan, J.; Karger, D.R. Tackling the poor assumptions of naive bayes text classifiers. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 616–623.
43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
44. Müller-Putz, G.; Scherer, R.; Brunner, C.; Leeb, R.; Pfurtscheller, G. Better than random: A closer look on BCI results. *Int. J. Bioelectromagn.* **2008**, *10*, 52–55.
45. Coplan, J. Three pitfalls in the early diagnosis of mental retardation. *Clin. Pediatr.* **1982**, *21*, 308–310. [\[CrossRef\]](#)
46. Jones, K.; Jones, M.; Campo, M. *Smith's Recognizable Patterns of Human Malformation*, 8th ed.; Elsevier: Amsterdam, The Netherlands, 2021.
47. Ulovec, Z.; Šošić, Z.; Škrinjar, I.; Atović, A.; Čivljak, M.; Szivovics, L. Prevalence and significance of minor anomalies in children with impaired development. *Acta Paediatr.* **2004**, *93*, 836–840. [\[CrossRef\]](#)
48. Marden, P.M.; Smith, D.W.; McDonald, M.J. Congenital anomalies in the newborn infant, including minor variations. *J. Pediatr.* **1964**, *64*, 357–371. [\[CrossRef\]](#)
49. Nordberg, A.; Miniscalco, C.; Lohmander, A.; Himmetlmann, K. Speech problems affect more than one in two children with cerebral palsy: Swedish populationbased study. *Acta Paediatr.* **2013**, *102*, 161–166. [\[CrossRef\]](#)
50. Leung, A.K.C.; Kao, C.P. Evaluation and management of the child with speech delay. *Am. Fam. Physician* **1999**, *59*, 3121. [\[PubMed\]](#)
51. Bleszynski, J.J. Longitudinal research on the development of speech of people with mild mental retardation. *US-China Educ. Rev.* **2016**, *6*, 249–256.
52. Dodd, B.; Thompson, L. Speech disorder in children with Down's syndrome. *J. Intellect. Disabil. Res.* **2001**, *45*, 308–316. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.