

1 Cover Page

2

3 Title: Automated Phonological Error Scoring for Children with Language and Hearing

4 Impairment

5

6 Authors:

7 ¹Simon Sundström and ¹Charalambos Themistocleous

8 ¹Department of Special Needs Education, Faculty of Educational Sciences, University of Oslo,

9 Oslo, Norway

10

11 Corresponding author: Simon Sundström, Department of Special Needs Education,

12 Faculty of Educational Sciences, University of Oslo, Oslo, Norway,

13 simon.sundstrom@isp.uio.no

14

15 Conflict of Interest: None

16 **Automated Phonological Error Scoring for Children with Language and Hearing**
17 **Impairment**

18 ¹Simon Sundström and ¹Charalambos Themistocleous

19 ¹Department of Special Needs Education,

20 Faculty of Educational Sciences, University of Oslo, Oslo, Norway

21

22 **Abstract**

23 **Purpose:** Phonological production impairments are prevalent in children with developmental
24 language disorder (DLD) and hearing impairment (HI). This study aims to quantify and compare
25 phonological errors in Swedish-speaking children using a novel automated assessment tool and
26 provide an automatic machine learning classification algorithm of children with DLD and HI to
27 age-matched controls based on phonological errors.

28 **Methods:** 72 Swedish-speaking children (29 with DLD, 14 with HI, and 29 typically developing)
29 participated. Phonological production was elicited using a 72-item confrontation naming task. A
30 novel tool was developed to calculate a composite phonological error score and specific scores for
31 different phonological errors (deletions, insertions, substitutions, and transpositions) from written
32 speech productions. This tool leverages the International Phonetic Alphabet (IPA) and a form of
33 the Normalized Damerau–Levenshtein Distance for accurate error analysis.

34 **Results:** The composite score successfully differentiated between children with DLD and
35 typically developing children, highlighting its sensitivity in detecting phonological impairment.
36 Machine learning models can accurately differentiate between children with and without language
37 disorders. However, children with DLD and HI differed in the phonemic deletion errors, which
38 suggests that their phonemic production is relatively similar.

39 **Conclusions:** Children with DLD and HI exhibit significantly higher phonological error rates
40 compared to typically developing peers. Children with HI and DLC are comparably impaired in
41 phonology (as manifested by the compositive phonological score). These findings highlight the
42 potential of machine learning for early identification and targeted intervention in language
43 disorders, improving outcomes for affected children and demonstrated the potential of a
44 multilingual tool for scoring phonological errors.

45

46

47 **1 Introduction**

48 Children with Developmental Language Disorder (DLD) experience difficulties in language
49 production and perception, often appearing early in childhood and potentially persisting into
50 adulthood. While the exact cause remains unknown, these challenges are primarily linked to
51 neurocognitive factors. Similarly, children with hearing impairment (HI) often encounter
52 difficulties in speech and language development. However, their challenges are directly related to
53 sensory input limitations, referred to as language disorder associated with hearing loss (Bishop et
54 al., 2017). Nevertheless, assuming the same prevalence of DLD as in the general pediatric
55 population (at least 7%; Norbury et al. (2016)), a substantial proportion of children with HI can
56 also be assumed to have language problems that are independent of their hearing ability, i.e., they
57 have DLD, or would have had if not for their HI. This means that assessment procedures used with
58 children with HI should be sensitive to whether language problems are related to their hearing
59 impairment or to other factors (Hardman et al., 2023).

60 Both children with DLD (Lancaster & Camarata, 2019) and children with HI (Lederberg
61 et al., 2013) demonstrate a wide range of individual differences in the severity and particular areas
62 of language and communication affected. DLD can manifest as difficulties in expressive language
63 (producing language) and/or receptive language (understanding language), impacting one or more
64 linguistic domains, such as phonology (speech sounds), grammar (sentence structure), semantics
65 (word meaning), or pragmatics (social use of language) (Bishop, 2017). Research has indicated
66 that a substantial proportion (up to 40%) of children with DLD experience speech sound
67 difficulties severe enough to warrant a diagnosis of speech sound disorder, with the majority (90%)
68 of these cases being phonological in nature (Ttofari Eecen et al., 2019).

69 While many children with HI develop spoken language on par with typically developing
70 peers, any degree of hearing loss poses a risk for delayed or atypical language development
71 (Tomblin et al., 2015), especially in the domains of phonology and grammar, where impaired
72 performance has been found in around 50% of children (Tuller & Delage, 2014). For
73 example, both children with DLD and HI may experience difficulties in language
74 perception, impacting their understanding of conversations and instructions. These language
75 challenges can have a cascading effect on a child's overall language abilities, impacting their verbal
76 expression, vocabulary acquisition, and narrative skills. Frustration arising from communication
77 difficulties can also lead to behavioral challenges in both DLD and HI populations. Not only does
78 insufficient assessment and management of language difficulties in childhood and adolescence
79 have consequences for communication but they likely also contribute to a higher prevalence of
80 psychosocial difficulties and lower academic outcomes in both children with DLD (Botting et al.,
81 2016; Ziegenfusz et al., 2022) and HI (de Jong et al., 2023; Sarant et al., 2015), hindering their
82 education and impeding their academic success.

83 With respect to Swedish, both children with DLD and children with HI have been found to
84 perform below their typically hearing peers on measures of phonological production accuracy. In
85 a cross-sectional study of Swedish-speaking four–six-year-old children, Sundstrom et al. (2019)
86 showed that children with DLD produced phonemes and stress patterns significantly less
87 accurately compared to controls with typical language development. This was found for immediate
88 repetition of real words and non-words, as well as for confrontation naming of pictures.
89 Interestingly, no significant correlation was found between the repetition of stress and tonal word
90 accents and measures like phonological production or receptive vocabulary Sundstrom et al.

91 (2019). However, a significant correlation emerged between the ability to repeat stress in real
92 words and expressive grammar skills.

93 Regarding Swedish-speaking children with HI, Nakeva Von Mentzer et al.
94 (2015) investigated non-word repetition ability in five–seven-year-olds with severe bilateral HI
95 who wore cochlear implants. Their findings suggested a lower ability of children with cochlear
96 implants to repeat the consonants, vowels, primary stress and number of syllables in nonwords,
97 compared to age-matched controls. Similar results for non-word repetition performance were
98 obtained by Sundström et al. (2018b) in a study comparing four- to six-year-old children with
99 moderate to severe HI who used cochlear implants or hearing aids to typically hearing age peers.
100 In addition, they found that consonants, vowels and stress patterns were significantly harder for
101 the children with HI to produce in a picture naming task (Sundström et al., 2018b).

102 Few studies have compared the phonological production ability of Swedish-speaking
103 children with DLD and children with HI. Ibertsson et al. (2008) investigated nonword repetition
104 ability in children aged 5–9 years. Consonant production accuracy were not found to differ
105 between children with mild–moderate HI who used hearing aids and children with SLI, while
106 children with cochlear implants performed significantly below the other groups. In a more recent
107 study, Sundström et al. (2018a) found no difference between children with DLD and children with
108 HI in the ability to correctly produce consonants and vowels when repeating words and nonwords,
109 or when naming pictures.

110 In summary, previous studies indicate that Swedish-speaking children with DLD and
111 children with HI may face considerable challenges in accurately producing consonants, vowels,
112 stress and tonal word accents. However, differentiating between these groups based on

113 phonological production performance may require more in-depth error analyses, rather than coarse
114 judgments of speech sounds as either correct or incorrect.

115 Consequently, screening and assessment procedures for children with HI need to be
116 capable of differentiating between language difficulties stemming from their hearing impairment
117 and those arising from other underlying factors, such as DLD (Hardman et al., 2023). This
118 distinction is crucial because it has implications for both diagnosis and treatment. If a child's
119 language difficulties are primarily attributable to their hearing impairment, interventions focused
120 on improving auditory access and providing appropriate language models will be most
121 beneficial. However, if DLD is also present, additional therapeutic approaches targeting specific
122 language skills and cognitive processes may be necessary. Furthermore, recognizing the potential
123 coexistence of DLD and HI highlights the importance of comprehensive assessment practices for
124 children with hearing loss. By employing tools and techniques sensitive to both auditory and
125 language processing, clinicians can gain a more accurate understanding of a child's unique needs
126 and develop tailored intervention plans that address their specific challenges, ultimately promoting
127 optimal language development outcomes. The present study focuses on an aspect of language use
128 that is often problematic for children with DLD and children with HI, namely phonological
129 production, defined as the ability to correctly produce speech sounds. Studying phonological errors
130 in these groups may yield important insights into not only what characterizes phonological
131 production ability in each group, but also concerning similarities and differences.

132

133 **The present study**

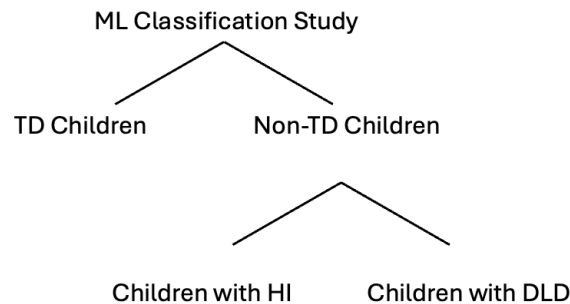
134 The manual scoring of phonological errors presents significant challenges within clinical and
135 research settings. It is a labor-intensive and time-consuming process, often susceptible to human

136 error and variability. Accurate and reliable analyses are contingent upon clinicians possessing
137 specialized expertise in phonetic transcription and a nuanced understanding of the speaker's
138 dialectal variations (Kent, 1996). These demands can strain resources and potentially introduce
139 inconsistencies in assessment outcomes. To address these limitations earlier research has
140 employed distance metrics to provide an Error Frequency Analysis in individuals with acquired
141 language disorders, such as apraxia of speech and aphasia with phonemic paraphasia (Smith et al.,
142 2019) and to score spelling differences in individuals with PPA (Themistocleous et al., 2020). For
143 example, in our previous study, Themistocleous et al. (2020) compared and evaluated several
144 distance metrics and showed that the *Normalized Damerau–Levenshtein Distance* replicates the
145 manual scoring process. The advantage of the *Normalized Damerau–Levenshtein Distance* over
146 other distance metrics is that it accounts for the transpositions of phonemes or graphemes, whereas
147 the simpler *Levenshtein Distance* only for deletion, insertion, and substitution errors.

148 To enhance the efficiency and objectivity of phonological assessment, this study proposes
149 an automated Composite Phonological Score (CPS) specifically designed for evaluating the
150 phonological performance of children with DLD. The CPS algorithm leverages a string distance
151 metric, the Normalized Damerau–Levenshtein Distance (Jurafsky & Martin, 2009), to compute
152 the phonemic dissimilarities between a child's production and the target utterance. This approach
153 enables the quantification of phonological accuracy at a phonemic level. We have refined the
154 algorithm to generate not only a composite score but also discrete scores for specific categories of
155 phonetic errors, namely deletions, insertions, substitutions, and transpositions. This metric
156 compares the response word and the target word. To account for phonemic processes that appear
157 at the surface level, we transcribe these both words in the International Phonetic Alphabet (IPA).
158 The IPA transcription involves both the representation of phonemes into words and phonological

159 processes involved in the production of the surface phonemic forms. In Swedish, for example, the
160 velar /k/ undergoes a change to become alveopalatal /ç/ when it appears before /i/ or /y/, /ø/, /ä/.
161 This can be observed in words like *kille* (/ˈçil:ɛ/) or *köpa* (/ˈçø:pa/). The phonemic distance is
162 calculated using the *Normalized Damerau–Levenshtein Distance*, which accounts for four types
163 of phonemic errors, namely, deletions, insertions, substitutions, and transpositions of phonemes
164 (Themistocleous et al., 2020). A multilingual web tool, utilizing the *Phonemic Distance Algorithm*,
165 offers support for scoring phonological errors in over 65 languages and language varieties, and is
166 available at Open Brain AI (<http://openbrainai.com>) (Themistocleous, 2024). The final score is a
167 distance metric from 0 when the words are identical to 1 when they are dissimilar. This scoring
168 system provides clinicians and researchers with a more comprehensive understanding of a child's
169 phonological profile, facilitating targeted intervention planning and progress monitoring.

170 This study has two primary objectives. First, we aim to quantify phonological production
171 differences in Swedish-speaking children with DLD and HI compared to Typically Developing
172 (TD) children. By establishing measurable benchmarks, we hope to improve early diagnosis and
173 intervention for these conditions. Second, we intend to develop a classification model to
174 differentiate between children with DLD, HI, and aged matched typically developing children
175 based on their phonological production patterns. Our primary hypothesis is that Swedish-speaking
176 children with DLD and HI will exhibit distinct phonological production profiles compared to TD
177 children and differ from each other. To evaluate this, we have designed a two-step machine
178 learning approach (Figure 1):



179

180 Figure 1 Two-fold ML Classification Approach: (1) distinguishing between typically developing
181 children (TD) and non-TD children; subsequently (2) classifying between children with HI and
182 children with DLS.

183 (1) *Classify TD vs. non-TD children*: This initial step differentiates typically developing
184 children from those with either DLD or HI.

185 (2) *Classify DLD vs. HI*: The second step focuses on distinguishing between children with
186 DLD and HI.

187 If successful, this automated system has the potential to expedite diagnosis, allowing clinicians to
188 quickly identify and address phonological difficulties, which is particularly vital for early and
189 effective intervention in language impairment cases.

190 **2 Methodology**

191 **2.1 Participants**

192 72 children, 29 with DLD, 14 children with HI and 29 age-matched, typically developing children,
193 participated in the study. All children were between 4 and 6 years old, speakers of the Central
194 Swedish dialect, and had no known neurodevelopmental or psychiatric disorders. Children with
195 DLD received a DLD diagnosis by a speech and language therapist and had mainly phonological
196 and grammatical, excluding pronounced lexical, semantic, and pragmatic deficits. The children
197 with HI all had congenital or early acquired bilateral mild to profound sensorineural hearing loss.

198 Three of the children had bilateral cochlear implants, nine used bilateral conventional air
 199 conduction hearing aids, and two had a bimodal solution with a hearing aid in one ear and a
 200 cochlear implant in the other.

201

Table 1 Demographic information of the participants for each group children with DLD, HI, and TD children for age, maternal education and their scores in neurolinguistic testing.

	Condition	Age	Mat. Edu	Word Rep	Non-Word Rep	Expres. Gramm	Recept. Vocab	Non-Verbal
Mean	DLD	59.7	14.9	0.57	0.485	28.6	77	24.4
	HI	60.3	14	0.579	0.467	23.1	56.8	22.9
	TD	61.2	16	0.828	0.73	34.2	90.6	25.8
Median	DLD	58	16	0.547	0.492	29	73	24
	HI	59	13.8	0.625	0.458	25.5	55.5	23
	TD	62	15	0.856	0.748	34	84	26
Mode	DLD	52	16	0.344	0.248	32	72	26
	HI	64	12	0.0949	0.0859	26	49	18
	TD	49	15	0.926	0.748	31	71	28
Standard deviation	DLD	7	2.25	0.11	0.12	5.31	19.1	6.05
	HI	7.47	2	0.22	0.208	7.29	17.8	4.12
	TD	8.04	2.22	0.114	0.134	4.75	20.7	3.68

202 Note: DLD: children with Developmental Language Disorder; TD: typically developing control
203 children; SD: Standard Deviation; CI: Confidence Intervals. HC and children with DLD differed
204 significantly in Age, Maternal Education (years), Word Repetition, Nonword Repetition,
205 Expressive Grammar, Receptive Vocabulary, and Non-Verbal Cognitive Ability. A Mann-
206 Whitney U test showed significant difference between children with DLD and HC for Nonword
207 Repetition, Expressive Grammar, Receptive Vocabulary at $p < .05$.

208 2.1.1 Age in Months

209 Given the rapid cognitive and linguistic changes in early life stages such as infancy and early
210 childhood, age in months allows for a more nuanced understanding of development.

211 2.1.2 Maternal Education Years

212 Maternal Education Years is a proxy for the home educational environment or socioeconomic
213 status. This metric aims to control for environmental variables that could influence cognitive or
214 linguistic development.

215 2.1.3 Non-Verbal Cognitive Ability

216 The Cognitive Ability score is a metric of non-verbal intelligence that attempts to encapsulate an
217 individual's general cognitive functions. It is assessed through the block design subtest of the
218 Swedish version of the Wechsler Preschool and Primary Scale of Intelligence, Third Edition
219 (WPPSI-III) (Wechsler, 2005) a standardized intelligence test.

220 2.1.4 Word and Nonword Repetition

221 Word and Nonword Repetition evaluates phonological memory and language processing abilities.
222 The individual is tasked with repeating real or made-up words, and the proportion of accurately
223 repeated phonemes is calculated. This measure is often used to diagnose children with DLD and
224 other language disorders.

225 2.1.5 Expressive Grammar

226 Expressive Grammar was assessed with Gramba (Hansson & Nettelbladt, 2004). The ability to use
227 noun morphology, verb morphology, and syntax are probed through a sentence completion test.
228 During this assessment, children are prompted to complete sentences, allowing researchers to
229 gauge their grasp of these fundamental aspects of language. To ensure accuracy and enable detailed
230 analysis, all responses were audio-recorded and later transcribed. The transcriptions were then
231 scored, with a maximum achievable score of 44. This scoring system provides a quantifiable
232 measure that enables the comparison of linguistic abilities across individuals or groups.

233 2.1.6 Receptive Vocabulary Raw

234 Receptive Vocabulary Raw scores aim to quantify the breadth of words an individual can
235 understand. Receptive Vocabulary was assessed using The Peabody Picture Vocabulary Test, or
236 PPVT–III, adapted in Swedish. The examiner pronounces a word, and the child is asked to point
237 to one of four pictures that best represents the meaning of that spoken word. Given the absence of
238 standardized norms for Swedish children between the ages of four and six, the study relies on the
239 raw max scores for analysis.

240

241 2.2 Procedures and Phonological Scoring

242 2.2.1 72-item confrontation naming task

243 Children were assessed for language impairment during one or two sessions (total duration 90–
244 120 minutes including suitable breaks). The testing took place in a quiet room at Linköping
245 University hospital, at the preschool, or in the child’s home. All testing was administered by the
246 first author. Phonological production was elicited with a 72-item confrontation naming task. In
247 this task, researchers assessed participants' ability to produce sounds and words accurately by
248 presenting them with pictures of objects and asking them to name the object depicted (Sundstrom
249 et al., 2019).

250 2.2.2 Automated Phonological Scoring

251 The Normalized Damerau–Levenshtein Distance is a distance metric, which allows the
252 comparison of two strings and calculates a composite score of the differences between the two
253 strings. A string can be a sequence of letters (graphemes), phonemes or another type of elements.
254 The Damerau–Levenshtein function ($dlev_{a,b}(i,j)$) is the distance between *two strings of*
255 *phonemes' words' a and b* by estimating the distance between the *i*–symbol of an initial substring
256 of the word *a* (prefix) and a *j*–symbol prefix of word *b*. The *restricted distance* is calculated
257 recursively as:

$$258 \quad dlev_{a,b}(i,j) = \min \begin{cases} 0 & \text{if } i = j = 0 \\ dlev_{a,b}(i-1,j) + 1 & \text{if } i > 0 \\ dlev_{a,b}(i,j-1) + 1 & \text{if } j > 0 \\ dlev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} & \text{if } i, j > 0 \\ dlev_{a,b}(i-2,j-2) + 1 & \text{if } i, j > 1 \text{ and } a[i] = b[j-1] \text{ and } a[i-1] = b[j] \end{cases}$$

259

260 where $1_{(a_i \neq b_j)}$ is equal to 0 when $a_i = b_j$ and is equal to 1 otherwise. The phonemic scoring

261 requires the match one of the following processes:

262

263 1. $dlev_{a,b}(i-1, j) + 1$: a deletion from word a to word b.

264 2. $dlev_{a,b}(i, j-1) + 1$: an insertion from word a to word b

265 3. $dlev_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)}$: a (mis-)match from word a to word b.

266 4. $dlev_{a,b}(i-2, j-2) + 1$: transposition

267

268 Typically, a score of 0 indicates perfect phonological production (i.e., no difference between the
269 target and the response), and higher scores indicate more errors.

270 For the purposes of this study, we have adapted the Normalized Damerau–Levenshtein
271 Distance algorithm to automatically compute the *Composite Phonological Score of Errors (CPS)*
272 metric from written speech productions. This adaptation enables the identification and
273 categorization of specific speech errors, *deletions* (i.e., the absence of phonemes that should be
274 present in the target production), *insertions* (i.e., the presence of extra phonemes not found in the
275 target production; the *substitutions* (i.e., the replacement of one phoneme with another), and
276 *transpositions* (i.e., the swapping of the order of two adjacent phonemes).

277 Our implementation involves a two-step process. The first step involves the Phonetic
278 Transcription. The written speech productions are first converted into their corresponding
279 representations using the *International Phonetic Alphabet (IPA)*. This standardized phonetic
280 notation allows for a precise comparison at the sound level, regardless of the original language or
281 orthography. The second step involves the Error Calculation. The IPA transcription of the target
282 production is then compared to the IPA transcription of the response. The modified Normalized

283 Damerau–Levenshtein Distance algorithm analyzes the differences between these two phonetic
284 sequences, identifying and classifying the specific error types (deletions, insertions,
285 substitutions, and transpositions). This information is then used to calculate the CPS
286 metric, providing a quantitative assessment of the accuracy of the speech production. The
287 Automated Phonological Scoring is approach was developed and implemented in *Open Brain AI*
288 (<http://openbrainai.com>) (Themistocleous, 2024) and provides a Multilingual Composite
289 Phonological Score of Errors and the individual phonemic errors adapted to several languages.

290 2.2.3 Statistical Analysis

291 Pearson and Spearman rank-order correlations were computed between CPS, phonemic errors, and
292 the other numerical variables of interest. Statistical tests were conducted to assess the differences
293 between the groups. The choice of statistical test (independent *t*-test or Mann-Whitney U) was
294 determined based on the Levene's test for equality of variances.

295

$$296 \quad DV \sim condition + (1 | child\ id) + (1 | task) \quad (1)$$

297

298 We employed linear mixed-effects models to evaluate the effects of children with DLD vs.
299 typically developing children on the phonological distance. The dependent variable was the
300 phonological distance and the *condition* (children with DLD vs. typically developing children)
301 were the predictors. The *participant* was modeled as a random slope to model participants'
302 differences. We computed the Linear Mixed-Effects Models in R, a programming language for
303 statistical analysis (R Core Team, 2020), using the “lme4” package (Bates et al., 2015) and
304 extracted the p values with the LmerTest package (Kuznetsova et al., 2016). Finally, post-hoc
305 contrasts were provided using the R package emmeans (EMMs, also known as least-squares

306 means), which provides estimated marginal means (Russell, 2020). A *t*-test was performed to
307 compare the duration of vowels and consonants produced by patients and clinicians.

308

309 2.3 Machine Learning

310 From our statistical models, we wanted to assess the implications of our measures for identifying
311 a child with DLD from children with TD. Therefore, we designed a robust classification approach
312 to work with a small amount of data. To assess the diagnostic value of phonological scores as
313 predictors of hearing Impairment and DLD, we implemented two machine learning approaches.
314 The first approach focused on differentiating typically developing children from those with hearing
315 Impairment or DLD based on their phonological errors. Subsequently, we explored the ability to
316 differentiate within the latter group, specifically between children with DLD and those with
317 hearing Impairment.

318 2.3.1 Model Comparison

319 For the classification we evaluated three classifiers: Decision Tree Classifier, Random Forest
320 Classifier, and Gradient Boosting Classifier. Each of these classifiers has unique characteristics
321 and is suited for several types of data and classification problems. The primary motivation for
322 using these models is that they are able to support smaller datasets and also provide feature
323 importance, which was critical for us because we wanted to assess the phonological measures
324 developed for this study.

325

326 *1. Decision Tree Classifier (DT)* is a simple, yet powerful model used in supervised
327 learning. For the classification it creates a tree-like graph of decisions by splitting the data
328 into branches at decision nodes, which are formed based on feature values. Each decision

329 node in the tree represents a feature in an instance to be classified, and each branch
330 represents a value that the node can assume. Instances are classified starting from the root
331 node and sorted based on their feature values down the tree along the branches until they
332 reach a leaf node, which provides the classification (Hastie et al., 2009).

333

334 *2. Random Forest Classifier (RF)* this is an ensemble learning method for classification
335 that operates by constructing a multitude of DTs at training time and outputting the class
336 that is the mode of the classes of the individual trees (Breiman, 2001). In other words, RFs
337 create a set of decision trees from randomly selected subsets of the training set, then
338 aggregate their individual predictions to make a final decision. It introduces additional
339 randomness when growing trees; instead of searching for the most important feature while
340 splitting a node, it searches for the best feature among a random subset of features. Because
341 of its ensemble nature, the random forest can achieve higher accuracy than a single DT and
342 is better at avoiding overfitting. Random forests handle unbalanced and missing data and
343 maintain accuracy for missing data.

344

345 *3. Gradient Boosting Classifier (GBC)* is an ensemble approach, like RFs, that builds
346 models sequentially, each new model attempting to correct errors made by the previous
347 ones (Bentéjac et al., 2021). Models are added one at a time, and existing models in the
348 sequence are not changed. It involves three parts: a Impairment function to be optimized,
349 a weak learner to make predictions, and an additive model to add weak learners to minimize
350 the Impairment function. New learners are created to be maximally correlated with the
351 negative gradient of the Impairment function associated with the whole ensemble.

352 2.3.2 Preprocessing and Addressing Data Imbalance

353 For the ML classification, we employed the automated phonological error scores for each word
354 (CPS and error types), demographic data (age and maternal education), and psychological
355 assessments (nonverbal IQ, word and nonword repetition abilities, and vocabulary skills) (Table
356 1). The individual scores of each word items of the test were consolidated into a single score
357 performance for each child by calculating median values for each child. Following data
358 preparation, we addressed potential imbalances in how frequently each condition occurs within
359 their dataset. Balancing a dataset is crucial for developing a reliable and fair predictive model as
360 class imbalance can lead the machine learning model to develop a bias toward the majority class,
361 often at the expense of poorly predicting the minority class. To counter this imbalance, we
362 employed SMOTE and works by creating samples by over-sampling with replacement (Chawla et
363 al., 2002). During *five*-fold cross-validation, SMOTE is applied within each training fold
364 separately, ensuring that each fold used for validation remains a true representation of the original
365 dataset. This approach provides a robust estimate of how the model will perform on unseen data.
366 We set up a machine learning pipeline to manage missing data (using median values to fill in gaps),
367 standardize measurements (to give them equal weight in the analysis), and apply the three
368 classifications classification algorithms to predict the condition based on the input data.

369 2.3.3 Training the models

370 The training of the models was conducted using a five-fold grouped cross-validation, which
371 considers the child id to ensure that there is no data leakage from the training test into the test set.
372 To tune the models and select the best one employed GridSearchCV, an approach implemented in
373 scikit-learn that automate the process of tuning parameters of a model by evaluating multiple

374 combinations of parameter tunes, cross-validate the results to find the best set of parameters that
375 offer the most accurate model.

376 A 5-fold cross-validation was used to split into 5 different subsets; each fold serves as a test
377 set once, with the remaining parts used as the training set, thus rotating through all combinations.
378 The performance of the final model is evaluated through standard statistical measures, namely
379 their accuracy, F1 score, and by generating a confusion matrix and a classification report that
380 explain how well the model predicts conditions and where it might be making errors. We then
381 extracted the feature importance from the final ML model. Each classifier was trained using the
382 training data from the group five-fold and then validated on the testing data. Post prediction, each
383 model's performance was assessed using standard evaluation metrics, namely the *accuracy*, which
384 is the proportion of correctly predicted classification; the F1 Score, which is the harmonic mean
385 of precision and recall; the precision, which is the proportion of positive identifications that were
386 indeed correct; the recall, which is the proportion of actual positives that were identified correctly;
387 the Area Under the Receiver Operating Characteristic Curve (ROC/AUC), and the Cohen's Kappa
388 (Hastie et al., 2009).

389 **3 Results**

390 In this section, first we discuss the differences between children with DLD, HI, and TD. Table 2
391 shows the Mean and Standard Deviation (SD) scores. Subsequently, we present the statistical
392 analysis for each of the outcome phonological measures.

393

394 Table 2 Mean and Standard Deviation (SD) scores.

Condition	CPS		Deletions		Insertions		Substitutions		Transpositions	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD

DLD	0.27	0.22	0.47	0.87	0.15	0.47	1.55	1.64	0.009	0.11
HI	0.28	0.26	0.63	1.2	0.15	0.45	1.49	1.78	0.007	0.092
TD	0.11	0.16	0.17	0.49	0.10	0.41	0.70	1.19	0.005	0.069

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development, CPS composite phonological score.

395
 396 Across all three groups (DLD, HI, and TD), substitutions are the most frequent type of phonemic
 397 error. Children with DLD and HI exhibit a similar rate of substitutions, which is notably higher
 398 than that of TD children. Deletions are the next most frequent error type, with the HI group
 399 showing the highest rate, followed by DLD, and then TD. Insertions occur at a relatively low rate
 400 across all groups, with DLD and HI children showing similar rates, slightly higher than TD
 401 children. Transpositions are the least common error, with minimal occurrences in all groups.
 402 Overall, children with DLD and HI demonstrate considerably higher rates of phonemic errors
 403 compared to TD children, suggesting greater challenges in phonological development.

404 405 **3.1 Phonological Errors: CPS, Deletions, Insertions, Substitutions, Transpositions**

406 Table 3 shows the comparative effects of the three groups, namely DLD, HI, and TD on the CPS.

407
 408 Table 3 Statistical analysis of the effect of the group on the CPS.

<i>Effect</i>	<i>Estimate</i>	<i>SE</i>	95% Confidence Intervals		<i>df</i>	<i>t</i>	<i>p</i>
			<i>Lower</i>	<i>Upper</i>			
(Intercept)	0.2648	0.0549	0.1573	0.3724	2.2	4.83	0.033

HI - DLD	0.0311	0.0311	-0.0298	0.092	69.33	1	0.32
TD - DLD	-0.1628	0.025	-0.2118	-0.1139	69.15	-6.52	<.001

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development.

409
 410 TD compared to children with DLD and HI differed significantly in their composite phonological
 411 score. Children with TD produced phonological errors closer to 0 whereas both children with DLD
 412 and HI had a composite phonological score closer to 30%. The difference in the productions was
 413 also shown in the post-hoc Bonferroni analysis, which showed significant differences between
 414 children with DLD and TD ($p < .001$) and HI and TD ($p < .001$) but no difference between children
 415 with DLD and HI.

416 3.1.1 Deletions

417 Table 4 shows the results of the statistical analysis of the effects of children with DLD, HI, and
 418 TD on the phonemic deletion errors.

419

420 Table 4 Statistical analysis of the effect of the group on the deletions.

<i>Effect</i>	<i>Estimate</i>	<i>SE</i>	95% Confidence Intervals		<i>df</i>	<i>t</i>	<i>p</i>
			<i>Lower</i>	<i>Upper</i>			
(Intercept)	0.551	0.1543	0.249	0.853	2.23	3.57	0.06
HI - DLD	0.185	0.0944	1.40E-04	0.37	66.88	1.96	0.054
TD - DLD	-0.315	0.0758	-0.464	-0.167	66.51	-4.16	<.001

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development.

421
422 Children with HI and DLD differed significantly from the TD children regarding deletions, which
423 is also evidenced by the higher number of deletions in children with DLD and HI compared to TD
424 as shown in Table 4. Specifically, we find that deletions occur more frequently in children with HI
425 compared to those with DLD. This difference is statistically significant ($p = 0.05$). More
426 pronounced variations are observed between children with HI and those with TD, as well as
427 between children with DLD and TD. These comparisons reveal significant differences, indicating
428 that deletions are relatively rare in children with TD.

429 3.1.2 Insertions

430 Table 5 shows the comparative effects of the three groups, namely DLD, HI, and TD on the
431 phonemic insertion errors.

432

433 Table 5 Statistical analysis of the effect of the group on the insertions.

<i>Effect</i>	<i>Estimate</i>	95% Confidence Intervals			<i>df</i>	<i>t</i>	<i>p</i>
		<i>SE</i>	<i>Lower</i>	<i>Upper</i>			
(Intercept)	0.17624	0.0651	0.0487	0.3038	2.04	2.708	0.111
HI - DLD	0.00253	0.0204	-0.0374	0.0424	67.59	0.125	0.901
TD - DLD	-0.04733	0.0162	-0.0792	-0.0155	65.77	-2.915	0.005

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development.

434

435 The post-hoc Bonferroni analysis showed that children with DLD differed significantly compared
436 to those with TD in insertions ($p = 0.015$), primarily because they tended to produce more insertion
437 errors. Similarly, children with HI also differed significantly from those with TD ($p = 0.052$).
438 However, there was no significant difference in the frequency of insertion errors between children
439 with DLD and those with HI.

440 3.1.3 Substitutions

441 Table 6 shows the comparative effects of the three groups, namely DLD, HI, and TD on the
442 phonemic insertion errors.

443

444 Table 6 Statistical analysis of the effect of the group on the phonemic substitution errors.

<i>Effect</i>	<i>Estimate</i>	<i>SE</i>	95% Confidence Intervals		<i>df</i>	<i>t</i>	<i>p</i>
			<i>Lower</i>	<i>Upper</i>			
(Intercept)	1.6132	0.513	0.608	2.618	2.05	3.146	0.085
HI - DLD	0.0498	0.148	-0.24	0.339	69.03	0.337	0.737
TD - DLD	-0.8347	0.119	-1.067	-0.602	68.61	-7.033	< .001

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development.

445

446 The post-hoc Bonferroni analysis showed that children with DLD differed significantly from those
447 with TD ($p < .001$) as children with DLD produced more substitution errors than TD children.
448 Children with HI also showed significant differences from those with TD ($p < .001$). However,
449 there was no significant difference in the incidence of insertion errors between children with DLD
450 and those with HI.

451 3.1.4 Transpositions

452 **Error! Reference source not found.** shows the comparative effects of the three groups, namely
453 DLD, HI, and TD on the phonemic transposition errors.

454

455 Table 7 Statistical analysis of the effect of the group on the phonemic transposition errors.

Effect	Estimate	SE	95% Confidence Intervals		df	t	p
			Lower	Upper			
(Intercept)	0.00887	0.0035	0.00201	0.01574	2.21	2.533	0.115
HI - DLD	-0.00111	0.00319	-0.00736	0.00514	68.65	-0.349	0.728
TD - DLD	-0.00372	0.00253	-0.00868	0.00124	66.22	-1.47	0.146

Notes: DLD = developmental language disorder, HI = hearing impairment, TD = typical development.

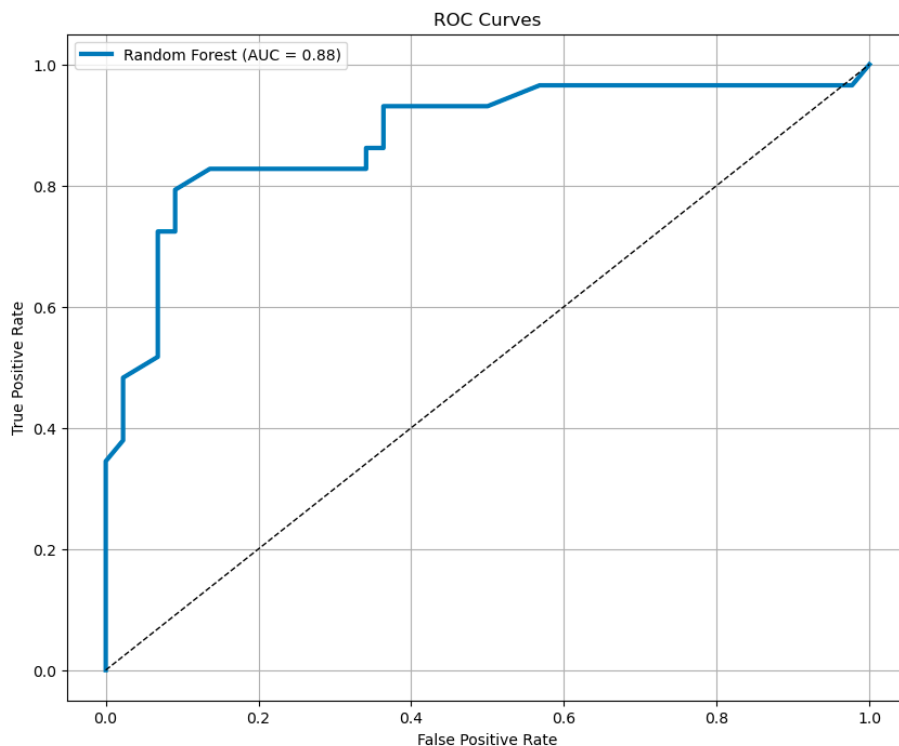
456

457 We found no errors with regards to transposition errors between the three groups.

458

459 **3.2 Classification 1: TD vs non-TD children**

460 We evaluated three classification models: Decision Trees, Random Forests, and Gradient
461 Boosting. The classification models were able to classify between TD children and non-TD
462 children. The best model, a Random Forests classifier, was able to perform the classification with
463 a relatively high classification accuracy and F1-score, i.e., 93%.



464

465 Figure 2 ROC / AUC Curve of the best performing model (AUC = 88%).

466 The ROC Curve of the best performing model, namely a RF model, which was 88%. Precision
467 informs the model's accuracy of positive predictions. The model's accuracy, precision and F1-
468 score approximated a mean 93% across folds, which indicates good overall performance across
469 both classes. We run a feature importance, which showed that the CPS had substantial contribution.

470

471 Table 2 The ranking of the first 5 features and their importance.

Ranking	Feature	Feature Importance
1	CPS	0.60
2	Expressive grammar	0.17
3	Non-Verbal	0.08
4	Non-Word Repetition	0.07

5	Receptive Vocabulary	0.04
---	----------------------	------

472

Note: Feature importance is total score of 1.

473

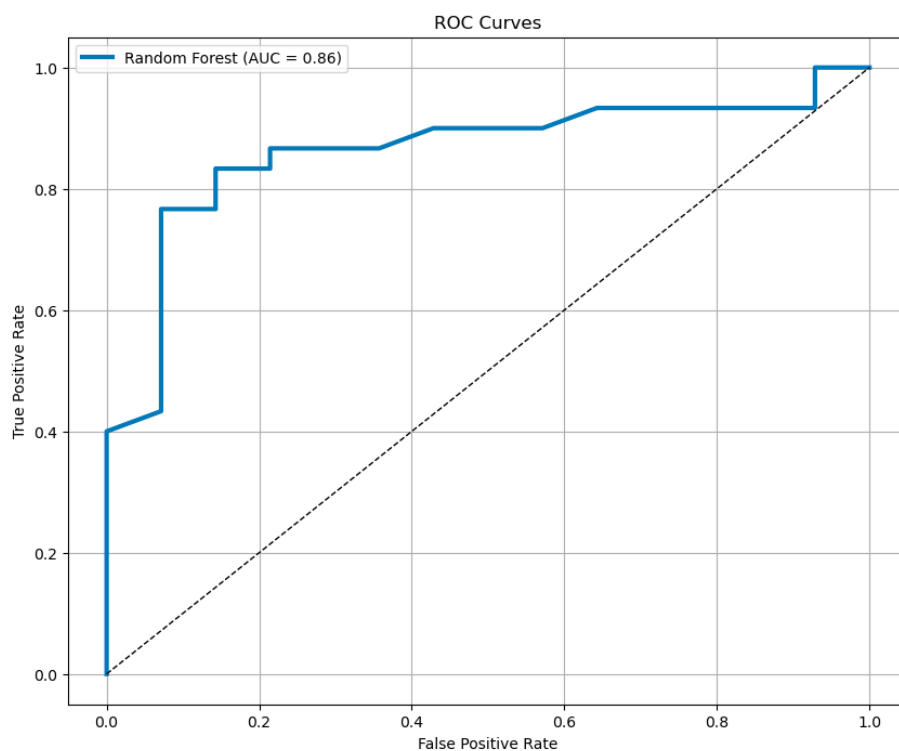
474 3.3 Classification 2: Children with DLD vs. HI

475 We followed the same model classification approach as in classification 1 and evaluated three

476 classification models: Decision Trees, Random Forests, and Gradient Boosting. The best model in

477 this case was a Decision Tree classifier and classified children with DLD and HI with 91%

478 classification accuracy and F1-score (Table 2) and ROC/AUC was 86% (Figure 2).



479

480 Figure ROC / AUC Curve of the best performing model (AUC = 86%).

481

482 Table 8 Classification results of the best performing model, a Random Forest model.

Label	Precision	Recall	F1-score
Accuracy			0.91
Macro Avg	0.89	0.91	0.90
Weighted Avg	0.92	0.91	0.91

483

484 Table 8 shows the precision, recall, F1-score, and support for each label, along with the
485 overall accuracy, macro average, and weighted average for the model. The model is highly
486 effective, especially in predicting class 1 with high precision and good recall. Class 0, while having
487 a lower precision, still had excellent recall. The overall accuracy and weighted metrics suggest that
488 the model performs robustly, making it quite effective for tasks where both classes are crucial. The
489 high F1-scores across both classes also indicate a strong balance between precision and recall,
490 crucial for many practical applications. The feature importance is shown in Table 9.

491

492 Table 9 The ranking of the first 5 features and their importance.

Ranking	Feature	Feature Importance
1	Receptive Vocabulary'	0.43
2	Word Repetition	0.39
3	Non-Verbal	0.09
4	Expressive Grammar'	0.08

493

494 The most importance feature was Receptive Vocabulary and Word Repetition. Unlike the previous
495 case, the role of the CPS to distinguish between individuals with DLD and HI was negligible.

496

497 **4 Discussion**

498 Traditional phonological assessments rely heavily on manual scoring, which is time-consuming,
499 labor-intensive, and prone to human error and variability. This study addresses the challenges of
500 manually scoring phonological errors by proposing an automated system that calculates phonemic
501 errors and introduces a novel approach by leveraging computational methods to standardize and
502 streamline this process (Heeringa et al., 2009; Hixon et al., 2011; Riches et al., 2011; Smith et al.,
503 2019). The new algorithm calculates a Compositive Phonological Score (CPS) from the
504 Normalized Damerau–Levenshtein Distance for phonemic error calculation and, at the same time,
505 introduces specific scores for individual errors like deletions, insertions, substitutions, and
506 transpositions (Themistocleous, 2024; Themistocleous et al., 2020). The proposed automated
507 system ensures consistency across assessments by applying the same criteria and calculations to
508 each case, eliminating the subjectivity inherent in manual scoring. This consistency is crucial for
509 reliable diagnosis and tracking of progress over time. Also, the automated scoring enhances
510 precision in phonemic error analysis, as the system can measure phonemic differences between a
511 child’s production and the target pronunciation, providing detailed insights that are difficult to
512 achieve manually.

513 Specifically, we have applied this algorithm to assess phonological performance in Swedish-
514 speaking children with DLD and HI, compared to TD children, and provided a detailed analysis
515 of phonological errors among Swedish-speaking children with DLD, HI, and TD children,
516 demonstrating distinct patterns of phonological errors across these groups (Sundström et al.,
517 2018a; Sundstrom et al., 2019). This study had two primary objectives. First, we aimed to quantify
518 phonological production differences in Swedish-speaking children with DLD and HI compared to
519 TD children. The study demonstrated that children with DLD and HI exhibited significantly higher

520 CPS than TD children, indicating more frequent phonological errors. DLD and HI groups showed
521 a CPS of around 0.27-0.28, while TD children had a lower CPS of 0.11. The findings suggest that
522 phonological errors are considerably more common in children with language impairments than
523 in typically developing peers (Andreou et al., 2024).

524 However, the CPS did not significantly differ between the DLD and HI groups, suggesting
525 similar levels of overall phonological difficulty in both groups. The lack of significant difference
526 between DLD and HI groups underscores the complexity of diagnosing and treating language
527 disorders that manifest similar phonological difficulties (Law et al., 1996). It highlights the need
528 for comprehensive and multifaceted approaches to diagnosis and intervention, the importance of
529 understanding these conditions' shared and unique aspects, and the potential challenges in clinical
530 practice. Also, the phonological difficulties in DLD and HI groups underscore that interventions
531 targeting phonological errors might need to be similar for both groups. However, since the
532 underlying causes of these errors differ (DLD being a language disorder and HI related to hearing
533 loss), interventions should still be tailored to address each group's specific needs and underlying
534 issues. For children with HI, interventions can focus more on improving auditory processing and
535 maximizing residual hearing through assistive technologies (e.g., hearing aids and cochlear
536 implants). In contrast, for children with DLD, therapy might emphasize enhancing language
537 processing and production skills (Hoover, 2019). The similarity in CPS underscores the need for
538 interventions to go beyond phonological errors and address the broader context of each disorder.

539 Concerning the specific types of Phonological Errors, substitutions were the most common
540 error across all groups, with DLD and HI children showing significantly higher rates than TD
541 children. Thus, replacing one phoneme with another is a prevalent issue in DLD and HI. Deletions
542 were more frequent in the HI group than in the DLD and TD groups, with the HI group showing a

543 trend towards more deletion errors than the DLD group. Both DLD and HI groups had significantly
544 more deletions compared to TD children. Insertions were relatively rare across all groups, but both
545 DLD and HI children exhibited higher rates than TD children. The differences between DLD and
546 TD were significant, highlighting that children with language disorders may add extra phonemes
547 do not present in the target word. Transpositions were the most minor common type of error and
548 did not show significant differences between the groups, indicating that swapping adjacent
549 phonemes is not a prominent issue in these populations. These findings indicate that despite
550 different etiologies, children with DLD and HI experience comparable difficulty levels in
551 phonological production, which suggests that phonological errors in these conditions may stem
552 from broader, shared difficulties in language processing, whether due to impaired auditory input
553 (in HI) or impaired linguistic processing (in DLD). Understanding these shared challenges can
554 inform the development of more effective cross-cutting intervention strategies.

555 Second, the study intended to develop a classification model to differentiate between
556 children with DLD, HI, and aged matched typically developing children based on their
557 phonological production patterns. We hypothesized that Swedish-speaking children with DLD and
558 HI will exhibit distinct phonological production profiles compared to TD children and differ from
559 each other. The hypothesis was confirmed, and the Classification of TD vs. Non-TD Children
560 resulted in a high classification accuracy. Specifically, the Random Forest classifier was the most
561 effective, achieving high classification accuracy (93%) and a strong F1 score. The CPS emerged
562 as the most crucial feature in distinguishing TD children from those with language impairments
563 (DLD and HI), underscoring the utility of the CPS metric in clinical diagnostics. The classification
564 results of DLD vs. HI Children showed that the Decision Tree classifier had the best performance,
565 with an accuracy of 91% and a high F1 score.

566 Interestingly, in this classification, the CPS was not a significant feature. Instead, Receptive
567 Vocabulary and Word Repetition were the most important predictors, which suggests that while
568 CPS is crucial for identifying general language impairment, other linguistic measures are more
569 critical in differentiating between specific disorders like DLD and HI.

570 The study offers a more objective and consistent method for assessing phonological
571 performance by quantifying phonological errors with the CPS. The method can reduce reliance on
572 labor-intensive manual scoring and minimize the risk of human error, making diagnostics more
573 precise and accessible. The study provides valuable insights into the phonological profiles of
574 children with DLD and HI. The several types of phonological errors highlight the specific
575 challenges faced by these groups, such as a higher prevalence of substitutions and deletions. These
576 findings can inform targeted intervention strategies that address the needs of children with these
577 language impairments. The study's application of machine learning models to classify children
578 based on their phonological errors is another significant contribution as it demonstrates the
579 potential that the combination of metrics employed in this study can corroborate existing screening
580 approaches for DLD and HI (see Bao et al. (2024) for a recent review). The ability to accurately
581 distinguish between TD, DLD, and HI children using phonological metrics and other linguistic
582 measures demonstrates the potential of AI in enhancing early diagnosis and intervention. This
583 approach can be particularly beneficial in resource-limited settings where access to specialized
584 clinicians is scarce.

585 **4.1 Limitations and Future Research**

586 While this study contributes significantly to phonological assessment and language disorder
587 diagnosis, several limitations should be noted. First, the study involved a relatively small sample
588 size of 72 children, which may limit the generalizability of the findings. Although the study

589 included three distinct groups (DLD, HI, and TD), the small number of participants in each group
590 may not fully capture the variability in the broader population of children with these conditions.
591 The study is also limited to Swedish-speaking children, specifically those speaking the Central
592 Swedish dialect. This focus restricts the applicability of the findings to other languages or dialects,
593 where phonological rules and error patterns may differ. Finally, the proposed method capture only
594 phonemic level differences, thus, domains such as lexical and post-lexical prosody, which might
595 be affect in children with HI and TD require a separate scoring algorithm (Sundstrom et al., 2019;
596 Themistocleous, 2016). Despite these limitations, the study offers valuable insights and tools for
597 assessing phonological errors in children with language disorders.

598 Future research should aim to address these limitations by including larger and more diverse
599 populations, exploring additional linguistic factors, and refining the machine learning models and
600 error analysis methods to enhance the robustness and applicability of the findings. By establishing
601 measurable benchmarks for phonological errors, the study contributes to the early diagnosis and
602 treatment of language disorders. Early identification of children with DLD or HI can lead to timely
603 interventions, potentially mitigating the long-term impacts of these disorders on language
604 development and academic achievement (Davidson et al., 2019). The study's implementation of
605 the phonological scoring algorithm in Open Brain AI (<http://openbrainai.com>), with multilingual
606 support, underscores its broad applicability (Themistocleous, 2024). This makes the tool feely
607 accessible to a wide range of clinicians and researchers, facilitating its integration into various
608 linguistic and cultural contexts. This generalizability enhances the tool's utility in global clinical
609 practice and research.

610 4.2 Conclusion

611 The study represents a significant advancement in the automated assessment of phonological
612 errors, offering a reliable, scalable tool for diagnosing and understanding language disorders. Its
613 integration of machine learning further enhances its diagnostic power, making it a valuable
614 resource for clinicians and researchers aiming to improve outcomes for children with language
615 impairments.

616 Data Availability Statement

617 The data are available upon request to the first author (SS). The Phonological Scoring Application,
618 developed by the second author, is freely accessible at Open Brain AI (<http://openbrainai.com>).

619 5 References

- 620 Andreou, G., Aslanoglou, V., Lymperopoulou, V., & Vlachos, F. (2024). Morphological and
621 phonological errors in the written language production of children with DLD.
622 *European Journal of Special Needs Education*, 39(2), 235-248.
623 <https://doi.org/10.1080/08856257.2023.2207055>
- 624 Bao, X., Komesidou, R., & Hogan Tiffany, P. (2024). A Review of Screeners to Identify Risk of
625 Developmental Language Disorder. *American Journal of Speech-Language
626 Pathology*, 33(3), 1548-1571. https://doi.org/10.1044/2023_AJSLP-23-00286
- 627 Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects
628 Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
629 <https://doi.org/10.18637/jss.v067.i01>
- 630 Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient
631 boosting algorithms. *Artificial Intelligence Review*, 54, 1937-1967.
- 632 Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of
633 developmental language disorder (DLD). *Int J Lang Commun Disord*, 52(6), 671-680.
634 <https://doi.org/10.1111/1460-6984.12335>
- 635 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & and the, C.-c. (2017).
636 Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study
637 of problems with language development: Terminology. *Journal of Child Psychology
638 and Psychiatry*, 58(10), 1068-1080.
639 <https://doi.org/https://doi.org/10.1111/jcpp.12721>

- 640 Botting, N., Toseeb, U., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2016). Depression and
641 Anxiety Change from Adolescence to Adulthood in Individuals with and without
642 Language Impairment. *PLoS One*, 11(7), e0156678.
643 <https://doi.org/10.1371/journal.pone.0156678>
- 644 Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
645 <https://doi.org/10.1023/A:1010933404324>
- 646 Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic
647 minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-
648 357.
- 649 Davidson, L. S., Geers, A. E., Hale, S., Sommers, M. M., Brenner, C., & Spehar, B. (2019).
650 Effects of Early Auditory Deprivation on Working Memory and Reasoning Abilities in
651 Verbal and Visuospatial Domains for Pediatric Cochlear Implant Recipients. *Ear and*
652 *Hearing*, 40(3). [https://journals.lww.com/ear-](https://journals.lww.com/ear-hearing/fulltext/2019/05000/effects_of_early_auditory_deprivation_on_working.8.a_spx)
653 [hearing/fulltext/2019/05000/effects_of_early_auditory_deprivation_on_working.8.a](https://journals.lww.com/ear-hearing/fulltext/2019/05000/effects_of_early_auditory_deprivation_on_working.8.a_spx)
654 [spx](https://journals.lww.com/ear-hearing/fulltext/2019/05000/effects_of_early_auditory_deprivation_on_working.8.a_spx)
- 655 de Jong, T. J., van der Schroeff, M. P., Stapersma, L., & Vroegop, J. L. (2023). A systematic
656 review on the impact of auditory functioning and language proficiency on
657 psychosocial difficulties in children and adolescents with hearing loss. *International*
658 *Journal of Audiology*, 1-11.
- 659 Hansson, K., & Nettelbladt, U. (2004). *Gramba. Grammatiktest för barn [Grammatical*
660 *assessment for children]*. . Pedagogisk Design.
- 661 Hardman, G., Herman, R., Kyle, F. E., Ebbels, S., & Morgan, G. (2023). Identifying
662 Developmental Language Disorder in Deaf Children with Cochlear Implants: A Case
663 Study of Three Children. *Journal of clinical medicine*, 12(17), 5755.
664 <https://doi.org/10.3390/jcm12175755>
- 665 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data*
666 *Mining, Inference, and Prediction*. (2 ed.). Springer.
- 667 Heeringa, W., Johnson, K., & Gooskens, C. (2009). Measuring Norwegian dialect distances
668 using acoustic features. *Speech Communication*, 51(2), 167-183.
669 <https://doi.org/10.1016/j.specom.2008.07.006>
- 670 Hixon, B., Schneider, E., & Epstein, S. L. (2011). Phonemic similarity metrics to compare
671 pronunciation methods. *12th Annual Conference of the International Speech*
672 *Communication Association, INTERSPEECH 2011 Florence*.
- 673 Hoover, J. R. (2019). Phonological treatment options for children with expressive language
674 impairment. *Seminars in Speech and Language*,
- 675 Ibertsson, T., Willstedt-Svensson, U., Radeborg, K., & Sahlén, B. (2008). A methodological
676 contribution to the assessment of nonword repetition—a comparison between
677 children with specific language impairment and hearing-impaired children with
678 hearing aids or cochlear implants. *Logopedics Phoniatrics Vocology*, 33(4), 168-178.
679 <https://doi.org/10.1080/14015430801945299>
- 680 Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to*
681 *natural language processing, computational linguistics, and speech recognition* (2
682 ed.). Pearson Prentice Hall.

- 683 Kent, R. D. (1996). Hearing and Believing. *American Journal of Speech-Language Pathology*,
684 5(3), 7-23. <https://doi.org/10.1044/1058-0360.0503.07>
- 685 Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *lmerTest*:
686 *Tests in Linear Mixed Effects Models*. In R Foundation for Statistical Computing.
687 <https://CRAN.R-project.org/package=lmerTest>
- 688 Lancaster, H. S., & Camarata, S. (2019). Reconceptualizing developmental language
689 disorder as a spectrum disorder: issues and evidence. *Int J Lang Commun Disord*,
690 54(1), 79-94. <https://doi.org/10.1111/1460-6984.12433>
- 691 Law, J., Garrett, Z., Nye, C., Cochrane Developmental, P., & Group, L. P. (1996). Speech and
692 language therapy interventions for children with primary speech and language delay
693 or disorder. *Cochrane Database of Systematic Reviews*, 2015(5).
- 694 Lederberg, A. R., Schick, B., & Spencer, P. E. (2013). Language and literacy development of
695 deaf and hard-of-hearing children: Successes and challenges. *Developmental*
696 *Psychology*, 49(1), 15-30. <https://doi.org/10.1037/a0029558>
- 697 Nakeva Von Mentzer, C., Lyxell, B., Sahlén, B., Dahlström, Ö., Lindgren, M., Ors, M.,
698 Kallioinen, P., Engström, E., & Uhlén, I. (2015). Segmental and suprasegmental
699 properties in nonword repetition – An explorative study of the associations with
700 nonword decoding in children with normal hearing and children with bilateral
701 cochlear implants. *Clinical Linguistics & Phonetics*, 29(3), 216-235.
702 <https://doi.org/10.3109/02699206.2014.987926>
- 703 Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., &
704 Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical
705 presentation of language disorder: evidence from a population study. *Journal of Child*
706 *Psychology and Psychiatry*, 57(11), 1247-1257.
707 <https://doi.org/https://doi.org/10.1111/jcpp.12573>
- 708 R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. In R
709 Foundation for Statistical Computing. <https://www.R-project.org/>
- 710 Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition
711 in adolescents with specific language impairment and autism plus language
712 impairments: a qualitative analysis. *J Commun Disord*, 44(1), 23-36.
713 <https://doi.org/10.1016/j.jcomdis.2010.06.003>
- 714 Russell, L. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. In R
715 Foundation for Statistical Computing. [https://CRAN.R-](https://CRAN.R-project.org/package=emmeans)
716 [project.org/package=emmeans](https://CRAN.R-project.org/package=emmeans)
- 717 Sarant, J. Z., Harris, D. C., & Bennet, L. A. (2015). Academic Outcomes for School-Aged
718 Children With Severe–Profound Hearing Loss and Early Unilateral and Bilateral
719 Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, 58(3), 1017-
720 1032. https://doi.org/10.1044/2015_JSLHR-H-14-0075
- 721 Smith, M., Cunningham, K. T., & Haley, K. L. (2019). Automating Error Frequency Analysis via
722 the Phonemic Edit Distance Ratio. *J Speech Lang Hear Res*, 62(6), 1719-1723.
723 https://doi.org/10.1044/2019_JSLHR-S-18-0423
- 724 Sundström, S., Löfkvist, U., Lyxell, B., & Samuelsson, C. (2018a). Phonological and
725 grammatical production in children with developmental language disorder and

- 726 children with hearing impairment. *Child Language Teaching and Therapy*, 34(3), 289-
727 302. <https://doi.org/10.1177/0265659018805202>
- 728 Sundström, S., Löfkvist, U., Lyxell, B., & Samuelsson, C. (2018b). Prosodic and segmental
729 aspects of nonword repetition in 4- to 6-year-old children who are deaf and hard of
730 hearing compared to controls with normal hearing. *Clinical Linguistics & Phonetics*,
731 32(10), 950-971. <https://doi.org/10.1080/02699206.2018.1469671>
- 732 Sundstrom, S., Lyxell, B., & Samuelsson, C. (2019). Prosodic aspects of repetition in
733 Swedish-speaking children with developmental language disorder. *Int J Speech Lang*
734 *Pathol*, 21(6), 623-634. <https://doi.org/10.1080/17549507.2018.1508500>
- 735 Themistocleous, C. (2016). Seeking an Anchorage. Stability and Variability in Tonal
736 Alignment of Rising Prenuclear Pitch Accents in Cypriot Greek. *Language and*
737 *Speech*, 59(4), 433-461. <https://doi.org/10.1177/0023830915614602>
- 738 Themistocleous, C. (2024). Open Brain AI and language assessment. *Frontiers in Human*
739 *Neuroscience*, 18. <https://doi.org/10.3389/fnhum.2024.1421435>
- 740 Themistocleous, C., Neophytou, K., Rapp, B., & Tsapkini, K. (2020). A tool for automatic
741 scoring of spelling performance. *Journal of Speech, Language, and Hearing*
742 *Research*, 63, 4179-4192. [https://doi.org/https://doi.org/10.1044/2020_JSLHR-20-](https://doi.org/10.1044/2020_JSLHR-20-00177)
743 [00177](https://doi.org/10.1044/2020_JSLHR-20-00177)
- 744 Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P.
745 (2015). Language Outcomes in Young Children with Mild to Severe Hearing Loss. *Ear*
746 *and Hearing*, 36. [https://journals.lww.com/ear-](https://journals.lww.com/ear-hearing/fulltext/2015/11001/language_outcomes_in_young_children_with_mild_to_8.aspx)
747 [hearing/fulltext/2015/11001/language_outcomes_in_young_children_with_mild_to_](https://journals.lww.com/ear-hearing/fulltext/2015/11001/language_outcomes_in_young_children_with_mild_to_8.aspx)
748 [8.aspx](https://journals.lww.com/ear-hearing/fulltext/2015/11001/language_outcomes_in_young_children_with_mild_to_8.aspx)
- 749 Ttofari Eecen, K., Eadie, P., Morgan, A. T., & Reilly, S. (2019). Validation of Dodd's Model for
750 differential diagnosis of childhood speech sound disorders: A longitudinal
751 community cohort study. *Developmental Medicine and Child Neurology*, 61, 689 -
752 696. [https://doi.org/s://doi.org/10.1111/dmcn.13993](https://doi.org/10.1111/dmcn.13993)
- 753 Tuller, L., & Delage, H. (2014). Mild-to-moderate hearing loss and language impairment:
754 How are they linked? *Lingua*, 139, 80-101.
755 [https://doi.org/https://doi.org/10.1016/j.lingua.2013.10.009](https://doi.org/10.1016/j.lingua.2013.10.009)
- 756 Wechsler, D. (2005). Wechsler preschool and primary scale of intelligence (Swedish
757 version). *Stockholm: Psykologiförlaget AB*.
- 758 Ziegenfusz, S., Paynter, J., Flückiger, B., & Westerveld, M. F. (2022). A systematic review of
759 the academic achievement of primary and secondary school-aged students with
760 developmental language disorder. *Autism & Developmental Language Impairments*,
761 7, 239694152210993-23969415221099397.
762 <https://doi.org/10.1177/23969415221099397>
- 763