Diagnosing Specific Language Impairment in Children

Abstract

Specific Language Impairment is one of the most common developmental disorders in children; however, it has a complicated and labor-intensive diagnosis process. Improving the diagnosis process can allow children to access the necessary help sooner, leading to better outcomes. In this study, we run a logistic regression model on a dataset with 1163 observations and predictors relating to the language skills and linguistic complexity of the children. Our final model predicts Specific Language Impairment in children in the validation set with a 83.9% accuracy.

Background and Significance

Specific Language Impairment (SLI), commonly referred to as speech delay, is one of the most prevalent developmental disorders that affect 7-10% of children in kindergarten [1]. It is characterized by a difficulty in acquiring and expressing language, despite the absence of any apparent cognitive or physical impairments. Early diagnosis is vital, as untreated SLI can have negative long-term effects in academics, social-life, and overall well-being [1]. However, the cause of SLI has yet to be identified, and SLI has historically lacked a standardized definition with no set diagnostic criteria [2, 3]. Current diagnostic practices are also labor-intensive and highly variable, relying on speech pathologists to compare a child's language skills to those of their peers [3]. Given the critical need for timely intervention, the inconsistency in diagnostic criteria, and the variability in evaluations, this study aims to examine patterns in characteristics of children with and without SLI to identify more reliable and efficient methods for detecting SLI.

Research Question: To what extent can various measures of language skills and linguistic complexity predict Specific Language Impairment in children?

Data

Data Description

Our data comes from three independent studies that employ narrative retellings prompted by wordless picture books—a widely validated method for identifying language impairments. The first study compared narrative abilities in adolescents with SLI with low versus high non-verbal IQ [4]. The second study examined whether story grammar units predicted SLI diagnosis [5]. The third study assessed how storytelling task complexity and various evaluation measures impact SLI identification [6]. Participant characteristics were recorded in each study and entered into the CHILDS project (an open-source repository of child language development data), which was compiled and later shared on Kaggle in 2017 [7]. This dataset includes 1163 total observations, 62 predictor variables (various measures of language ability), and a binary variable for whether the participant had SLI or not.

Data Cleaning

Before starting our modeling process, we cleaned the data. We found 119 missing values in the "sex" predictor. These missing values come from the first study, and since the female to male proportion differed across the 3 studies, it was not reasonable to impute these values, so we dropped these rows. Next, we removed x-variables that were not useful, such as "filename", "group", and "corpus". Finally, we handled multicollinearity issues in our dataset. We examined variables with a correlation of 1 and removed standardized versions of existing predictors for better interpretation. We also conducted a Variance Inflation Factor (VIF) analysis with a threshold of 9, identified 14 variables with multicollinearity issues, and removed them from the dataset. Our final dataset consisted of 1034 observations and 37 predictors.

Methods and Results

Fitting First Order Models

First, we split our data into training and validation sets, with about 75% of the data being used to fit the models, where 789 observations were in the train set, and 255 were in the validation set. We fit a logistic regression model on the training data using all the predictors, and then conducted stepwise regression for variable selection. We used stepwise regression with both AIC and BIC criteria, which suggested two different models (Model 1 and Model 2, respectively). Model 1 has 13 predictors whereas Model 2 has 11 predictors, and Model 2 is nested under Model 1.

To compare the performance of the two models, we calculated multiple performance metrics on the validation data (see Table 1). Since the two models were also nested, we conducted a Likelihood Ratio Test for model comparison and got a p-value of 0.0434. This suggests that Model 2 does not fit the dataset as well as Model 1. As such, we decided to use Model 1 as our final first order model, since it performed better on more metrics and was suggested by the Likelihood Ratio Test.

Considering Interactions

Since our final first order model (Model 1) only achieved an accuracy of 81.2% and sensitivity of 62.5%, we explored interaction terms in hopes of improving performance. Using stepwise regression on Model 1 with all interaction terms, we developed Model 3 (AIC criterion, 31 predictors) and Model 4 (BIC criterion, 11 predictors). We also created Model 5 by combining non-interaction terms from Model 3 and including only age-related interaction terms, totaling 16 predictors. We are interested in this model because the numerous interactions in Model 3 would be hard to interpret for a diagnostician, and our goal is to make diagnosing SLI easier. Moreover, the *age* variable alone is not meaningful–it reflects participants' age when they are assessed for their language abilities rather than the age of diagnosis. For instance, a positive correlation between *age* and *SLI* indicates that the researchers recruited more older than younger children with SLI, not that *age* predicts *SLI*. Furthermore, examining interactions between age and language measures provides diagnosticians with specific characteristics to focus on when evaluating children in particular age groups.

To determine the best model, we calculated performance metrics for each interaction model and found that Model 5 outperformed all models on 5 of the 6 metrics (see Table 1). While its sensitivity was slightly lower than Model 3, Model 5's greater parsimony makes it preferable. In addition, although Model 5 shows only modest improvement over our best first-order model (Model 1), a Likelihood Ratio Test indicates that Model 1 does not fit the dataset as well as Model 5, $p = 0.0145$. As such, our final model is Model 5.

Table 1

Performance metrics of the first-order (model 1-2) and interaction (model 3-5) models.

	AUC	Accuracy	Sensitivity	Specificity	Precision	F measure
Model 1	0.8781	0.812	0.625	0.8743	0.625	0.625
Model 2	0.8753	0.812	0.5938	0.8848	0.6333	0.6129
Model 3	0.8745	0.8314	0.6719	0.8848	0.6615	0.6667
Model 4	0.879	0.8078	0.5469	0.8953	0.6364	0.5882
Model 5	0.884	0.839	0.6563	0.9005	0.6885	0.672

Note. Model 1 is the best first order model based on the AIC criterion. Model 2 is the best first order model based on BIC criterion. Model 3 is the best interaction model using AIC criterion based on Model 1. Model 4 is the best interaction model using BIC criterion based on Model 1. Model 5 is like Model 3, but only includes interaction terms with age. These models are fitted and tested on the validation dataset.

Model Diagnostics

Before finalizing our model, we plotted the residuals of Model 5 and identified 5 outliers (see Figure 1 in Appendix). Next, to identify any influential observations or outliers, we calculated the delta deviance of every observation, plotted the deviance by index plot, and visually identified a threshold of 10 (see Figure 2 in Appendix). We found no influential outliers, so no outliers were removed, but we identified two influential observations above the threshold of 10. Upon examination, these two observations do not have special characteristics that limits the generalizability of this study. After removing these two influential observations, our final data included 1042 observations, and we refitted our final model with the updated dataset.

Discussion and Other Considerations

Conclusion

Our final model consisted of 16 predictors. Of the 16 predictors, only 11 of them were significant (see Table 1 in Appendix). Overall, our model indicates that as measures of lexical diversity and linguistic ability increase, the log odds of having a SLI decrease. For example, *freq_ttr* is a predictor that divides the number of unique words a child said by the total number of words the child said and provides a measure of lexical diversity. When increasing the *freq_ttr* by 1 unit, the log odds of having an SLI decreases by 9.06, holding all else constant. In addition, the coefficient for the interaction between *age* and *word errors* is 0.3292, which suggests that as a child's age increases, the impact of word errors on log odds of having an SLI increases. As such, we recommend speech pathologists to measure the relevant predictors to aid them in diagnosing children with SLI.

Limitations and Future Directions

Although our model may provide insights on diagnosing SLI, there are two particular limitations in our study. First, our final model indicated that sex predicts SLI (though not significant), where male children are more likely to have SLI than female children. This aligns with older studies that reported a 2:1 ratio of language impairment in male than female children [8]. However, recent research found no significant sex differences in SLI, suggesting that past diagnostic criterias may be biased or estimates of general language impairment is not applicable for SLI [9]. Since our data, collected from 2004-2006, may reflect these biases, it is possible that sex may no longer be a reliable predictor of SLI. In addition, efforts to improve diagnostics have led to replacing SLI with the term "developmental language disorder" that has a stricter diagnostic criteria [3]. Future research using more recent data and this updated diagnostic measure could provide more insight on language disorders relevant today.

Second, our findings are generalizable only to children between the ages of 4 to 12. The original dataset included children ages 4 to 15, but removing 119 observations with missing *sex* values excluded one study that contained all the adolescent data from the ages 13 to 15. The purpose of this study in particular was to address the gap in research, where most studies focus on children of younger ages and not of adolescents [4]. However, with the exclusion, our model is not able to predict SLI in the older age group. As such, future research could explore whether our predictors remain applicable or differ for this older age group, as supporting older individuals remains crucial and in cases where some may not be diagnosed early.

Finally, although our model contains the variable *age* for the age of the participant at the time of assessment, our model does not capture whether one's language abilities may be affected by their age of SLI diagnosis and levels of support the child receives. In the future, we are interested in exploring how diagnosis and treatments may affect the management of symptoms of developmental language disorders, which would be more applicable and useful for pediatricians and speech pathologists in addressing language disorders for children.

References

[1] National Institute on Deafness and Other Communication Disorders. (2019, July). *Specific language impairment*. Retrieved from www.nidcd.nih.gov/sites/default/files/Documents/health/voice/specific-language-impairment.pdf.

[2] Bishop, D. V. M. (2006). What Causes Specific Language Impairment in Children? *Current Directions in Psychological Science, 15*(5), 217-221. doi.org/10.1111/j.1467-8721.2006.00439.x.

[3] Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. N. & Wake, M. (2014). Specific language impairment: a convenient label for whom? *International Journal of Language & Communication Disorders, 49*(4), 416-451.

[4] Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative skills in adolescents with a history of SLI in relation to non-verbal IQ scores. *Child Language Teaching and Therapy, 23*(1), 95–113. doi.org/10.1177/0265659007072322.

[5] Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the *Edmonton Narrative Norms Instrument*. Edmonton Narrative Norms Instrument.

[6] Gillam, R., & Pearson, N. (2004). *Test of Narrative Language*. Pro-Ed Inc.

[7] Diagnose specific language impairment in children. Kaggle. (2017). https://www.kaggle.com/datasets/dgokeeffe/specific-language-impairment/data.

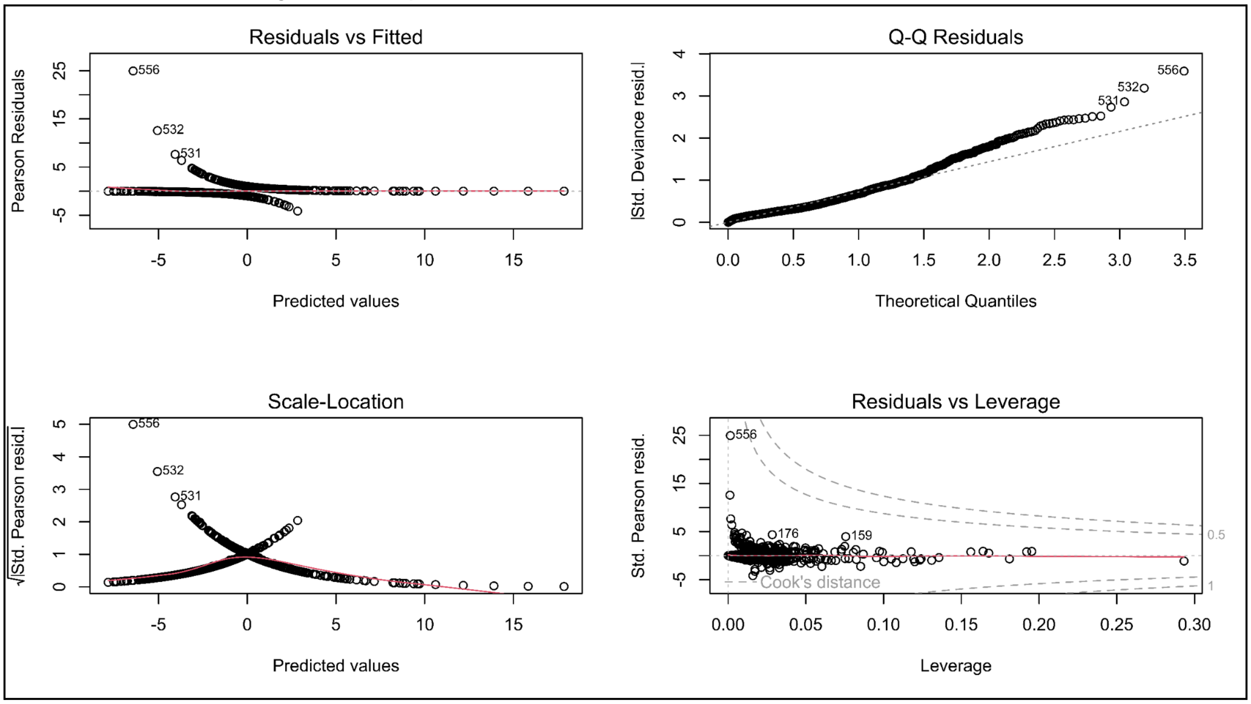
[8] Stevenson, J., & Richman, N. (1976). The prevalence of language delay in a population of three‐year‐old children and its association with general retardation. *Developmental Medicine & Child Neurology, 18*(4), 431-441.

[9] Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research, 40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245.

Appendix

Figure 1

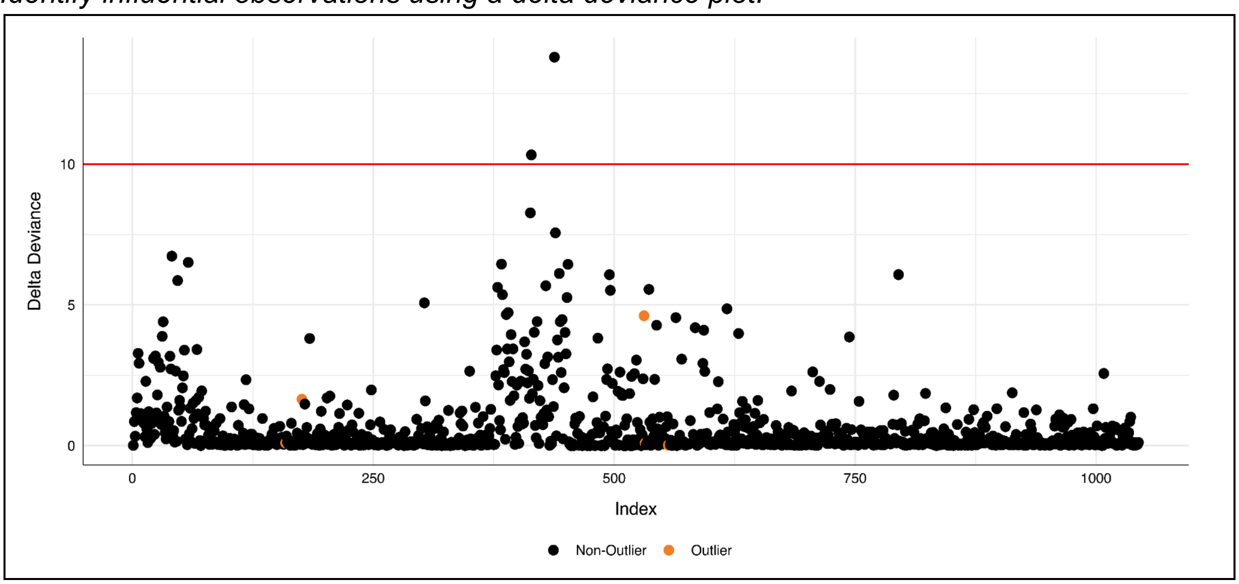
Outlier detection using residual plots.



Note. 4 residual plots for the final model with 16 predictors fitted with all the data (*N* = 1044).

Figure 2

Identify influential observations using a delta deviance plot.



Note. We set a threshold of delta deviance = 10 and removed the 2 observations above 10.

Table 1
Coefficients and significance of our final model predictors.

Predictor	Estimates	*SE*	*z-value*
(Intercept)	8.3810 **	2.87	2.917
sex = male	-0.2148	0.21	-1.025
age_years	0.8601 **	0.31	2.757
freq_ttr	-9.0556 ***	1.82	-4.975
r_2_i_verbs	1.0634 ***	0.31	3.441
retracing	-0.0312	0.05	-0.578
average_syl	-4.1206 ***	1.23	-3.345
mlu_morphemes	-0.5031	0.29	-1.707
ipsyn_total	-0.0279 *	0.01	-2.168
present_progressive	-0.0618 **	0.02	-3.146
regular_past_ed	-0.1254 ***	0.03	-4.930
regular_3rd_person_s	-0.0872 ***	0.02	-4.746
irregular_3rd_person	0.0326 *	0.02	1.978
word_errors	-1.3132 *	0.61	-2.144
age_years:word_errors	0.3292 ***	0.09	3.608
age_years:mlu_morphemes	-0.0200	0.04	-0.542
age_years:retracing	0.0099	0.001	1.511

Significance codes: * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.