BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2026

http://excel.fit.vutbr.cz

Detection of dyslexia from child's read speech

Richard Martinec*

Abstract

The aim of this paper is to experiment with how dyslexia can be classified from read speech using machine learning. The primary approach focuses on extracting sound features using the HuBERT speech model. Various techniques are used in feature processing, so that the features can be used to train a support vector machine (SVM) classifier. A maximum detection accuracy of 96.2% was achieved in one approach and certain experiments were carried out, revealing a bias present in the dataset, skewing the accuracy. A different approach discarding this bias yielded an accuracy of 85%. In general, the results suggest that these approaches could aid in child dyslexia diagnosis in the real world.

*xmartir00@stud.fit.vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Today, dyslexia is diagnosed through various standardized exercises, such as reading fluency or writing exercises, which might reveal relevant symptoms, such as difficulty in **word recognition**, **word decoding** or **spelling** [1].

Popular methods involving machine learning (ML) in the detection of dyslexia usually require specialized hardware [2], making them difficult in terms of data collection and usage. An approach for the detection of dyslexia from speech recordings would make data collection easier and improve the classifier's accessibility.

Various studies have focused on detecting not only dyslexia, but also speech fluency disorders from speech recordings, which are relevant to detecting dyslexia, considering certain similarities in the symptoms. One study ([3]) compared spectrograms, MFCCs and Wav2Vec 2.0 features in the classification of dysarthria using SVMs, showing highest accuracy when using the Wav2Vec embeddings. Another study ([4]) classified dysfluency in English speech using MFCCs and an SVM and k-NN classifier. One dyslexia-specific study ([5]) used a CNN to classify dyslexia from spectrogram images of Arabic speech.

In this work, the supplied dyslexia dataset was first pre-processed. Multiple kinds of features (audio embeddings, forced alignments, ASR transcriptions) were then extracted. These features were then processed in various ways and SVM classifiers were trained and validated using the leave-one-out approach to determine which method is best. All of this work was done in **Python**.

Multiple classification methods (varying in input features and their pre-processing) were tested. Since a dataset bias was discovered that skewed the accuracies of some classifiers, other methods were also tested with satisfactory results.

To showcase the classification methods, a web demo enabling the user to classify their own speech at `https://dyslex.rickmt.com` was created.

2. Dataset and feature extraction

The dataset consisting of 46 dyslexic and 92 non-dyslexic (intact) Czech speech recordings of children was provided by the **Faculty of Arts (MUNI)**. The recordings were captured in conjunction with eye-tracking data, used in a related study ([6]). Since the dataset was initially collected for the sole purposes of this study, the recordings could not be transferred or used outside of MUNI, therefore, all work involving the raw recordings had to be planned in advance and was carried out during a visit.

The recordings consisted of two exercises and the instructor's guidance speech[1]; the texts of both exercises were identical throughout the recordings. The audio recordings were segmented before feature extraction using a semi-automated approach, and while features

[1] In this pattern: instructions, warm-up exercise, more instructions, main exercise.

were extracted for all segments, only the main exercise speech was used in training.

The **hubert-large-ls960-ft** model was used to extract feature embeddings and the **parakeet-tdt-0.6b-v3** ASR model was used to collect transcriptions. **NeMo Forced Aligner** was also used to extract alignments.

3. Classification using feature embeddings

A script was written to perform training with leave-one-out cross-validation per each embedding layer (25 in total). Since the input feature tensors had to be shortened to vectors of reasonable dimensions, multiple pre-processing methods were tested, primarily mean-pooling and averaging over separate words[2].

3.1 Training

SVM kernels such as **RBF** or **sigmoid** were also tested, however, **linear** yielded the best results.

Figure 1 shows the chosen pipeline for the best model, whereas **figure 2** presents its balanced accuracies for each embedding layer.

A maximum accuracy of **96.2%** was achieved when using the 6th embedding layer.

3.2 Dataset bias

Considering the unusually high accuracy, other validation methods were tried. One included switching the main exercise features in each sample for features extracted from the instructor's guidance speech, while keeping the ground-truth classifications unchanged. Although the resulting accuracy was expected to be low, an accuracy of approximately 90% was obtained. Later, it was discovered that the dyslexics and intacts were recorded in separate classrooms[3], mostly corresponding to their diagnoses, with the different reverb being the possible culprit.

A different test suggested that the classifier is still trained to recognize dyslexic features. This time, only the validation samples had their features swapped in order to see whether the model trained on actual dyslexic/intact data would be able to classify the instructor's speech as before. High accuracy would suggest that the SVM only classifies based on the bias. However, the actual accuracy was low, suggesting that even with the bias present, this model is still trained to recognize other features, possibly those of dyslexia.

[2]Prior to averaging, the feature tensor was split into subtensors, each representing a single word, based on the forced alignments. These subtensors were mean-pooled before being averaged.

[3]It should be noted that the primary focus during data collection was the eye-tracking data, therefore separating the dyslexics and intacts in this way was not seen as an issue.

4. Classification using ASR/alignments

Since the output features of a model like HuBERT have no specific meaning or interpretation, it is difficult to determine whether any features that portray dyslexia symptoms are present. In this approach, feature vectors were assembled using **known algorithms**, with properties such as mean word duration or silence-to-speech ratio, determined based on either the ASR transcripts or the forced alignments. The features were chosen based on known dyslexia symptoms, as well as observations when working with the recordings. **Figure 3** shows the difference in the dyslexic and intact samples when it comes to total run time, suggesting a slower pace of speech.

4.1 Training

An accuracy of **86%** was achieved when using either ASR transcripts or forced alignments to build feature vectors.

It should be noted that while this classifier might score well when detecting dyslexia, symptoms such as misspelling or reordering words are not being considered in the features due to insufficient data extracted from the recordings.

The DET curves in **figure 4** belong to the HuBERT embedding model (blue) and ASR features model (orange).

5. Web demo

A web application demo was created to showcase some of the classifiers trained in this work. Its secondary purpose is to collect speech samples for further research.

The **PHP Symfony framework** was used when implementing the backend. The feature extraction and classification are done by a background Python script invoked upon sample submission. The frontend implements logic that allows the user to observe the lengthy feature extraction process in real-time, before showing a summarized classification (average of decision function values), as well as the detailed classifications.

6. Conclusion

Several SVM classifiers were trained to detect dyslexia with balanced accuracies of around 90%. Although a bias was discovered in the dataset, experiments show that dyslexic features are still considered during classification.

The models assessed in this work would benefit from validation using broader datasets, which were not available at the time. Such assessment could determine the model's real-world applicability.

References

[1] G. Reid Lyon, Sally E. Shaywitz, and Bennett A. Shaywitz. A definition of dyslexia. *Annals of Dyslexia*, 53(1):1–14, 2003.

[2] Yazeed Alkhurayyif and Abdul Rahaman Wahab Sait. A review of artificial intelligence-based dyslexia detection techniques. *Diagnostics*, 14(21):2362, October 23 2024. Review.

[3] Farhad Javanmardi, Saska Tirronen, Manila Kodali, Sudarsana Reddy Kadiri, and Paavo Alku. Wav2vec-based detection and severity level classification of dysarthria from speech. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 1–5. IEEE, June 2023.

[4] Mahesha P. An approach for classification of dysfluent and fluent speech using k-nn and svm. *International Journal of Computer Science, Engineering and Applications*, 2(6):23–32, December 2012.

[5] Alia Hussein, Ahmed Talib Abdulameer, Ali Abdulkarim, Husniza Husni, and Dalia Al-Ubaidi. Classification of dyslexia among school students using deep learning. *Journal of Techniques*, 6(1):85–92, Mar. 2024.

[6] Dostalova N. Sedmidubsky J. Svaricek, R. and A. Cernek. Insight: Combining fixation visualisations and residual neural networks for dyslexia classification from eye-tracking data. online, 2025.