UNIVERSITAS
GADJAH MADA
**Analisis Gerakan Mata sebagai Biomarker Disleksia Menggunakan Machine Learning dengan Explainable
Artificial Intelligence**
Annisa Nabila, Prof. Dr. Eng. Ir. Sunu Wibirama, S.T., MEng.,IPM. ; Ir. Noor Akhmad Setiawan, S.T., M.T., Ph.D., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Disleksia merupakan gangguan belajar spesifik yang memengaruhi kemampuan membaca dan dapat ditandai oleh karakteristik gerakan mata yang berbeda dibandingkan individu tanpa disleksia. Teknologi *eye tracking* berpotensi digunakan sebagai biomarker objektif karena mampu merekam karakteristik gerakan mata selama proses membaca. Penelitian ini bertujuan untuk membandingkan performa beberapa algoritma *machine learning* dalam klasifikasi disleksia berdasarkan fitur numerik *eye tracking* serta mengidentifikasi fitur yang paling berkontribusi terhadap hasil klasifikasi menggunakan metode *Explainable Artificial Intelligence* (XAI) SHAP. Dataset yang digunakan adalah ETDD70 yang terdiri atas data 70 partisipan dengan kelompok disleksia dan non-disleksia. Penelitian menggunakan 13 fitur numerik yang merepresentasikan karakteristik *fixation*, *saccade*, *progression*, *regression*, dan *transition*. Seleksi fitur dilakukan menggunakan SelectKBest berbasis uji ANOVA untuk memilih fitur yang memiliki kemampuan terbaik dalam membedakan kedua kelompok, sedangkan klasifikasi dilakukan menggunakan *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Gaussian Naive Bayes* (GNB), *Extreme Gradient Boosting* (XGBoost), dan *Decision Tree* (DT). Evaluasi dilakukan menggunakan *Repeated Stratified 5-Fold Cross Validation* sebanyak 10 pengulangan pada tiga tugas membaca, yaitu *Syllables*, *Meaningful Text*, dan *Pseudo Text*. Selain klasifikasi pada masing-masing tugas, dilakukan penggabungan prediksi menggunakan metode *ensemble majority voting* dan *weighted average*. Hasil penelitian menunjukkan bahwa LR memberikan performa terbaik pada tugas *Syllables* dengan *balanced accuracy* sebesar 84,6% dan pada tugas *Meaningful Text* sebesar 85,9%, sedangkan GNB memberikan performa terbaik pada tugas *Pseudo Text* sebesar 86,3%. Penggabungan ketiga tugas menggunakan *ensemble majority voting* menghasilkan performa terbaik secara keseluruhan, yaitu SVM dengan *balanced accuracy* sebesar 87,9%. Analisis SHAP menunjukkan bahwa fitur yang berkaitan dengan jumlah dan durasi fiksasi pada kunjungan pertama ke AOI dominan pada tugas *Syllables*, sedangkan jumlah *revisits* dan durasi fiksasi dominan pada tugas *Meaningful Text*. Pada tugas *Pseudo Text*, fitur yang berkaitan dengan durasi fiksasi dan perilaku pada kunjungan pertama memberikan kontribusi terbesar. Hasil tersebut menunjukkan bahwa kombinasi seleksi fitur menggunakan ANOVA, klasifikasi berbasis SVM, dan metode *ensemble* mampu meningkatkan kinerja klasifikasi disleksia serta memberikan interpretasi terhadap karakteristik gerakan mata yang dimanfaatkan model dalam proses pengambilan keputusan.

Kata kunci: disleksia, *eye tracking*, *machine learning*, SVM, SHAP.

UNIVERSITAS
GADJAH MADA
**Analisis Gerakan Mata sebagai Biomarker Disleksia Menggunakan Machine Learning dengan Explainable
Artificial Intelligence**
Annisa Nabila, Prof. Dr. Eng. Ir. Sunu Wibirama, S.T., MEng.,IPM. ; Ir. Noor Akhmad Setiawan, S.T., M.T., Ph.D., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Dyslexia is a specific learning disorder that affects reading ability and can be characterized by differences in eye movement patterns compared with non-dyslexic individuals. Eye tracking technology has the potential to be used as an objective biomarker because it can record eye movement characteristics during reading. This study aims to compare the performance of several machine learning algorithms for dyslexia classification based on numerical eye tracking features and to identify the features that contribute most to classification results using Explainable Artificial Intelligence (XAI) with SHAP. The dataset used in this study was ETDD70, consisting of eye tracking data from 70 participants divided into dyslexic and non-dyslexic groups. Thirteen numerical features representing fixation, saccade, progression, regression, and transition characteristics were used. Feature selection was performed using SelectKBest with the ANOVA test, while classification was conducted using Support Vector Machine (SVM), Logistic Regression (LR), Gaussian Naive Bayes (GNB), Extreme Gradient Boosting (XGBoost), and Decision Tree (DT). Model evaluation was performed using Repeated Stratified 5-Fold Cross Validation with 10 repetitions across three reading tasks, namely Syllables, Meaningful Text, and Pseudo Text. Prediction results were also combined using majority voting and weighted average ensemble methods. The results showed that LR achieved the highest performance on the Syllables task with a balanced accuracy of 84.6% and on the Meaningful Text task with a balanced accuracy of 85.9%, while GNB achieved the highest performance on the Pseudo Text task with a balanced accuracy of 86.3%. Combining the three tasks using the majority voting ensemble resulted in the best overall performance, with SVM achieving a balanced accuracy of 87.9%. SHAP analysis showed that features related to the number and duration of fixations during the first visit to an AOI were dominant in the Syllables task, while the number of revisits and fixation duration were dominant in the Meaningful Text task. In the Pseudo Text task, features related to fixation duration and reading behavior during the first visit contributed most to the classification. These results indicate that the combination of feature selection, SVM classification, and ensemble learning can improve dyslexia classification performance while providing an interpretable representation of the eye-movement characteristics utilized by the model.

Keywords: dyslexia, *eye tracking*, *machine learning*, SVM, SHAP.