Robust Ensemble Feature Selection for Dyslexia Prediction

Using eye movements and explainable AI to detect dyslexia early with 96.4% accuracy

GLOBAL IMPACT

The Dyslexia Challenge

15%

Global Population

Affected by dyslexia worldwide

10-12%

School Students

Experience reading difficulties

96.4%

Detection Accuracy

Achieved by our model

Early detection is crucial for specialized intervention to improve academic and emotional outcomes. Eye movement patterns during reading provide reliable markers for dyslexia screening.

Eye Movement as a Dyslexia Marker

Typical Patterns in Dyslexia

- Increased fixation duration
- Higher fixation count
- Irregular saccades with regressions
- Shorter saccade amplitudes

These distinctive patterns make eye tracking a powerful non-invasive screening tool.

DATASET

Research Foundation

Participants

185 children aged 9-10 years

- 97 high-risk dyslexia (HR-D)
- 88 low-risk dyslexia (LR-D)

Technology

Ober-2 infrared eye tracker

- 2D tracking (X and Y axis)
- Millisecond precision
- Both eyes recorded

Data Split

70:30 ratio for validation

- Training dataset: 70%
- Testing dataset: 30%

Feature Extraction Methods

Statistical Features

30 features including means, standard deviations, and ranges of eye movements for both eyes.

Dispersion-Based

16 features capturing eye movement dispersion during fixation using specialized algorithms.

Velocity-Based

11 features measuring temporal aspects like speed and frequency of saccades and fixations.

Total: 57 features extracted from eye-tracking data, requiring optimal selection for efficient prediction.

Three-Method Feature Ordering Strategy

PCA

Unsupervised dimensionality reduction capturing feature variance. 15 principal components identified.

RF-FI

Random Forest feature importance using Gini index. Dispersion and velocity features ranked highest.

XAI-SHAP

Explainable AI providing transparent feature contribution analysis. Scan path and fixation duration most impactful.

Each method offers unique perspectives: PCA focuses on variance, RF-FI on predictive power, and SHAP on interpretable contribution.

INNOVATION

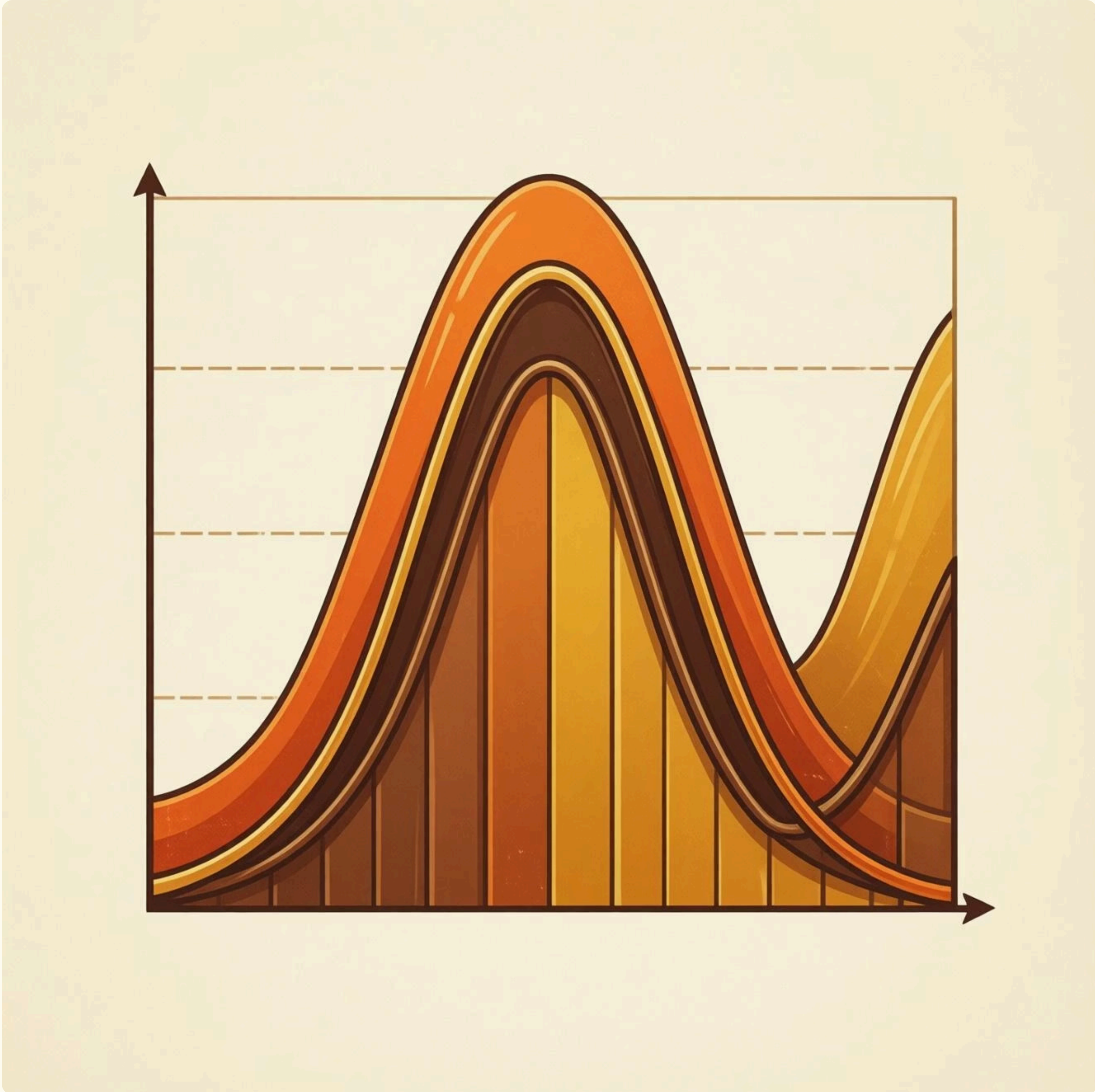
Robust Rank Aggregation (RRA)

Why RRA?

Statistical technique minimizing noise and outlier influence by evaluating significance across multiple ranking methods.

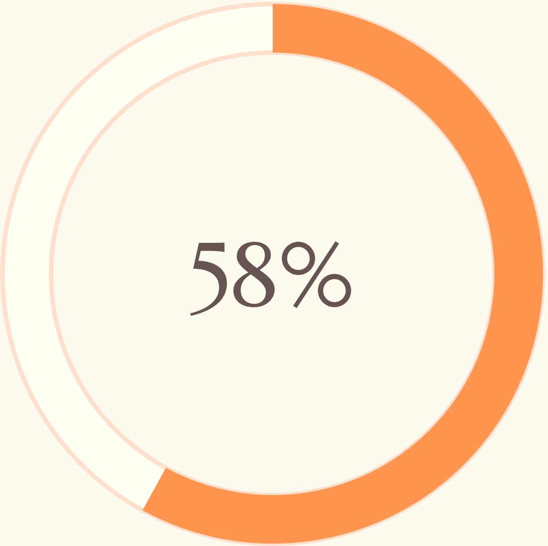
Key advantages:

- Probabilistic model-based
- Identifies consistently ranked features
- Beta distribution for robustness
- Low p-values = high importance



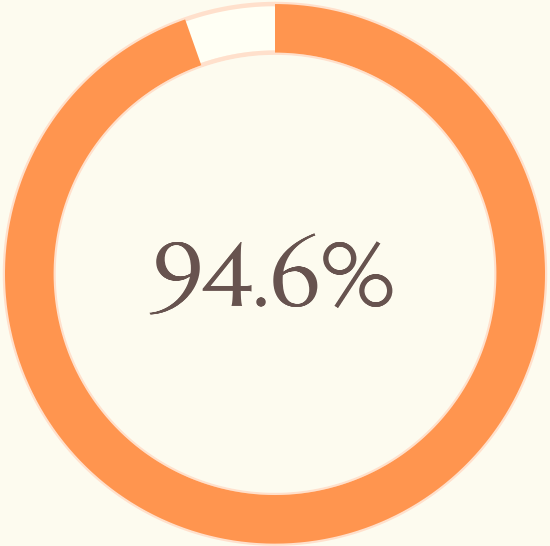
RRA achieved **stronger alignment** with PCA (0.57), RF-FI (0.85), and SHAP (0.88) compared to other aggregation methods.

Accumulation Effect Analysis



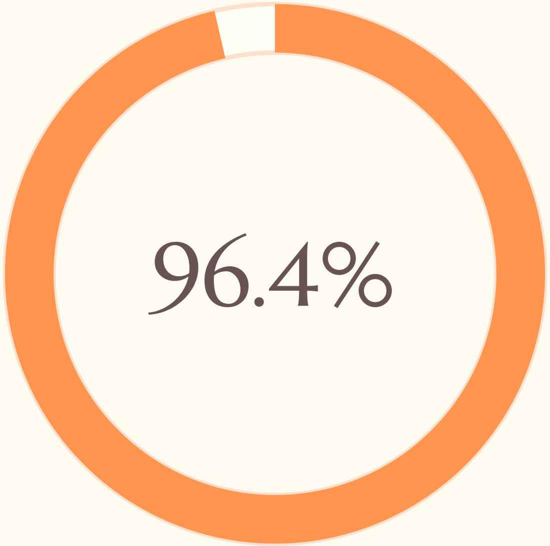
Feature Reduction

From 57 to 24 optimal features



RF Accuracy

Using baseline classifier

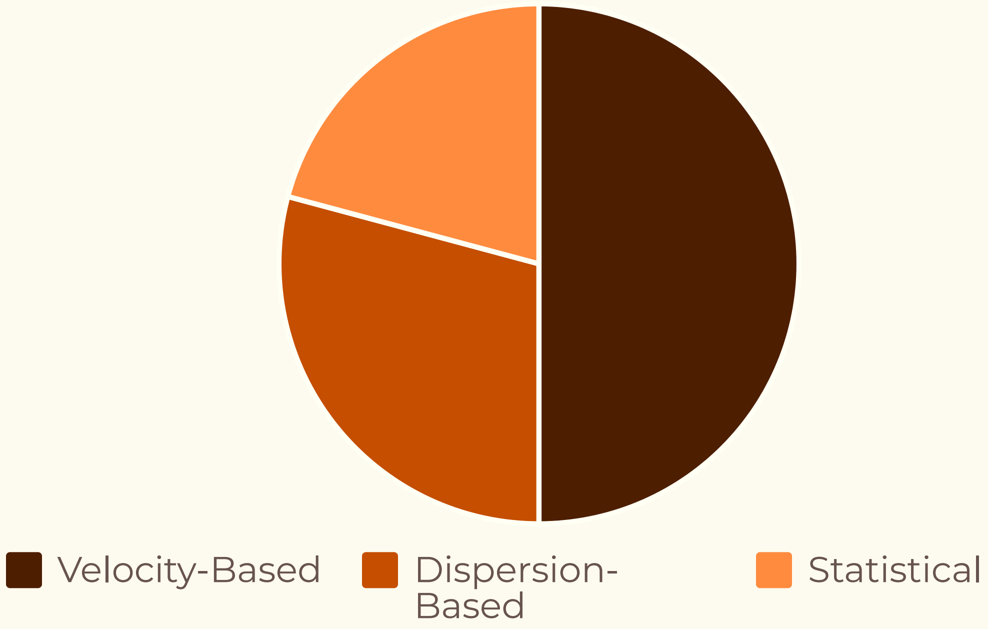


Stacking Accuracy

Best ensemble performance

RRA outperformed individual methods: PCA (38 features, 83.7%), RF-FI (55 features, 92.9%), and XAI-SHAP (49 features, 91.3%).

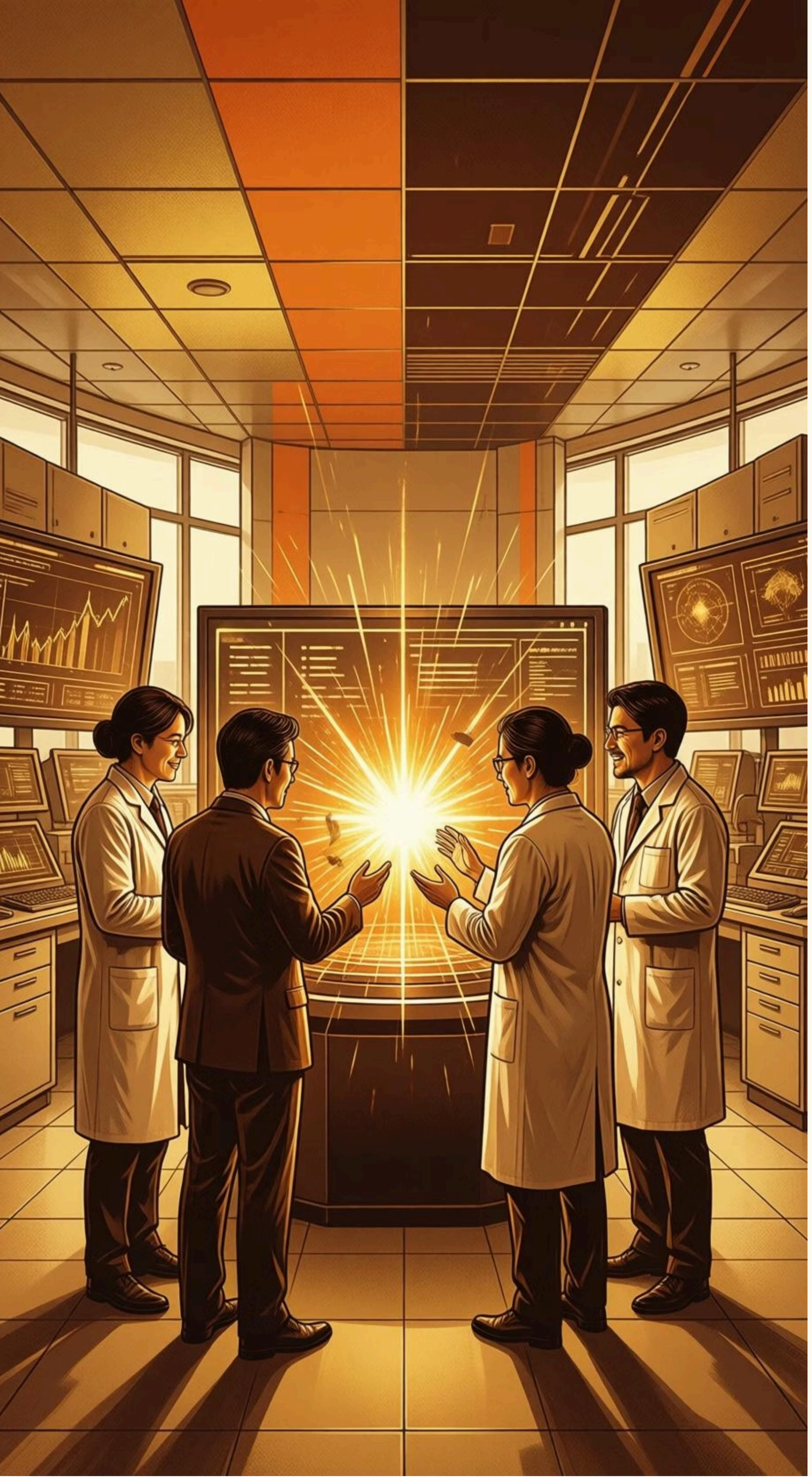
Feature Type Distribution



Velocity Features Dominate

The selected 24 features show velocity-based measurements are most reliable for dyslexia detection.

This validates that **eye movement speed patterns** are trustworthy indicators of reading difficulties.



Breakthrough Results

96.4% Accuracy

Stacking classifier (RF + SVM + KNN with LR meta-classifier) achieved highest performance

58% Reduction

From 57 to 24 features while improving accuracy over individual methods

Robust Method

RRA effectively integrates multiple ranking techniques for optimal feature selection

Future Directions

Integration with EEG, neuroimaging, and behavioral datasets for holistic analysis. Adaptive feature selection with confidence-driven approaches for dynamic model improvement.