SageMaker Autopilot Accuracy and Feature Engineering Runtime Trade-offs Across Varying Tabular Dataset Sizes and Complexities

Assignee Research

June 12, 2026

Abstract

Epilepsy is a neurological disease characterized by recurrent seizures caused by abnormal electrical activity in the brain. One of the methods used to diagnose epilepsy is through electroencephalogram (EEG) analysis. EEG is a non-invasive medical test for quantifying electrical activity in the brain. Applying machine learning (ML) to EEG data for epilepsy diagnosis has the potential to be more accurate and efficient. However, expert knowledge is required to set up the ML model with correct hyperparameters. Automated machine learning (AutoML) tools aim to make ML more accessible to non-experts

1 Introduction

This paper examines: Comparison of Automated Machine Learning (AutoML) Tools for Epileptic Seizure Detection Using Electroencephalograms (EEG). Research question: What is the impact of varying the size and complexity of tabular datasets (e.g., from OpenML-CC18) on the trade-off between model accuracy and feature engineering runtime in SageMaker Autopilot, compared to static feature selection methods?.

2 Methodology

Systematic literature search across multiple databases yielded 6 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.8/10.

3 Results

6 papers retrieved. 11 claims extracted; 11 independently verified. Quality review score: 7.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Epilepsy is a neurological disease characterized by recurrent seizures caused by abnormal electrical activity in the bra	✓	0.30
Electroencephalogram (EEG) is a non-invasive medical test for quantifying electrical activity in the brain.	✓	0.32
Applying machine learning (ML) to EEG data for epilepsy diagnosis has the potential to be more accurate and efficient.	✓	0.31
Expert knowledge is required to set up the ML model with correct hyperparameters.	✓	0.23
Automated machine learning (AutoML) tools aim to make ML more accessible to non-experts and automate many ML processes t	✓	0.40
The study compares the performance of three different AutoML tools, AutoGluon, Auto-Sklearn, and Amazon Sagemaker, on th	✓	0.45
Performance measures used for evaluation include accuracy, F1 score, recall, and precision.	✓	0.25
The results show that all three AutoML tools were able to generate high-performing ML models for the diagnosis of epilep	✓	0.33
The generated ML models perform better when the training dataset is larger in size.	✓	0.26
Amazon Sagemaker and Auto-Sklearn performed better with smaller datasets.	✓	0.28
This is the first study to compare several AutoML tools and shows that AutoML tools can be utilized for epilepsy diagnos	✓	0.30

References

- https://doi.org/10.1007/s10462-024-10726-1
- https://doi.org/10.3390/computers12100197
- https://doi.org/10.1016/j.jiixd.2024.01.002