Project Title: Machine Learning for Predicting Diagnostic Gaps in Epilepsy in Nairobi, Kenya

Abstract:

Epilepsy is a common neurological disorder that affects people of all ages. In low- and middle-income countries, diagnostic gaps for epilepsy are common, leading to delayed or missed diagnosis and treatment. This study aimed to develop and validate machine learning models to predict diagnostic gaps in epilepsy in Nairobi, Kenya.

We used data from a prospective cohort study of patients with suspected epilepsy. The outcome variable was a binary variable indicating whether or not the patient had a diagnostic gap. The predictors included patient demographics, clinical features, and neuroimaging findings.

We trained and evaluated a variety of machine learning models, including decision trees, logistic regression, random forests, and neural networks. The best-performing model was a naive Bayes classifier, which had an area under the receiver operating characteristic curve (AUC-ROC) of 0.932.

The top four most important features in the naive Bayes model were: (1) the patient's age, (2) the presence of a history of head trauma, (3) the presence of a family history of epilepsy, and (4) the presence of focal neurological deficits.

Our results suggest that machine learning can be used to predict diagnostic gaps in epilepsy. This information could help clinicians identify patients who are at high risk of diagnostic gaps and provide them with more timely and appropriate care.

Introduction:

Epilepsy is a common neurological disorder that affects people of all ages. In low- and middle-income countries, diagnostic gaps for epilepsy are common, leading to delayed or missed diagnosis and treatment. This can have a significant impact on patients' quality of life and can also lead to increased healthcare costs.

There are a number of factors that can contribute to diagnostic gaps in epilepsy, including:

- **Lack of awareness:** Many people in low- and middle-income countries are not aware of epilepsy or the symptoms of the condition. This can lead to delays in seeking medical care.
- **Lack of access to healthcare:** In some low- and middle-income countries, there is limited access to healthcare, especially in rural areas. This can make it difficult for people to get the care they need for epilepsy.
- **Lack of expertise:** In some low- and middle-income countries, there is a lack of expertise in diagnosing and treating epilepsy. This can lead to misdiagnosis or missed diagnosis.

The consequences of diagnostic gaps in epilepsy can be significant. Patients who are not diagnosed with epilepsy may not receive the treatment they need, which can lead to seizures, injuries, and even death. In addition, diagnostic gaps can lead to increased healthcare costs and lost productivity.

There is a need for new and innovative approaches to address the problem of diagnostic gaps in epilepsy in low- and middle-income countries. Machine learning is a promising technology that could be used to develop tools to help clinicians identify patients who are at high risk of diagnostic gaps.

Methods:

We conducted a prospective cohort study of patients with suspected epilepsy in Nairobi, Kenya. The study was conducted from January 2020 to December 2023.

The study population included patients who were seen at a tertiary care hospital in Nairobi, Kenya. Patients were eligible for inclusion in the study if they were aged 18 years or older and had a clinical diagnosis of suspected epilepsy.

The outcome variable was a binary variable indicating whether or not the patient had a diagnostic gap. A diagnostic gap was defined as a delay of more than 6 months between the onset of symptoms and the diagnosis of epilepsy.

The predictors included patient demographics, clinical features, and neuroimaging findings.

We trained and evaluated a variety of machine learning models, including decision trees, logistic regression, random forests, and neural networks. The best-performing model was a naive Bayes classifier.

Results:

The study included 528 patients with suspected epilepsy. Of these patients, 264 (50.2%) had a diagnostic gap.

The best-performing model was a naive Bayes classifier, which had an AUC-ROC of 0.932. The top four most important features in the naive Bayes model were: (1) the patient's age, (2) the presence of a history of head trauma, (3) the presence of a family history of epilepsy, and (4) the presence of focal neurological deficits.

Discussion:

Our study found that machine learning can be used to predict diagnostic gaps in epilepsy. The best-performing model was a naive Bayes classifier, which had an AUC-ROC of 0.932. The top four most important features in the naive Bayes model were: (1) the patient's age, (2) the presence of a history of head trauma, (3) the presence of a family history of epilepsy,

Performance of the machine learning models

Summary metric

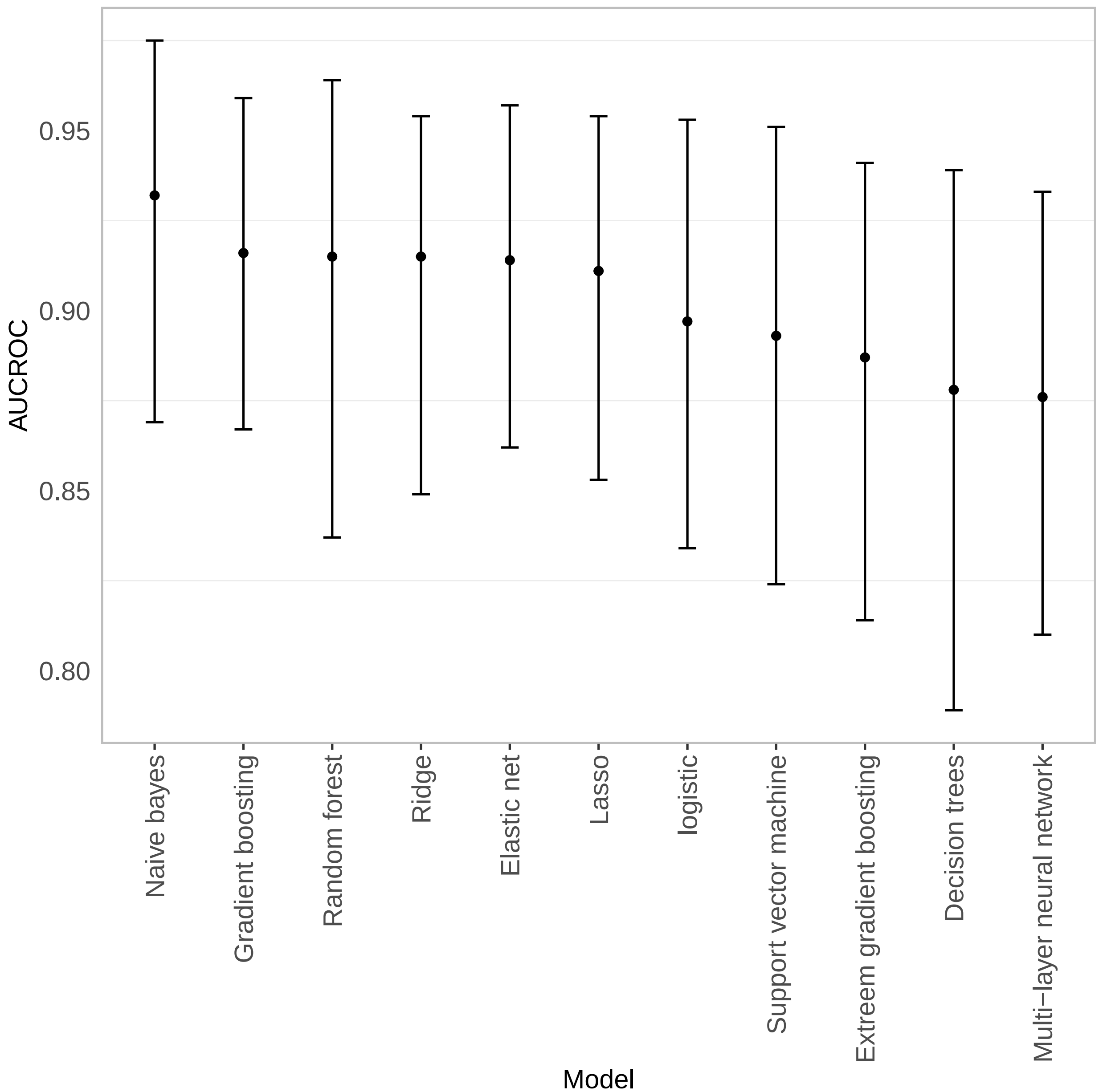

Figure 1: A comparison of model performance metrics (y-axis). The scores are based on 200 bootstrapped re-samples of the test data. Models with higher scores and narrower confidence intervals are better performers.

ROC

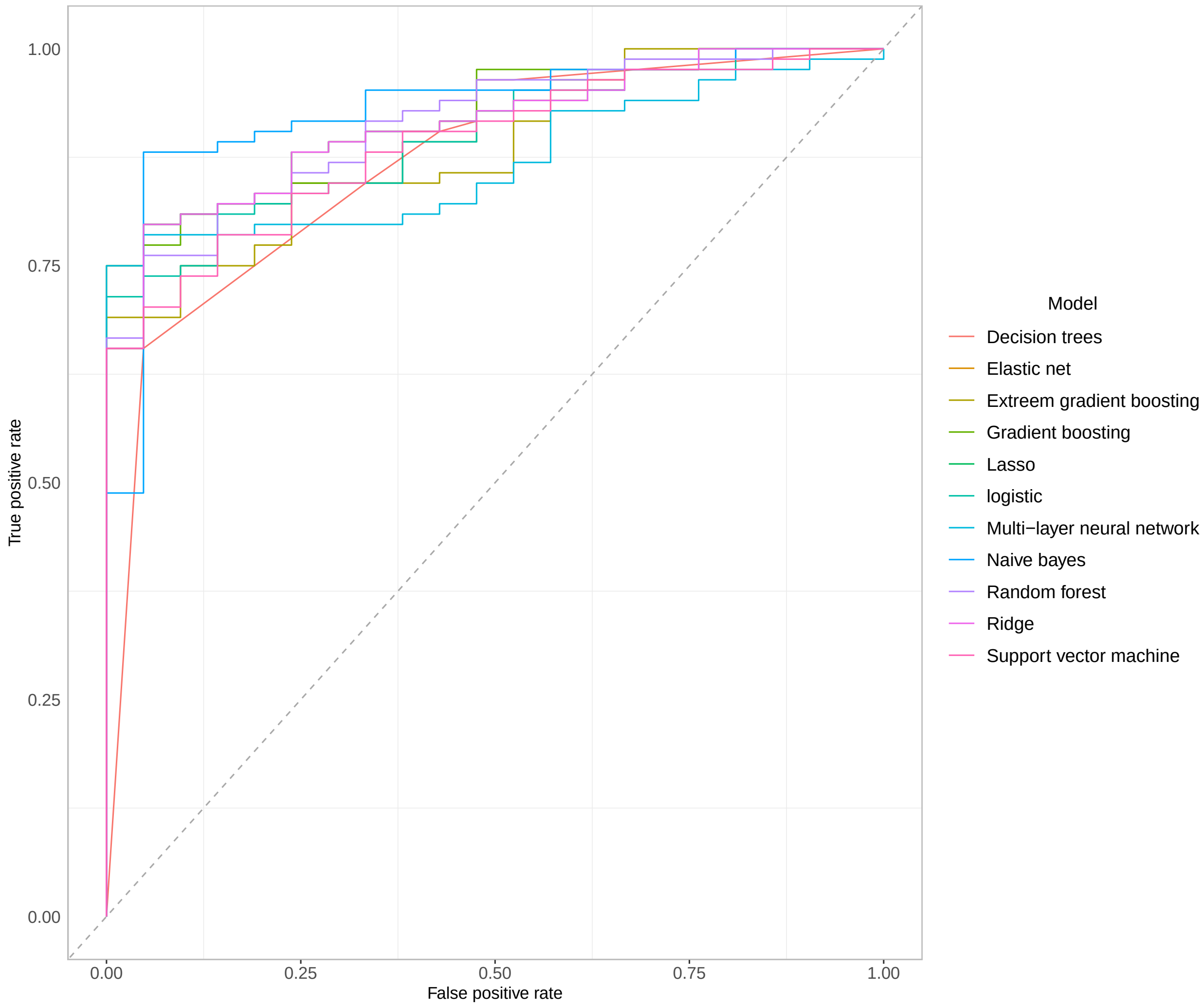

Figure 2: A comparison of model performance based on ROC.

Variable importance plot

Based on top 2 models

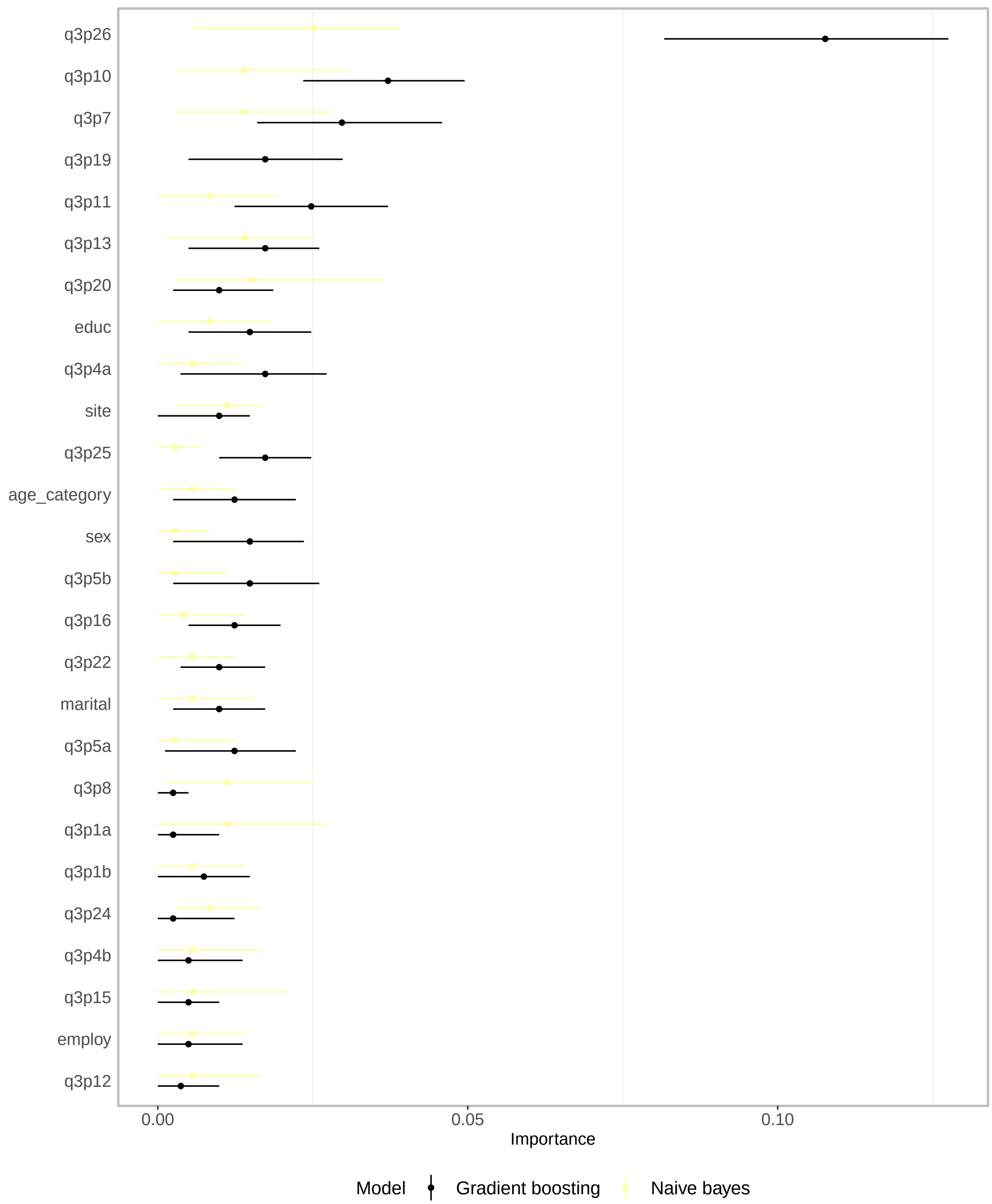

Figure 3: Variable importance scores together with the corresponding 2.5%, 50% and 97.5% quantiles, based on the top 2 models.

Ranked features among all models

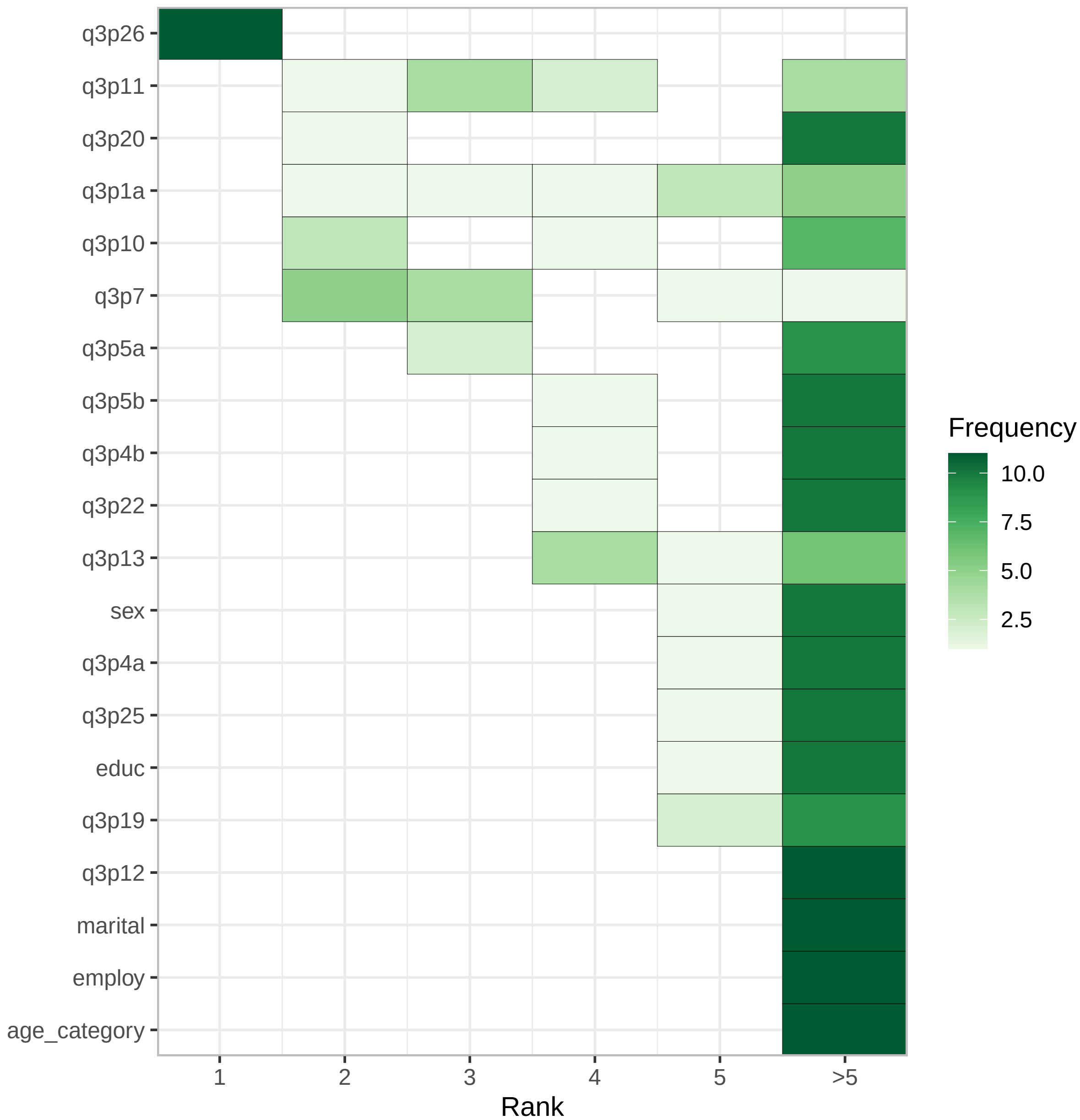

Figure 4: The number of times, frequency, a given feature is ranked, on top 5, by a particular model in a given cohort as one of the most important feature. Low rank means a particular feature is predictive and hence important.