



UMEÅ UNIVERSITY

Statistical studies no. 61

# Machine learning for predicting diverse stroke outcomes: binary, multi-class, and time-to-event

Josline Adhiambo Otieno

## Academic dissertation

Which, with the due permission of the Vice-Chancellor of Umeå University for the examination for the Degree of Doctor of Philosophy, is presented for public defence in HUM.D.210 - Hummelhonung on Friday, 20 March, 2026 at 09:30.

The thesis will be defended in English.

Faculty opponent:

Professor Paul Lambert, Department of Medical Epidemiology and Biostatistics, KI, Sweden, and the Cancer Registry of Norway, Norwegian Institute of Public Health, Norway

---

Department of Statistics,  
Umeå School of Business, Economics, and Statistics

**Organisation**

Umeå University  
Department of Statistics,  
Umeå School of Business,  
Economics, and Statistics

**Document type**

Doctoral thesis

**Date of publication**

27 February 2026

**Author**

Josline Adhiambo Otieno

**Title**

Machine learning for predicting diverse stroke outcomes: binary, multi-class, and time-to-event

**Abstract**

This thesis aimed to improve the clinical prediction of short- and long-term outcomes after stroke by developing, evaluating, and comparing classical statistical and modern machine learning models. Using large, high-quality national stroke registries, specifically the Swedish national stroke register (Riksstroke) and the Sentinel Stroke National Audit Programme (SSNAP), the thesis investigates whether advanced machine learning models offer added value in predicting clinical outcomes compared to traditional models, and importantly in which contexts such improvements occur and are clinically meaningful. It addresses several key gaps in the literature, including the lack of external validation for many prediction models across different health systems, the limited research on predicting multi-class functional outcomes, few comprehensive simulation-based evaluations of prediction models under different realistic data conditions, and the need for multiple-horizon evaluations of competing risk prediction models to support fair model selection. The first paper of the thesis evaluated models for predicting short-term mortality after stroke using large national registries from two European countries (Riksstroke and SSNAP). The results showed that machine learning models offered only modest performance gains over well-specified logistic regression models, demonstrating that traditional approaches remain competitive, especially when predictors are limited and the dataset is structured. The second paper performed multi-class prediction of functional outcomes three months after stroke, a clinically important, yet methodologically challenging outcome. All models demonstrated similar overall accuracy. However, machine learning, particularly neural networks and gradient-boosting models, indicated clearer advantages over multinomial logistic regression in distinguishing the functional dependence category. Using explainability approaches such as SHapley Additive exPlanations, the study demonstrated that complex models can still provide interpretable insights into the contribution of risk factors in predictions. The third paper comprehensively evaluated the classical Cox proportional hazards model and machine learning models for predicting time-to-event outcomes using both simulation and real-world registry data. The Cox regression model performed better when its assumptions were satisfied or when the violations of the assumptions were minimal, while tree-based models demonstrated better performance in the presence of non-linearity, misspecification, or large number of noise variables. The final paper compared multiple modeling frameworks for predicting competing risks at multiple evaluation time points (horizons). The results showed that the performance of the models depended on the dataset and the evaluation time point, and no model consistently performed the best. Tree-based and deep-learning models achieved better discrimination when events were common, while pseudo-observation-based and Fine-Gray models showed better calibration, especially at longer horizons. In summary, the thesis demonstrated that model choice should be guided not by popularity but by data structure, clinical context, and evaluation using different metrics and at multiple time horizons. Traditional and machine learning models each have strengths and rigorous validation, calibration assessment, and explainability are crucial for trustworthy clinical prediction.

**Keywords:** Predictive modeling, Machine learning, Survival analysis, Stroke.

**Language**

English

**ISBN**

978-91-8070-890-6 (print)  
978-91-8070-891-3 (pdf)

**ISSN**

1100-8989

**Number of pages**

32 + 4 papers