

## Classification models for predicting the source of gastrointestinal bleeding in the absence of hematemesis

### ABSTRACT

Management of acute gastrointestinal bleeding necessitates the identification of the source of bleed. The source of bleeding which is clear in patients presenting with hematemesis, is unclear in the absence of it. Logistic regression, decision tree, naïve Bayes, LogitBoost and KNN models were constructed from non endoscopic data of 325 patients admitted via the emergence department (ED) for GIB without hematemesis. The performance of the models in predicting the source of bleeding into upper gastrointestinal bleeding or lower gastrointestinal bleeding was compared. Overall the models demonstrate good performance with regards to sensitivity specificity, PPV, NPV and classification accuracy on the simulated data. On the GIB data, the naive Bayes model performed best with a prediction accuracy and sensitivity of 86%, specificity of 85% and area under curve of 92%. Classification models can help to predict the source of gastrointestinal bleeding for patients presenting without hematemesis and may generally be useful in decision support in the ED. The models should be explored further for clinical relevance in other settings.

**Keyword:** Upper gastrointestinal bleeding; Emergency department; Hematemesis; Classification models; Naive Bayes; Weka; Naive Bayes classifier; Random Forest; Decision tree; Logit boost; Logistic regression.