Beyond the classroom: A comparative analysis of machine learning learning techniques for predicting atrial fibrillation

Carleton University
School of Mathematics and Statistics

Antonio Duran, Graham Gee, Hoang-Nam Chu, Matthew Lunn
Supervisor: Dr. Shirley Mills
Carleton University, Department of Mathematics and Statistics, Ottawa, Ontario

Société statistique du Canada
Statistical Society of Canada

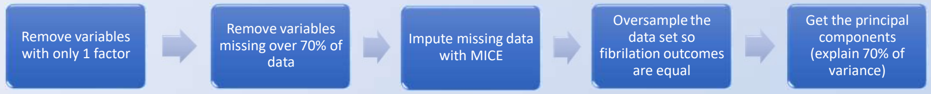
Introduction

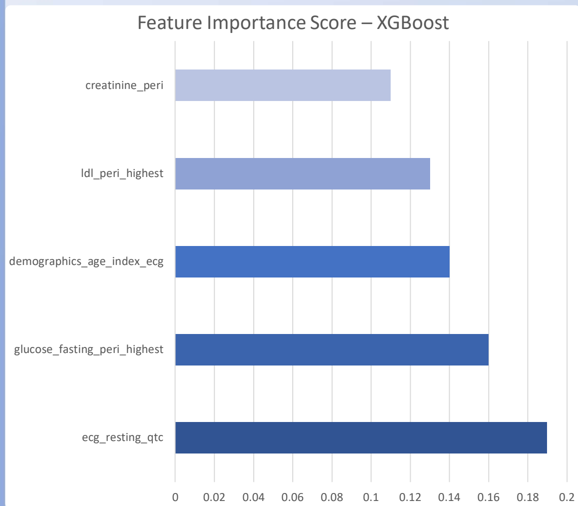
- Atrial Fibrillation (AF) is a cardiovascular disease marked by an irregular heartbeat, or heart rhythms, in the upper chambers of the heart.
- The consequences of AF are several, including blood clots which could dislodge and lead to complications such as stroke and heart failure.
- Early detection is vital as oral anticoagulants are one of the only consistent and definitive therapeutic interventions.
- Risk scores have been developed that use traditional statistics to identify patients at risk (CHADS2, CHA2DS2-VASc). These models have had moderate success with AUC scores of 0.59–0.73.
- More general models need to be developed which can predict AF risk across inputs under different clinical conditions and with a variety of patient populations.

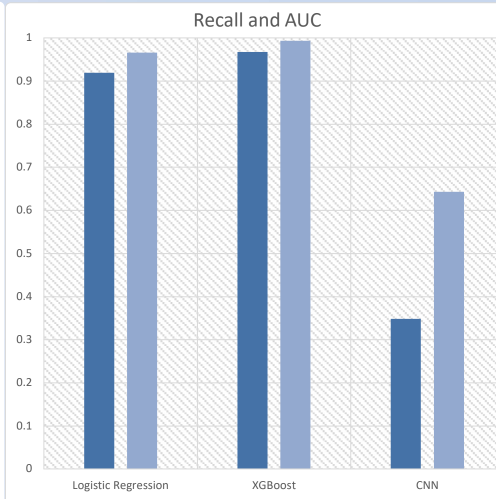
Objective

- Given a synthetic dataset containing electronic health record and ECG markers of roughly 100,000 patients, a comparative analysis of several supervised learning approaches was performed.

Pre-Processing







Results

Model	Recall	AUC	Accuracy
Logistic Regression	0.9193	0.9663	0.9084
XGBoost	0.9673	0.9938	0.9651
CNN	0.3485	0.6433	0.7385

Methods

- We focused on three types of classifiers to perform our analysis: logistic regression, extreme gradient boosting (XGBoost), and convolutional neural networks (CNNs).
- We chose these classifiers based on the supervised learning techniques that we learned in our data mining course—logistic regression and boosting—yet extended our learning by selecting a contemporary machine learning model used by the latest research—CNNs, where it has seen success in predicting atrial fibrillation in patients through analysis of ecg signals.
- To measure our performance, we considered performance metrics that were important in a clinical health setting: AUC and Recall.

Conclusion

- XGBoost had the best performance metrics when analyzing the data over training, test, and validation sets, producing the most consistent results, and
- Convolutional Neural Networks while very good at image analysis, did not perform as well when analyzing the data, as the convolutional operations performed are meant to capture local structure in the data, something this dataset lacks; adjusting values for kernel and filters produced little meaningful result.
- Logistic regression suffered from overfitting because of the class imbalance in the dataset, after using oversampling, it gave us better results, though still underperforming compared to XGBoost.
- XGBoost worked well for several reasons. It diminishes the risk of overfitting by penalizing the model with regulation techniques if it becomes too large and overfit the data, and it allows the ability to fine tune the hyperparameters for this specific problem to optimize the model's ability to predict atrial fibrillations.

References

Slide 1

HC0 I thought this opening statement from your proposal would be a good would be a good hook
Hoang-Nam Chu, 2025-02-12T12:43:12.646

HV(0 0 I agree
Hoffmann, Vivian (IFPRI), 2025-02-13T21:30:09.692

HC1 “Shifting consumer decisions from high carbon to low carbon choices can have a mitigative effect on global GHG emissions.
”

I thought this would tie nicely into the aim of the research.
Hoang-Nam Chu, 2025-02-12T12:45:54.712

HV(1 0 yes that's right. but should be "mitigating" not "mitigative"
Hoffmann, Vivian (IFPRI), 2025-02-13T21:33:55.582

HC2 Whatever I miss here I could cover in the main graphic of the poster.
Hoang-Nam Chu, 2025-02-12T12:50:26.297

HC3 I’m not completely done the main graphic yet but the entire is to capture the methodology. I could add more about the GHG calculator, just not quite sure how to include it yet.
Hoang-Nam Chu, 2025-02-12T12:51:17.904