Classification of Treadmill Running Fatigue With Machine Learning

Dhrasti Dalal[1], Alizee Leleu[3], Yumi Lamansky[4], Mohit Malu[2], Tomas Ward[3], Andreas Spanias[2]

[1]ASU School of Biological and Health Systems Engineering, [2]ASU School of Electrical, Computer and Energy Engineering, [3]Dublin City University, [4]ASU School of Computing and Augmented Intelligence

Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

NATIONAL SCIENCE FOUNDATION

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

INTRODUCTION

- Running has increased in popularity resulting in more injuries especially amongst amateur runners [1]
- **Fatigue** - reduction of maximal force/power production
  - results in biomechanical form changes to compensate [2]
  - strain on the tendons, joints, and ligaments of the lower body and back
  - **detection of fatigue can reduce injuries**

fatigue detection

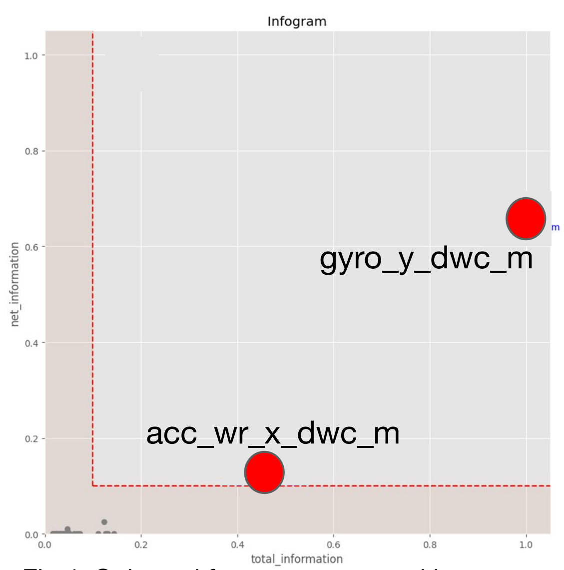
- velocity, force, and EMG analysis in biomechanics lab - **expensive and intrusive**
  - using IMU is cheaper and can be used in the field
- **support vector machines** and **random forest** classification typically used [2,3]

OBJECTIVES

1. feature extraction and selection on treadmill data using automatic ML algorithms
2. exploration of multiple difference machine learning models for fatigue classification of treadmill data
3. cross-training between treadmill and track fatigue classification models

FEATURE SELECTION

H20 selected features



MLJAR selected features

Acc_WR_Z_mean
Acc_WR_Y_mode
Acc_WR_Y_max
Acc_WR_Z_range
Acc_WR_X_energy
Gyro_X_25%
Acc_WR_X_dwc_m
Acc_WR_Z_dwc_m
Gyro_X_dwc_m
Gyro_Y_dwc_m
Mag_Y_dwc_v

Fig 1. Selected features outputted by automatic ML python algorithms. H20 infogram displays total information (a measure of how much the variable drives the response) plotted against net information (a measure of how unique the variable is) for treadmill data

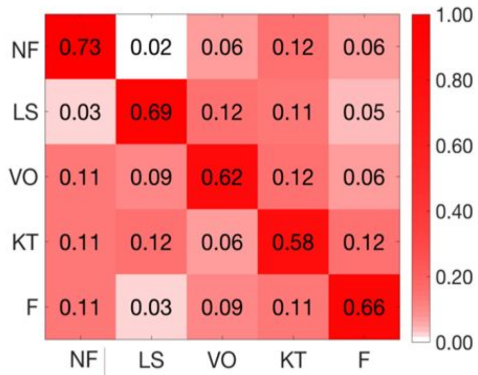
Algorithm	Neural Network	Random Forest	Xgboost	Catboost
f1 score	0.3852	0.606604	0.581303	0.612859

Fig 2. MLJAR supervised output suggests that the CatBoost algorithm will perform best for fatigue classification and neural networks will perform significantly worse

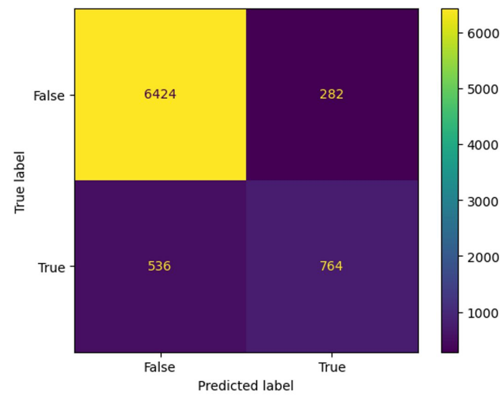
ALGORITHM EXPLORATION

RANDOM FOREST

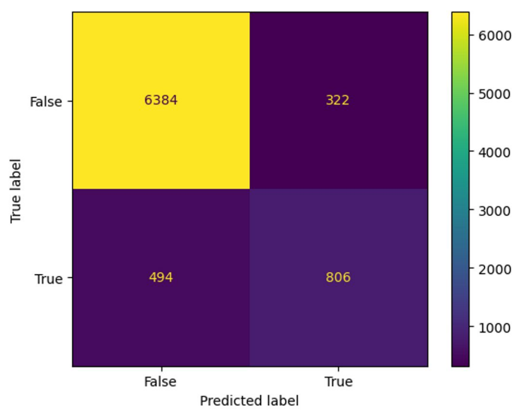
track data [4]



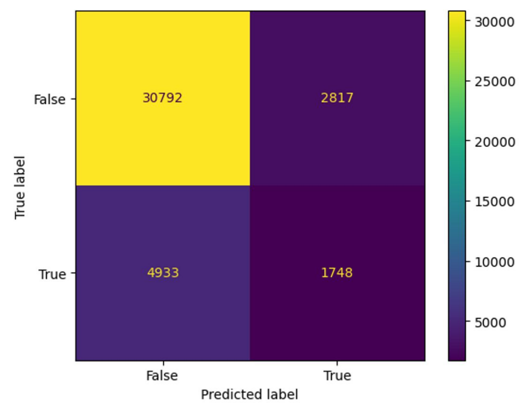
treadmill data



CATBOOST



NEURAL NETWORK



DISCUSSION & CONCLUSION

- H20 and autoML allow for optimized feature reduction for track and treadmill data independently
  - for treadmill data - length, gyro_y_dwc_m, and acc_wr_x_dwc_m are the only admissible features for classification
- Random forest model predicts fatigue most accurately
- cross-training improves results

future work

- account for class imbalances/collect more balanced data
- expand cross-training to all models
- explore more data preprocessing

ACKNOWLEDGEMENTS, REFERENCES, & KEY

HOST INSTITUTIONS

PARTNER INSTITUTIONS

This project is funded by the National Science Foundation (NSF) Award 1953745

FUNDED BY:

SenSIP Sensor, Signal and Information Processing Center

ASU Ira A. Fulton Schools of Engineering Arizona State University