Random Forest Classifier on Biomechanical Measures of 181,909 Steps

Natalie (Kramer) Kupperman MEd, ATC
& Jay Hertel PhD, ATC

PURPOSE

Assess if a random forest classifier could be applied to biomechanical measures captured by wearable sensors to predict whether running steps were performed outdoors or on a treadmill.

DATA

RunScribe Pod:
caputures gait kinematics in-situ

181,909 steps
(131,971 outdoor, 49,938 treadmill)
from 6 runners
(4F, 2M)
over 32 runs
(22 outdoors, 10 treadmill)

PROCESSING

scikits learn
machine learning in Python

scikit learn:
package used to implement random forest

METRICS

Feature	Outdoor: Mean (SD)	Treadmill: Mean (SD)	t-statistic (p-value)
Max Pronation Velocity	668.35 (436.19)	393.16 (262.60)	0.54 (0.61)
Pronation Excursion: foot strike - max pronation	-11.37 (11.39)	-10.75 (7.92)	1.32 (0.24)
Pronation Excursion: max pronation - toe off	8.5 (12.62)	8.04 (10.30)	0.37 (0.72)
Step Length (m)	1.27 (0.23)	1.37 (0.29)	-1.06 (0.34)
Footstrike Type	5.59 (3.1)	4.7 (2.6)	-0.44 (0.68)
Stride Pace (m/s)	3.66 (0.69)	3.76 (0.89)	-0.65 (0.55)
Cycle Time (ms)	702.51 (72.32)	738.02 (71.63)	-0.60 (0.58)
Step Rate (steps/min)	172.21 (13.36)	163.82 (12.86)	0.73 (0.50)
Braking g's	10.84 (3.05)	9.31 (3.04)	2.03 (0.10)
Contact Time (ms)	277.48 (79.53)	303.75 (92.15)	-0.26 (0.80)
Contact Ratio	66.88 (31.28)	72.35 (35.72)	-0.01 (0.99)
Flight Ratio	22.94 (9.21)	20.72 (10.29)	-0.26 (0.81)
Flight Time (ms)	73.53 (50.53)	65.01 (61.01)	-0.03 (0.98)
Impact g's	11.14 (3.43)	9.99 (3.36)	-0.03 (0.98)

* Footstrike type is a unitless metric in the RunScribe™ algorithm that is on a scale from 0 to 15 with lower values representing rearfoot strike, middle values representing midfoot strike, and higher values representing forefoot strike.

CLASSIFIER MODELS

Logistic Regression

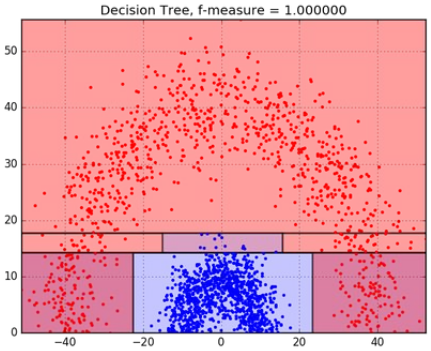
Fits a line to split data exactly into two sections

Can use coefficients to understand variable importance

Decision Tree Classifier

Can create multiple decision boundaries & *does not* need to fit a single line

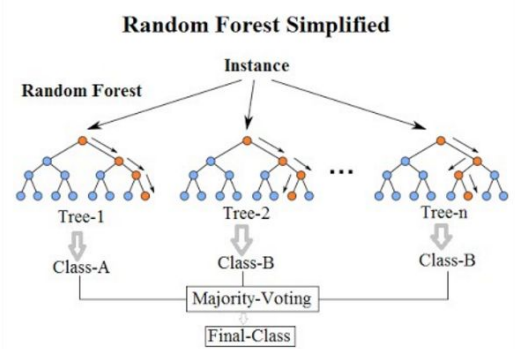
Can use feature importance to understand splits in the tree



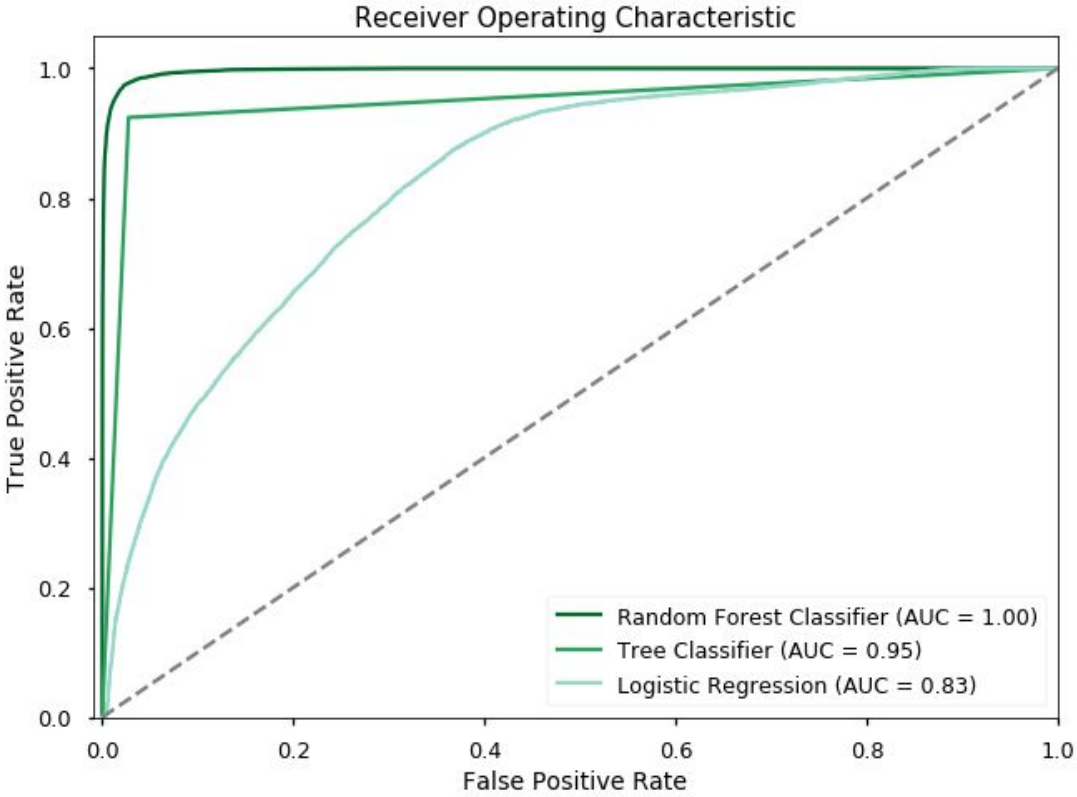
Random Forest Classifier

Splits the training data into smaller data sets and creates smaller decision trees (ensemble method)

Each tree than "votes" & majority vote is the prediction



Graphics fro BigML Blog (https://blog.bigml.com/2016/09/28/logistic-regression-versus-decision-trees/) & Towards Data Science (https://towardsdatascience.com/predicting-success-in-online-education-2b5979fa7016)



MODEL SCORES

These 'accuracy' scores are model specific & are calculated using the test data. The modules are included in for each classifier within the scikits Learn module

Logistics Regression

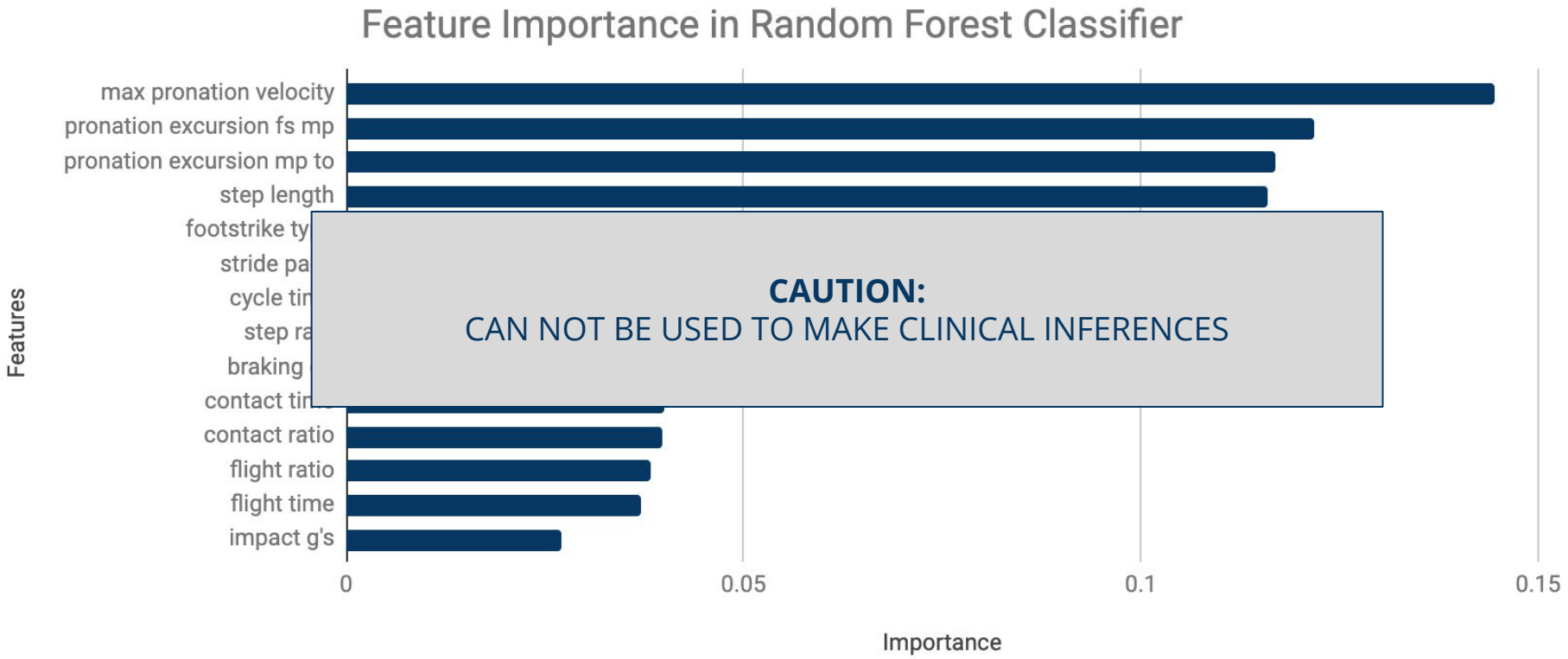
0.73

Decision Tree Classifier

0.96

Random Forest Classifier

0.98


Feature Importance in Random Forest Classifier
max pronation velocity
pronation excursion fs mp
pronation excursion mp to
step length
footstrike ty
stride pa
cycle tir
step ra
braking
contact tir
contact ratio
flight ratio
flight time
impact g's
Features
0
0.05
0.1
0.15
Importance
CAUTION:
CAN NOT BE USED TO MAKE CLINICAL INFERENCES

CONCLUSIONS

On this data set, a random forest classifier
out-performs both a logistic regression and decision tree.

The use of non-linear statistical techniques
should be considered in future research.

Code for the classifiers can be found at bit.ly/runscribeRFclassifier