# A Predictive Model for Next-Day Fatigue in Professional Canadian Women's Rugby Players

**The Knitting Club - Xuan Yu, Hannah Yan, Kate Coulter, Avani Gupta, & Danny Witt**

## Objective

The goal of our work was to predict the circumstances under which members of the Canadian national women's Rugby team transition into a maximum level of self-reported fatigue. To make these predictions, we utilized longitudinal data measuring reported psychological and physiological variables from players during the 2017-2018 season. Our prediction model would be valuable for determining training regimens for players prior to the season and in real-time throughout the season. The project was motivated by the fact that prediction of fatigue can be ambiguous and currently relies on a 5-part "Monitoring Score", which may or may not actually predict player's reported fatigue on subsequent training days.

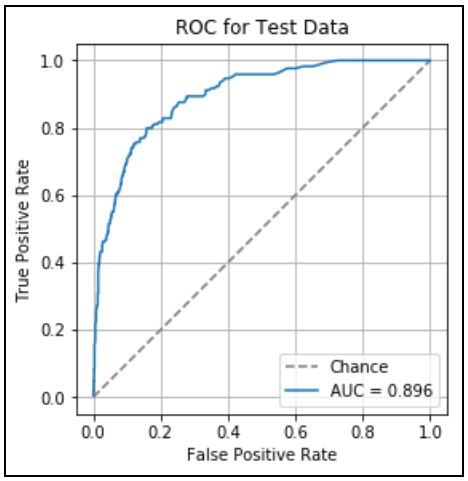
## Methodology

***Data Preprocessing***

- Initial aggregated dataset consisted of "wellness" and "exertion" (i.e., RPE) data.
- Isolate training data only: removed data collected on dates with a game.
- Continuous variables (p = 11): normalized by player to be consistently in the 0 to 1 scale.
- Categorical variables (p = 5): converted to dummy variables.
- Missing data:
    - Removed rows with all predictors = NaN
    - Removed predictors with > 50% missing data
    - Imputed *median data per player* for other predictor variables

***Predictive Analysis***

- ***Multivariate time series*** sliding window (lag) method for next day fatigue
- ***Random Forest Ensemble Method***
    - Validation: k-Fold Cross Validation (k=20)
    - Response variable: player reported next-day high fatigue state (i.e., > 80% reported fatigue state)
- ***Model Metrics***
    - ROC AUC (train) = 0.884
    - ROC AUC (test) = 0.896




## Visualization of App Mock-Up


*References: Activity Monitor by Jardson Almeida*