**Relative performance of three machine learning techniques for classification of accelerometer-based behaviors in white storks (*Ciconia ciconia*)**

Stephanie Cunningham

**Abstract**

Over the last twenty years, biologging devices have shown considerable decreases in size, allowing for a greater range of applications and inference on wildlife. Yet the improvements in technology have increased the amount of data coming from these devices, requiring and expansion beyond traditional methods for data analysis. Machine learning methods have commonly been used to classify these data, though further investigation into improving these methods will benefit conservation and management. I tested K-nearest neighbors (KNN), random forests, and gradient boosting machines (GBM) in classifying five behaviors (flapping flight, soaring, standing, sitting, and walking) in white stork (*Ciconia ciconia*) accelerometer data. KNN had the lowest overall accuracy of 0.87, and random forest had the highest overall accuracy at 0.90, but was only slightly higher than GMB (0.89). AUC values were 0.97, 0.87, and0.98 for KNN, random forest, and GBM, respectively.

**Introduction**

Improvements in technology over the last two decades have greatly increased the information researchers can gather from free-ranging wildlife (Sequeira et al 2021). A primary example is the accelerometer, which has been increasingly included in wildlife tracking collars and generates substantial amounts of information on animal energy expenditure and behavior. Generally, animals fitted with tracking devices are directly observed (either in the wild or captivity) for a period of time in order to match behaviors to acceleration signatures, and then prediction methods are used to classify acceleration data from unobservable periods (Brown et al. 2013). Wildlife researchers are keen to explore models that can provide the most accurate predictions to improve inference of subsequent analyses of behavioral patterns, which could then be used to inform conservation and management of a variety of species.

My objective was to compare the relative performance of three machine learning methods (K-nearest neighbors [KNN], random forest, and gradient boosting machines [GBM]) in classifying five behaviors from animal-borne accelerometers, using white storks (*Ciconia ciconia*) as a case study. Given an unlabeled set of summary statistics, the goal is to predict a behavior class. While KNN and random forests have received much attention from wildlife biologists, and random forests especially have been used with much success to classify behaviors in animal-borne accelerometers (Dentinger et al. 2022, Nathan et al. 2012), a general examination of the literature suggests that only recently (e.g., since ~2020) have more complex methods such as GBMs been employed by wildlife researchers. Therefore, I was interested in examining the performance of this method in particular compared to KNN and random forest, which have been used more extensively.

**Methods**

*Data Preparation*

I used accelerometer data from white storks in 2013-2014 (Flack et al. 2015), which were generated from a study on stork migration costs (Flack et al. 2016). These data were available from the AcceleRater web tool (Rescheff et al. 2014). Stork-borne accelerometers (contained in back-pack transmitters) recorded data bursts every 5 min at 10.54 Hz for 3.8 s over 3 axes (X, Y, Z), which yielded 40 data points per axis per burst. There were 1746 bursts in the data set. The data were assigned one of five behavioral classes via direct observation, active flight (i.e., flapping), passive flight (i.e., soaring), sitting, standing, and walking.

From the raw acceleration data, I calculated static (positional) and dynamic (from movement) acceleration (Shepard et al. 2008) and developed a suite of summary statistics to use as predictors (Rescheff et al. 2014; Table 1). I used functions from the ‘caTools’ (Tuszynski 2021) and ‘moments’ packages (Komsta and Novomestky 2015) for calculating summary statistics. There was a total of 28 predictor variables (including variations for each axis pairs) used in the models. Eighty percent of the data were randomly selected to be included in the training set, while the remaining 20% was used for model validation.

*Analyses*

I first used K-Nearest Neighbors (KNN) to predict behavior and implemented this model using the ‘caret’ package (Kuhn 2021). I used 5-fold cross-validation for tuning the model, and created a tuning grid of length 50 between 1 and 401 to select the optimal value of k. In the final model, k, the number of neighbors to be examined, was equal to 9.

I implemented the random forest model using the ‘ranger’ package (Wright and Ziegler 2017). I tuned the model by examining a sequence of 1 to 10 for mtry values, minimum node size of 1, 3, or 5, and sample fraction of 0.4, 0.7, and 1. The final model included 800 trees, had an mtry of 6, minimum node size of 5, and a sample fraction of 1.

I used the package ‘xgboost’ (Chen et al. 2021) for building a gradient boosting machine. The multinomial classification option is broken in the ‘gbm’ package (Greenwell et al. 2020), so I used a different package to accommodate the five behavior classes in the response. I tuned the model using a random search (Bergstra and Bengio 2012) with 1,000 iterations. The final model was fit with a maximum depth of 8, eta of 0.22, subsample of 0.74, and 0.63 columns sampled by tree.

For each model, I used the ‘pROC’ package (Robin et al. 2011) to calculate AUC and create ROC curves. The confusionMatrix function from the ‘caret’ package (Kuhn 2021) was used to generate overall accuracy, sensitivity, and specificity for each model. I used the ‘tidyverse’ (Wickham et al. 2019) and ‘recipes’ (Kuhn and Wickham 2021) packages for data formatting and wrangling. All data management and analyses were conducted in R version 4.1.1 (R Core Team 2021).

**Results**

All three methods predicted stork behavior well. KNN had the poorest performance, with an overall accuracy of 0.87 and area under the curve (AUC) of the receiver operating characteristic (ROC) curve was 0.97 (Figure 1). Sensitivity (Table 2) was greatest for SITTING (0.96) and lowest for P\_FLIGHT (0.5). Specificity (Table 3) was highest for A\_FLIGHT and P\_FLIGHT (1.0), and lowest for STAND (0.92). The random forest model had an overall accuracy of 0.91 and AUC of 0.87 (Figure 2). Sensitivity was greatest for WALK (0.95) and lowest for P\_FLIGHT (0.61). Specificity was highest for P\_FLIGHT (1.00) and lowest for WALK (0.94). GBM had an overall accuracy of 0.89 and AUC of 0.98. The full set of sensitivities and specificities are shown in Tables 2 and 3.

**Discussion**

As expected, KNN was outperformed by both the random forest and GMB models based on overall accuracy. However, AUC of the random forest model was less than both of the other models. Given previous usage of random forests in classifying accelerometer data, I found the AUC somewhat surprising. The random forest had the best performance in predicting behavior classes from accelerometer data, but was only slightly better than the GBM. Bursts identified as standing comprised 49% of the total data set. Perhaps greater representation of the other behavior classes could have improved performance of all three methods. It is possible that the random grid tuning procedure for GBM was not run for enough iterations to provide the best parameter estimates for this model. Because this process was time-consuming and computationally-demanding, it was only run for 500 iterations, but for more intensive applications the tuning grid might include a greater number of iterations (e.g., 1,000 or 5,000) or include changes to the random distributions used for drawing the parameters. Further, all tuning had a somewhat subjective feel to it, so it seems possible that the range of values supplied to the tuning functions could be sub-optimal.

Investigation into efficient methods for classifying accelerometer data into behaviors could greatly improve understanding of animal ecology and life cycles, especially for species such as migratory birds that travel great distances or inhabit remote locations and are thus difficult to monitor directly. Additionally, further work could be undertaken to understand the difference between backpack-mounted devices such as those used in this analysis, compared to neck collars that might rotate on an animal, and the associated performance of machine learning models.

**Acknowledgements**

I would like to acknowledge J. Ratschat for his article on tuning xgboost models in R (https://towardsdatascience.com/getting-to-an-hyperparameter-tuned-xgboost-model-in-no-time-a9560f8eb54b).

**Literature Cited**

Bergstra, J. and Bengio, Y. 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13: 281-305.

Brown, D.D., Kays, R., Wikelski, M., Wilson, R. and Klimley, A.P. 2013. Observing the unwatchable through acceleration logging of animal behavior. Animal Biotelemetry 1: 20. https://doi.org/10.1186/2050-3385-1-20

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Yutian. 2021. Xgboost: Extreme Gradient Boosting. R package version 1.5.0.2. https://CRAN.R-project.org/package=xgboost

Dentinger, J.E., Börger, L., Holton, M.D., Jafari-Marandi, R., Norman, D.A., Smith, B.K., Oppenheimer, S.F., Strickland, B.K., Wilson, R.P., and Street, G.M. 2022. A probabilistic framework for behavioral identification from animal-borne accelerometers. Ecological Modelling 464: 109818. https://doi.org/10.1016/j.ecolmodel.2021.109818

Flack, A., Fiedler, W., Blas, J., Pokrovski, I., Kaatz, M., Mitropolsky, M., Aghababyan, K., Fakriadis, Y., Makrigianni, E., Jerzak, L., Shamina, C., Azafzaf, H., Rotics, S., Nathan, R., and Wikelski, M. 2015. Data from: Migration costs of eight Eurasian white stork populations; available at dx.doi.org/10.5441/001/1.78152p3q.

Flack, A., Fiedler, W., Blas, J., Pokrovsky, I., Kaatz, M., Mitropolsky, M., Aghababyan, K., Fakriadis, I., Makrigianni, E., Azafzaf, H., Feltrup-Azafzaf, C., Rotics, S., Mokotjomela, T.M., Nathan, R., and Wikelski, M. 2016. Costs of migratory decisions: A comparison across eight white stork populations. Science Advances 2:e1500931.

Greenwell, B., Boehmke, B., Cunningham, J., and GBM Developers. 2020. Package ‘gbm’: Generalized Boosted Regression Models. R package version 2.1.8. https://cran.r-project.org/web/packages/gbm/index.html

Komsta, L. and Novomestky, 2015. moments: Moments, cumulants, skewness, kurtosis

and related tests. R package version 0.14. https://CRAN.R-project.org/package=moments

Kuhn, M. 2021. Caret: Classification and regression training. R package version 6.0-90. https://CRAN.R-project.org/package=caret

Kuhn, M., and Wickham, H. 2021. Recipes: Preprocessing and feature engineering steps for modeling. R package version 0.1.17. https://CRAN.R-project.org/package=recipes

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., and Getz, W.M. 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. Journal of Experimental Biology 215(6): 986-996. https://doi.org/10.1242/jeb.058602

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria. https://www.R-project.org/

Resheff, Y.S., Rotics, S., Harel, R., Spiegel, O. and Nathan, R. 2014. AcceleRater: a web application for supervised learning of behavioral modes from acceleration measurements. Movement Ecology 2:27.

Robin, X., Turck, N. Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77.

Sequeira, A.M.M., O’Toole, M., Keates, T.R., et al. 2021. A standardization framework for bio-logging data to advance ecological research and conservation. Methods in Ecology and Evolution 12: 996-1007. https://doi.org/10.1111/2041- 210X.13593

Shepard, E.L.C., Wilson, R.P., Quintanta, F., Gómez Laich, A., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A., Morgan, D.T., Myers, A.E., Newman, C., and Macdonald, D.W. 2008. Identification of animal movement patterns using tri-axial accelerometry. Endangered Species Research 10: 47-60. DOI: 10.3354/esr00084

Sur, M., Suffredini,T., Wessells, S.M., Bloom, P.H., Lanzone, M., et al. 2017. Improved supervised classification of accelerometry data to distinguish behaviors of soaring birds. PLOS ONE 12(4): e0174785. https://doi.org/10.1371/journal.pone.0174785

Tuszynski, J. 2021. caTools: Moving window statistics, GIF, Base64, ROC AUC, etc. R package version 1.18.2. https://CRAN.R-project.org/package=caTools

Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. 2019. Welcome to the tidyverse. Journal of Open Source Software 4(43): 1686. DOI: 10.21105/joss.01686

Wright, M.N., and Ziegler, A. 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77(1): 1-17. DOI: 10.18637/jss.v077.i01

**Tables**

Table 1. Summary statistics calculated from raw acceleration data.

|  |  |
| --- | --- |
| **Measure** | **Definition** |
| ODBA | Overall dynamic body acceleration; proxy for energy expenditure |
| Mean of difference in dynamic acceleration in two axes | Average difference in values of two axes. |
| Standard deviation of difference in two axes | Standard deviation of difference in two axes. |
| Skewness | Measure of skew in distribution of all values from each axis |
| Kurtosis | “Tailedness” of distribution of all values from each axis. |
| Covariance between two axes | Covariance between all pairs of two axes |
| Correlation between two axes | Pearson’s correlation coefficient between all pairs of two axes |
| Minimum dynamic acceleration | Minimum dynamic acceleration for each axis |
| Maximum dynamic acceleration | Maximum dynamic acceleration for each axis |
| Amplitude | Difference between minimum and maximum for each axis |

**Table 2.** Sensitivity for each behavior class for each model (K-nearest neighbors, random forest and gradient boosting machines).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Behavior** | | | | |
| **Model** | A\_FLIGHT | P\_FLIGHT | SITTING | STAND | WALK |
| KNN | 0.91 | 0.5 | 0.96 | 0.87 | 0.86 |
| Random Forest | 1.00 | 0.61 | 0.93 | 0.90 | 0.94 |
| GBM | 0.88 | 1.00 | 0.88 | 0.95 | 0.81 |

**Table 3.** Specificity for each behavior class for each model (K-nearest neighbors, random forest and gradient boosting machines).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Behavior** | | | | |
| **Model** | A\_FLIGHT | P\_FLIGHT | SITTING | STAND | WALK |
| KNN | 1.00 | 1.00 | 0.96 | 0.92 | 0.93 |
| Random Forest | 0.99 | 1.00 | 0.98 | 0.95 | 0.94 |
| GBM | 1.00 | 0.97 | 0.99 | 0.89 | 0.98 |

**Figures**

Chart

Description automatically generated**Figure 1.** ROC curve for KNN classification of five behavioral categories from white stork accelerometer data.

**Chart

Description automatically generatedFigure 2.** ROC curve for random forest classification of five behavioral categories from white stork accelerometer data. The AUC for this model was 0.87.

Chart

Description automatically generated**Figure 3.** ROC curve for GBMclassification of five behavioral categories from white stork accelerometer data. The AUC for this model was 0.97.