Can footwear satisfaction be predicted from mechanical properties?

Matthew Q. Salzano\textsuperscript{a}, Gillian Weir\textsuperscript{a}, Jessica Thompson\textsuperscript{b}, Matthieu B. Trudeau\textsuperscript{b}, Christopher Ertel\textsuperscript{b}, Kayley Dear\textsuperscript{b}, Steffen Willwacher\textsuperscript{c} and Joseph Hamill\textsuperscript{a}

\textsuperscript{a}Department of Kinesiology, University of Massachusetts Amherst, Amherst, MA, USA; \textsuperscript{b}Brooks Running Research Laboratory, Brooks, Inc, Seattle, WA, USA; \textsuperscript{c}Biomechanics and Fundamentals of Engineering, Offenburg University of Applied Sciences, Offenburg, Germany

ABSTRACT
Research is often conducted to investigate footwear mechanical properties and their effects on running biomechanics, but little is known about their influence on runner satisfaction, or how well the shoe is perceived. A tool to predict runner satisfaction in a shoe from its mechanical properties would be advantageous for footwear companies. Data in this study were from a database (n = 615 subject-shoe pairings) of satisfaction ratings (gathered after participants ran on a treadmill), and mechanical testing data for 87 unique subjects across 61 unique shoes. Random forest and elastic net logistic regression models were built to test if footwear mechanical properties and subject characteristics could predict runner satisfaction in 3 ways: degree-of-satisfaction on a 7-point Likert scale, overall satisfaction on a 3-point Likert scale, and willingness-to-purchase the shoe (yes/no response). Data were divided into training and validation sets, using an 80–20 split, to build the models and test their accuracy, respectively. Model accuracies were compared against the no-information rate (i.e. proportion of data belonging to the largest class). The models were not able to predict degree-of-satisfaction or overall satisfaction from footwear mechanical properties but could predict runner’s willingness to purchase with 68–75% accuracy. Midsole Gmax at the heel and forefoot appeared in the top five of variable importance rankings across both willingness-to-purchase models, suggesting its role as a major factor in purchase decisions. The negative regression coefficient for both heel and forefoot Gmax indicated that softer midsoles increase the likelihood of a shoe purchase. Future models to predict satisfaction may improve accuracy with the addition of more subject-specific parameters, such as running goals or foot proportions.

ARTICLE HISTORY
Received 16 December 2021
Accepted 10 May 2022

KEYWORDS
Footwear mechanical properties; footwear satisfaction; predictive model; random forest; logistic regression

Introduction
Footwear research and development is often focused on how the designs of various shoe properties alter running biomechanics in order to investigate a link to injury prevention or performance enhancement. An often-overlooked aspect of this research is how changes to footwear mechanical properties influence runner satisfaction in the shoe, or, in other words, how well a shoe meets a runner’s requirements in terms of fit, cushioning, comfort, performance, and other relevant aspects. Here, satisfaction refers to the initial satisfaction felt in a shoe that compels a runner to purchase a shoe. While the benefits of innovations to reduce injury occurrence or aid performance to runners are obvious, the likelihood of a runner buying the shoe will also depend on the satisfaction they feel with the shoe.

Footwear manufacturers often highlight how innovations to shoe features will benefit the runner. This strategy implies runners seek shoes with features that improve performance or reduce injury. Although the literature is inundated with biomechanical studies on footwear characteristics, conclusions surrounding their effects on injury and performance remain equivocal. For example, midsole thickness appears to alter running biomechanics (Chambon et al., 2014; Law et al., 2019; TenBroek et al., 2014), but results remain mixed on how it influences other parameters such as shock attenuation (impact peak, loading rate, tibial acceleration) (Chambon et al., 2014; Hamill et al., 2011;
Law et al., 2019) and running performance (metabolic efficiency) (Burns & Tam, 2020; Hoogkamer, 2020). The story is similar for midsole hardness and impact characteristics (Nigg et al., 2012; Sterzing et al., 2013). More resilient midsole foams seem to improve running economy (Nigg et al., 2012; Sterzing et al., 2013), while heavier shoes increase oxygen consumption by about 1% per 100 grams (Franz et al., 2012; Frederick et al., 1984). Optimal longitudinal stiffness values may exist to benefit running performance (Ortega et al., 2021; Roy & Stefanyshyn, 2006), but these likely vary between individuals and may be speed-dependent (Day & Hahn, 2021; McLeod et al., 2020). Notably, a shoe perceived as more comfortable shows slight benefits to metabolic demand compared to shoes rated as uncomfortable (Luo et al., 2009). This litany of ambiguous results across biomechanical studies provides little insight into which footwear characteristics may be important to individual runners.

While biomechanical studies dominate the literature, footwear manufacturers also collect data on footwear perception through wear testing. Mechanical properties are thought to be necessary for perceived comfort (Miller et al., 2000), but their relationship to overall satisfaction is unclear. In surveys of runners worldwide, cushioning is often cited as a critical feature in footwear (Clifton et al., 2011; Schubert et al., 2011) although the importance of specific features may differ between gender (Clifton et al., 2011) and age (Schubert et al., 2011). Runners can perceive differences in midsole hardness (Milani et al., 1997) and tend to rate softer shoes (i.e. decreased hardness) as having better cushioning (Sterzing et al., 2013, 2015). However, it is not as simple as ‘softer is better’ as hard shoes have been found to be more comfortable than soft shoes when combined with high torsional stiffness (Miller et al., 2000).

It would be helpful for shoe manufacturers to know which mechanical properties are the most important in determining the satisfaction of a shoe to a runner. Furthermore, the ability to predict if a shoe will be well-received based simply on its mechanical properties would be invaluable to footwear manufacturers. Therefore, the purpose of this study was to determine the relationship between the mechanical properties of footwear and runner satisfaction. In addition, we investigated the relationship between perceived cushioning and the willingness of the runners to purchase the footwear.

Methods

Data collection

Data used in this study were obtained from a database of shoe mechanical properties and footwear perception data compiled from multiple studies conducted at the Brooks Run Research Laboratory. The database contained responses ($n = 615$) from 86 unique participants (24 males, 62 females; $66.0 \pm 10.3$ kg; $1.68 \pm 0.07$ m; $36.3 \pm 8.9$ years-old; see Figure 1 for distribution of participant characteristics), in which they rated footwear based on cushioning and satisfaction. Responses were taken after participants ran in the shoes for two minutes on a treadmill. Also noted within the database were mechanical testing data on 61 shoe models across seven brands. This research was conducted according to all applicable regulations, and informed consent was obtained from each participant prior to testing.

The mechanical properties of each shoe were assessed using a custom gravity-driven impact test machine, using the ASTM F1976-13 standard. Each shoe was impacted 30 times at two locations, in the

![Figure 1](image_url)

Figure 1. A simple histogram shows the distribution of males and females in the dataset. Violin plots show the distribution, across all subjects, of age, body mass, and height. The cross and circle represent the median and mean, respectively, of the data.
heel and forefoot (FF), with an 8.5 kg mass. Drop height was varied between shoes to maintain an energy input of 5J. The mean of the last 10 impacts was used to calculate mechanical variables. Mechanical variables derived were: Gmax, energy return (ER), loading rate (LR), time-to-peak (TTP). Midsole stiffness on the Asker-C scale (Duro) was also assessed using a hand-held durometer (Shore C, Instron, Norwood, MA, USA). Other measures related to the mechanical properties of each shoe were heel and FF stack height (including sock liner), longitudinal bending stiffness (Flex), and shoe weight. A summary of the mechanical properties across all shoes can be seen in Figure 2.

Footwear perceptual data were gathered after participants ran on an instrumented force treadmill (Bertec Corporation, Columbus, OH, USA) for one minute at 3.35 m/s and for one minute at their preferred pace. Participants ran in different models of shoes from several manufacturers but were blinded to the brand by taping over any visible logos. After running for a total of two minutes in each shoe, runners were asked a series of questions rating the shoes in terms of heel cushioning, forefoot cushioning, longitudinal flexibility, transition, energy return, stability, and overall satisfaction. Subjects were given extra time on the treadmill if they felt two minutes was not enough time to form an opinion on the shoe. The focus of the models in this paper is on their overall satisfaction, which can also be viewed as their point-of-purchase satisfaction. Subjects were asked to rate their overall satisfaction with the shoe and whether or not they would be willing to purchase it. A definition of satisfaction was not provided to the runners as we wanted them to use their own personal criteria. We did ask that participants try to only consider factors that influence the feel and performance of the shoe (i.e.
cushioning, weight, flexibility, energy return, etc.), and to ignore aesthetics (i.e. colour). Overall satisfaction was rated on a 7-point Likert scale, with very dissatisfied rated as 1 and very satisfied rated as a 7, while willingness-to-purchase the shoe was answered on a Yes/No basis. Satisfaction questions asked to participants, including answer choices, can be seen in Supplementary Material 1. Figure 3 presents the distribution of satisfactory responses across all shoes as grouped by brand.

Data analysis

In pre-processing, 7-point overall satisfaction scores were transformed into a 3-point scale to allow for a general satisfaction score. Here, all dissatisfied responses (score of 1–3 on a 7-point scale) were given a score of 1, all neutral scores (score of 4) were re-scored as 2, and all satisfied responses (score of 5–7) were re-scored as a 3. This removed the level of satisfaction as a confounding factor while also maintaining a distinction between satisfaction and willingness to purchase.

High collinearity amongst predictors can influence the performance and stability of some predictive models (Kuhn & Johnson, 2013). Therefore, a correlation matrix of the mechanical properties was created, and all Pearson Product Moment Correlations (Figure 4) above 0.8 and below −0.8 were evaluated to determine which variables to remove from the models. Correlations between values at the heel and FF were ignored, as previous research has shown different cushioning between the two locations can impact perception (Sterzing et al., 2013). LR, TTP, and Gmax at each location all showed the highest correlations with each other. Since LR is calculated from TTP, and TTP is derived from Gmax, random forest models were run with LR and TTP removed to see if predictive performance improved (see ‘Model building’ section).
Random Forest

Random forest models, decision tree-based algorithms designed for regression or classification, were chosen as one of the types of models to test the predictive power of shoe mechanical properties on overall satisfaction. In essence, a decision tree is a series of ‘if-then’ statements that try to describe the response variable (Kuhn & Johnson, 2013). Decision trees start with a root node, which is the predictor that best splits the data and creates two more nodes (Starmer, 2021). At each subsequent node, another variable is chosen that best splits the data based on a metric of the lowest impurity (e.g., Gini impurity to measure the probability of misclassification using the selected variable) (Starmer, 2021). This process repeats until no parameters remain or the addition of a parameter does not reduce impurity (Starmer, 2018). If a parameter does not reduce impurity, then it is eliminated from the model (Starmer, 2018). In random forest models, a collection, or forest, of trees (on the order of hundreds or thousands) is built and a final prediction is made based on the average or most common prediction of all tree models in the forest (Kuhn & Johnson, 2013). This prediction via aggregation reduces variance in the overall model (Kuhn & Johnson, 2013). Random forests also reduce the correlation between trees by choosing amongst a random subset of predictors at each node, instead of all available predictors (Hastie et al., 2009; Kuhn & Johnson, 2013). The drawback to random forest models is that the relationships between predictors and the outcome remain unknown, but they do provide each predictor’s importance to the model (Kuhn & Johnson, 2013).

The algorithms for each model also contain hyperparameters that can be tuned and optimised to increase accuracy. For the random forest algorithm, the number of variables randomly selected at each node (called \( m_{\text{try}} \)) is the most tuneable hyperparameter, but a default value of \( \sqrt{p} \), where \( p \) is the number of predictors, is typically used (Hastie et al., 2009; Kuhn & Johnson, 2013). Other hyperparameters can also be tuned and optimised (Kuhn & Johnson, 2013), but Probst et al. (2019) found random forest models are robust to the

![Figure 4. A correlation matrix with overlaid heatmap showing pairwise correlations between mechanical property variables. Correlations above 0.8 or below −0.8 were evaluated to determine if they should be removed from the models.](image-url)
values used. Therefore, default values were used for the number of trees in the forest (1000) and the minimum number of samples in the terminal node (1). The ordinal forest algorithm, a variation of the random forest model that handles ordinal outcome variables, also has hyperparameters (number of score sets tried prior to calculation of optimal score set, size of the forests grown for each score set, and size of the final forest) that can be tuned, although default values are also robust to dataset size (Hornung, 2020).

Logistic regression

Logistic regression models, which handle binary responses (i.e. Yes/No) by modelling the probability of belonging to a specified class, were also chosen as one of the predictive modelling methods in this study. While tree-based models (e.g., random forest model) are better able to handle complex data relationships, logistic regressions provide the relationship between the predictors and the outcome (Hastie et al., 2009). An extension of the logistic regression, typically called an ordered logistic regression, handles ordinal data with >2 classes.

Elastic net regularisation can be applied to logistic regression to handle collinearity and provide feature selection (Kuhn & Johnson, 2013; Zou & Hastie, 2003). This method is a combination of two regularisation methods, ridge regression (L2 regularisation) and LASSO (least absolute shrinkage and selection operator, or L1 regularisation), both of which add a penalty to large parameter coefficients if they do not reduce the sums of squared errors (Kuhn & Johnson, 2013; Zou & Hastie, 2003). The elastic net algorithm (Friedman et al., 2010) used in this study has two hyperparameters that require tuning. The first hyper-parameter, \( \alpha \), is the mixing percentage for the two regularisation methods, so that \( \alpha = 0 \) is pure ridge regression and \( \alpha = 1 \) is pure LASSO (Friedman et al., 2010; Kuhn & Johnson, 2013). The second hyperparameter, \( \lambda \), is the strength of the penalty and can be found automatically by the algorithm (Friedman et al., 2010).

Model building

Three models were built using one of the satisfaction scores as the outcome variable. The degree-of-satisfaction model used the original 7-point Likert scale; the overall satisfaction used the 3-point transformed scale; the final model used the willingness-to-purchase responses. The predictors for the models included all of the shoes’ mechanical properties and their classification as a stability shoes (Yes/No), in addition to participant characteristics such as body mass, age, and sex. Two versions of each model were also built: an elastic net logistic regression model using the full set of predictors and a random forest model using a reduced set of predictors (LR and TTP removed due to collinearity). Ordinal forests and ordered logistic regressions were used for the degree-of-satisfaction and overall satisfaction models due to the ordinal nature of the responses. Outliers were removed (one participant and two shoe models) and data were centred and scaled for the logistic regression models. Descriptions of models are presented in Table 1, while the list of predictors used for each model can be seen in Table 2.

The data were randomly divided into a training set (80% of data), used to build and train the models, and a validation set (20% of data), used for validation of the final model. Data splitting was performed using a stratified method based on the shoe model to preserve proportions in the training and validation sets. From the training set, 10 repeats of 5-fold cross-validation were used to build the model. Hyperparameters tuned in the models were: \textit{mtry} (range: 1 to \( \sqrt{p} \)) for the random forest models, the number of trees in the final forest (# trees in the final forest: 1000, 2000, 3000, 4000,
5000) for the ordinal forest models, and a tuning length of 10 for the logistic regression models. Setting the tuning length to 10 tells the algorithm to tune over 10 evenly spaced values of \( \alpha \) between 0.1 and 1, as well as 10 random values of \( \lambda \) for each \( \alpha \). Due to their insensitivity to the dataset size (Hornung, 2020; Probst et al., 2019), other hyper-parameters were kept at their defaults for both model algorithms (Friedman et al., 2010; Hornung, 2020; Kuhn, 2021; Liaw & Wiener, 2002). When all repeats were completed, the model with the highest accuracy was chosen. The accuracy of this model with new data was then tested with the original validation set.

**Statistical analyses**

Model performance was evaluated by comparing its accuracy against the no information rate (NIR). The accuracy is calculated as the percent of correct predictions, and the NIR is the percent of the sample that belongs to the largest group, i.e. the model’s accuracy if it classified every observation in the validation set as belonging to the largest group (Kuhn, 2021). An exact test with an alpha of 0.05 was performed to determine if the accuracy and NIR were significantly different. Sensitivity and specificity were only calculated for models where accuracy was significantly greater than NIR. All statistical analyses were run in R (R Core Team, 2020). The R code used for this study can be found in Supplementary Material 2.

**Results**

Overall, the ordinal forest models to predict degree-of-satisfaction and overall satisfaction performed poorly, with prediction accuracies not significantly different than the NIR (Table 3). However, the random forest model to determine willingness-to-purchase showed 75% accuracy, which was significantly better than the NIR \( (p = 0.001; \text{Table 3}) \). In the random forest version of the willingness-to-purchase model, age and body mass were the most important predictors overall, while Heel Gmax was the most important mechanical property (Table 2). The runner’s sex was the least important factor, and stability shoe classification was eliminated from the model.

The accuracies of the degree-of-satisfaction and overall satisfaction models were not significantly different from the NIR when modelled as ordered logistic regressions (Table 3). The logistic regression model for predicting whether a shoe would be bought had an accuracy of 68%, which was significantly different from the NIR. The \( \alpha \) for the final model was 0.6, indicating an almost even mixture of ridge and LASSO regularisation. Heel ER was the mechanical property with the largest (absolute

---

Table 2. Variable importance, by rank, for each version of the willingness-to-purchase model.

| Rank | Random forest | Elastic net (coef) |
|------|---------------|-------------------|
| 1    | Age           | Stability \(-0.488\) |
| 2    | Body Mass     | Heel ER \(-0.357\) |
| 3    | Heel Gmax     | FF Gmax \(-0.273\) |
| 4    | Heel Stack    | Heel Gmax \(-0.209\) |
| 5    | FF Gmax       | FF Stack \(-0.202\) |
| 6    | FF Duro       | Flex \(-0.173\) |
| 7    | Heel Duro     | Sex \(0.134\) |
| 8    | FF ER         | FF Duro \(-0.091\) |
| 9    | Heel ER       | Age \(0.072\) |
| 10   | Shoe Weight   | Shoe Weight \(-0.046\) |
| 11   | FF Stack      | FF ER \(-0.044\) |
| 12   | Flex          | Heel LR \(-0.021\) |
| 13   | Sex           | Body Mass \(0.008\) |
| 14   | Stability     | Heel Stack \(-0.001\) |
| 15   | Heel Gmax     | Heel Duro |
| 16   | Heel TTP      | FF TTP |
| 17   | FF LR         | FF LR |
| 18   | Intercept     | Intercept \(0.100\) |

Elastic Net refers to the elastic net logistic regression. Italicised variables were eliminated from the models. Regression coefficients (coef) are shown for each variable for the elastic net model.

Table 3. Summary of results for each model.

|                   | Degree-of-satisfaction | Overall satisfaction | Willingness-to-purchase |
|-------------------|-------------------------|----------------------|-------------------------|
|                   | RF          | EN        | RF          | EN        | RF          | EN        |
| Accuracy          | 22.13       | 24.58     | 63.93       | 59.32     | 74.59       | 67.80     |
| 95% CI            | [15.12, 30.54] | [17.12, 33.35] | [54.75, 72.43] | [49.89, 68.27] | [65.91, 82.04] | [58.57, 76.10] |
| NIR               | 26.23       | 23.73     | 60.66       | 57.63     | 50.00       | 55.08     |
| p-value           | \(0.072\)   | \(<0.001\) | \(<0.001\)  | \(<0.001\) | \(<0.001\)  | \(<0.001\) |
| Sensitivity       | –           | –         | –           | –         | 73.77       | 69.81     |
| Specificity       | –           | –         | –           | –         | 75.41       | 66.15     |

Accuracy refers to the model’s accuracy with the overall validation set while NIR is the ‘no information rate’. ‘RF’ and ‘EN’ refer to the random forest and elastic net logistic regression models, respectively. Bold denotes a model with accuracy significantly greater than NIR \( (p < 0.05) \).
value) coefficient, indicating its importance to the model (Table 2). TTP and LR at the FF were eliminated from the model, in addition to the heel durometer.

**Discussion**

The purpose of this study was to determine if the overall satisfaction (i.e. how well a shoe meets a runner’s requirements in terms of fit, cushioning, comfort, performance, and other relevant aspects of a shoe) can be predicted from its mechanical properties. Overall satisfaction ratings, based on comfort, cushioning, fit, performance, etc., for each shoe, were obtained in three ways: on a 7-point Likert scale (i.e. degree-of-satisfaction), said scale transformed into a 3-point scale (i.e. overall satisfaction), and whether or not a person would buy the shoe. Our results suggest that mechanical properties can perform better than random guessing (i.e. NIR), but are still limited in their ability to predict satisfaction. Our models could not predict degree-of-satisfaction, but they were able to predict willingness to buy with approximately 65% accuracy.

Variables describing midsole hardness were consistently ranked among the most important mechanical properties for determining a runner’s willingness to purchase a shoe. Gmax at the heel and forefoot ranked as two of the top three mechanical properties in both random forest and logistic regression versions (Table 2). The negative regression coefficient for both Gmax values suggests that softer shoes increase the likelihood a shoe will be purchased. This finding is supported by previous research that reported the importance of shoe cushioning (Clifton et al., 2011; Schubert et al., 2011), and the preference for softer shoes (Sterzing et al., 2013, 2015).

The importance of mechanical properties implicated in running performance was mixed across the random forests and logistic regression. Lighter shoes (Franz et al., 2012) and more resilient midsole foams (Worobets et al., 2014) are thought to improve the running economy, so it is reasonable to expect shoe weight and energy return to be important to a runner choosing a shoe. However, the models showed that shoe weight was of low importance to the runners when they considered a possible purchase. It is possible that runners could not perceive the differences in shoe weight (Mohr et al., 2016). The importance of energy return in the midsole was quite mixed across models. Heel energy return was the most important mechanical property in the logistic regression (with forefoot energy return ranked quite low), while energy return at the heel and forefoot were of middling importance in the random forest. As a result, energy return as an important feature for footwear satisfaction remains inconclusive, although Clifton et al. (2011) found that it was not a requirement for an ideal shoe in either men or women. However, the negative coefficient for heel energy return in the logistic regression suggests runners prefer shoes with less energy return.

Our results show that mechanical properties are able to predict a runner’s willingness to purchase between two-thirds and three-fourths of the time. However, these prediction accuracies are still not ideal and thus, some improvements to models are required. It is possible that the addition of other mechanical properties not quantified here, such as torsional stiffness (Miller et al., 2000), may improve the accuracies of our models. While we utilised standard metrics typically reported in the footwear industry and scientific literature, non-standard metrics, like the rate of stiffness during loading and unloading, could influence runner perception and are a topic of future exploration. However, Miller et al. (2000) noted that mechanical properties are important for comfort perception but are not enough. Our prediction accuracies agreed with this conclusion, despite trying to control for differences in preference between ages (Schubert et al., 2011) and sexes Clifton et al., 2011). Other subject-specific characteristics are likely needed to create models with better prediction accuracies. Anatomical factors, such as skeletal alignment and foot proportions, may impact perceived fit and comfort (Miller et al., 2000; Mündermann et al., 2001). Controlling for a runner’s goals may also be needed, as leisure runners may have different shoe preferences to goal- or performance-oriented runners. Alternatively, runners could be grouped together based on which features are most important to them when searching for a shoe.

Despite the similar predictive performance, variable importance rankings were quite inconsistent amongst the two versions of the
willingness-to-purchase model (Table 2) with signif-

icant prediction accuracies. For example, the 
classification of a stability shoe was eliminated 
from the random forest but was the most impor-
tant predictor for the logistic regression. This result 
can be explained by the difference in how random 
forests and logistic regressions form relationships 
between predictors. Logistic regressions form linear 
relationships amongst the predictors, while random 
forests form more complex relationships by using 
predictor combinations to create regions within the 
data (Kuhn & Johnson, 2013). The performance of 
random forests can be adversely affected if the data 
does not fit into these regions (Kuhn & Johnson, 
2013), which could explain the similar accuracies 
between the two models; although the random for-
est did perform better than the logistic regression.

There are several limitations associated with this 
study. Likert scales were chosen as the methodo-
logical approach to assess degree-of-satisfaction. 
Visual analogue scales, which use a continuous 
scale as opposed to an ordinal scale are a popular 
alternative. A study comparing the two methodolo-
gies found that visual analogue scales were gen-
ernally more stable across running trials and sessions 
(Mills et al., 2010). However, for overall comfort, 
the Likert scale scores were stable across sessions 
but not trials, while the opposite was found for the 
visual analogue scale (Mills et al., 2010). Thus, it is 
possible that the choice of scale influenced our 
results. The shoes used in this study came primar-
ily (two-thirds) from one footwear company. The 
shoes from one company were a mix of production 
and prototype shoes while shoes from the other 
brands were production-only. However, runners 
were blinded to shoe brands and models, and all 
shoe properties were within the range of what is 
available on the market. Finally, satisfaction scores 
were based on running in the shoes for a short 
period of time on a treadmill. Although this meth-
ethodology reflects the limited experience in a shoe at 
the time of purchase, it is possible that longer runs 
and/or different surfaces could change a runner’s 
perception of a shoe.

Conclusion

In conclusion, it does seem feasible to predict a 
runner’s overall satisfaction with a shoe from 
mechanical properties alone. Our models were not 
able to predict a runner’s degree of satisfaction in a 
shoe and, while models predicting willingness-to-
purchase outperformed random guessing (i.e. no 
information rate), prediction accuracy was still 
relatively low. Future statistical models to predict 
runner satisfaction should incorporate more sub-
ject-specific characteristics, such as whether a run-
ner is a leisure- vs. performance-oriented, the 
distance of a typical run, and various foot dimen-
sions and proportions.

Disclosure statement

The authors declare no conflict of interest.

Funding

This study was funded by a grant from Brooks 
Running, Inc.

ORCID

Matthew Q. Salzano http://orcid.org/0000-0003-
4000-982X
Gillian Weir http://orcid.org/0000-0001-5147-8213
Steffen Willwacher http://orcid.org/0000-0002-1303-3165

References

Burns, G. T., & Tam, N. (2020). Is it the shoes? A simple 
proposal for regulating footwear in road running. British 
Journal of Sports Medicine, 54(8), 439–440. https://doi.
org/10.1136/bjsports-2018-100480
Chambon, N., Delattre, N., Guéguen, N., Berton, E., & Rao, 
G. (2014). Is midsole thickness a key parameter for the 
running pattern? Gait and Posture, 40(1), 58–63. https://
doi.org/10.1016/j.gaitpost.2014.02.005
Clifton, P., Burton, M., Subic, A., Perret-Ellena, T., Bedford, 
A., & Schembri, A. (2011). Identification of performance 
requirements for user-centered design of running shoes. 
Procedia Engineering, 13, 100–106. https://doi.org/10.
1016/j.proeng.2011.05.058
Day, E. M., & Hahn, M. E. (2021). Does running speed 
affect the response of joint level mechanics in non-rear-
foot strike runners to footwear of varying longitudinal 
bending stiffness? Gait and Posture, 84, 187–191. https://
doi.org/10.1016/j.gaitpost.2020.11.029
Franz, J. R., Wierzbinski, C. M., & Kram, R. (2012). 
Metabolic cost of running barefoot versus shod: Is lighter 
better? Medicine and Science in Sports and Exercise, 44(8), 
1519–1525. https://doi.org/10.1249/MSS.0b013e3182514a88
Frederick, E. C., Daniels, J. T., & Hayes, J. W. (1984). The 
effect of shoe weight on the aerobic demands of running.
In N. Bachl, L. Prokop, & R. Suckert (Eds.), *Current topics in sports medicine* (pp. 616–625). Urban & Schwarzenberg.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Hamill, J., Russell, E. M., Gruber, A. H., & Miller, R. (2011). Impact characteristics in shod and barefoot running. *Footwear Science*, 3(1), 33–40. https://doi.org/10.1080/19424280.2010.542187

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. (2nd ed.). Springer-Verlag.

Hoogkamer, W. (2020). More isn’t always better. *Footwear Science*, 12(2), 75–77. https://doi.org/10.1080/19424280.2019.1710579

Hornung, R. (2020). Ordinal forests. *Journal of Classification*, 37(1), 4–17. https://doi.org/10.1007/s00357-018-9302-x

Kuhn, M. (2021). caret: Classification and Regression Training. R package version 6.0-91, <https://CRAN.R-project.org/package=caret>.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (1st ed.). Springer. https://doi.org/10.1007/978-1-4614-6849-3

Law, M. H. C., Choi, E. M. F., Law, S. H. Y., Chan, S. S. C., Wong, S. M. S., Ching, E. C. K., Chan, Z. Y. S., Zhang, J. H., Lam, G. W. K., Lau, F. O. Y., & Cheung, R. T. H. (2019). Effects of footwear midsole thickness on running biomechanics. *Journal of Sports Sciences*, 37(9), 1004–1010. https://doi.org/10.1080/02640414.2018.1538066

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22. https://cran.r-project.org/doc/Rnews/

Luo, G., Stergiou, P., Worobets, J., Nigg, B., & Stefanyshyn, D. (2009). Improved footwear comfort reduces oxygen consumption during running. *Footwear Science*, 1(1), 25–29. https://doi.org/10.1080/19424280902993001

McLeod, A. R., Brunning, D., Johnson, A. W., Ward, J., & Hunter, I. (2020). Improving running economy through altered shoe bending stiffness across speeds. *Footwear Science*, 12(2), 79–89. https://doi.org/10.1080/19424280.2020.1734870

Milani, T. L., Hennig, E. M., & Lafortune, M. A. (1997). Perceptual and biomechanical variables for running in identical shoe constructions with varying midsole hardness. *Clinical Biomechanics*, 12(5), 294–300. https://doi.org/10.1016/S0268-0033(97)00008-9

Miller, J. E., Nigg, B. M., Liu, W., Stefanyshyn, D. J., & Nurse, M. A. (2000). Influence of foot, leg and shoe characteristics on subjective comfort. *Foot and Ankle International*, 21(9), 759–767. https://doi.org/10.1177/10711007002100908

Mills, K., Blanch, P., & Vicenzino, B. (2010). Identifying clinically meaningful tools for measuring comfort perception of footwear. *Medicine and Science in Sports and Exercise*, 42(10), 1966–1971. https://doi.org/10.1249/MSS.0b013e3181dbc8

Mohr, M., Trudeau, M. B., Nigg, S. R., & Nigg, B. M. (2016). Increased athletic performance in lighter basketball shoes: Shoe or psychology effect? *International Journal of Sports Physiology and Performance*, 11(1), 74–79. https://doi.org/10.1123/ijssp.2014-0538

Mündermann, A., Stefanyshyn, D. J., & Nigg, B. M. (2001). Relationship between footwear comfort of shoe inserts and anthropometric and sensory factors. *Medicine and Science in Sports and Exercise*, 33(11), 1939–1945. https://doi.org/10.1097/00005768-200111000-00021

Nigg, B. M., Baltich, J., Maurer, C., & Federolf, P. (2012). Shoe midsole hardness, sex and age effects on lower extremity kinematics during running. *Journal of Biomechanics*, 45(9), 1692–1697. https://doi.org/10.1016/j.jbiomech.2012.03.027

Ortega, J. A., Healey, L. A., Swinnen, W., & Hoogkamer, W. (2021). Energetics and biomechanics of running footwear with increased longitudinal bending stiffness: A narrative review. *Sports Medicine*, 51, 873–894. https://doi.org/10.1007/s40279-020-01406-5

Frobst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, 20(53), 1–32.

R Core Team. (2020). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. https://www.r-project.org/

Roy, J. P. R., & Stefanyshyn, D. J. (2006). Shoe midsole longitudinal bending stiffness and running economy, joint energy, and EMG. *Medicine & Science in Sports & Exercise*, 38(3), 562–569. https://doi.org/10.1249/01.mss.0000193562.22001.e8

Schubert, C., Oriwol, D., & Sterzing, T. (2011). Gender and age related requirements of running shoes: A questionnaire on 4501 runners. In Footwear Science: Vol. 3:sup1. https://doi.org/10.1080/19424280.2011.575850

Starmer, J. (2018). StatQuest: Decision Trees, Part 2 – Feature Selection and Missing Data. YouTube. https://www.youtube.com/watch?v=wpNl-JwwplA&list=PLbh5jK0oLUlCTaGLRoHQD5uF7qGfuJF&index=43

Starmer, J. (2021). Decision and classification trees, clearly explained!!! YouTube. https://www.youtube.com/watch?v=_L39rN6gz7Y

Sterzing, T., Custoza, G., Ding, R., & Cheung, J. T. M. (2015). Segmented midsole hardness in the midfoot to forefoot region of running shoes alters subjective perception and biomechanics during heel-toe running revealing potential to enhance footwear. *Footwear Science*, 7(2), 63–79. https://doi.org/10.1080/19424280.2015.1008589

Sterzing, T., Schweiger, V., Ding, R., Cheung, J. T. M., & Brauner, T. (2013). Influence of rearfoot and forefoot midsole hardness on biomechanical and perception variables during heel-toe running. *Footwear Science*, 5(2), 71–79. https://doi.org/10.1080/19424280.2012.757810
TenBroek, T. M., Rodrigues, P. A., Frederick, E. C., & Hamill, J. (2014). Midsole thickness affects running patterns in habitual rearfoot strikers during a sustained run. *Journal of Applied Biomechanics, 30*(4), 521–528. https://doi.org/10.1123/jab.2012-0224

Worobets, J., Wannop, J. W., Tomaras, E., & Stefanyshyn, D. (2014). Softer and more resilient running shoe cushioning properties enhance running economy. *Footwear Science, 6*(3), 147–153. https://doi.org/10.1080/19424280.2014.918184

Zou, H., & Hastie, T. (2003). Regression shrinkage and selection via the elastic net, with applications to microarrays. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67*(2), 301–320. http://www.jstor.org/stable/3647580