Immediate effects of EVA midsole resilience and upper shoe structure on running biomechanics: a machine learning approach

Andrea N. Onodera[1,2,*], Wilson P. Gavião Neto[3], Maria Isabel Roveri[1], Wagner R. Oliveira[2] and Isabel CN Sacco[1,*]

[1] Physical Therapy, Speech and Occupational Therapy Department, University of São Paulo, School of Medicine, São Paulo, Brazil

[2] Dass Nordeste Calçados e Artigos Esportivos Inc, Ivoti, Rio Grande do Sul, Brazil

[3] School of Engeneering & IT, Centro Universitário Ritter dos Reis, Porto Alegre, Rio Grande do Sul, Brazil

* These authors contributed equally to this work.

ABSTRACT

Background. Resilience of midsole material and the upper structure of the shoe are conceptual characteristics that can interfere in running biomechanics patterns. Artificial intelligence techniques can capture features from the entire waveform, adding new perspective for biomechanical analysis. This study tested the influence of shoe midsole resilience and upper structure on running kinematics and kinetics of non-professional runners by using feature selection, information gain, and artificial neural network analysis.

Methods. Twenty-seven experienced male runners (63 ± 44 km/week run) ran in four-shoe design that combined two resilience-cushioning materials (low and high) and two uppers (minimalist and structured). Kinematic data was acquired by six infrared cameras at 300 Hz, and ground reaction forces were acquired by two force plates at 1,200 Hz. We conducted a Machine Learning analysis to identify features from the complete kinematic and kinetic time series and from 42 discrete variables that had better discriminate the four shoes studied. For that analysis, we built an input data matrix of dimensions 1,080 (10 trials $\times$ 4 shoes $\times$ 27 subjects) $\times$ 1,254 (3 joints $\times$ 3 planes of movement $\times$ 101 data points + 3 vectors forces $\times$ 101 data points + 42 discrete calculated kinetic and kinematic features).

Results. The applied feature selection by information gain and artificial neural networks successfully differentiated the two resilience materials using 200(16%) biomechanical variables with an accuracy of 84.8% by detecting alterations of running biomechanics, and the two upper structures with an accuracy of 93.9%.

Discussion. The discrimination of midsole resilience resulted in lower accuracy levels than did the discrimination of the shoe uppers. In both cases, the ground reaction forces were among the 25 most relevant features. The resilience of the cushioning material caused significant effects on initial heel impact, while the effects of different uppers were distributed along the stance phase of running. Biomechanical changes due to shoe midsole resilience seemed to be subject-dependent, while those due to upper structure seemed to be subject-independent.

INTRODUCTION

Sports shoes have many roles in running; among them, providing adequate impact-force absorption (*Clarke, Frederick & Cooper, 1983*; *Hennig, 2011*), stability for foot/ankle movements (*Cheung, Wong & Ng, 2011*) and comfort (*Nigg, 2010*). These roles have been the most studied in running and shoe biomechanics so far. Running shoes are basically constituted by upper, midsole and sole. Among different possible combinations of these three elements in the shoe construction, the upper is definitely the most prone to variations in its construction, such as color, model design, added elements and materials, and certainly, the last two factors will have a particular influence on running biomechanics. Runners select a comfortable running shoe using their own comfort criteria (*Nigg et al., 2015*) and, because the shoe upper maintains a large contact area with the foot, it would have a stronger influence over fit and comfort, which in turn would impact in runner's kinematic and kinetic strategies during practice and competitions. It has been demonstrated that a firmer foot contact with a shoe resulted in lower loading rates due to a better coupling of foot-footwear, which optimizes the use of the midsole impact absorption technology by favoring a better foot positioning inside the shoe (*Hagen & Hennig, 2009*). To investigate the isolated influence of upper types in running biomechanics would help runners to have a more comprehensive and efficient approach in the shoe construction process, as well as in the choice of the running shoe by runners.

Nevertheless, the most manipulated and studied shoe part in biomechanics is still the midsole (*Maclean, Davis & Hamill, 2009*; *Milani, Hennig & Lafortune, 1997*; *Nigg et al., 2003*; *Worobets et al., 2014*). The majority of shoe companies invests a large amount of time, effort and money on development of damping materials technologies, such as gels, air, and springs for supposedly improving sports performance. Ethylene-vinyl acetate (EVA) is a copolymer of ethylene and vinyl acetate highly elastic sintered to form a porous material similar to rubber, yet with excellent toughness. Its porous elastomeric characteristic is much more flexible as low-density polyethylene, commonly used in shoes construction, and because of its properties of resistance, flexibility, temperature toughness, it has been one of the most used copolymer in the shoe midsole construction (*Verdejo & Mills, 2004*; *Wang, Hong & Li, 2012*).

The midsole hardness is the most explored physical characteristic of the midsole in biomechanical studies (*Clarke, Frederick & Cooper, 1983*; *Hennig, Valiant & Liu, 1996*; *Kersting & Brüggemann, 2006*; *Maclean, Davis & Hamill, 2009*; *Milani, Hennig & Lafortune, 1997*; *Nigg et al., 2012*). Running with hard shoes resulted in same peak magnitude vertical GRF (ground reaction force) as running with soft ones (*Clarke, Frederick & Cooper, 1983*; *Kersting & Brüggemann, 2006*; *Nigg et al., 1987*) and faster time to achieve the first peak (*Clarke, Frederick & Cooper, 1983*). Therefore, midsole hardness affected the loading rate but not in a proportional rate (*Milani, Hennig & Lafortune, 1997*). Most

of running kinematic changes due to the midsole hardness occurs at the ankle joint (*Clarke, Frederick & Cooper, 1983*; *Hardin, Van den Bogert & Hamill, 2004*; *Maclean, Davis & Hamill, 2009*) and some authors state that these different midsole hardness lead to different impact perception by runners (*Hennig, Valiant & Liu, 1996*; *Milani, Hennig & Lafortune, 1997*), which in turn causes distinct alteration on running mechanics (*Kersting & Brüggemann, 2006*; *Nigg et al., 2012*) and may mislead the real impact damping by shoe midsoles (*Hennig, Valiant & Liu, 1996*; *Milani, Hennig & Lafortune, 1997*).

Apart from midsole hardness, resilience is also an important mechanical property of midsole that has been seldom studied (*Sinclair et al., 2014*; *Sinclair et al., 2016*; *Worobets et al., 2014*). It represents the energy restored by the cushioning material after an applied force ceases. Managing resilience while maintaining the hardness of a polymeric foam is possible by adding different kinds of compounds to its formula. EVA added to elastomers could have ideal softness and high resilience characteristics, would have a full-recovery capacity for the next foot step after a heel strike, while a less resilient (more viscous) material would have the capacity of attenuating more energies at initial loading cycles, easily achieving compression flattening after some cycles. It is expected that different resilience materials would mainly reflect different initial impact forces, because more resilient materials will quickly restore the cushioning property while less resilient materials will take a little longer to restore the cushioning property (*Sun et al., 2008*). *Sinclair et al. (2014)* have shown that running with shoes with energy return component resulted in greater tibial acceleration peak, calcaneous eversion and internal tibia rotation compared to conventional running shoes. In a later study, they have shown lower oxygen consumption and respiratory exchange ratio with more resilient model of shoes (*Sinclair et al., 2016*). *Worobets et al. (2014)* manipulated only the midsole materials, maintaining the upper structure, and also reported lower oxygen consumption when running with a more compliant/resilient midsole condition. However, the isolated effect of resilience changing in shoe midsole is still unknown in running biomechanics.

The majority of biomechanics studies vary the shoe model as a whole to investigate the effects of various structural shoe properties and elements while running (*Azevedo et al., 2012*; *Braunstein et al., 2010*; *Dixon, 2008*; *McNair & Marshall, 1994*). Such an approach deeply interferes with an appropriate differentiation and interpretation of which shoe characteristics most influence the kinetic and kinematic changes during running. The novelty of this study proposal was to manipulate selectively the upper structure and the cushioning material resilience and to investigate the effects of this manipulation in the biomechanics of running. Identifying more precisely which shoe characteristics really matter for impact attenuation and lower limb kinematic adaptation would help runners to choose more wisely the running shoes regardless the brand or model and direct further running training regimes based on that choice.

Nonetheless, individual's mechanical and neuromuscular adaptations to changes in shoes are influenced by mechanical, neurophysiological, anatomical and even psychological factors and, therefore, is likely to observed different individuals using different strategies in response to changes in running shoes (*Kersting & Brüggemann, 2006*; *Nigg*

et al., 2003). Thereby, one may conclude that regardless the type of shoe modification, the biomechanical responses observed may be subject-dependent.

We proposed to identify the relevant biomechanical features which are most affected by different shoes conditions during running, two midsole resiliencies and two upper structures. We adopted an approach based on machine learning (ML), which has been used in the literature to identify crucial features and relevant patterns for classifying and predicting locomotor patterns as being a result of a given health condition (*Hoerzer et al., 2015*; *Muniz et al., 2010b*; *Schöllhorn et al., 2002*), but it has never been used before to identify effects of different running shoes on biomechanics. In contrast to the related literature, our ML strategy is entirely supervised, and it consisted of using Information Gain (IG) to select important features and Artificial Neural Networks (ANN) to classify the different shoe resiliencies and upper types. Our assumptions were: (h1) low versus high resilient cushioning effects on running kinematics and ground reaction forces are classifiable by using a ML approach; (h2) structured versus minimalist upper effects on running kinematics and ground reaction forces are classifiable by using a ML approach; and (h3) that there are biomechanical changes due to shoe midsole resilience or upper structure that are subject-independent.

METHODS

Subjects

Twenty-seven experienced non-professional male runners (36.0 ± 7.3 years old, 1.72 ± 0.05 m, 73.9 ± 6.2 kg, 62.9 ± 43.8 km/week run, 7.5 ± 7.1 years of practice) with a rearfoot strike pattern and with no experience in minimalist shoe participated in this study. All runners were free of injury or musculoskeletal pain according to the definition of *Macera et al. (1989)*; had no major foot or ankle postural alterations or deformities, excessive static pronation or supination of the foot and ankle complex according to the Foot Posture Index (*Redmond, Crosbie & Ouvrier, 2006*); and did not present leg length discrepancy greater than 1 cm. All athletes that participated in the Porto Alegre's Marathon in 2012 were invited by electronic media, and 158 athletes answered positively to participate in the study; however, only 27 effectively matched the eligibility criteria and came to the lab to be evaluated. Then, a telephone call was made to select runners that fit to the inclusion criteria and scheduled the biomechanical evaluation in the laboratory. All subjects agreed to participate in the study approved by the Ethics Committee of the School of Medicine of the University of Sao Paulo (Ethical Application CEP-FMUSP: protocol #054/14) and signed a written consent form.

Tested running shoes

Four running shoes were especially developed by a local sportive shoe manufacturer. The final masses of the 4 constructions were equivalent to avoid negative effects due to mass differences (*Frederick, 1984*). All shoes were constructed using the same last, the same design of upper pieces, midsole geometry, and outsole. The hardness of cushioning materials was fixed at 40 Asker C, measured by a durometer (GoTechAskerC, Taichung,

Figure 1 **Image of shoes prototypes used in the experiment.** Illustration of testing shoes. (A) Structured Upper condition (SU). (B) Minimalist Upper condition (MU). (C) Low Resilience cushioning material condition (LR). (D) High Resilience cushioning material condition (HR).

Taiwan). The two shoe uppers had the same design and shape, but the different structure and materials:

(1) **SU—structured upper**: 15 mm of soft foam in the heel collar and tongue, hard heel cup involving the medial, lateral and posterior parts of the heel, synthetic pieces sewed in the vamp and doubled fabric over the whole shoe (Fig. 1A).

(2) **MU—minimalist upper**: light-weight mesh, tongue without foam, without heel cup, and almost all pieces of the upper were connected by means of heat fusion (Fig. 1B).

Both cushioning materials were made of ethylene-vinyl acetate (EVA); were inserted in the same rearfoot area within the midsole; and had an oval shape of 10 mm thickness, 50 mm width, and 70 mm length. Resilience was assessed by vertical resiliometer (GoTech GT7042-V1, Taichung, Taiwan). The midsoles were: (1) **LR—low resilience**—5% of resilience (±3%) (Fig. 1C), and (2) **HR—high resilience**—55% of resilience (±3%) (Fig. 1D).

The first tested condition (condition 1—upper SU and cushioning material LR) was the same for all runners; the other three testing conditions were randomized for each subject using simple draw. The other three conditions were: condition 2—upper SU and cushioning material HR condition 3—upper MU and cushioning material LR, and condition 4—upper MU and cushioning material HR. The subjects were asked to lace their shoes tightly and comfortably, in the same way they typically lace during their running practice.

Experimental protocol

Kinematic data were acquired by six infrared cameras (VICON T-40, Oxford, UK). Sixteen passive-reflexive markers (14 mm diameter) were fixed on both lower limbs (two anterior superior iliac spines; two posterior iliac spines; two lateral epicondyles of the knees; two markers over the lower lateral 1/3 surface of the thighs; two lateral malleolus; two markers over the lower 1/3 of the shank; two second metatarsal heads; two posterior surface of calcaneous at the same height above the plantar surface of the foot as the toes markers) according to Plugń Gait marker set (*Kadaba, Ramakrishnan & Wootten, 1990*). The two foot markers were fixed on the shoes (second metatarsal heads and calcaneous) after deep palpation of bone prominences. The laboratory coordinate system was established at one corner of one force plate, and all initial calculations were based on this global coordinate system. In Nexus software (Vicon Nexus 1.7, Oxford, UK), each data sample from each lower limb segment (foot, shank, and thigh) was modeled as a rigid body with a local coordinate system that coincided with anatomical axes. Translations and rotations of each segment were reported relative to neutral positions defined during the static standing trial.

The program calculates the joint angles by means of a decomposition matrix based on Cardan sequences and six degrees of freedom model. The decomposition matrix describes the relationship between two local coordinate systems, one for each segment between which the relative angle is determined. The joint kinematics was considered as the movement of the distal segment in relation to the proximal; e.g., for determining the knee angle, the thigh was the proximal segment and the shank the distal one. The movements occur around 3 different axes which describe two definition of movement each: flexion/extension, abduction/adduction, and internal/external rotation (*Hamill, Selbie & Kepple, 2014*).

Ground reaction forces were acquired at 1200 Hz by two force plates (AMTI BP600600, Watertown, USA) embedded in the center of a 25 m walkway. Acquisitions of kinematic and force data were synchronized by a 64 multichannel Vicon MX Giganet Lab and A/D converter.

Running velocity was kept between 9.5 and 10.5 km/h (mean 10.1 ± 0.5 km/h), monitored by 2 photoelectrical sensors (Tecsistel Speed View, Novo Hamburgo, Brazil). Ten trials per subject for each shoe condition were collected, resulting in 40 trials on the dominant limb. The limb dominance was defined as the leg used to kick a soccer ball (*Greenberger & Paterno, 1995*).

Biomechanical data analysis

In accordance with many studies in biomechanics (*Muniz et al., 2010b*; *Nigg et al., 2012*), we adopted a Butterworth filter (implemented using the original code from MATLAB) to minimize noise, and the marker coordinates were filtered using a 12 Hz zero-lag fourth-order low-pass Butterworth filter. Force data was filtered with a 300 Hz zero-lag fourth-order low-pass Butterworth filter also implemented in a MATLAB code. The angular and force data from initial contact to take-off were normalized in stance time (interpolated 0–100%) and in magnitude by the body weight.

The 30 discrete kinematic features analyzed were: peak angles (degrees), angles at the beginning of stance phase (degrees), instant of peak angle (seconds), range of motion from the beginning of stance phase to peak angle (degrees), and final angle of stance phase (degrees); for ankle, knee and hip joints; for sagittal and frontal planes of movement ($5 \times 3 \times 2$).

For vertical and antero-posterior forces, the magnitude of first vertical force peak (1VFP) (body weight), time of 1VFP (milliseconds), magnitude of second vertical force peak (2VFP)(body weight), time of 2VFP (milliseconds), loading rate (slope of 20%–80% of 1VFP) (body weight/second), time of minimal vertical force in midstance (milliseconds), propulsion rate (slope of curve between minimal vertical force in midstance and the 2VFP) (body weight/second), minimum breaking antero-posterior force (body weight), breaking antero-posterior impulse (body weight*second), median frequency of 1VFP (Hz), time of stance phase (milliseconds), and decay rate (slope of curve from 2VFP to the end of stance phase) (body weight/ second) were calculated.

The whole interpolated time-series of all three planes of motion (sagittal, frontal and transversal) and forces (vertical, antero-posterior and medio-lateral) were also analyzed. Usually, cross-sectional studies that investigate shoe effects in running biomechanics involve high-dimensional and redundant datasets (*Maurer et al., 2012*; *Nigg et al., 2012*), and feature selection techniques have been used to help identifying the biomechanical parameters that is most influenced by shoe characteristics (*Hoerzer et al., 2015*; *Maurer et al., 2012*; *Nigg et al., 2012*). This was the main reason why we chose to include both discrete and whole time-series points in the analysis. As explained in the next sections, we adopted an approach based on ANN, which can receive large numbers of data simultaneously and the pieces of data do not have to be isolated from each other (*Barton & Lees, 1997*).

Machine learning approach

To assess the effects of shoe interventions, many studies have generated high-dimensional and redundant datasets (*Maurer et al., 2012*; *Nigg et al., 2012*), which impose challenges for understanding an underlying phenomenon of interest (*Guyon et al., 2006*; *Yu & Liu, 2003*). To overcome these challenges, Machine Learning (ML) techniques have been adopted to find patterns on biomechanical data (*Begg & Kamruzzaman, 2005*; *Hoerzer et al., 2015*; *Maurer et al., 2012*; *Muniz et al., 2010b*; *Nigg et al., 2012*). ML aims to learn from data. In a typical classification scenario, we have a categorical outcome (like low vs high shoes resilience) that we wish to predict or classify based on a set of features or variables (like ground reaction forces). On the basis of a training set of data, we observe the outcome and feature measurements for a set of instances (like the subject's trials) (*Hastie, Tibshirani & Friedman, 2001*). Using this data, we build a classification model, which will enable us to classify the outcome for new unseen instances. A good model is one that accurately classify such an outcome. This scenario characterizes a supervised learning problem since each observed instance involves the outcome variable (i.e., the desired output value) to guide the learning process.

A typical supervised classification approach consists of two parts: (i) *variable or feature selection* and (ii) *classification*. Information Gain (IG) and Artificial Neural Networks

(ANN) are ML techniques that have been successfully used for feature selection and classification in many areas. Whereas ANN have been used to classify patterns in biomechanics studies (*Hastie, Tibshirani & Friedman, 2001*; *Muniz et al., 2010b*; *Rupérez et al., 2012*; *Witten, Frank & Hall, 2011*), IG has not been explored in biomechanics. Instead of IG, Principal Components Analysis (PCA) has been a popular technique to select/construct features on walking and running biomechanical data (*Hoerzer et al., 2015*; *Maurer et al., 2012*; *Muniz et al., 2010b*; *Nigg et al., 2012*). Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are techniques that have been used in classification tasks, like in distinguishing effects of velocity (*Joo et al., 2014*), aging (*Begg & Kamruzzaman, 2005*; *Fukuchi et al., 2011*; *Wu, Wang & Liu, 2007*), gender (*Baltich, Maurer & Nigg, 2015*; *Maurer et al., 2012*), diseases (*Muniz et al., 2010a*; *Muniz et al., 2010b*; *Muniz & Nadal, 2009*; *Nüesch et al., 2012*) and footwear conditions (*Trudeau et al., 2015*). As a result, age, gender (*Maurer et al., 2012*; *Nigg et al., 2012*) and the inter-subject's movement variability (*Federolf, Boyer & Andriacchi, 2013*; *Von Tscharner, Enders & Maurer, 2013*) are intrinsic factors that have shown more influence than shoes characteristics on running biomechanics.

Therefore, when the problem involves assessing the effects of shoes interventions, the feature selection stage has to be carefully conducted since it plays a critical role in minimizing bias and the influence of such intrinsic factors. In this context, even PCA has limitations (*Von Tscharner, Enders & Maurer, 2013*), and a subject-independent analysis on shoes interventions is still an open issue in the literature. In contrast to PCA, IG is a supervised method that ranks variables individually without applying data transformations, and it has the potential to facilitate the interpretation of the influence of a single variable on the underlying classification task (*Begg & Kamruzzaman, 2005*; *Muniz & Nadal, 2009*).

As many studies in the literature (*Joo et al., 2014*; *Oh, Choi & Mun, 2013*; *Schöllhorn, 2004*), we adopted ANN for classification, even though SVM has been also successfully used for the same task (*Begg & Kamruzzaman, 2005*; *Maurer et al., 2012*; *Nigg et al., 2012*; *Trudeau et al., 2015*). Both SVM and ANN have been producing state-of-the-art results, even though some comparative studies (*Begg & Kamruzzaman, 2005*; *Fischer et al., 2011*; *Muniz et al., 2010b*; *Yang et al., 2012*) have indicated a slight advantage in favor of SVM. However, SVM were originally designed for binary classification, which involves only two classes (e.g., low versus high resilient), and an effective way of extend it for multiclass classification is a research issue (*Hsu & Lin, 2002*). On the other hand, an ANN maps straightforwardly into multiclass classification without requiring any further adjustments to approach the problem. In this study, although we have approached shoes resilience and upper characteristics as two binary and independent classification problems, a natural extension of this study will focus on the fact that the effects on running may be related to not only single shoes characteristics, but also to the combination of them, resulting in a classification problem with more than two classes. In this context, ANN represents a more stable scenario for further comparisons between binary and multiclass classification results.

Input variables and feature selection by IG

The 3D joint angular displacement time series was vectorized to a 1,254 dimensional vector (3 lower limb joints $\times$ 3 planes of angular displacement $\times 101$ interpolated data points $+ 3$ vectors of ground reaction forces $\times 101$ interpolated data points $+ 42$ discrete calculated kinetic and kinematic features). An input data matrix M was then created (10 trials $\times$ 4 shoe condition $\times$ 27 subjects), resulting in a matrix dimension of $1,080 \times 1,254$. The 1,080 lines of M represented each subject trial in terms of 1,254 input variables, some of which may be more affected by specific characteristics of shoe design. In this context, finding a small subset of input variables is a desired result (*Rupérez et al., 2012*) because that may indicate a more discriminative and less redundant subset of features that would improve the results of the classification.

It was not practical to test all subsets of the 1,254 input variables/columns available, then we used IG to rank the variables in a decreasing order of relevance. As a supervised method, IG ranks an input variable X according to their discriminative power to separate the subject's trials in terms of a target variable C, like shoes resilience or upper. Usually, a distinc t value c_i of a target variable C is known as a class. In our study, $C = \{low, high\}$ for resilience and $C = \{minimalist, structured\}$ for upper. IG is a correlation measure based on the information-theoretical concept of *entropy*, and the entropy of a variable C is defined as in Eq. (1) (*Yu & Liu, 2003*):

$$H(C) = -\sum_{i} P(c_i) \log_2(P(c_i)) \tag{1}$$

and the entropy of C after observing values of a variable X is defined as Eq. (2)

$$H(C|X) = -\sum_{j} P(x_j) \sum_{i} P(c_i|x_j) \log_2(P(c_i|x_j)) \tag{2}$$

where $P(c_i)$ is the prior probabilities for the values of C, and $P(c_i|x_j)$ is the posterior probabilities of C given the values of X. The amount by which the entropy of C decreases reflects additional information about C provided by X, which is called *information gain* and it is given by Eq. (3)

$$IG(C|X) = H(C) - H(C|X). \tag{3}$$

In our experiments, diverse subsets of variables were tested; they contained an increasing number of variables (25, 50, 100, 150 and 200) and bigger subsets were systematically formed by aggregating less relevant variables according to the IG criteria. This step was crucial to determine the smallest number of variables that achieved accurate discrimination of the resiliencies and upper structures. In this context, the greatest subset of variables involved in our experiments included 200 variables, since our analysis indicated that 200 variables were enough to evaluate the hypotheses of this study.

Classification Procedure

As described previously, lines in the matrix M are instances of subject's trials, and each trial belongs to a class c_i of resilience (*low* or *high*) and a class of upper (*minimalist* or *structured*). As a supervised classification approach, a subset of subject's trials was used

for training and fitting a classification model, and the rest of the subject's trials were used for validating the model. To estimate how accurately the classification model will perform in practice, we adopted the standard k-fold-cross-validation, which divides the subject's trials into k mutually exclusive folds of nearly equal size: $k-1$ folds are used for training, and the remaining fold for testing. The procedure repeats k times, such that each fold is used once for validation. For this reason, the validation results are usually averaged over the k rounds. For each training round, a subset of variables was selected by using IG, their values were scaled from -1 to 1, and then an ANN learning model was trained.

The classification accuracy of the resulting ANN model was then computed on the test fold. Because we are dealing with binary classification problems (i.e., two classes of resilience and upper), accuracy is given in terms of the entries of a confusion matrix for positive and negative classes (*Han, Pei & Kamber, 2011*), as shown in Eq. (4). Given two classes (e.g., *low* and *high* resilience), we can express in terms of positive trials (trials of one class, e.g., *resilience = low*) versus negative trials (e.g., *resilience = high*). True positives (TP) refer to the positive trials that were correctly classified by the ANN model, while true negatives (TN) are the negative trials that were correctly classified by the ANN model. Similarly, false positives (FP) and false negatives (FN) are the negative and positive trials, respectively, that were incorrectly classified by the ANN model.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}. \tag{4}$$

We used a traditional feed-forward multi-layer perceptron network with 3 layers (*Fischer et al., 2011*): an input, an output and a single hidden layer (*Hastie, Tibshirani & Friedman, 2001*). The number of entries in the input layer corresponds to the number of selected variables in the feature selection stage, and the output layer consists of 2 neurons, which corresponds to each class c_i in our target variables, i.e., $C = \{low, high\}$ for resilience or $C = \{minimalist, structured\}$ for upper. As an ANN requires a parameter setting, which is still a research issue, we perform an exhaustive searching through a subset of parameters values: the number of neurons in the hidden layer was selected from the set $\{10, 25, 50, 75\}$; the learning rate $\in \{0.05, 0.15, 0.25\}$ and the number of training cycles $\in \{300, 700, 1,100, 1,500\}$. To reduce the risk of overfitting, we adopted the decay procedure (*Hastie, Tibshirani & Friedman, 2001*), as implemented in the RapidMiner software.

We conducted the classification in two contexts:

I. a 4-fold-cross-validation for *each subject* (40 trials) to assess the existence of effects from shoes conditions, finding one accuracy value to discriminate the shoe condition for each subject and for each subset of features. The purpose of considering trials of one subject only is to conclude about our assumptions h1 and h2 in a context that avoids the influence of subject's intrinsic factors.

II. a standard 10-fold-cross-validation involving *all subjects*' trials to assess the existence of subject-independent changes induced by the shoes interventions, finding one accuracy value for each subset of features.

By comparing classification accuracies between the contexts I and II, it was possible to analyze the subject-dependency of the results and provisionally evaluate a pattern

Table 1 Number of variables of each feature subset according to IG rank for midsole resilience comparison.

Resilience comparison	25 most relevant variables	50 most relevant variables	100 most relevant variables	150 most relevant variables	200 most relevant variables
Medio-lateral GRF	9	11	14	17	18
Sagittal Ankle	6	9	16	18	25
Transversal Hip	3	12	17	33	39
Vertical GRF	2	4	6	12	16
Antero-posterior GRF	2	2	7	11	17
Discrete Force Variables	2	3	4	4	4
Discrete Kinematic Variables	1	1	1	2	2
Frontal Knee		7	12	12	14
Sagittal Hip		1	19	35	38
Transversal Knee			4	5	12
Frontal Ankle				1	5
Frontal Hip					6
Transversal Ankle					4
% of total features (1,254)	2%	4%	8%	12%	16%

induced by different resilience and upper conditions. Classification accuracy of higher than 80% was considered good (*Hoerzer et al., 2015*), and we reported results when the ANN method achieved the best classification accuracy. The machine learning procedures were conducted in the software RapidMiner (v.5.3.015, Dortmund, Germany).

RESULTS

All 1,254 variables were involved in the experiments of resiliencies and uppers. In both cases, the accuracy of discriminating the effect of shoe conditions indicated that context I outperformed context II.

The composition of the most relevant features subsets according to IG are detailed in Tables 1 and 2 for the midsole resilience and upper structures comparisons, respectively. Although forces have been the most discriminative variables for both midsole and upper, the top five features for upper and resilience came from different components, respectively vertical and medio-lateral forces.

Resilience effect

In context II, 200 variables were sufficient to distinguish midsole resiliencies with an accuracy of 84.8% (Fig. 2, red line). The accuracies indicated that context I, which considers only trials of a single subject, outperformed context II for all subsets of features (Fig. 2, blue line). A mean accuracy of 89.4% ($\pm$8.3%) to classify the two resiliencies was reached by considering only the 25 most relevant features, while the best accuracy of 93.9% ($\pm$5.0) to classify the two resiliencies was reached with the 150 most relevant features.

Table 2 Number of variables of each feature subset according to IG rank for upper structures comparison.

Upper structure comparison	25 most relevant variables	50 most relevant variables	100 most relevant variables	150 most relevant variables	200 most relevant variables
Antero-posterior GRF	10	19	33	38	40
Vertical GRF	8	12	16	19	21
Discrete Force Variables	4	4	4	6	9
Sagittal Ankle	2	10	20	22	28
Medio-lateral GRF	1	4	12	18	22
Transversal Knee		1	7	19	25
Frontal Ankle			5	13	26
Transversal Ankle			2	9	10
Discrete Kinematic Variables			1	2	2
Sagittal Knee				4	7
Transversal Hip					6
Frontal Knee					4
% of total features (1,254)	2%	4%	8%	12%	16%
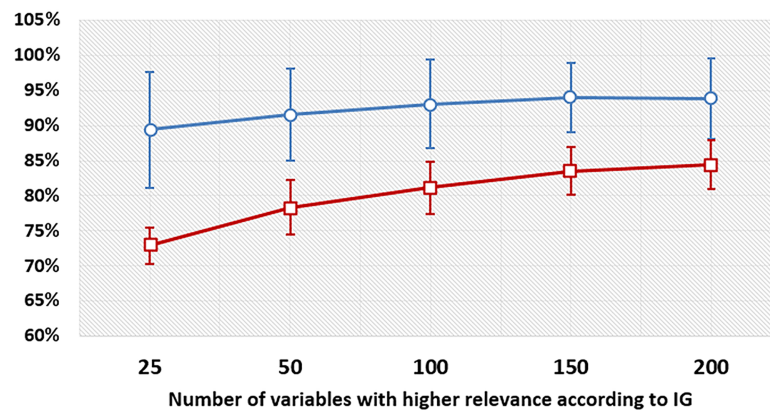
Figure 2 Accuracy levels to discriminate midsole resilience materials in various contexts considering different subsets of variables. Mean accuracy and standard deviation for each subset of input variables with the highest IG values to discriminate resilience materials. Red line represents the context II and considers all subjects together. Blue line represents the context I and considers each subject in isolation.

Among the 200 most relevant variables to discriminate between low and high resilience cushioning materials, six of them were discrete biomechanical variables (4 ground-reaction force variables and 2 kinematic variables) (Fig. 3). The 5 most relevant variables came from medio-lateral force, between 6% and 10% of stance phase.

Upper structure effect

The results indicated that the upper structures effects were less complex than the cushioning materials ones. In the context II, accuracy higher than 85% was achieved by considering only 25 variables to differentiate upper structures (Fig. 4, red line). As in the case of resiliencies, results on uppers shown that context I outperformed context II; it
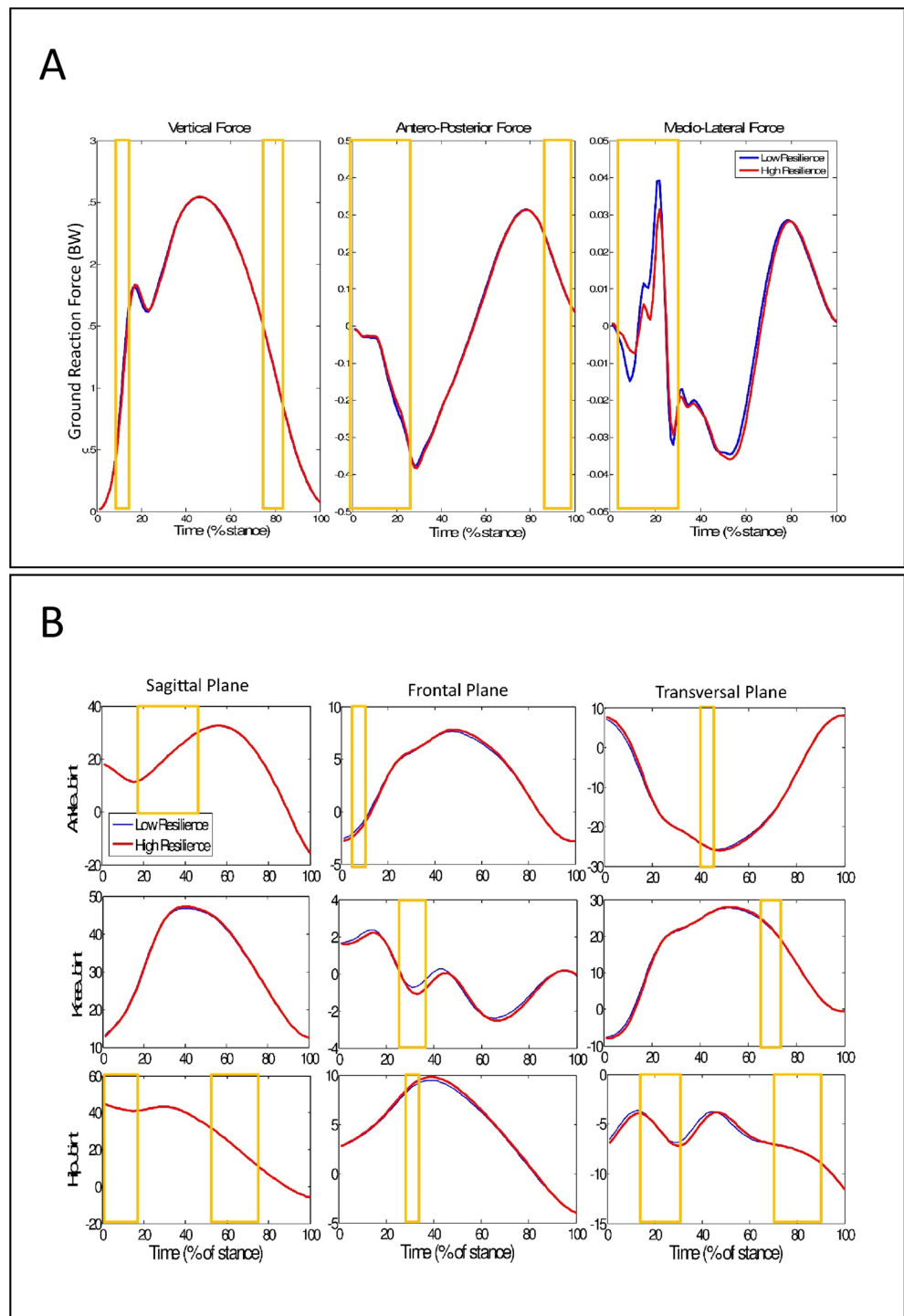
Figure 3 Ground reaction force and Kinematics time-series during running with different resilience midsoles. (A) Mean time series of ground reaction force for different resilience of cushioning materials. (B) Mean time series of joints kinematics in all planes of motion for different resilience materials. Blue lines represent the low resilience cushioning condition and red dotted lines represent the high resilience cushioning condition. The 200 highest IG variables are highlighted in the yellow boxes.
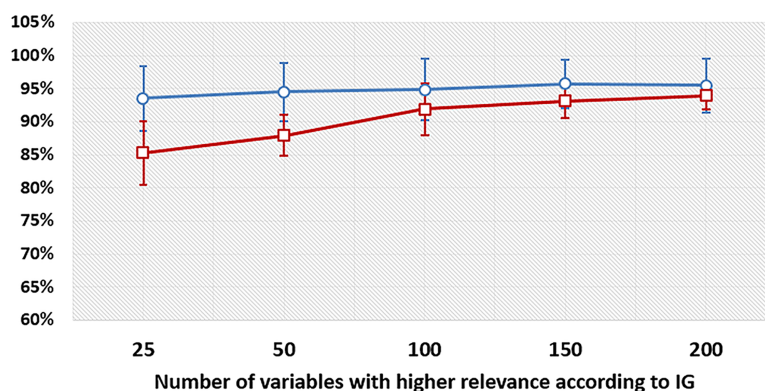
Figure 4 Accuracy levels to discriminate upper structures in various contexts considering different subsets of variables. Mean accuracy and standard deviation for each subset of input variables with the highest IG values to classify upper structures. Red line represents the context II and considers all subjects together. Blue line represents the context I and considers each subject in isolation.

was possible to obtain a mean accuracy of 93.4% (±4.8%) with 25 variables, and 95.6% (±3.8) with 150 variables (Fig. 4, blue line).

Among the 200 most relevant variables to discriminate structured and minimalist uppers 11 of them came from the discrete biomechanical variables (nine force variables and two kinematic variables) (Fig. 5). The five most relevant variables to discriminate between uppers were vertical forces from 11% to 14% of stance phase and the first peak.

DISCUSSION

We proposed an entirely supervised approach based on ANN to distinguish the effects of different midsole resiliencies and upper structures of shoes on running biomechanics. The results confirmed our first and second hypotheses because it was possible to observe the effects on running kinematics and kinetics caused by low and high resilience cushioning midsoles and structured and minimalist uppers by the adopted ML approach. IG was efficient in selecting important features, as was confirmed by the proportionally slower increase in classification accuracy with respect to increasing numbers of input features (Fig. 4). The top biomechanical variables in the IG rank may be considered the most responsible for distinguishing the effects of the upper structures and midsole resiliencies.

When analyzing all subjects together, the methodology successfully differentiated the two resiliencies with 84.8% accuracy using 200 variables and the two shoe uppers with 85.3% accuracy using 25 variables, which is higher than the classification rate threshold of 80% chosen by *Hoerzer et al. (2015)*. Intra-subject analysis increased the classification accuracy for resiliencies to a mean of 89.4% and for uppers using just 25 variables to a mean of 93.4%. This indicated that the adopted ML analysis achieved more accuracy to identify different conditions (cushioning materials and shoe uppers) within a given subject than within the set of all subjects, which is consistent with the higher inter-subject variability.

Among the variables in the interpolated time series, the 25 most relevant features for discriminating midsole resilience were mainly forces, ankle flexion-extension, and hip
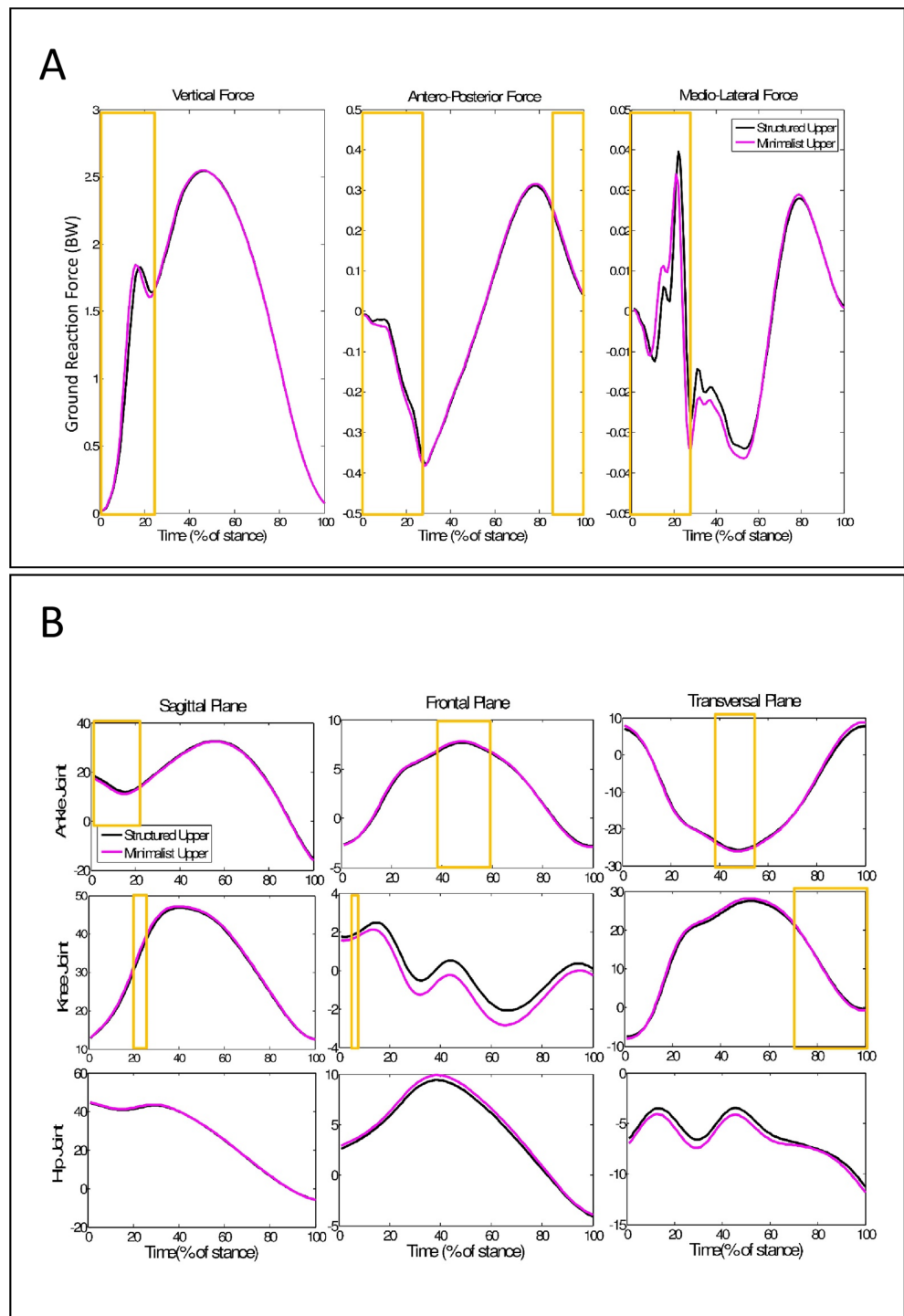
Figure 5 **Ground reaction force and kinematics time-series during running with different shoe upper structures.** (A) Mean time series of ground reaction force for different shoe upper structures. (B) Mean time series of joints kinematics in all planes of motion for different shoe upper structures. Black lines represent the structured upper condition and Pink dotted lines represent the minimalist upper condition. The 200 highest IG variables are highlighted in the yellow boxes.

rotation variables. *Sinclair et al. (2014)* reported higher tibial acceleration, calcaneus eversion and internal rotation of tibia when running with high resilience midsole shoes. However, due to the different shoe brands, part of these results could be attributed to the midsole resilience, and part of them to other parts of the shoe (differences in the upper, for example). According to our results, the main biomechanical alterations caused by different midsole resilience will be on ground reaction forces, sagittal plane of ankle and transversal plane of hip. Considering the discrete variables that differentiated the resiliencies, four were related to vertical force (first peak and minimal force) and two were related to ankle kinematics (dorsiflexion and eversion). Resilience causes significant effects on the initial impact of the heel with the ground while running. The cushioning materials were inserted only under the heel part of the shoe; considering we focused our study on rearfoot strikers, it was acceptable that the resilience especially influenced the variables related to this first part of the stance phase.

The 200 most relevant features that discriminated the two resiliencies were distributed in short time windows spread over all cycle periods of the kinematic and kinetic time series. This does not mean, however, that these 200 features were equally relevant for all 27 subjects. These short and sparse windows corresponded to individual patterns of biomechanical responses, or a group of individuals with the same biomechanical responses, which leads us to refute part of the third hypothesis and conclude that changes due to shoe midsole resilience seemed to be subject-dependent. This means that is possible to discriminate the two resiliencies with a fewer number of variables and with a higher accuracy if we analyze individualized biomechanical data. In summary, each runner responded differently to resilience in the cushioning materials, changing a different pool of biomechanical variables that represent distinct motor strategies. According to the ''muscle tuning paradigm'' proposed by *Nigg & Wakeling (2001)*, the individualized ability to modify the muscle tonus in response to impact stimuli is one of the causes of different adaptabilities to shoes observed among subjects. Additionally, the most discriminative kinematic features were not found at the same time periods as the most discriminative kinetic features in the stance phase, suggesting that kinematics adjustments in the lower limbs caused by shoe changes might not be influencing impact attenuation. This differs from what *Hennig, Valiant & Liu (1996)* and *Milani, Hennig & Lafortune (1997)* suggested in their studies.

The upper structures classification had higher discriminative power than midsole resiliencies, as was reflected in the higher accuracy levels of upper structure classification. The 25 most relevant features for discriminating upper structures were composed mainly of force variables (both discrete variables and time series features) and sagittal ankle variables. From the 200 features, we found that the most relevant biomechanical variables for the classification of uppers were concentrated in the first and last third of the stance phase for all three-force components, sagittal plane of ankle, and for all planes of knee. The exceptions were for the frontal and transverse planes of the ankle, which had the central third as the most relevant part for classification. Therefore, changes in shoe uppers seem to be more subject-independent.

Eleven discrete biomechanical features were among the most relevant for classifying upper structures. Nine were force variables related to first and second vertical peaks, minimal vertical force, and breaking antero-posterior force; two were from ankle kinematics. According to the relevant discrete parameters, it seemed that force features had highest discriminative power, so were more relevant for differentiating upper structures than kinematic variables and they were temporally distributed across the stance phase. The flexibility of the upper (due to different applied materials) could affect the flexibility of the foot inside the shoe, its own capability to absorb impact by stretching the foot arches and generate the propulsion force, so it is reasonable to assume that the force features across the entire stance phase are relevant variables for distinguishing upper differences. The differences in structure probably affect sensitivity to the ground, and consequently how runners modulate applied force to the ground. As seen by *Hagen & Hennig (2009)*, firmer foot contact with a shoe could result in lower loading rates due to a better coupling between the foot and shoe, which in turn facilitates the use of the impact-absorbing technology of the midsole.

In a training context, it would be ideal if each runner performed a biomechanical assessment of his or her running shoes. Our sample was constituted by rearfoot strikers, which were habituated to classic running shoes (not minimalist). This running strategy is the most prevalent among runners (*De Almeida et al., 2015*; *Hasegawa, Yamauchi & Kraemer, 2007*) and according to *Lieberman et al. (2010)*, rearfoot strikers who grew up wearing shoes are more prone to be influenced by shoes conditions, so for this type of runners it is especially important to understand how this external factor (running shoe) changes (or not) each runner's particular running mechanics. Moreover, only few variables among the 1,254 available (about 12% for different cushioning materials and 4% for different shoe uppers) are evidently necessary to detect alterations of running biomechanics with high accuracy (93.9% ± 5.0% and 94.3% ± 4.5%, respectively). This result is plausible with the theory of "functional groups" (*Hoerzer et al., 2015*; *Nigg et al., 2003*) and it is appropriate to understand particular cases. Our results showed that runners have different responses to the materials used in the shoes and a general conclusion arising from a heterogeneous sample may lead to wrong outcomes. Further studies could analyze the influence of smaller sub-groups with similar biomechanical responses to shoe characteristics. This procedure would be useful for understanding how different resilience of cushioning material affects the running mechanics for specific "types" of athletes. It was also possible to demonstrate that a biomechanical study of sports shoes isolating characteristics to be tested provided more specificity in the comprehension of the influence each part on running biomechanics. Future studies could explore other types of feature analysis from biomechanical data such as complexity (*Sejdić et al., 2014*), or sample entropy (*Rathleff et al., 2010*; *Søndergaard et al., 2010*) to distinguish running shoe properties and constructions from the perspective of signal variability.

CONCLUSION

The applied methodology based on feature selection by IG and classification by ANN successfully differentiated the high and low resilience materials and the structured and

minimalist uppers with accuracies of higher than 85% using 200 features (16% of 1,254 available features). The classification of upper structures presented higher accuracy levels than that of midsole resilience probably due to higher inter-individual variability, but in both cases the forces are among the 25 most relevant features subset. The ground reaction forces are the most important features to differentiate midsole resilience and the resilience caused valuable effects on initial heel impact while running. The different patterns of biomechanical response chosen by runners to adapt to different resilience probably led to lower accuracy levels for this classification. We can therefore conclude that biomechanical changes due to shoe midsole resilience seem to be subject-dependent and changes due to upper structure seem to be subject-independent.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The research has been co-funded by Brazilian Agency for Research funding (CAPES) and Dass Nordeste Calçados e Artigos Esportivos S/A. Wagner Oliveira works in Dass Nordeste Calçados e Artigos Esportivos S/A. Andrea N Onodera works in Dass Nordeste Calçados e Artigos Esportivos S/A and conducts her PhD supervised by Dr Isabel CN Sacco. Mrs Maria Isabel Roveri is part of the primary authors PhD, but is not associated to Dass Nordeste Calçados e Artigos Esportivos S/A and is funded by a CAPES scholarship. Data was collected, analyzed and the paper written with no influence from the funding agencies or shoe company, and no author will receive anything of value from the commercial products included in this paper.

The authors, therefore, have no competing interests.

Author Contributions

- Andrea N. Onodera conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Wilson P. Gavião Neto conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Maria Isabel Roveri and Wagner R. Oliveira conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.
- Isabel CN Sacco conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The School of Medicine at the University of Sao Paulo granted Ethical approval to carry out this study (Ethical Application CEP-FMUSP: Protocol #054/14).

Data Availability

The following information was supplied regarding data availability:

Figshare: https://figshare.com/s/6530932d478b54ff5d6c.

REFERENCES

Azevedo AP, Brandina K, Bianco R, Oliveira VH, Souza JR, Mezencio B, Amadio AC, Serrão JC. 2012. Effects of replica running shoes upon external forces and muscle activity during running. *Journal of Sports Sciences* **30**(9):929–935 DOI 10.1080/02640414.2012.682080.

Baltich J, Maurer C, Nigg BM. 2015. Increased vertical impact forces and altered running mechanics with softer midsole shoes. *PLOS ONE* **10**(4):e0125196 DOI 10.1371/journal.pone.0125196.

Barton JG, Lees A. 1997. An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait & Posture* **5**(1):28–33 DOI 10.1016/S0966-6362(96)01070-3.

Begg R, Kamruzzaman J. 2005. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics* **38**:401–408 DOI 10.1016/j.jbiomech.2004.05.002.

Braunstein B, Arampatzis A, Eysel P, Brüggemann GP. 2010. Footwear affects the gearing at the ankle and knee joints during running. *Journal of Biomechanics* **43**:2120–2125 DOI 10.1016/j.jbiomech.2010.04.001.

Cheung RT, Wong MY, Ng GY. 2011. Effects of motion control footwear on running: a systematic review. *Journal of Sports Sciences* **29**(12):1311–1319 DOI 10.1080/02640414.2011.591420.

Clarke TE, Frederick EC, Cooper LB. 1983. Effects of shoe cushioning upon ground reaction forces in running. *International Journal of Sports Medicine* **4**:247–251 DOI 10.1055/s-2008-1026043.

De Almeida MO, Saragiotto BT, Yamato TP, Lopes AD. 2015. Is the rearfoot pattern the most frequently foot strike pattern among recreational shod distance runners? *Physical Therapy in Sport* **16**:29–33 DOI 10.1016/j.ptsp.2014.02.005.

Dixon SJ. 2008. Use of pressure insoles to compare in-shoe loading for modern running shoes. *Ergonomics* **51**:1503–1514 DOI 10.1080/00140130802239562.

Federolf PA, Boyer KA, Andriacchi TP. 2013. Application of principal component analysis in clinical gait research: identification of systematic differences between healthy and medial knee-osteoarthritic gait. *Journal of Biomechanics* **46**:2173–2178 DOI 10.1016/j.jbiomech.2013.06.032.

Fischer A, Do M, Stein T, Asfour T, Dillmann R, Schwameder H. 2011. Recognition of individual kinematic patterns during walking and running—a comparison of artificial neural networks and support vector machines. *International Journal of Computer Science in Sport* **10**:63–67.

Frederick EC. 1984. Physiological and ergonomics factors in running shoe design. *Applied Ergonomics* **15**:281–287 DOI 10.1016/0003-6870(84)90199-6.

Fukuchi RK, Eskofier BM, Duarte M, Ferber R. 2011. Support vector machines for detecting age-related changes in running kinematics. *Journal of Biomechanics* **44**:540–542 DOI 10.1016/j.jbiomech.2010.09.031.

Greenberger HB, Paterno MV. 1995. Relationship of knee extensor strength and hopping test performance in the assessment of lower extremity function. *Journal of Orthopaedic and Sports Physical Therapy* **22**:202–206 DOI 10.2519/jospt.1995.22.5.202.

Guyon I, Gunn S, Nikravesh M, Zadeh LA. 2006. *Feature extraction: foundations and applications.* New York: Springer-Verlag Berlin Heidelberg.

Hagen M, Hennig EM. 2009. Effects of different shoe-lacing patterns on the biomechanics of running shoes. *Journal of Sports Sciences* **27**:267–275 DOI 10.1080/02640410802482425.

Hamill J, Selbie WS, Kepple TM. 2014. Three-dimensional kinematics. In: *Research methods in biomechanics.* Human kinetics, 2nd edition.

Han J, Pei J, Kamber M. 2011. *Data mining: concepts and techniques.* Boston: Morgan Kaufmann.

Hardin EC, Van den Bogert AJ, Hamill J. 2004. Kinematic adaptations during running: effects of footwear, surface, and duration. *Medicine and Science in Sports and Exercise* **36**:838–844.

Hasegawa H, Yamauchi T, Kraemer WJ. 2007. Foot strike patterns of runners at the 15-km point during an elite-level half marathon. *Journal of Strength and Conditioning Research* **21**:888–893.

Hastie T, Tibshirani R, Friedman J. 2001. *The elements of statistical learning.* New York: Springer New York Inc.

Hennig EM. 2011. Eighteen years of running shoe testing in Germany—a series of biomechanical studies. *Footwear Science* **3**:71–81 DOI 10.1080/19424280.2011.616536.

Hennig EM, Valiant GA, Liu Q. 1996. Biomechanical variables and the perception of cushioning for running in various types of footwear. *Journal of Applied Biomechanics* **12**:143–150 DOI 10.1123/jab.12.2.143.

Hoerzer S, Von Tscharner V, Jacob C, Nigg BM. 2015. Defining functional groups based on running kinematics using Self-Organizing Maps and Support Vector Machines. *Journal of Biomechanics* **48**:2072–2079 DOI 10.1016/j.jbiomech.2015.03.017.

Hsu CW, Lin CJ. 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* **13**:415–425 DOI 10.1109/72.991427.

Joo S, Oh SE, Sim T, Kim H, Choi CH, Koo H. 2014. Prediction of gait speed from plantar pressure using artificial neural networks. *Expert Systems with Applications* **41**:7398–7405 DOI 10.1016/j.eswa.2014.06.002.

Kadaba MP, Ramakrishnan HK, Wootten ME. 1990. Measurement of lower extremity kinematics during level walking. *Journal of Orthopaedic Research* **8**:383–392 DOI 10.1002/jor.1100080310.

Kersting UG, Brüggemann GP. 2006. Midsole material-related force control during heel-toe running. *Research in Sports Medicine* **14**:1–17 DOI 10.1080/15438620500528158.

Lieberman DE, Venkadesan M, Werbel WA, Daoud AI, D'Andrea S, Davis IS, Mang'eni RO, Pitsiladis Y. 2010. Foot strike patterns and collision forces in habitually barefoot versus shod runners. *Nature* **463**:531–535 DOI 10.1038/nature08723.

Macera CA, Pate RR, Powell KE, Jackson KL, Kendrick JS, Craven TE. 1989. Predicting lower-extremity injuries among habitual runners. *Archives of Internal Medicine* **149**:2565–2568 DOI 10.1001/archinte.1989.00390110117026.

Maclean CL, Davis IS, Hamill J. 2009. Influence of running shoe midsole composition and custom foot orthotic intervention on lower extremity dynamics during running. *Journal of Applied Biomechanics* **25**:54–63 DOI 10.1123/jab.25.1.54.

Maurer C, Federolf P, Von Tscharner V, Stirling L, Nigg BM. 2012. Discrimination of gender-, speed-, and shoe-dependent movement patterns in runners using full-body kinematics. *Gait Posture* **36**:40–45 DOI 10.1016/j.gaitpost.2011.12.023.

McNair PJ, Marshall RN. 1994. Kinematic and kinetic parameters associated with running in different shoes. *British Journal of Sports Medicine* **28**:256–260 DOI 10.1136/bjsm.28.4.256.

Milani TL, Hennig EM, Lafortune MA. 1997. Perceptual and biomechanical variables for running in identical shoe constructions with varying midsole hardness. *Clinical Biomechanics* **12**:294–300 DOI 10.1016/S0268-0033(97)00008-9.

Muniz AM, Liu H, Lyons KE, Pahwa R, Liu W, Nadal J. 2010a. Quantitative evaluation of the effects of subthalamic stimulation on gait in Parkinson's disease patients using principal component analysis. *International Journal of Neuroscience* **120**:609–616 DOI 10.3109/00207454.2010.504904.

Muniz AM, Liu H, Lyons KE, Pahwa R, Liu W, Nobre FF, Nadal J. 2010b. Comparison among probabilistic neural network, support vector machine and logistic regression

for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait. *Journal of Biomechanics* **43**:720–726 DOI 10.1016/j.jbiomech.2009.10.018.

Muniz AM, Nadal J. 2009. Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait. *Gait Posture* **29**:31–35 DOI 10.1016/j.gaitpost.2008.05.015.

Nigg BM. 2010. *Biomechanics of sport shoes.* Alberta: Topline Printing Inc.

Nigg BM, Bahlsen HA, Luethi SM, Stokes S. 1987. The influence of running velocity and midsole hardness on external impact forces in heel-toe running. *Journal of Biomechanics* **20**:951–959 DOI 10.1016/0021-9290(87)90324-1.

Nigg BM, Baltich J, Hoerzer S, Enders H. 2015. Running shoes and running injuries: mythbusting and a proposal for two new paradigms: 'preferred movement path' and 'comfort filter'. *British Journal of Sports Medicine* **49**:1290–1294 DOI 10.1136/bjsports-2015-095054.

Nigg BM, Baltich J, Maurer C, Federolf P. 2012. Shoe midsole hardness, sex and age effects on lower extremity kinematics during running. *Journal of Biomechanics* **45**:1692–1697 DOI 10.1016/j.jbiomech.2012.03.027.

Nigg BM, Stefanyshyn D, Cole G, Stergiou P, Miller J. 2003. The effect of material characteristics of shoe soles on muscle activation and energy aspects during running. *Journal of Biomechanics* **36**:569–575 DOI 10.1016/S0021-9290(02)00428-1.

Nigg BM, Wakeling JM. 2001. Impact forces and muscle tuning: a new paradigm. *Exercise and Sport Sciences Review* **29**:37–41 DOI 10.1097/00003677-200101000-00008.

Nüesch C, Valderrabano V, Huber C, Von Tscharner V, Pagenstert G. 2012. Gait patterns of asymmetric ankle osteoarthritis patients. *Clin Biomech (Bristol, Avon)* **27**:613–618 DOI 10.1016/j.clinbiomech.2011.12.016.

Oh SE, Choi A, Mun JH. 2013. Prediction of ground reaction forces during gait based on kinematics and a neural network model. *Journal of Biomechanics* **46**:2372–2380 DOI 10.1016/j.jbiomech.2013.07.036.

Rathleff MS, Olesen CG, Moelgaard CM, Jensen K, Madeleine P, Olesen JL. 2010. Non-linear analysis of the structure of variability in midfoot kinematics. *Gait Posture* **31**:385–390 DOI 10.1016/j.gaitpost.2010.01.008.

Redmond AC, Crosbie J, Ouvrier RA. 2006. Development and validation of a novel rating system for scoring standing foot posture: the Foot Posture Index. *Clin Biomech (Bristol, Avon)* **21**:89–98 DOI 10.1016/j.clinbiomech.2005.08.002.

Rupérez MJ, Martín-Guerrero JD, Monserrat C, Alcañiz M. 2012. Artificial neural networks for predicting dorsal pressures on the foot surface while walking. *Expert Systems with Applications* **39**:5349–5357 DOI 10.1016/j.eswa.2011.11.050.

Schöllhorn WI. 2004. Applications of artificial neural nets in clinical biomechanics. *Clin Biomech (Bristol, Avon)* **19**:876–898 DOI 10.1016/j.clinbiomech.2004.04.005.

Schöllhorn WI, Nigg BM, Stefanyshyn DJ, Liu W. 2002. Identification of individual walking patterns using time discrete and time continuous data sets. *Gait Posture* **15**:180–186 DOI 10.1016/S0966-6362(01)00193-X.

Sejdić E, Lowry KA, Bellanca J, Redfern MS, Brach JS. 2014. A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**:603–612 DOI 10.1109/TNSRE.2013.2265887.

Sinclair J, Franks C, Fau-Goodwin J, Naemi R, Chockalingam N. 2014. Influence of footwear designed to boost energy return on the kinetics and kinematics of running compared to conventional running shoes. *Comparative Exercise Physiology* **10**:199–206 DOI 10.3920/CEP140010.

Sinclair J, Mcgrath R, Brook O, Taylor PJ, Dillon S. 2016. Influence of footwear designed to boost energy return on running economy in comparison to a conventional running shoe. *Journal of Sports Sciences* **34**:1094–1098 DOI 10.1080/02640414.2015.1088961.

Søndergaard KH, Olesen CG, Søndergaard EK, De Zee M, Madeleine P. 2010. The variability and complexity of sitting postural control are associated with discomfort. *Journal of Biomechanics* **43**:1997–2001 DOI 10.1016/j.jbiomech.2010.03.009.

Sun PC, Wei HW, Chen CH, Wu CH, Kao HC, Cheng CK. 2008. Effects of varying material properties on the load deformation characteristics of heel cushions. *Medical Engineering & Physics* **30**:687–692 DOI 10.1016/j.medengphy.2007.07.010.

Trudeau MB, Von Tscharner V, Vienneau J, Hoerzer S, Nigg BM. 2015. Assessing footwear effects from principal features of plantar loading during running. *Medicine and Science in Sports and Exercise* **47**:1988–1996 DOI 10.1249/MSS.0000000000000615.

Verdejo R, Mills NJ. 2004. Heel-shoe interactions and the durability of EVA foam running-shoe midsoles. *Journal of Biomechanics* **37**:1379–1386 DOI 10.1016/j.jbiomech.2003.12.022.

Von Tscharner V, Enders H, Maurer C. 2013. Subspace identification and classification of healthy human gait. *PLOS ONE* **8**(7):e65063 DOI 10.1371/journal.pone.0065063.

Wang L, Hong Y, Li JX. 2012. Durability of running shoes with ethylene vinyl acetate or polyurethane midsoles. *Journal of Sports Sciences* **30**:1787–1792 DOI 10.1080/02640414.2012.723819.

Witten IH, Frank E, Hall MA. 2011. *Data mining: practical machine learning tools and techniques.* San Diego: Morgan Kaufmann.

Worobets J, Wannop JW, Tomaras E, Stefanyshyn D. 2014. Softer and more resilient running shoe cushioning properties enhance running economy. *Footwear Science* **6**:147–153 DOI 10.1080/19424280.2014.918184.

Wu J, Wang J, Liu L. 2007. Feature extraction via KPCA for classification of gait patterns. *Human Movement Science* **26**:393–411 DOI 10.1016/j.humov.2007.01.015.

Yang M, Zheng H, Wang H, McClean S, Hall J, Harris N. 2012. A machine learning approach to assessing gait patterns for Complex Regional Pain Syndrome. *Medical Engineering & Physics* **34**:740–746 DOI 10.1016/j.medengphy.2011.09.018.

Yu L, Liu H. 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: *Twentieth international conference on machine learning*, 856–863.