PREDICTION OF MILITARY COMBAT CLOTHING SIZE USING DECISION TREES AND 3D BODY SCAN DATA

Kolose, S., Stewart, T., Hume, P.A., Tomkinson, G. Applied Ergonomics. April 2021.

ABSTRACT

Aim: To determine how well decision tree models can predict tailor-assigned uniform sizes using anthropometry data from the New Zealand Defence Force Anthropometry Survey (NZDFAS). This information may inform automatic sizing systems for military personnel.

Methods: Anthropometric data from two separate samples of the New Zealand Defence Force military were used. Data on Army personnel from the NZDFAS (n = 583) were used to develop a series of shirt- and trouser-size prediction models based on decision trees. Different combinations of physical, automatic, and post-processed measurements (the latter two derived from a 3D body scan) were trialled, and the models with the highest cross-validation accuracy were retained. The accuracy of these models were then tested on an independent sample of Army recruits (n = 154).

Results:

The automated measurement method (measurements derived automatically by the body scanner software) were the best predictors of shirt size (58.1% accuracy) and trouser size (61.7%), with body weight and waist girth being the strongest predictors. Clothing sizes that were incorrectly predicted by the model where generally one size above or below the tailor-predicted size.

Conclusions:

Anthropometry measurements, when used with decision tree models, show promise for classifying clothing size. Methodological changes such as fitting gender-specific models, using additional anthropometry variables, and testing other data mining techniques are avenues for future work. More research is required before fully automated body scanning is a viable option for obtaining fast and accurate clothing sizes for military clothing and logistics departments.

KEY WORDS

Anthropometry, 3-D body scanning, Decision tree, CART, clothing size.

1. Introduction

Correct fitting military uniforms and equipment are important for the survival and effectiveness of military personnel (Choi, 2016; Sparks, 2012). These uniforms must be suitable for specific occupational tasks; for example, loose fitting uniforms can get caught in machinery and cause accidents (Traumann et al., 2019), while tight fitting uniforms can severely restrict soldiers movement such as preparing to run from a prone position. The use of anthropometry in the design of apparel and equipment should contribute towards improved fit, better integration and compatibility among apparel items and equipment, and maximise mobility. Developing a clothing size range for a specific population can be challenging due to differences in body size and body proportions across different age, sex, and ethnic groups. Recent increases in the body size of military personnel (e.g. in Australia *(Tomkinson et al., 2009)*, Canada *(Keefe et al., 2015)*, and the U.S. (Knapik et al., 2018) appear to have coincided with changes in body proportions (Tomkinson et al., 2010; Tomkinson et al., 2017), which collectively create additional challenges for maintaining a current and effective uniform sizing scheme *(Keefe et al., 2017)*. An increasing number of women in Army combat roles also has implications for integrating the female body shape into clothing and equipment that was originally designed for male operators (Coltman et al., 2020; Keefe et al., 2015; Toma et al., 2016). Accurate and up-to-date anthropometric measurements are required to maintain optimal sizing systems.

Traditional anthropometric measurement techniques involve sizing the body using physical methods such as callipers, rulers, or measuring tapes (Simmons and Istook, 2003; Vinué, 2017). This approach has limitations including individual sizing being time-consuming (depending on the number of measures) and error-prone (e.g. measurement fatigue) (Vinué, 2017) and heavily dependent on the measurer's judgement (Liu et al., 2017). Furthermore, the precise location of body landmarks and how these are measured can be subjective (Apeagyei, 2010). These limitations can all result in inaccurate anthropometric dimensions that will subsequently lead to unfit garments (Liu et al., 2014).

More recently, the introduction of three-dimensional (3D) body scanner measurement systems have revolutionised the way anthropometric data are collected, assessed, and updated. This technology can provide highly detailed, accurate and reproducible anthropometric data (Istook and Hwang 2001; Lerch, MacGillivray, and Domino 2007; Wang, Wu, Lin, Yang, and Lu 2007; D'Apuzzo 2009) that can be stored and re-used in the future. The number of anthropometric variables that can be derived from a single body scan is almost limitless, and allows for the extraction of 1D (e.g. girths, lengths, breadths), 2D (e.g. cross-sectional areas) and 3D (e.g. surface areas and body volumes) measures (Daanen and Van de Water, 1998). There have been several large-scale military anthropometric surveys conducted worldwide (da Silva et al., 2017; Gordon et al., 2013; Hart et al., 1967; Keefe et al., 2015; Pringle et al., 2011; Tomkinson et al., 2012) that have amassed large amounts of 3D anthropometric data. With these high dimensional datasets, it can be difficult to determine the best combination of variables required for developing a sizing system or determining the optimal clothing size category for an individual. These difficulties can be associated with the volume of variables (e.g. too many variables would require greater processing power) or the variables themselves are not related or specific to clothing design (e.g. dactylion length).

To overcome this problem, various data mining techniques have been used to understand anthropometry variables for the purposes of classifying clothing size and developing sizing systems. Esfandarani and Shahrabi (2012) developed a suit sizing system based on principal component analysis (PCA) which identified height and chest circumference as the main components. Laing et al. (1999) used the k-means clustering to establish size charts for protective clothing used by New Zealand firefighters. Some studies used a combination of data mining techniques; for example, Hsu and Wang (2005b) used a two-step process combining both PCA and decision trees to establish a sizing system for Taiwanese soldiers' trousers. The PCA identified two factors (waist girth and outside leg length) that were utilised by a decision tree to classify clothing size. The resulting sizing system achieved 95% population coverage. Bagherzadeh et al. (2010) used a three-method approach (PCA, two-step cluster analysis, and decision trees) to determine a sizing system for the lower body of Iranian males. These methods show that it is possible to develop and predict clothing sizes from high dimensional anthropometric data as opposed to using more traditional (physical measurement) methods (Hsu, 2009).

Several studies have compared the accuracy of traditional versus 3D anthropometry measurements (Glock et al., 2017; Koepke et al., 2017; Kuehnapfel et al., 2016), but few studies have compared the efficiency and accuracy of tailor versus automated body scan-based clothing sizing in military combat uniforms. In 2019, the New Zealand Army (NZ Army) were interested in how 3D body scanning technology could optimise their recruit sizing process (Kolose et al., 2019). Sizing large numbers of recruits (up to 90 recruits a day) is time consuming, especially when conducted in parallel with other activities such as medical and dental checks. The use of automated measurements derived by 3D body scanning technology, has the potential to improve the efficiency of the NZ Army recruit sizing process. It has been successfully used in the Dutch military (Daanen et al., 2014) but has never been tested in the NZDF.

This study specifically examined the NZ Army in-service combat uniform worn at the time and known as the Multi-Camouflage Uniform (MCU) (Figure 1). The MCU was available in 12 shirt and 12 trouser sizes (each ranging from 2XS to 7XL). All references to size in this paper refer to this (2XS to 7XL) size range.

Figure 1. NZ Army MCU shirt and trouser (courtesy of New Zealand Defence Force).

The NZ Army recruit garment sizing process is conducted by trained tailors employed by an external equipment supplier. When sizing the MCU, the tailor estimates the initial shirt size by observing the recruit's chest circumference and estimates initial trouser size by observing waist circumference and height. From this assessment, different garment sizes are trialled until the right fit is obtained. No measurement tape is used in the sizing process. The average processing time per individual ranges from 10 to 15 minutes (Kolose et al., 2019).

Between 2016 and 2018 the New Zealand Defence Technology Agency (DTA) conducted the NZDF tri-service anthropometric survey (NZDFAS) of 1,003 uniformed personnel. During the survey, 84 different anthropometric measurements were assessed using one of three methods: physical measurements (using traditional anthropometry); 3D body scanner automatic measurements (generated automatically by the body scanner software, Anthroscan); or post-processed measurements (processed after the data collection using third-party software called CySlice). Post-processed measurements could not be measured 'practically' by physical means (e.g., crotch height) or automatically (not recognized as an automated measure in Anthroscan). These require user input to identify landmarks and utilise digital tools (e.g., tapes, rulers, coordinates).

The purpose of this study was to investigate how accurate decision trees are at predicting well-fitting MCUs and to improve the selection of MCU shirt and trouser sizes. More specifically, the aims of this study were to: (1) determine how well decision tree models predict tailor-assigned uniform sizes using anthropometry data from the NZDFAS; and (2) examine which combination of automatic, post-processed, and physical measurements are the most accurate predictors. It is hoped that this information can inform automatic sizing systems for military personnel.

2. Method

2.2 Design

Anthropometric data from two separate samples of the NZDF military were used in this study. The first sample were those who participated in the New Zealand Defence Force Anthropometry Survey (NZDFAS). The second (smaller) sample consisted of New Zealand Army recruits. The NZDFAS dataset was used to develop a series of uniform size prediction models, while the recruit dataset was used to test the accuracy of each model. Both data collection activities had ethical approval from the Auckland University of Technology Ethics

Committee (AUTEC #14/126 NZDF anthropometry Survey: Variations in kinanthropometry and implications for the New Zealand Defence Force).

2.3 Participants

NZDFAS

The NZDFAS dataset was collected between February 2016 and June 2018 and consisted of 1,003 uniformed participants from the Royal New Zealand Navy (n = 131), Royal New Zealand Air Force (n = 289) and the NZ Army (n = 583). The participants were recruited across the five major military establishments throughout New Zealand using a mixture of stratified and purposeful sampling. The participants were from various service backgrounds with the majority from engineering and technical (25%), combat (19%), and logistics and administration (11%). The ethnic composition of participants was mainly European (41%), New Zealand European (37%), and Māori/Pacific (20%). This study focussed on the NZ Army participants only (n = 583; 97 female and 485 male).

Recruits

The second sample consisted of 154 NZ Army recruits (19 females and 134 male) from the Waiouru Military Camp recruitment course (June 2019 intake). Testing occurred during their assigned medical examination and clothing sizing assessment week. These recruits were not involved in the 2016–2018 NZDFAS dataset. NZ Army recruits were chosen as the test population as the purpose of the study was to optimise the sizing process during recruitment intake.

In both samples (NZDFAS and recruits) the senior commanding officer for each regiment requested personnel to attend an initial brief with the research team where they were given information about the survey (e.g., purpose, scan procedure, rights to withdraw, informed consent). After the brief, the recruits were free to withdraw at any time without any consequences to their employment or recruitment course. This was supported by the senior leadership team.

2.4 Procedures

NZDFAS

In the NZDFAS sample, participants provided demographic information (gender, age, service) and self-reported ratings of their current uniform fit in terms of comfort ("Very good", "Good", "Acceptable", "Poor" and "Very poor", for both shirt and trousers). The tailor-assigned sizes for each participant's current uniform were obtained from the military's clothing store records.

Participants were then scanned inside a Vitus Smart XXL® body scanner (Human Solutions, Kaiserlauten, Germany) wearing tight-fitting underwear (e.g. briefs for males, and sports bra and underwear for females) and a swim cap (to standardise head-related measures such as standing height); whilst they were positioned in the 'standard' posture as shown in Figure 2.

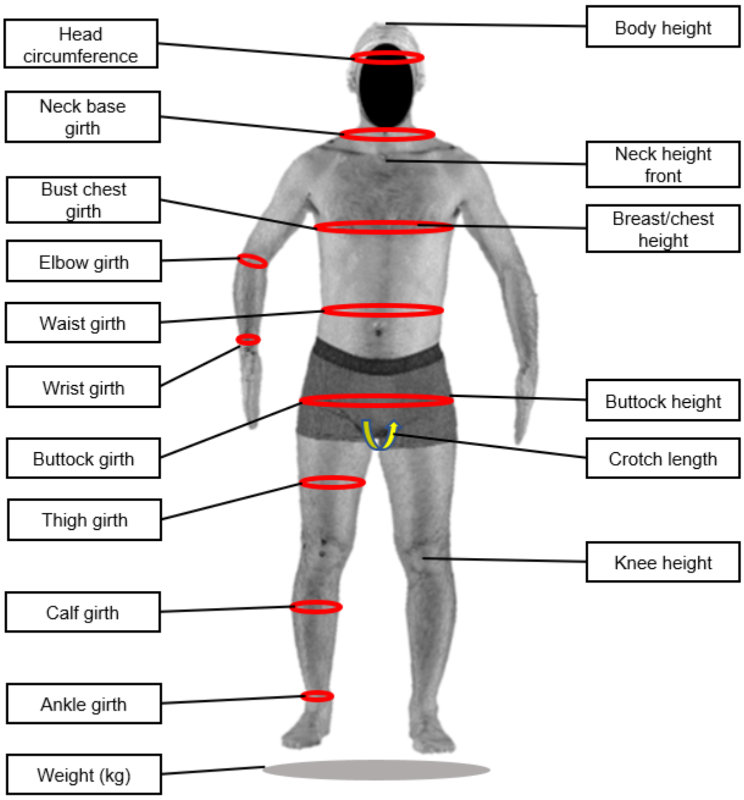

Figure 2. Body scan posture (standard posture) used in the NZDFAS. The labels identify the 17 automatic measurements that were captured.

The body scan data were processed in the *Anthroscan*© Version 6.0 software (Human Solutions, Kaiserlauten, Germany) to automatically detect 17 anthropometric measurements that were previously validated against physical measurements (Kolose et al., 2020). The Anthroscan automated measurements are derived from ISO 7250 and ISO 8559 (Human Solutions, 2015). These measurements (Figure 2) were extracted and checked for scan errors; for example, incorrect scan posture, the presence of lighting artefacts in the scan, and checking that the hair bun did not interfere with head circumference measurements. The NZDFAS dataset also consisted of 25 physical measurements that were taken on each participant using traditional methods, and 42 post-processed measurements that were derived from the 3D scan data using the third-party software *CySlice* (v3.4, Headus, Perth, Australia). The measurements were split across these three measurement methods (automatic, physical and post-processed) for the purpose of optimizing participant throughput while limiting individual processing times at each data collection activity (Kolose et al., 2020). The three methods are summarised in

Table **1** and the full measurement list is provided in Supplementary file 1.

Table 1. NZDFAS measurement methods description.

Method (n)	**Description**	**Equipment**	**Processing time per individual (min)**
Automatic (17)	Measurements are captured using an automatic landmark algorithm. Requires minimum operator effort except for final checking of measurement placement. The software extracts the measurements.	Vitus XXL 3-D body scanner and Anthroscan© software. Human Solutions Ltd.	<1 (fastest)
Physical (25)	Traditional anthropometric measurements conducted by accredited anthropometrists. High operator input (physical palpation of skin surface etc) is required. High accuracy dependant on the skill of the anthropometrist.	Traditional anthropometry equipment (callipers, scales, stadiometer).	30–50 (slowest)
Post-processed (42)	The operator extracts the measurements from each body scan using a suite of digital tools (e.g. tapes, rulers). Requires high operator input with specialist skills in digital manipulation.	CySlice by Headus Ltd.	20

The number in parentheses refer to the number of measurements in each method.

Recruits

The body scanning procedure for the NZ Army recruits was identical to the NZDFAS. No physical or post-processed measurements were derived in the recruit’s dataset as this sample was collected for the purpose of validating prediction models that were fit using automatic measurements only (see analysis section). As the recruit sample had not yet been issued with their MCU uniform, they did not complete questions related to clothing fit. However, they were each measured by the clothing tailor and assigned an MCU trouser and shirt size. Like the NZDFAS sample, these sizes were obtained from the military's clothing store records and were treated as the optimal clothing size for each individual.

2.5 Analyses

In the first instance, key demographic and body size characteristics were compared between the NZDFAS and recruit samples using Fisher's exact test or the Mann-Whitney U test. The primary analysis focused on developing decision tree models for classifying tailor-assigned clothing sizes (MCU shirt and trouser) from the NZDFAS anthropometry variables. Several models were fit to test five different combinations of measurements (automatic, physical, post-processed, automatic + post-processed, and automatic + physical + post-processed). Prior to model fitting, shirt and trouser sizes with a low frequency (e.g., 3XL and 4XL) were combined into the next largest category (e.g., 2XL). Any individuals who rated the fit of the current uniform as 'Poor', 'Very poor', or provided no rating (n = 135) were removed as the intention was to predict an optimal-fitting uniform size. This meant the final analysis sample contained 437 individuals. Variables that were not conceptually related to shirt (e.g., ankle, calf, or crotch girth) or trouser (e.g., head circumference, radiale-stylion length) pattern sizing were removed to minimise model complexity. The final input variables for shirt were body height, breast height, bust chest girth horizontal, elbow girth, head circumference, neck at base girth, neck height front (supersternale), waist girth, weight, and wrist girth. The input variables for trousers were ankle girth, body height, bust chest girth horizontal, buttock girth, buttock height, calf girth right, crotch length, knee height, thigh girth right horizontal, waist girth, and weight. All decision trees were fit using the CART (classification and regression trees) algorithm in IBM SPSS Statistics© (v24, IBM Corp., Armonk, NY, USA) and validated using 10-fold cross-validation. CART was chosen because the output (a decision tree) is straightforward to interpret, the first nodes (or levels) of the tree are the most important variables, and there is no need for implicit assumptions (e.g., any preconceived knowledge of how different body measurements are related). The CART algorithm determined the optimal cut-off point for each explanatory variable. Several candidate hyperparameter values were trialled, and those that maximised model accuracy were selected as the optimal hyperparameters for the final model. These were set to: maximum tree depth = 4, minimum samples in the parent node = 20, minimum samples in the child node = 7, and maximum surrogates = 5.

Variables from the measurement combination with the greatest accuracy were used to develop the final shirt and trouser size models. These models were then applied to the recruit's dataset to test the models' generalisability in an independent sample. The statistical classification metrics (Lavrač, 1999) computed from the correct and incorrect predictions were: (1) overall accuracy (the proportion of all cases correctly identified); and (2) Cohen's Kappa score which measures the degree of agreement between the true values and the predicted values (clothing size), which is an indicator of each model's performance (interpreted as ≤0 = no agreement, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement); and for each size category (McHugh, 2012) (3) sensitivity (the proportion of positive or true cases correctly identified, e.g. proportion of XS sizes predicted as XS); (4) specificity (the proportion of negative or false cases correctly identified, e.g. proportion of non-XS sizes predicted as non-XS); and (5) balanced accuracy (the mean of sensitivity and specificity).

These metrics are presented in addition to overall accuracy to account for the unequal number of individuals across size categories (McHugh, 2012; Narayanan et al., 2020). To inspect where the model was correct and incorrect, a comparison of the model-predicted sizes and the tailor-assigned sizes were presented as a confusion matrix table. A summary of the methods and analyses are provided in Figure 3.

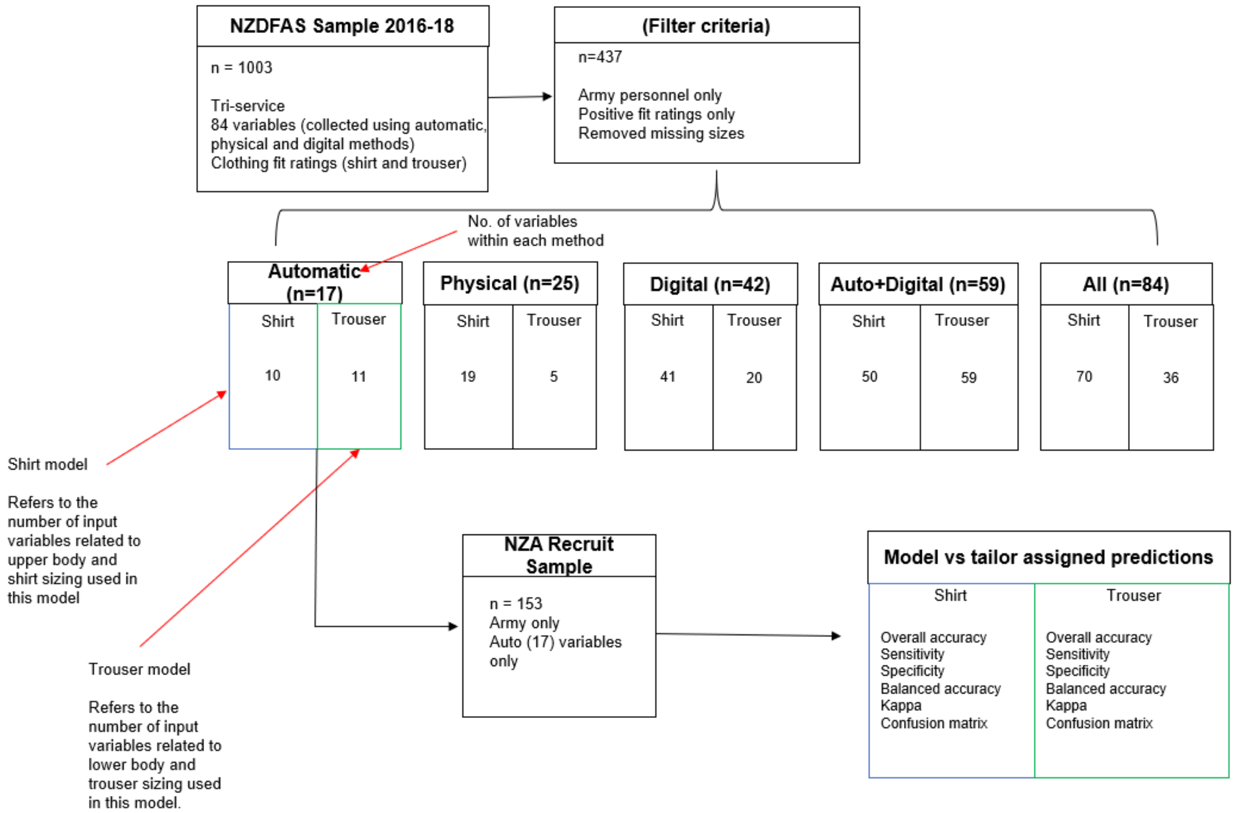

Figure 3. Overview of methods and analyses. Decision trees models were developed from the NZDFAS 2016-18 dataset and tested on the NZ Recruit datasets.

3. Results

The demographic characteristics of both samples (NZDFAS and NZ Army recruits) are presented in

Table **2**. The mean and standard deviation for age within these samples was 32 ± 10.4 and 20 ±2.1 years, respectively. The most common tailor-assigned shirt and trouser sizes were medium (39% and 44%) and small (37% and 40%) for each sample, respectively. Two NZDFAS participants possessed shirt sizes greater than 2XL, and three had trouser sizes greater than 2XL. No recruits had shirt and trouser sizes greater than 2XL. Key clothing-related anthropometry measures are also presented in Table 2. Although body height was similar for the samples, there were significant differences between the two samples for chest girth, buttock girth, crotch length, neck girth, waist girth, and body weight. The results of the fit satisfaction survey showed that a higher proportion of those experiencing a 'Poor' or 'Very poor' fit were female (31% vs. 7% of males).

Table 2. Descriptive characteristics of the NZDFAS and NZ Army recruit samples.

Variable	**Level**	**NZDF (n = 437)**	**Recruits (n = 153)**	**Mean diff%**	**P-value[a]**
Gender	Male (n = 381)	297 (87%)	135 (88%)		0.96
	Female (n = 56)	39 (13%)	18 (12%)		
Shirt Size (n = 408)	XS	42 (10.3%)	33 (21.6%)		**<0.001**
	S	50 (14.9%)	56 (36.6%)		
	M	132 (39.3%)	49 (32%)		
	L	85 (25.3%)	10 (6.5%)		
	XL	31 (7.6%)	1 (0.7%)		
	2XL	20 (4.9%)	0		
Trouser Size (n = 337)	XS	17 (5.1%)	21 (13.7%)		**<0.001**
	S	64 (19%)	61 (39.9%)		
	M	146 (43.5%)	51 (33.3%)		
	L	85 (25.2%)	18 (11.8%)		
	XL	24 (7.1%)	2 (1.3%)		
Measures	Body height (mm)	1772 ± 74	1758 ± 72	13.4	0.08
	Chest girth (mm)	1041 ± 80	993 ± 67	47.9	**<0.001**
	Waist girth (mm)	880 ± 99	822 ± 64	58.0	**<0.001**
	Body weight (kg)	84 ± 12.5	76 ± 9.6	8.6	**<0.001**

Data are presented as mean ±SD or n (%) where appropriate; [a]P value of difference between NZDFAS and recruits (Fisher's exact test or Mann-Whitney U test where appropriate). Bold items represent statistically significant results (p < 0.05).

Table 3 shows the results of the CART analysis for the five different measurement combinations in the NZDFAS sample. For shirt size, classification was most accurate for automatic measurements followed by all categories (automatic + physical + post-processed), (58.1% and 55.1% cross-validation accuracy, respectively). For trousers, the automatic measurements were the most accurate (61.7%) followed by automatic + post-processing (57.3%). In general, models that used measurements obtained using the body scanner (e.g., automatic, and post-processed) demonstrated higher accuracy than models using only the physical measurements (Table 3).

Table 3: Classification results for shirt and trouser size for measurements obtained from cross-validation.

	Automatic	**Physical**	**Post-processed**	**Automatic + Post-processed**	**Automatic + Physical + Post-processed**
Shirt	58.1	43.4	54.9	57.4	55.1
Trouser	61.7	45.7	51.6	57.3	56.1

Data presented as overall accuracy (%) from 10-fold cross validation.

The shirt and trouser models that used the set of automatic measurements were deemed the final models, given their overall accuracy and simplicity. Figure 4 shows the final decision tree model for predicting shirt size, while Figure 5 shows the final decision tree model for predicting trouser size. The variable at each node and the split criteria are presented. The number and percent within each node represent the frequency of participants within each size category that pass through the node.

For shirt size classification, the root node (Node 0) was split according to body weight. The subsequent variables deemed important by the model were waist girth, neck height front, and body height. The final size classification in these nodes is based on the size with the highest frequency of cases. For trouser size, the root node was also split according to body weight. The subsequent variables deemed important by the model were waist girth, buttock girth, buttock height, thigh girth, and crotch length.

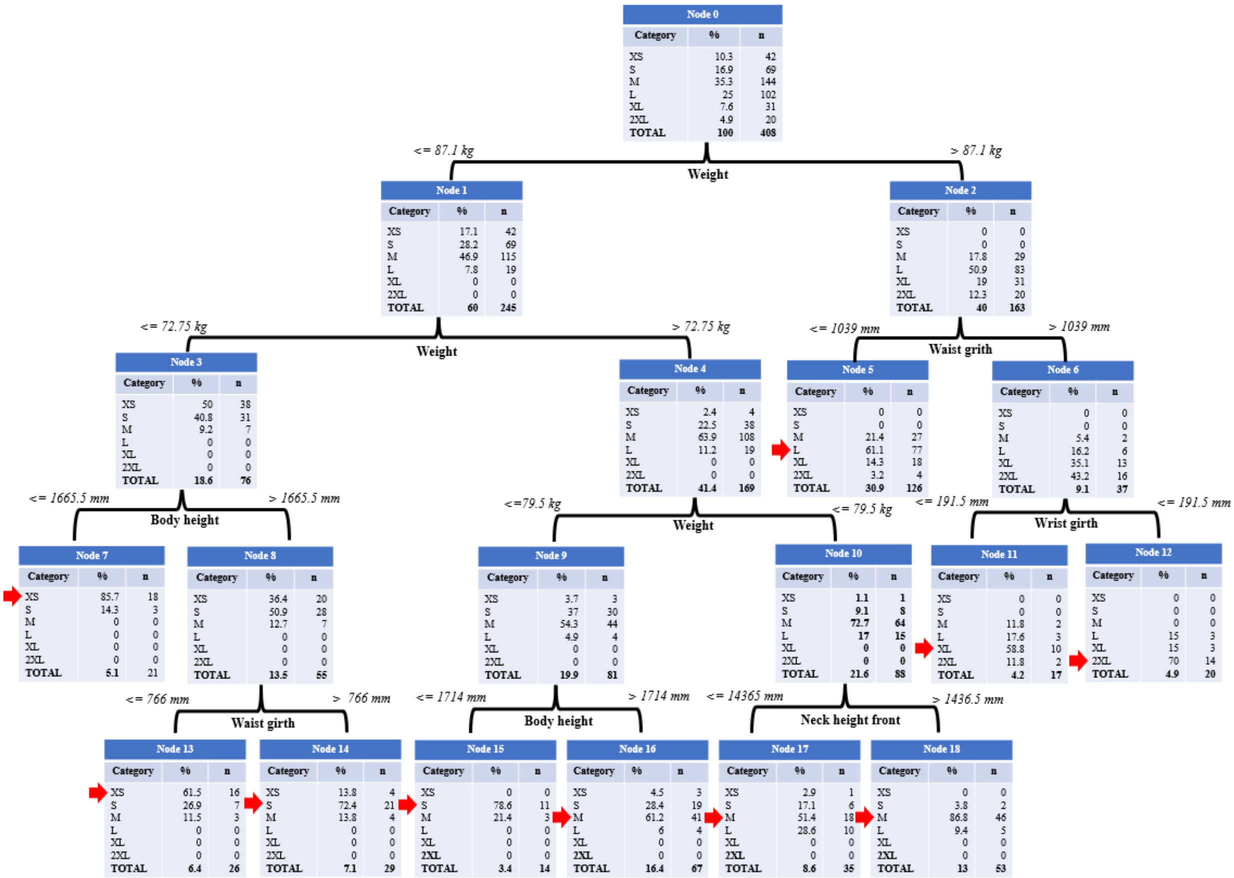

Figure 4. Shirt size classification for the NZ Army.

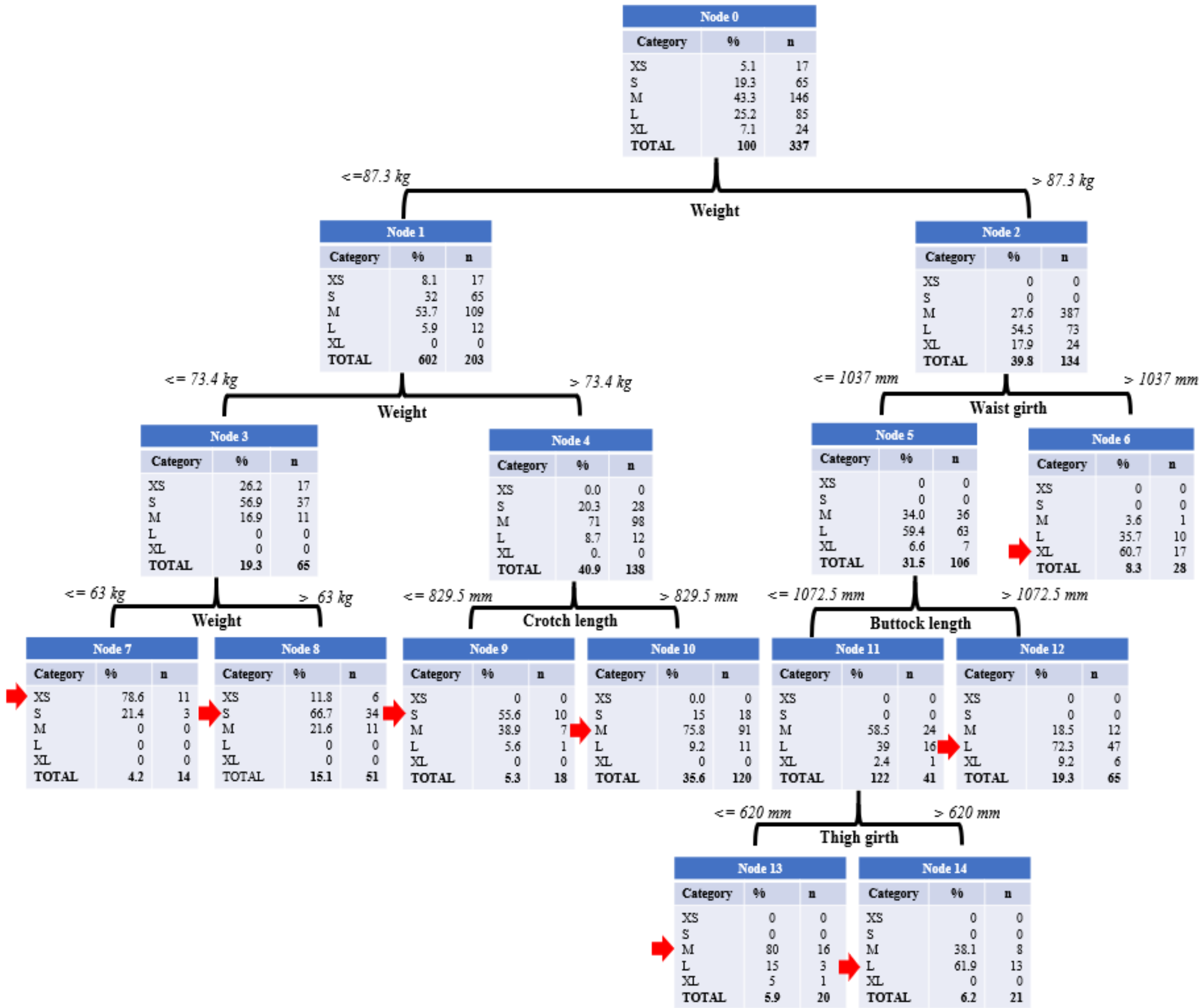

Figure 5. Trouser size classification for the NZ Army.

The models presented in Figure 4 and Figure 5 were tested in the recruit dataset. A greater overall accuracy and Kappa coefficient were observed for trouser size classification (*OA = 62.8%; 95% CI [55, 70], κ = 0.45*) compared to shirt size classification (*OA = 54.9%; 95% CI [47, 63], κ = 0.38*). The shirt model can be interpreted as having 'slight' agreement, while the trouser model achieved 'moderate' agreement.   shows the shirt and trouser model performance for each clothing size. The shirt model was most accurate at classifying sizes L (balanced accuracy = 86.2%), XS (80.5%) and M (69.2%). For trouser sizes, the model performed best for XS (80.2%), and XL (75%).

Table 4: Classification results for shirt and trouser in the recruit dataset.

	Shirt						Trouser				
Metric	**2XS**	**XS**	**S**	**M**	**L**	**XL**	**XS**	**S**	**M**	**L**	**XL**
Sensitivity	0.0	72.7	35.7	65.3	80.0	0.0	61.9	65.6	66.7	44.4	50.0
Specificity	100	88.3	84.5	73.1	92.3	99.3	98.5	75.0	73.5	96.3	100
Balanced Accuracy	50.0	80.5	60.1	69.2	86.2	49.7	80.2	70.3	70.1	70.4	75.0

Data are presented as percent (%).

Table 5 shows the confusion matrix comparing the model predictions to the tailor-assigned sizes. The individuals that were predicted by the model were generally predicted as one size above (e.g., medium instead of small for shirt) or one size below the correct size.

Table 4: Confusion matrix for shirt and trouser predictions in the recruit dataset.

		Tailor assigned size										
		Shirt						**Trouser**				
		2XS	**XS**	**S**	**M**	**L**	**XL**	**XS**	**S**	**M**	**L**	**XL**
Model predicted size	**2XS**	**0**	0	0	0	0	0	-	-	-	-	-
	XS	4	**24**	10	0	0	0	**13**	2	0	0	0
	S	0	8	**20**	7	0	0	8	**40**	13	2	0
	M	0	1	26	**32**	1	0	0	19	**34**	8	0
	L	0	0	0	10	**8**	1	0	0	4	**8**	1
	XL	0	0	0	0	1	**0**	0	0	0	0	**1**

Bold numbers along the diagonals indicate where tailor-assigned sizes (x-axis) and model prediction (y-axis) and were the same. For example, size XS, 24 (bold value) out of a possible 33 participants (sum of all values within the tailor assigned XS column) were predicted as size XS. However, eight of those participants were predicted as size S (and one size M).

4. Discussion

The aim of this study was to determine how well decision tree models predict tailor-assigned uniform sizes using anthropometry data from a military anthropometry survey, and to examine which combination of automatic, post-processed, and physical anthropometry measurements lead to the most accurate model. The classification of shirt and trouser size was more accurate when based on the NZDFAS automatic measurements (body scanner derived) as opposed to the set of physical (traditional) measurements, or a combination of measurement types. When the decision tree models were tested on the recruit dataset, the classification accuracy was highest for trousers (particularly for the large size) compared to shirts. Body weight was the most important variable in both the shirt and trouser decision trees.

4.2 Common variables with previous research

Our findings identified body weight and waist girth as the most important measurements for both shirt and trouser size classification. Body weight has traditionally been an important measurement for developing size charts (Emanuel et al., 1959) and it is also highly correlated with many anthropometric measurements (Behnke, 1961). Research using linear regression to estimate sizing suggest that garment dimensions in the height direction can be calculated purely based on body height (Liu et al., 2017). Previous studies also identified bust girth and waist girth as important variables for shirt sizing, and hip girth, height, and leg length as key measurements for trouser sizing (Bagherzadeh et al., 2010; Gupta and Gangadhar, 2004; Hsu and Wang, 2005a).

4.3 Automatic versus physical

Using the set of automated measurements collected as part of NZDFAS led to a more accurate prediction model than using the physical measurements or post-processed measurements. This could have been attributed to the type of variables used in the analysis and their relevance to clothing size. For example, the automatic measurement variables were more closely related to traditional clothing pattern dimensions (e.g., bust chest girth, waist girth, and height) compared to the physical measurement variables (e.g., arm span, knee height, tibiale laterale height). Physical and post-processed variables had minimal effect on increasing the accuracy of clothing size classification. For example, the shirt model fit using Automatic + Physical + Post-processed measurements had less accuracy than the model fit using only Automatic measurements (55.1% vs 58.1%). This trend was similar for trouser size prediction (56.1% vs 61.7%) which showed a 5.6% difference.

4.4 Classification accuracy

Although the overall classification accuracy for shirt and trousers was modest (between 58–62%), the confusion matrix demonstrated that the incorrect model predictions were generally one size above or below the correct size. This suggests that participants who are difficult to classify may fall on the border between two different size categories. Thus, prediction errors seem to be systematic (the correct size being one size up or down). This is in line with previous literature which has suggested that perfect accuracy when predicting a person's garment size from his or her body size is rarely achieved (Bradtmiller, 2015). The incorrect sizes that are one size above or below the assigned size, may present safety implications (e.g. one size up may result in 'loose' fit or 'snags' during an emergency egress from a Pinzgauer vehicle; one size down may restrict mobility that is required for running quickly from a stationary prone firing position). Similarly in the US, male soldiers reported that wearing body armour that was one size too big was associated with increased exposure at the neck and underarm regions when compared to the soldiers wearing body armour that was one size too small or the correct fit (Choi et al., 2018; Coltman et al., 2020). A small change in clothing size can have a detrimental effect on soldier safety.

The modest classification score may be attributed to several factors. First, it is possible that tailor-assigned sizes may have not been the optimal size for the participant's anthropometry. This may have arisen because the clothing tailors prioritise functional fit (e.g. slightly 'loose' fit to increase all round mobility) as opposed to a tailored fit (Kolose et al., 2019). There is also a small possibility that the clothing sizes obtained from the clothing store may have been out of date due to changes in anthropometry as a result of military training (i.e., participant sizes may have been assigned months prior to the NZDFAS data collection). Although excluding participants who rated their fit as 'Poor' or 'Very poor' may have alleviated some of these problems, these fit ratings were still subjective. Some participants may have preferred tight- or loose-fitting clothing, regardless of the functional fit. These issues may have caused a mismatch between a participant's anthropometry and the size they were allocated. This can be evidenced in Figure 4, where several participants with a tailor-assigned medium-sized shirt had a waist girth >1039 mm, yet other participants who were also assigned a medium-sized shirt had a waist girth <766 mm. The anthropometric measurements used in this study may not have been the most optimal variables for predicting clothing size. The NZDFAS dataset utilised 17 automatic measurements that were extracted using the Anthroscan software. It is possible that other automatic measurements may have increased the accuracy of the predictions, but the validity of these automatic measurements has not been tested.

4.5 Limitations and future implications

Future research in this area could focus on gender differences, particularly since female participants expressed greater dissatisfaction with uniform fit compared to male participants. This may be indicative of poor uniform design or sizing categories that do not align with the female form (given the MCU is unisex). As international military forces are looking to increase the number of females in the military, it may be more appropriate to have a gender-specific uniform and sizing system. A previous study (Fullenkamp et al., 2008) identified key anthropometric measurements that differentiate male and female personnel in the military (e.g. hip and shoulder breadth). This information may be useful for designing gender-specific uniforms. Cluster analysis has been applied to anthropometric datasets to develop distinct groups of military personnel based on their anthropometric characteristics (Chung et al., 2007; Hsu, 2009; Loker et al., 2005) which in turn, have been used to develop clothing sizing systems. If combined with universal military clothing standards (e.g., NATO sizing standard), it may be possible to develop gender-specific size categories and garment dimensions specific to a military population.

Although this study comprised participants across multiple trades, well-fitting uniforms will likely be more beneficial (from a safety perspective) to those on the 'front line' compared to sedentary (office work) roles. Various data mining techniques have been used to understand anthropometry variables for the purposes of classifying clothing size and developing sizing systems. Significant differences between each model could not be assessed as it is currently not possible to compare two or more decision tree models using the SPSS software**.** Furthermore, it is possible that surface area or volume measures would have improved model accuracy, but the NZDFAS dataset only contained data on lengths, breadths, depths, and circumferences.

5. Conclusions

Anthropometric measurements, when used with decision tree methods, show promise for supplying a correct fitting clothing size, with body weight and waist girth being the strongest predictors. Methodological changes such as fitting gender-specific models, using additional anthropometry variables, and testing other data mining techniques are avenues for future work. More research is required before fully automated body scanning is a viable option for obtaining fast and accurate clothing sizes for military clothing and logistics departments.

6. Authors' contributions

SK was responsible for study design and recruitment, and managed data collection with assistance from AUT and NZDF personnel. SK and PH were responsible for the ethics approval for the NZDF anthropometry survey. Data cleaning and processing were performed by SK, TS, PH and GT. SK performed the analysis (with assistance from TS) and drafted the manuscript. All authors contributed to the interpretation of results, editing and critical reviewing of the final manuscript, approved the final manuscript as submitted, and agreed to be accountable for all aspects of the work. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

7. Acknowledgements

The authors would like to thank all participants for their involvement in this study, members of the NZDF (NZDF Clothing and Logistics, The Army Depot), The Workwear Group (Waiouru clothing store) and the AUT kinanthropometry team for their assistance during data collection. Data collection for this study was funded by the NZDF. Participants were recruited as part of a Defence Technology Agency, Anthropometry project endorsed by the New Zealand Vice Chief of Defence and individual service chiefs.

8. Competing interests

The authors declare that they have no competing interests.

9. References

Apeagyei, P., 2010. Application of 3D Body Scanning Technology to Human Measurement for Clothing Fit. *International Journal of Digital Content Technolog.* 4, 58-68.

Bagherzadeh, R., Latifi, M., Faramarzi, A., 2010. Employing a Three-Stage Data Mining Procedure to Develop Sizing System. *World Applied Sciences Journal.* 8, 923-929.

Behnke, A.R., 1961. Anthropometric Fractionation of Body Weight. *Journal of Applied Physiology.* 16, 949-954.

Bradtmiller, B., 2015. Predicting Product Accommodation: The Role of the Anthropometric Fit Test. *Procedia Manufacturing.* 3, 4464-4471.

Choi, H.J., Garlie, T.N., Mitchell, K.B., 2018. Effects of Body Armor Fit on Encumbered Anthropometry Relative to Bulk and Coverage, International Conference on Applied Human Factors and Ergonomics. Springer, pp. 260-272.

Choi, T.M., 2016. Information Systems for the Fashion and Apparel Industry, in: Choi, T.-M. (Ed.). Woodhead Publishing, pp. xix-xx.

Chung, M.J., Lin, H.F., Wang, M.J., 2007. The Development of Sizing Systems for Taiwanese Elementary and High-School Students. *International Journal of Industrial Ergonomics.* 37, 707-716.

Coltman, C.E., Steele, J.R., Spratford, W.A., Molloy, R.H., 2020. Are Female Soldiers Satisfied With the Fit and Function of Body Armour? *Applied Ergonomics.* 89, 103197.

da Silva, G.V., Halpern, M., Gordon, C.C., 2017. Anthropometry of Brazilian Air Force pilots. *Ergonomics.* 60, 1445-1457.

Daanen, H.A., Van de Water, G.J., 1998. Whole Body Scanners. *Displays* 19, 111-120.

Daanen, H.A., Woering, A., Haar, F.B.T., Kuijpers, A.A.M., Haker, J.F., Reulink, H.G.B., 2014. Optimization of Military Garment Fit. *Ambience* 14, 7-9.

Emanuel, I., Alexander, M., Churchill, E., Truett, B., 1959. *A Height-Weight Sizing System for Flight Clothing.* Aero Med Lab, Ohio, USA.

Esfandarani, M.S., Shahrabi, J., 2012. Developing a New Suit Sizing System Using Data Optimization Techniques. *International Journal of Clothing Science and Technology.*

Fullenkamp, A.M., Robinette, K.M., Daanen, H.A.M., 2008. Gender Differences in NATO Anthropometry and the Implication for Protective Equipment. *Air Force Research Laboratory Human Effectiveness Directorate Biosciences and Protection Division.*, Wright-Patterson Air Force Base, Ohio, U.S.

Glock, F., Vogel, M., Naumann, S., Kuehnapfel, A., Scholz, M., Hiemisch, A., Kirsten, T., Rieger, K., Koerner, A., Loeffler, M., Kiess, W., 2017. Validity and Intraobserver Reliability of Three-Dimensional Scanning Compared with Conventional Anthropometry for Children and Adolescents from a Population-Based Cohort Study. *Pediatric Research* 81, 736-744.

Gordon, C.C., Blackwell, C.L., Bradtmiller, B., Parham, J.L., Hotzman, J., Paquette, S., Corner, B.D., Hodge, B.M., 2013. *2010 Anthropometric Survey of U.S. Marine Corps Personnel: Methods and Summary Statistics.* Anthrotech, Natick Soldier Research, Development of Engineering Center, Yellow Springs, Ohio, U.S.

Gupta, D., Gangadhar, B.R., 2004. A Statistical Model for Developing Body Size Charts for Garments. *International Journal of Clothing Science and Technology.* 16, 458-469.

Hart, G.L., Rowland, G.E., Malina, R., 1967. Anthropometric Survey of the Armed Forces of the Republic of Korea., New Jersey.

Hsu, C.H., 2009. Data Mining to Improve Industrial Standards and Enhance Production and Marketing: An Empirical Study in the Apparel Industry. *Expert Systems with Applications.* 36, 4185-4191.

Hsu, C.H., Wang, M.J.J., 2005a. Using Decision Tree-Based Data Mining to Establish A Sizing System for the Manufacture of Garments. *The International Journal of Advanced Manufacturing Technology* 26, 669-674.

Hsu, C.H., Wang, M.J.J., 2005b. Using decision tree-based data mining to establish a sizing system for the manufacture of garments., pp. 669-674.

Human Solutions, 2015. *Anthroscan User Guide Version 3.* Human Solutions GmbH, Kaiserslautern, Germany.

Keefe, A., Angel, H., Mangan, B., 2015. *2012 Canadian Forces Anthropometric Survey (CFAS).* Defense Research and Development Canada, Toronto, Canada.

Keefe, A., Kuang, J., Daanen, H., 2017. NATO Research Task Group: 3D Scanning for Clothing Fit and Logistics. *Proceedings of 3DBODY.TECH 2017. 8th International Conference and Exhibition on 3D Body Scanning and Processing Technologies.*, Montreal, Canada,, pp. 201-209.

Knapik, J.J., Redmond, J.E., Grier, T.L., Sharp, M.A., 2018. Secular Trends in the Physical Fitness of United States Army Infantry Units and Infantry Soldiers, 1976–2015. *Military Medicine.* 183, e414-e426.

Koepke, N., Zwahlen, M., Wells, J.C., Bender, N., Henneberg, M., Rühli, F.J., Staub, K., 2017. Comparison of 3D Laser-Based Photonic Scans and Manual Anthropometric Measurements of Body Size and Shape in a Validation Study of 123 Young Swiss Men. *PeerJ.*

Kolose, S., Hume, P.A., Stewart, T., Tomkinson, G.R., Stewart, A.D., Legg, S.J., 2020. Physique of New Zealand Defence Force personnel: Proformas (protocol and summary statistics) for the surface anthropometry and three-dimensional scanning survey. SPRINZ, Auckland University of Technology, Auckland, New Zealand.

Kolose, S., Stewart, T., Hume, P.A., 2019. NZDF Uniform Size Prediction Model and Application Development: Technical Report to NZDF. Auckland University of Technology., Auckland, New Zealand.

Kuehnapfel, A., Ahnert, P., Loeffler, M., Broda, A., Scholz, M., 2016. Reliability of 3D Laser-Based Anthropometry and Comparison with Classical Anthropometry. *Scientific Reports.* 6, 1-11.

Laing, R.M., Holland, E.J., Wilson, C.A., Niven, B.E., 1999. Development of Sizing Systems for Protective Clothing for the Adult Male. *Ergonomics.* 42, 1249-1257.

Lavrač, N., 1999. Selected Techniques for Data Mining In Medicine. Artificial Intelligence in Medicine 16, 3-23.

Liu, K., Wang, J., Kamalha, E., Li, V., Zeng, X., 2017. Construction of a Prediction Model for Body Dimensions Used in Garment Pattern Making Based on Anthropometric Data Learning. *The Journal of The Textile Institute.* 108, 2107-2114.

Liu, Z., Li, J., Chen, G., Lu, G., 2014. Predicting Detailed Body Sizes By Feature Parameters. International Journal of Clothing Science and Technology 26.

Loker, S., Ashdown, S.P., Schoenfelder, K., 2005. Size-Specific Analysis of Body Scan Data to Improve Apparel Fit. *Journal of Textile and Apparel, Technology and Management.* 4, 1-15.

McHugh, M.L., 2012. Interrater Reliability: The Kappa Statistic. Biochem Med (Zagreb) 22, 276-282.

Narayanan, A., Desai, F., Stewart, T., Duncan, S., Mackay, L., 2020. Application of Raw Accelerometer Data and Machine-Learning Techniques to Characterize Human Movement Behavior: A Systematic Scoping Review. 17, 360.

Pringle, R.H., Puxley, A.J., Puxley, K.P.M., Turner, G.M., Tyrell, A.K., 2011. Anthropometry Survey of UK Military Personnel 2006-7 (Issue 3). *QinetiQ Ltd*, Swindon, UK.

Simmons, K.P., Istook, C.L., 2003. Body Measurement Techniques. *Journal of Fashion Marketing and Management: An International Journal* 7, 306-332.

Sparks, E., 2012. Advances in Military Textiles and Personal Equipment. Woodhead Publishing.

Toma, D., Niculescu, C., Sălişteàn, A., Luca, D., Popescu, G., Popescu, A., Lăzăroaie, C., Său, C., Istrate, M., 2016. Improved Fit and Performance of Female Bulletproof Vests., *International Conference on Advanced Materials and Systems (ICAMS).* The National Research & Development Institute for Textiles and Leather-INCDTP, pp. 417-422.

Tomkinson, G., Clark, A., Blanchonette, P., 2009. Body Size Changes of Royal Australian Air Force Aircrew: 1971 - 2005., in: *Division., D.S.a.T.O.A.O.* (Ed.), Victoria, Australia, p. 28.

Tomkinson, G.R., Clark, A.J., Blanchonette, P., 2010. Secular Changes in Body Dimensions of Royal Australian Air Force Aircrew (1971-2005). *Ergonomic.s* 53, 994-1005.

Tomkinson, G.R., Daniell, N., Dale, M., Bowler, T., 2012. Australian Warfighter Anthropometry Survey (AWAS). Methods, Temporal Changes and Summary Statistics., in: Health and Use of Time Group, S.I.f.H.R. (Ed.). *University of South Australia*, Adelaide, Australia.

Tomkinson, G.R., Daniell, N., Fulton, A., Furnell, A., 2017. Time Changes in the Body Dimensions of Male Australian Army Personnel Between 1977 and 2012. *Applied Ergonomics* 58, 18-24.

Traumann, A., Peets, T., Dabolina, I., Lapkovska, E., 2019. Analysis of 3-D Body Measurements to Determine Trousers Sizes of Military Combat Clothing. *Textile Leather Review* 2, 6-14.

Vinué, G., 2017. Anthropometry: An R Package for Analysis of Anthropometric Data. *Journal of Statistical Software* 77, 39.