Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 289–300 DOI: 10.15439/2024F6463 ISSN 2300-5963 ACSIS, Vol. 39

Quality Control of Body Measurement Data Using Linear Regression Methods

Janis Bicevskis
0000-0001-5298-9859
University of Latvia
Raina boulevard 19,
Riga, LV-1586, Latvia
Email:
Janis.Bicevskis@lu.lv

Zane Bicevska
0000-0002-5252-7336
DIVI Grupa Ltd
Fridriha Candera str. 1,
Riga, LV-1046, Latvia
Email:
Zane.Bicevska@di.lv

Edgars Diebelis
0000-0002-5950-9915
DIVI Grupa Ltd
Fridriha Candera str. 1,
Riga, LV-1046, Latvia
Email:
Edgars.Diebelis@di.lv

Liva Purina
0009-0008-0339-5956
DIVI Grupa Ltd
Fridriha Candera str. 1,
Riga, LV-1046, Latvia
Email:
liva.purina@patterns.di.lv

***Abstract*—Body measurement data are inherently inaccurate and quite error-prone due to manual measurement and data collection. In this study, professionally collected and self-collected body measurement data were used to investigate to what extent potentially erroneous data can be identified during collection by utilizing the anthropologically given correlation of body measurements. The study specifically uses a dataset created within the framework of a project for made-to-measure pattern creation, consisting of data from 2053 female individuals with up to 52 recorded body measurements. Using linear regression, a method for validating the collected data is defined, wherein potentially inconsistent data are identified based on tolerance intervals. The tolerance intervals calculated within the study are specific to the particular application and the personal data used in the study. The outlined method is applicable to almost any set of manually collected body data in at least the triple-digit range, enabling the identification of probable data errors already during their collection.**

***Index Terms*—linear regression, body measurement assessment, data quality in pattern generation.**

I. Introduction

With the advancement of information technology capabilities, there are more and more specialized industry solutions that create, process, and use various types of graphic objects, such as photographs, drawn images, drawings, sketches, patterns, and so on. One type of these objects is technical drawings composed of polygons. The most typical examples are part drawings, geographic maps, and clothing patterns.

If polygon-based objects are created automatically (generated with the help of specialized programs and scripts), two types of problems arise: (A) How to ensure that the created object meets the established requirements (is correct)? (B) How to be confident that, even in the case of script changes, the resulting output is as correct as in the previous version (continuity is maintained)?

This work has been conducted within the research project "Competence Centre of Information and Communication Technologies" of The Recovery and Resilience Facility, contract No. 5.1.1.2.i.0/1/22/A/CFLA/008 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.7 “Technology for quality assessment of graphical objects described using polygons"

A. Problem Identification

Translating the problem statement to the specific application intended for testing the study results — automated generation of clothing patterns — the described problem statement is reduced to a series of derivative questions: (a1) How to ensure that the generated pattern fits the clothing intended for a specific person?, (a2) How to ensure that the components of the generated pattern are compatible with each other, i.e., that the respective garment can be sewn together properly?, (b1) How to detect systematic deviations from the standard if the pattern generation script has been modified?, (b2) How to ensure that the pattern generation script is functional in all its branches (completes the work as desired and produces the necessary result)?

Before the digitalization era, problems of type (a) were solved by sewing and trying on clothing prototypes for people of specific sizes. As long as patterns are generated for fixed measurement sets (standard sizes), sewing prototypes, though resource-intensive, would still be technically feasible. However, in cases where individual patterns are created for a wide variety of measurement sets, this is no longer feasible even theoretically — the number of potential measurement combinations is virtually infinite.

One solution could be finding equivalence classes where the included sets of measurements behave similarly for a specific type of clothing. Another solution could involve elements of image recognition and automated comparison to "spot" extreme and boundary cases. In this study, the main focus is on assessing the consistency of input data, namely, the interrelations of body measurements. If a mechanism has been developed that reliably identifies atypical deviations (errors) in the data, in the next steps, these data could be used for machine learning and evaluating the accuracy of image recognition.

Problems of type (b) primarily relate to regression testing in the field of software engineering—specific quality criteria need to be identified that, on one hand, correspond to the task at hand and, on the other hand, are automatable and verifiable. The need for automated testing also arises from the very large number of combinations to be tested—for instance, to test the correctness of 160 scripts on 150 body profiles, at least 25,000 test cases would be required, which is practically impossible to execute manually.

Regression testing is a well-researched topic in scientific literature. However, specifically regarding quasi-continuous objects (such as graphic objects composed of polygons), there is a need for a critical evaluation of existing concepts. This aspect will not be considered in this study; however, it should be noted that one of the solutions for regression testing is given in [1]. The essence of the solution is as follows. Unlike traditional systems regression testing where test cases are constructed according to system specifications, in this case it is proposed to accumulate use cases from previous system applications. Each use case accepted by a customer, or an expert may become a regression test case, it ensures stability of the system as all previous use cases must work correctly after changes in the system. This will not only save the resources needed for preparation of test cases. It also ensures that the system is maintained in good quality as all previous use cases are run repeatedly. Customer-accepted test cases differ from the, often unrealistic, test cases created by testers according to system requirements specification.

B. Main Idea of Solution

To address the first research question (a1) - " How to ensure that the generated pattern fits the clothing intended for a specific person?" - it is necessary to assess the quality of human body measurements that will be used for pattern generation. With qualitative body measurements, we understand those that reflect specific measurements of a person's body part with an accuracy of up to ±1 cm, and which, when used, can be used to automatically generate a pattern according to which clothing can be sewn to visually and functionally meet the requirements of the garment manufacturing industry — fitting snugly to the body, avoiding unwanted fabric gathers and protrusions, being symmetrical, allowing freedom of movement, and so on.

In the clothing manufacturing industry, automated individualized pattern creation for each client, without using a standard pattern base, is practically unused. Therefore, the issue of automating polygon control according to various constraints and conditions set by the client's body characteristics, as well as the appropriate clothing style, has not been addressed yet.

The quality of body measurements could be improved in various ways — by relying on anthropometric knowledge (statistical body proportions depending on gender, age, race, etc.), using 3D body scanning tools, visualizing clothing on digitally created avatars, and employing various statistical methods.

Each approach has its own advantages and drawbacks. For example, the use of 3D body scanning tools is limited by the relatively low availability of such equipment, specific requirements for the clothing worn by the measured person during scanning, and dependence on built-in (and unchangeable from outside) calculation algorithms.

In this study, statistical methods will be used, focusing on identifying potentially erroneous measurements before pattern making. The set of measurements for one person will be subjected to statistical comparison with historically accumulated measurements of other individuals (calculating mutual correlations). If the correlation coefficient is below a critical threshold, additional measurement verification is required. This approach can be applied in various ways—by controlling the historical data development of one individual, by comparing the mutual relationships of individual measurements, or by introducing additional indicators (such as lengths and areas of pattern lines, etc.).

Of course, correct body measurements still do not guarantee that the clothing made with the respective patterns will fit the specific individual. There may be an error in the pattern generation script, or perhaps the particular style is not suitable for the person's body type. Similarly, an incorrect result may arise due to issues in garment modeling or changes in the technical infrastructure.

In the case of individually designed patterns, the test objects are numerous and diverse — software used to construct individual clothing patterns (scripts), software in which scripts are executed (construction platform), interfaces with third-party solutions performing specific actions such as visualization, comparison, object placement, etc. Each aspect potentially requires different testing approaches. This publication describes an approach for assessing the quality of body measurements.

II. Quality of Body Measurements

The essence of automated quality control approaches for individually tailored clothing models is as follows: Designers create clothing models, program the garment generation algorithm (create scripts), and present them to clients. From various models, the client chooses the most suitable one and submits their body measurements, which can be numerous — sometimes even more than 50. The software should generate high-quality garment patterns for the client's specific measurements.

To verify the quality of generated patterns (both in terms of measurement compliance and software functionality), in real life, the selected clothing pieces (models) should be sewn from the patterns and fitted to the body, including comparing them with mass-produced analog model garments. Unfortunately, this approach requires immense time and material resources, especially if the measurement sets of different individuals are not analyzed for their similarity, which would allow for the creation of equivalence classes of measurement sets and the possibility of sewing one sample from each equivalence class. However, even using equivalence classes does not solve the problem because the required number of samples needs to be multiplied by the number of garment models, leading to the realization that a full quality check with conventional means is impossible.

In this study, it is proposed to perform an automated model verification by analyzing the mutual correlation of body measurements and additionally calculating integral measurements - the area and perimeter of the garment's element, which are closely related (high correlation coefficient). If these derived values differ from the values predicted by the statistical forecast, it indicates that the corresponding pattern algorithms need to be checked. This method can also reveal inaccuracies in the input of client measurements.

In this section, we will examine methods for obtaining body measurements and analyze the reliability of the obtained values. Two methods will be analyzed:

(1) automatic measurement determination from photographs and images obtained with the help of a 3D scanner;

(2) measurements taken by the clients themselves, followed by statistical analysis of the obtained values, which may show significantly different values from other clients, if such differences exist.

In the digital environment, measurements either need to be determined automatically (for example, from photographs or using 3D scanners) or rely on the client's (typically non-professional) self-measurements.

A. Anthropometry and 3D Scanning

Automated body measurement determination from 2D and 3D images has been widely researched, but unfortunately, the results are not widely used in practice. This is determined by various factors, such as the quality of the photographs, inaccurate posing, clothing worn during photography, limitations of image recognition algorithms, restrictions on the transmission of personal data, and others. Accurate and reliable recognition of body measurements is extremely important in various fields, such as fashion, healthcare, ergonomics, and virtual reality.

The ability to accurately perceive and analyze body measurements is crucial in personalized product design, optimizing suitability, and enhancing user experience. Traditional methods of obtaining body measurements often rely on manual measurement techniques performed by trained professionals. However, the introduction of digital technologies has paved the way for alternative approaches that can improve the accuracy, efficiency, and accessibility of measurements. Two notable technological areas that have significantly influenced body measurement recognition are 3D scanning and image recognition.

There are several works that provide insights into the latest techniques, algorithms, and challenges in image-based body measurement recognition [2], [3]:

1. an overview of various methods and approaches used for obtaining human body measurements from images [4],

2.the use of image processing methods for human body measurement and virtual clothing fitting including algorithms and methods for obtaining body measurements from images. [5],

3. a review of image processing methods used for automatic human body measurement, including various image analysis techniques, extraction algorithms, and measurement evaluation methods [3],

4. the use of deep learning methods to estimate human body measurements from images, including the use of Convolutional Neural Networks (CNN) and other deep learning architectures, to achieve accurate and stable measurement estimation [6].

Anthropometry, the measurement of human body dimensions, plays a crucial role in various fields such as ergonomics, clothing design, healthcare, and biometrics. Thanks to technological advancements, 3D body scanning has emerged as a powerful tool for capturing precise body measurements, offering a comprehensive and accurate alternative to traditional measurement method.

The dataset "IEEE IC 3DBP" [7] provides researchers with a valuable resource for comparative analysis and evaluation of various anthropometric methods in 3D body scanning. The study [3] compares various anthropometric measurement methods based on 3D body scanning. [8], [4] focuses on developing a population-specific anthropometric model based on 3D body scanning data. The findings emphasize the importance of considering population-specific anthropometric analysis variations for applications such as clothing design, ergonomics, and product development.

[5] investigates the use of machine learning algorithms in anthropometric analysis using 3D body scanning [9]. The study examines the application of machine learning models for automated measurement extraction, body segmentation, and anthropometric measurement prediction. The obtained data indicate the potential of machine learning methods to improve the efficiency and accuracy of anthropometric analysis in image datasets.

[3], [10], [11] proposes a comparative analysis of various 3D body scanning technologies and sensor technologies. It evaluates the performance, resolution, and accuracy of different scanning techniques such as structured light scanning, laser scanning, and depth sensing. The aim of [12], [13] is to confirm the accuracy and reliability of 3D body scanning measurements by comparing them with traditional anthropometric methods. The study provides a comparison between measurements obtained from 3D body scanning and manual measurements taken using calipers and measuring tapes. The collected data are used to validate the reliability and practicality of 3D body scanning as a reliable measurement method.

Unfortunately, practical pattern construction systems utilize these technologies to a very limited extent due to several unresolved issues. These include deficiencies in technical infrastructure and professional specialists, ethical considerations regarding client scanning, and other related factors.

B. Data Quality Assessment Using Correlation Methods

The focus of this study is on the quality analysis of measurements taken by the clients themselves or by their trusted persons. Several issues are observed when conducting these measurements:

1. measurements and techniques of taking them may vary depending on the pattern-making method. To standardize the process, visual materials and instructional videos can be used to guide clients on how to take accurate measurements for the specific method used in each case. However, practical experience shows that many clients tend to ignore instructions, either due to impatience or overconfidence in their skills, resulting in incorrect measurements,
2. individuals without sewing experience are unable to accurately measure the body, resulting in measurements that are too tight or too loose in the wrong places,
3. individuals cannot measure certain body dimensions themselves, thus they are forced to rely on assistance from others, which causes stress and additional errors.

If the obtained measurements are incorrect (they do not correspond to the specific body), the individual will end up with ill-fitting clothing pieces, even if the pattern construction algorithm is flawless.

If suspicious sets of measurements, which could arise due to erroneous actions, could be identified automatically, it would be possible to reduce the risk of creating inappropriate patterns. One method for identifying problematic sets of measurements could be to establish a mandatory relationship between measurement definitions and activating control mechanisms at the time of measurement registration. By using these relationships, clear errors could be filtered out, such as an impossible scenario where a woman's bust circumference is smaller than the underbust circumference. However, this approach does not help statistically identify combinations of measurements that are unlikely, as human bodies vary significantly. It is almost impossible to find universal measurement relationships based solely on experience. In this study, correlations of measurements could be analyzed using regression analysis and the capabilities of artificial intelligence on historically accumulated sets of body measurements.

In the human body, a series of measurements exhibit high correlation. For example, both arms or both legs are usually of practically equal length. The available body measurements can be used to calculate the mutual correlation for all pairs of measurements, and it can be observed that for some pairs of measurements, the correlation falls within the range of 0.8-0.9, while for others, it exceeds 0.9, sometimes even surpassing 0.97. For measurement pairs with a correlation above 0.9, it is advisable to create scatterplots of measurements for many clients and identify significant differences or "outliers". If outliers are detected, it indicates that the specific body data may be erroneous, but it does not justify concluding that the data is incorrect; in such cases, the user should be informed that there may be issues with the data. Additionally, more complex regression analysis of measurements is necessary, considering multiple measurements as independent variables and determining the value of the dependent measurement using regression methods (such as linear regression, etc.).).

C. Data Quality Assessment using Total Data Quality Methods

In this section, we will examine the extent to which the automated pattern-generating system can leverage the insights of the Total Data Quality Theory (TDQM).

In 2001, Redman proposed the following definition of data quality: "Data quality is the degree to which data satisfies the specific needs of a given customer." [14]. ISO 9000:2015 [15] provides the following definition of the concept of quality: "Quality is the degree to which consumer needs are satisfied; it represents all the features and characteristics of a product or service that meet customer demand." Quality is therefore a multidimensional concept and encompasses all aspects of how well data align with their purpose.

The notion from quality theory that data quality is relative and dependent on data usage is one of the central principles. When formalizing data quality requirements, it must be considered that they vary from one use case to another. For the same data, different requirements must be defined depending on its usage - distinct data quality specifications must be formulated. Data quality that satisfies all data uses, or in other words, "absolute" quality, is a goal to strive for but is rarely achieved.

The TDQM studies [16] explore a wide spectrum of data and information quality dimensions. As early as 1996, Wang and Strong [17] proposed 15 data quality dimensions, dividing them into 4 groups. In 2001, Redman [14] introduced 51 data quality dimensions, categorizing them into 9 quality groups. In 2013, the Data Management Association International UK Working Group [18] reduced the number of dimensions to six, thereby avoiding an overabundance of dimensions. This approach to data quality monitoring is currently used by the European Statistical Office - Eurostat. Six data quality dimensions are proposed, not tied to specific applications.: Completeness, Uniqueness, Timeliness, Validity, Accuracy, and Consistency.

Data quality is often associated with how accurate the data is. However, data quality is more than just accurate data. The dimensions are not tied to specific data applications, so they should be considered as universal requirements applicable to all applications where the substantive meaning of the dimensions is specified.

Previous research in the field of data quality has not resulted in a universally accepted common theory among researchers. The imprecision of the concept of dimensions serves as an obstacle to this. Most theoretical research is characterized by a wide range of data quality dimensions. Sometimes the number of data quality dimensions is not only too large (ranging from several dimensions to several tens with a list of additional criteria specified for each dimension [19], but also the difference between some of them is almost imperceptible.

In various proposals, the same attribute is often used to indicate semantically different dimensions and vice versa [20]. The main issue is that the precise meaning of each dimension is still being discussed, and there is no consensus on its meaning and how it should be evaluated. According to Batini and Scannapieco [20], theoretical studies on data quality have not yet provided a unified system of data quality concepts. Some authors [21] propose another solution for determining and evaluating data quality. The proposed data object-oriented quality model consists of three main components:

1. data object defines the data whose quality needs to be analyzed,
2. the data quality specification defines the conditions that must be met for the data to be considered qualitative,
3. the quality assessment process.

If data quality is associated with data objects without linking it to the concept of dimensions, it is possible to define an unlimited number of data objects with various structures. Depending on the use case, different data quality requirements can be formulated for the same data object. Therefore, the proposed solution corresponds to the relative nature of data quality and can be applied to quality management in pattern generation.

III. DATA QUALITY EVALUATION AND IMPROVEMENT

To demonstrate how linear regression methods can be applied to improve the quality of body measurement data, this section will describe step by step the process that was conducted in the study using a real dataset of body measurements.

A. Data Selection

Research question (III.A): is there available data to execute a data quality assessment using correlation methods (refer II.B)?

The authors utilized a dataset of body measurements collected as part of a clothing modeling project. The data was accumulated over an extended period from various sources, including self-measurement and professional tailors. From this comprehensive dataset, 2,053 unique measurement profiles were selected. These profiles were collected between 2019 and 2022 as part of a project focused on developing algorithms for customized pattern making based on individual body measurements (M2M). To ensure consistency, all measurements adhered to specific techniques customized for the pattern-making method employed in the project. Tailors involved in the project were already familiar with these techniques, but clients received additional resources, including instructions and video demonstrations, to ensure proper self-measurement if applicable.

Three pattern constructions — bodice base, sleeve, and trousers – were generated for the unique measurement profiles. These constructions were chosen because they utilize the widest and most frequently used body measurements. Only the profiles with successfully generated (withour errors) all three patterns in ,svg format were included in the next selection round. Errors typically occurred due to missing crucial measurements, but in rare cases, geometrical errors indicated illogical relationships within the measurements in a particular set.

Answer on research question (III.A): the authors of the study had access to a set of data collected during a M2M project with 2053 unique profiles. Additionally, they used (refer III.F) a dataset derived from Ansur II [24].

B. Data Classification

Research question (III.B): how to identify measurement profiles that are qualified for data quality assessment using correlation methods (refer II.B)?

The generated .svg files were manually reviewed, and the results were recorded in a table (Table I), noting the profile identifier, credibility (classification of measurements as determined by specialists), and measurements' source. Only profiles representing women's measurements were retained.

TABLE I.
CLASSIFICATION OF PROFILES BASED ON THEIR CREDIBILITY

1046	suspicious	measured by customer
1048	good	measured in house
1050	good	measured by customer
1051	suspicious	measured by customer
1053	good	measured by customer
1057	suspicious	measured by customer
1061	suspicious	measured by customer
1073	bad	measured by customer
1075	suspicious	measured by customer
1076	suspicious	measured by customer
1077	suspicious	measured by customer
1078	bad	measured by customer
1082	bad	measured by customer

The credibility of measurements was determined based on the following criteria:

Good - The generated pattern base constructions visually appear good and reliable; considering the measurements, they are proportionate and appropriate for the represented body size.

Suspicious - The generated pattern base constructions visually appear suspicious, which may mean that some measurements are disproportionately large or small compared to others in the set. For example, a disproportionately long shoulder slope may indicate an incorrectly entered shoulder length (Plg). Suspicious measurements were recorded in the comments.

Bad - The generated pattern base constructions visually appear bad and unreliable (such bases have not been sewn or tested on real people).

The source of the measurements was determined primarily based on the profile name and existing experience:

Measured in-house - Measurements for specific individuals were taken by a tailoring professional.

Measured by customer - Measurements for specific individuals were taken by the client or a tester following the instructions and videos provided.

Not sure - It is not possible to determine the source of the measurements, or there is uncertainty about which group it belongs to.

To create a set of measurement profiles that adequately represent real product users, the customer-measured profiles were divided into groups named “Plusminus” representing main chest circumference (Gkra) intervals (Fig 1), which historically correspond to specific construction methods upon which the programmed pattern constructions are built:

Plusminus1 (over Gkra = 63 cm),
Plusminus1.5 (over Gkra = 79 cm),
Plusminus2 (over Gkra = 95 cm),
Plusminus2.5 (over Gkra = 111 cm),
Plusminus3 (over Gkra = 127 cm),
Plusminus3.5 (over Gkra = 143 cm).

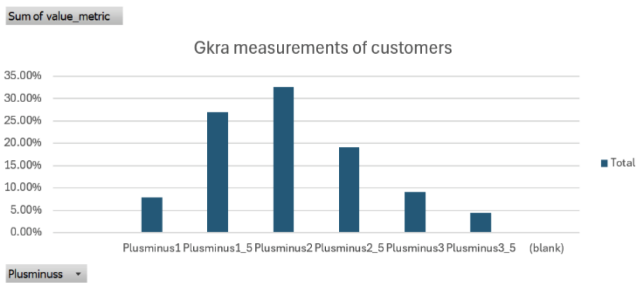


Fig 1. Division of customer profiles by Gkra.

Answer on research question (III.B): profiles where qualified by their origin and by evaluating the generated pattern base constructions for those datasets.

C. Acquiring the Minimum Data Set

Research question (III.C): how to evaluate the identified dataset selected in III.B regarding its completeness for data quality assessment using correlation methods (refer II.B)?

Based on the conclusions of the study [22], to accurately predict human measurements, 500 sets of individual measurements are sufficient. Taking this information into account, the number of profiles to be represented in each plusminus group was calculated (Table II), where first column represents the “Plusminus” group, the second column shows the percentage of this group within the sample dataset, and the third column indicates the number of individuals representing each group.

TABLE II.
ESTIMATION OF THE REQUIRED NUMBER OF MEASUREMENTS

Row Labels	Sum of value_metric	Desired profile count in each group
Plusminus1	7.92%	40
Plusminus1_5	27.00%	135
Plusminus2	32.57%	163
Plusminus2_5	19.04%	95
Plusminus3	9.08%	45
Plusminus3_5	4.40%	22
Grand Total	**100.00%**	**500**

A new table (Table III) was created with 469 profiles, which includes profile ID, name generated based on the profile’s main bust circumference (Gkra) value, “Plusminus” group, credibility.

Credibility was assessed on a scale from 1 to 3 according to the following criteria:

1 - most reliable input data, as they were measured by specialists who have mastered the specific clothing construction method, and the generated clothing base set is recognized as good.

2 - less reliable input data, as although the generated clothing base set is recognized as good, it is not possible to determine whether the measurements were taken correctly, as they were taken by clients or people from test groups.

3 - least reliable input data. The generated clothing base set is considered suspicious because some measurements seemed disproportionate or the base itself visually unusual but not obviously poor.

TABLE III.
UNIFIED DESCRIPTION OF PROFILE DATA

p_profiles_id	p_name	plusminus	credibility
190	66a	pm1	3
193	77a	pm1	3
261	86a	pm1_5	1
286	91a	pm1_5	1
288	84a	pm1_5	1
289	86b	pm1_5	1
291	101a	pm2	1
292	93a	pm1_5	1

The primary table was populated with all available data from credibility category 1. Unfortunately, only 230 profiles met these criteria. Therefore, the missing number of profiles in each “Plusminus” group of this category was identified (Table IV) to create an improved dataset in the future, targeting specialists who can measure individuals representing the missing size ranges. The last two columns provide information of how many credible measurement profiles are needed in each group with or without including persons that are measured several times.

TABLE IV.
ESTIMATION OF AVAILABLE/ MISSING MEASUREMENTS

Row Labels	Sum of value_metric	Desired profile count in each group	Profiles to be measured (w. duplicates)	Profiles to be measured (wo. duplicates)
Plusminus1	7.92%	40	23	30
Plusminus1_5	27.00%	135	8	23
Plusminus2	32.57%	163	92	98
Plusminus2_5	19.04%	95	61	65
Plusminus3	9.08%	45	39	41
Plusminus3_5	4.40%	22	7	13
Grand Total	100.00%	500	230	270

The missing table was filled with data from categories 2 and 3. Even including these less reliable data, the target of 500 profiles in the dataset has not been achieved (Table V).

TABLE V.
MISSING MEASUREMENTS PER PLUS-MINUS GROUP

	Plusminus1 (Gkra >= 63 cm)	Plusminus1.5 (Gkra >= 79 cm)	Plusminus2 (Gkra >= 95 cm)	Plusminus2.5 (Gkra >= 111 cm)	Plusminus3 (Gkra >= 127 cm)	Plusminus3.5 (Gkra >= 143 cm)
count right now	24	135	163	95	38	14
still needed	40	135	163	95	45	22
difference	-16	0	0	0	-7	-8

Interpreting the upcoming results, it's important to note that the missing 31 sets, needed to reach the minimum dataset size of 500 profiles, are spread across both very small and very large body size categories. In contrast, there were sufficient measurements in the average size groups.

To leverage existing resources, the authors opted to analyze readily available data for this initial phase. This would result in the development of a validation method for new measurements using an incomplete dataset. However, the method's design allows for seamless integration of a more robust dataset in the future, facilitating its adaptation to improved information.

Answer on research question (III.C): the amount of selected profiles is sufficient for mid-sized groups but insufficient for small and larger sized groups. The sample set is large enough to develop a system for validating measurement accuracy, but it is insufficient for making assumptions about underrepresented groups.

D. Determining Correlation between Measurements

Research question (III.D): which measurements in the selected dataset have high enough correlations with others to be forecasted and to help detect significant outliers, indicating poor data quality?

Using the prepared dataset of 469 profiles, correlation coefficients between measurements were determined (Fig 2).

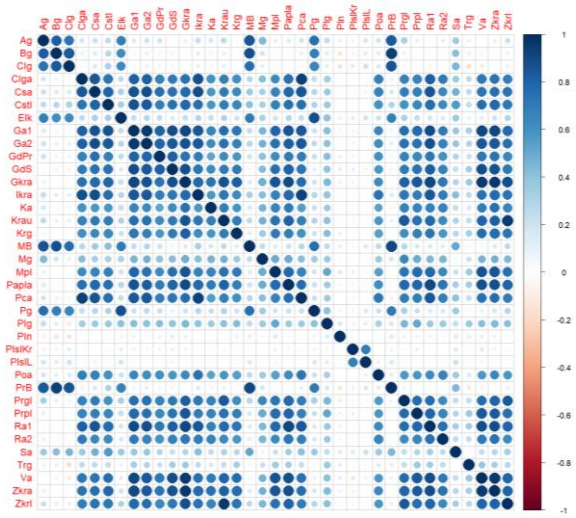


Fig 2. Correlation coefficients between measurements

37 of the most used measurements in the chosen pattern making method were selected. For each pair of measurements, an interactive scatterplot was created in RStudio [26], marking potential outliers located outside two standard deviations from the specific dataset's mean. Examples are in Fig 3 and Fig 4.

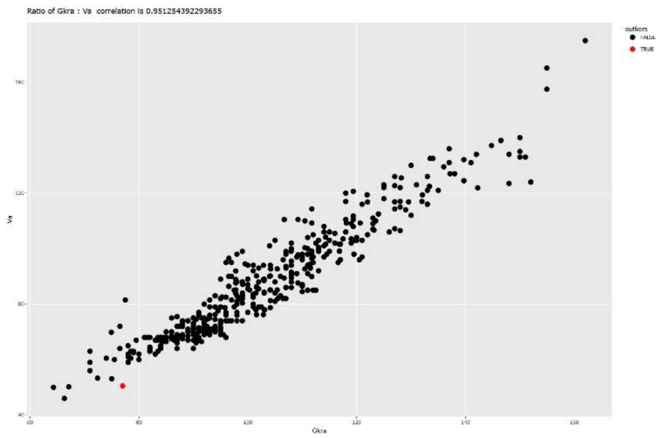


Fig 3. Correlation of Gkra and Va (0.951254392293655)

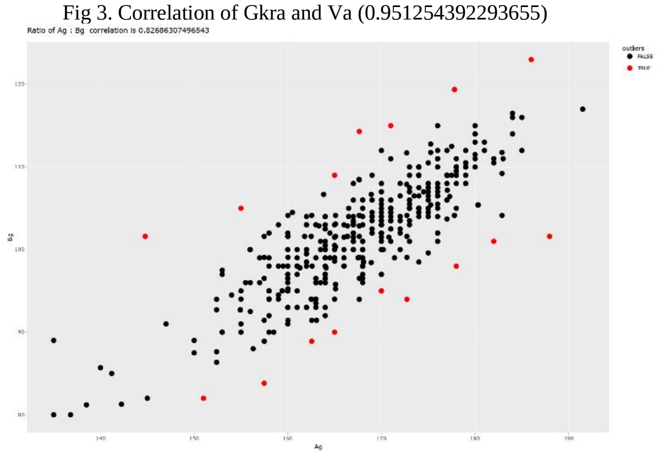


Fig 4. Correlation of Ag and Bg (0.8286307496543)

27 measurements show a high enough correlation (coefficient over 0.7) with others. Additional 4 measurements have moderate correlation (coefficient of $0.5 < |r| \leq 0.7$), the rest have low or negligible correlation.

For pairs of measurements exhibiting a correlation coefficient of at least 0.7, profile IDs and reliability were recorded. Since one of the authors is a tailor and was involved in measuring process of bodies from data set, it was easier to determine nature of the outliers in credibility group 1 and it was assumed that they most likely represent natural variation in body measurements, while outliers with profiles representing categories 2 and 3 could also be due to incorrectly taken measurements or input errors.

Results were documented (Table VI). The column "outlier_percent" indicates the percentage of outliers marked with credibility categories 2 and 3. The column "could be true outlier" records outliers that are distinctly far from the rest of the points in the scatter plot. Outliers with credibility category 3 predominated in this category.

TABLE VI.
ANALYSIS OF OUTLIERS

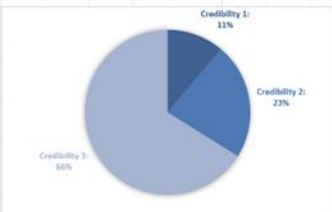
Measurement_1	Measurement_2	Correlation_Coefficient	outlier_percent	2 un 3 credibility	1 credibility	could be true outlier
Gkra	Zkra	0.97	91	1083, 2545, 2613, 2956, 3388, 3531, 190, 193, 2709,2781	368	
Ga1	Ga2	0.95	66	a lot	307, 347, 381, 381, 2236, 2238, 2240, 2241,2246, 2316, 2359, 2382, 2974, 3346, 3409	2982 extreme, 3145 and 2709 close to extreme, 3055, 507, 1076 on the fence
Va	Zkra	0.95	100	3301, 919, 738, 2505		
Krau	Zkrl	0.95	60	998, 3281, 190, 507, 939, 2600	2250, 2252, 2315, 2895	998
Gkra	Va	0.95	100	193		
Bg	PrB	0.94	97	a lot	3415	
Clga	Pca	0.93	43	3475, 193, 475,	324, 335, 2208,	324
Ikra	Pca	0.92	100	3282, 3367, 2600, 885, 3040		3282 very prominent outlier

Summarizing the results of the outliers' analysis (Table VII with pie chart within), it is evident that the majority of outliers occurred in profiles with low reliability. Although seemingly trivial, this aspect is significant for further data utilization.

TABLE VII.

DISTRIBUTION OF OUTLIERS ACCORDING TO THEIR CREDIBILITY

unique id's:	frequency	credibility		
3341	26	3	Credibility 1:	27
2982	16	3	Credibility 2:	55
2823	13	3	Credibility 3:	160
3145	12	2		
2709	12	3		
324	9	1		
2556	9	3		
3282	8	2		
1046	8	3		
3367	7	3		
412	6	1		
922	6	3		
3092	6	3		
2600	5	3		
2327	4	1		
3409	4	1		
2921	4	2		
2934	4	2		
834	4	3		
193	3	3		
507	3	3		



Answer on research question (III.D): using the given set of data, 27 out of 37 measurements show a high correlation (coefficient over 0.7) to be used in the initial outlier analysis.

E. Predicting Measurements from Existing Data

Research question (III.E): with what accuracy can each of the 37 measurements be predicted?

Profiles with significant outliers were excluded from the data set which was split into training and testing sets.

The training set contained 80% of the data, the remaining 20% of data was reserved for testing. For each measurement a predictive model was trained. Later the model was used on testing data set, and the results were compared.

In the initial study, a linear regression model was used due to its simplicity, interpretability, and the natural linear relationships among body measurements. A previous study [22] and [23] also recommended using linear regression for estimating body dimensions, noting that it tends to be more stable compared to tree-based models. The 'train()' function in R was utilized because it offers a variety of parameter options. For instance, 'preProcess = c("scale", "center")' standardizes the data by giving it zero mean and unit variance, and 'trControl = trainControl(method = "LOOCV")' specifies Leave-One-Out Cross-Validation (LOOCV) for model training. LOOCV is a robust cross-validation method where each observation is used once as a test set, with the rest serving as the training set, which helps in thoroughly evaluating the model's performance.

Results for the main chest circumference predictions compared to actual measurements are pictured in scatterplot below (Fig 5). There is a linear regression line fitted over all data points and ellipses with confidence level of 95% for each group of credibility. As expected, the data points that are the most credible (credibility 1, red) are closer to regression line and therefore predicted better. Data obtained from clients (credibility 2, green) tend to be more scattered away from the regression line, particularly for the least reliable entries (credibility 3, blue).

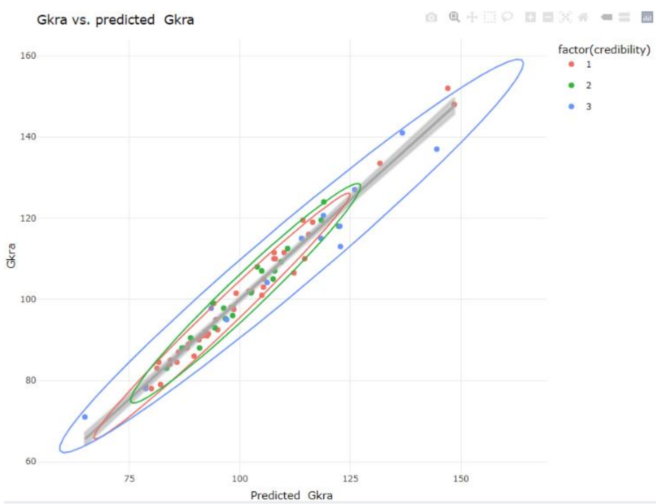


Fig 5. Gkra vs. predicted Gkra

Based on the predictive model's performance, body measurements can be predicted with an accuracy of +/- 1.7 cm for the middle 68% of the data within the testing set. Furthermore, within the middle 95% of the data, the accuracy extends to +/- 4 cm. To provide context, the average bust circumference measurement is 95 cm. Therefore, this level of accuracy ensures errors are limited to less than 3.8% of this particular measurement itself. This represents a significant improvement over the previous approach used in validating measurements before generating M2M patterns, which relied mainly on detecting unlikely scenarios (e.g., an ankle circumference exceeding a knee circumference), empirical formulas based on only a few other measurements and minimal and maximal values of each measurement across the population. Although the prediction of this trained model does not meet the goal of +-1 cm it still can be leveraged to validate user-provided measurements to avoid gross user data entry errors. If a measurement falls outside the predicted bounds, a warning notification would be triggered, prompting the user to re-check the entered value.

The performance of the models was further analyzed for each measurement individually using metrics such as R^2 (Fig 9), adjusted R^2 (Fig 10), coefficient of variation (Fig 6), F-statistic (Fig 7), and its p-value (Fig 8). The results may show that the rankings of the best-performing models differ based on these metrics. For instance, model Ag (body height) might have a lower coefficient of variation (CV) for its residuals (Fig 6), indicating more consistent prediction errors, while model Gkra (bust circumference) might have a statistically more significant fit based on its F-statistic and p-value (Fig 7). However, the worst-performing models consistently show poor results across all metrics, indicating a clear distinction in predictability between the best and worst-performing measurements.

The measurement Pln (forward projection of the shoulder) demonstrated the least predictability using linear regression methods, as evidenced by a CV of 34.7%, an R^2 value of 0.210, an adjusted R^2 of 0.110, an F-statistic of 2.094, and a p-value of 0.0004806754. This lack of predictability is likely due to the observational nature of the measurement, which makes it prone to misinterpretation.

This approach also struggles to accurately predict PlslL (shoulder slope right), PlslKr (shoulder slope left), and Trg (trapezoid muscle length), which is unsurprising given the potential imprecision of these measurements, even within the highest credibility data group. Alternative methods should be explored to achieve more accurate results for these measurements, particularly for PlslL and PlslKr, as shoulder slope significantly impacts garment fit.

In contrast, the remaining measurements appear to be fairly predictable, with an average CV of 4.3% and a median CV of 4%.

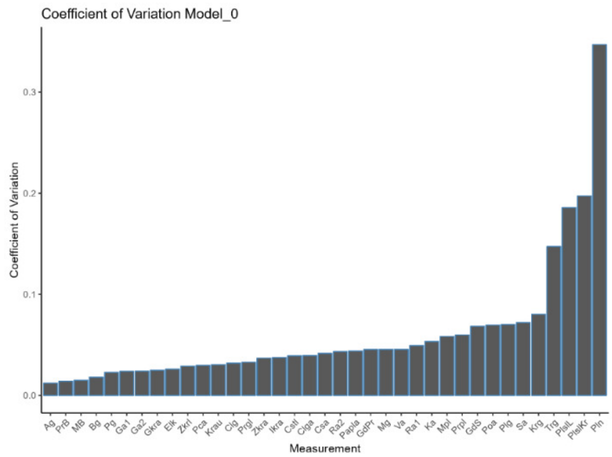


Fig 6. Coefficient of Variation for each measurement model

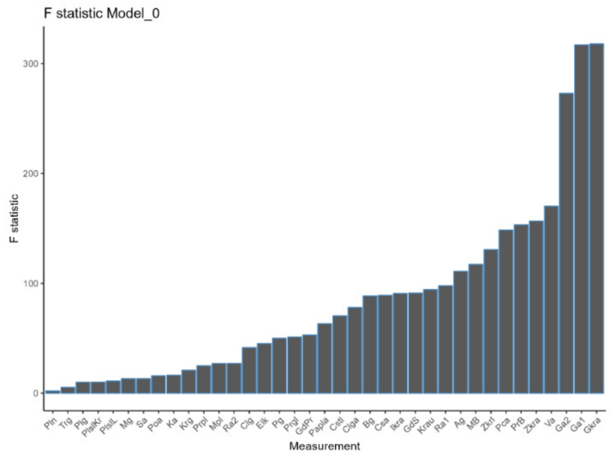


Fig 7. F statistic for each measurement model

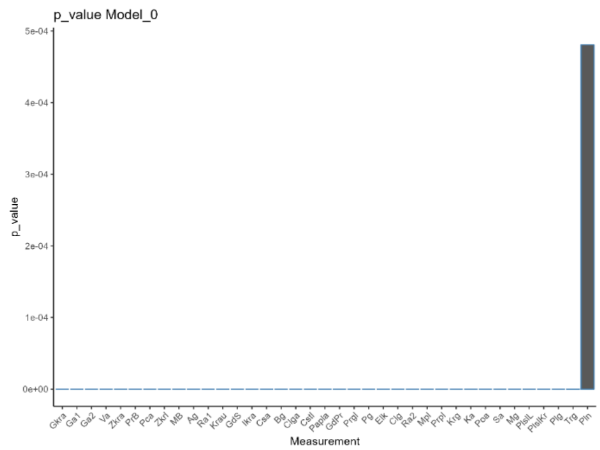


Fig 8. P values for each measurement model

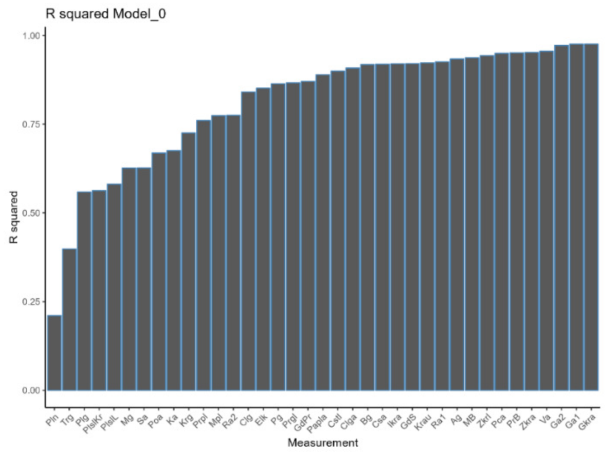


Fig 9. R^2 for each measurement model

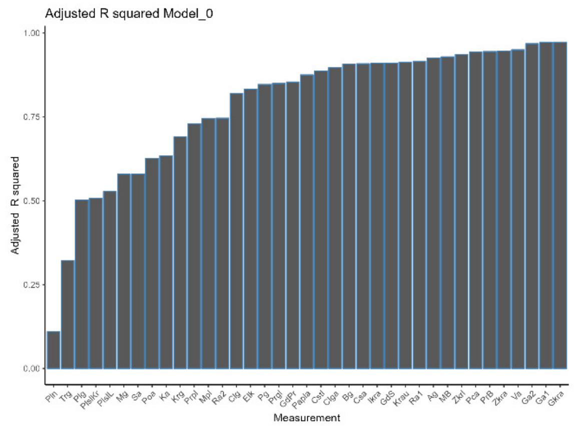


Fig 10. Adjusted R^2 for each measurement model

Answer on research question (III.E): the accuracy of predictions is within ±4 cm for majority of measurements.

F. Predicting the Impact of Additional Data

Research question (III.F): what impact would the addition of new, high-quality data have on the model?

To ensure that increasing the amount of high-quality measurement data would improve prediction accuracy, a dataset derived from Ansur II[24] data was created. Ansur II[24] data set consists of measurements of 1986 females. For the initial attempt, all measurements matching ours were selected, and the rest were calculated. For example, in our dataset, the Clg measurement is from the waist level to the middle of the kneecap. In Ansur II[24], knee height is measured from the middle of the kneecap to the floor [25]. Therefore, to obtain a comparable measurement to ours, the formula "Clg = Ansur II Iliodristal height [25] (similar to our PrB) – Ansur II Knee Height, Midpatella [25] " was used.

In cases where deriving the necessary measurements was not possible, they were assumed from our dataset. For instance, the Papla measurement correlates most with Gkra and Ra1. Since the Gkra measurement technique matches between our dataset and Ansur II [23], the Papla measurement for each person in the Ansur II [23] database was assigned based on the closest Gkra measurement in our dataset. This approach was used to introduce some natural variability as opposed to calculated values, which would be immediately captured by the regression model, leading to overfitting. These measurements were flagged to facilitate measuring the model's performance with and without them due to their questionable nature.

Linear regression models for each measurement were trained twice: once with all the calculated and assumed measurements, and once with only the original measurements. Models trained with original and added data had an average CV of 3.25% and a median CV of 3.2%. Models trained with fewer but only original data performed slightly better with an average CV of 3.08% and a median CV of 3.17%.

The results of the model performance using the dataset with more quality data showed an improvement in prediction precision, as evidenced by lower mean and median CVs compared to the results of the linear regression model of our dataset.

We must consider that the Ansur II dataset lacks information for the "Plusminus3" and "Plusminus3_5" groups. Consequently, profiles from these groups were temporarily excluded from our dataset, and the performances of our measurement models were reassessed. The CV decreased to a mean of 3.91% and a median of 3.79% which is still worse than the results of models trained with Ansur II[24] data set. In conclusion it's safe to say that if two datasets with similar variance is used to train the model the one with more quality data will provide better results.

Answer on research question (III.F): using only high-quality data to train the predictive models leads to more precise measurement predictions.

IV. PRACTICAL APPLICATIONS

The results presented here regarding the discussed body measurements and the calculated prediction accuracy are specific — they apply 1:1 only to the specifically used methodology for creating patterns and the specifically available sets of body measurements. For the specific M2M case, based on the conducted analyses, concrete tolerance intervals can be defined for 33 body measurements. When implemented in an application, these intervals generate warnings if an entered value for a body measurement is likely inconsistent with the other body measurements. This application thus signals to the user a likely, statistically justified data inconsistency, preventing avoidable errors in data collection. However, for 4 body measurements (Pln, PlslKr, PlslL, Pln) discussed in section III.E, no reasonably justifiable tolerance intervals can be calculated, and no corresponding warnings can be generated. As explained above, an improvement (i.e., a reduction of the corresponding tolerance intervals) would be achieved by expanding the available data set and repeating the calculations. It is credible — at least for the middle "Plusminus" groups — that the goal of 1cm tolerance for most body measurements is achievable with a three-digit number of data sets.

The specifically calculated 33 tolerance intervals cannot, of course, be directly transferred to other application areas. However, if we abstract from the specifically calculated values and interpret the applied approach as the investigation result, further statements about practical applicability can be made. The approach can fundamentally be transferred to any set of body measurements (both similar but differently defined ones, as well as completely different anthropological dimensions, such as body weight, which was not available to us). Wherever a set of (correlated) body measurements is manually recorded, and the data quantities are at least in the three-digit range, the outlined approach can be applied. Due to the anthropologically given correlation of body measurement data, it is possible to determine statistically justifiable tolerance intervals for most body measurements, and thus a practical method for avoiding errors in data collection can be defined. The outlined method is therefore not only applicable to the field of M2M pattern making used here, but also to the clothing industry in general and possibly even in completely different contexts (e.g., in the medical field).

V. Conclusions

The obtained results can be summarized as follows:

1. Even with a relatively small dataset of measurements (in this case, 469 profiles), it is possible to effectively identify potential data quality issues as long as a clear data quality assessment methodology is applied (refer the answer on research question III.D).
2. Using linear regression methods, it is possible to develop and train models that can reliably predict certain body measurements with high accuracy (refer the answers on research questions III.D and III.E).
3. The accuracy of measurement predictions could have been improved if more data from small and very large body sizes had been available (refer the answers on research questions III.C, III.D, III.E and III.F).
4. To get closer to the desired accuracy of ±1 cm, it is essential to obtain more reliable data measured by professionals with a credibility rating of 1. The current dataset is sufficient to significantly improve the existing solution for measurement validation, which relies on the minimal and maximal values of each measurement across the population and empirical formulas. However, a predictive model based on the existing data would primarily identify gross errors and mistypes. More reliable data, beyond the desired 500 profiles, should be obtained to test the model's accuracy and determine the percentage of accurate and inaccurate predictions available (refer the answers on research questions III.B, III.C, III.D, III.E and III.F).

The research is planned to be continued in two directions. Regarding the validation of body measurements, we intend to employ additional statistical methods for validation and compare them in terms of achievable statistical quality and practical applicability. Similar approaches can be found, among others, in [22].

Furthermore, we aim to investigate whether a similar use of statistical methods can also be employed to test M2M implementations (custom-made patterns). Assuming there is a pattern program that has been validated for N individuals, the question arises whether potentially faulty patterns for additional individuals can be identified without actually sewing and trying them on — the currently only practical way to validate an M2M pattern. Given the potentially avoidable costs involved, a machine-based, statistically driven method for identifying potential errors would be of very high practical relevance.

References

[1] J. Bicevskis, E. Diebelis, Z. Bicevska, A. Neimanis, "Regression Testing: Test Cases for Graphical Images." In Joint Proceedings of Baltic DB&IS'2022 Doctoral Consortium and Forum. https://ceur-ws.org/Vol-3158/paper6.pdf
[2] Ruiz N., Bueno M.B., Bolkart T., Arora, Lin M., Romero J., Bala R., "Human body measurement estimation with adversarial augmentation. International." Conference on 3D Vision, 2022 https://arxiv.org/abs/2210.05667
[3] Bartol K., Bojanić D., Petković T., Pribanic T., A "Review of Body Measurement Using 3D Scanning." IEEE Access, DOI: 10.1109/ACCESS.2021.3076595, 2021.
[4] Kuribayashi M., Nakai K., Funabiki N., "Image-Based Virtual Try-on System With Clothing-Size Adjustment." DOI: 10.48550/arXiv.2302.14197, 2023.
[5] Pleuss J.D., Talty K., Morse S., Kuiper P., Scioletti M., Heymsfield S.B., Thomas D.M., "A machine learning approach relating 3D body scans to body composition in humans." Eur J Clin Nutr. 2019 Feb; 73(2): pp. 200–208, published online 2018 Oct 12. doi: 10.1038/s41430-018-0337-1
[6] Ashmawi S., Alharbi M., Almaghrabi A., Alhothali A., "Fitme: Body Measurement Estimations using Machine Learning Method." Procedia Computer Science. Volume 163, pp. 209-217, 2019. https://doi.org/10.1016/j.procs.2019.12.102.
[7] "IEEE IC 3DBP" https://ieee-dataport.org/open-access/dataset-ieee-ic-3dbp-comparative-analysis-anthropometric-methods.
[8] Lu J., Wang M.J, "Automated anthropometric data collection using 3D whole body scanners." DBPL, Expert Systems with Applications 35(1-2):407-414, July 2008. DOI: 10.1016/j.eswa.2007.07.008
[9] Liu X., Wu Y., Wu H., "Machine Learning Enabled 3D Body Measurement Estimation Using Hybrid Feature Selection and Bayesian Search." Appl. Sci. 2022, 12(14), 7253; https://doi.org/10.3390/app12147253.
[10] Kus A., Unver E., Taylor A, "A Comparative Study of 3D Scanning in Engineering, Product and Transport Design and Fashion Design Education." Computer Applications in Engineering Education 17(3), pp. 263 –271, September 2009 DOI: 10.1002/cae.20213
[11] Seifert, E., Griffin, L, "Comparison and Validation of Traditional and 3D Scanning Anthropometric Methods to Measure the Hand. Paper presented at 11th Int. Conference and Exhibition on 3D Body Scanning and Processing Technologies. https://doi.org/10.15221/20.41 , 2020.
[12] Skorvankova, D., Riečický, A., Madaras, M, "Automatic Estimation of Anthropometric Human Body Measurements." 17th International Conference on Computer Vision Theory and Applications. (2021) DOI:

10.5220/0010878100003124, https://www.scitepress.org/PublishedPapers/2022/108781/108781.pdf
[13] Rumbo-Rodríguez, L., Sánchez-SanSegundo, M., Ferrer-Cascales, R., García-D'Urso, N., Hurtado-Sánchez, JA., Zaragoza-Martí, A., "Comparison of Body Scanner and Manual Anthropometric Measurements of Body Shape: A Systematic Review.". Int J Environ Res Public Health. 2021 Jun 8;18(12):6213. doi: 10.3390/ijerph18126213
[14] T.C. Redman, "Data Quality. The Field Guide", Digital Press, 2001.
[15] Quality management systems. https://www.iso.org/standard/62085.html
[16] European statistics Code of Practice — revised edition 2017. https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142
[17] R. Y., Wang, D. M., Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers", Journal of Management Information Systems, Springer, Vol.12., No.4, pp. 5-34, 1996.
[18] DAMA UK. https://www.dama-uk.org/
[19] A. Caro, C. Calero, M. Piattini, "A Portal Data Quality Model for Users And Developers", In ICIQ, pp. 462-476, 2007
[20] C. Batini, M. Scannapieco, "Methodologies for Information Quality Assessment and Improvement. Data and Information Quality Dimensions, Principles and Techniques", Springer International Publishing Switzerland, 2016.
[21] J. Bicevskis, Z. Bicevska, A. Nikiforova, I. Oditis, "An Approach to Data Quality Evaluation". In Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 196-201, IEEE, 2018.
[22] Meyer, P., Birregah, B., Beauseroy, P. et al., "Missing body measurements prediction in fashion industry: a comparative approach." Fash Text 10, 37 (2023). https://doi.org/10.1186/s40691-023-00357-5
[23] Janis Bicevskis, Edgars Diebelis, Zane Bicevska, Ivo Oditis, Girts Karnitis, Oskars Ozols. "Assessing the Accuracy of Body Measurements through Regression Analysis." In 18th Conference on Computer Science and Intelligence Systems, September 17–20, 2023. Warsaw, Poland : Position Papers Vol. 36, p.35-41. DOI: http://dx.doi.org/10.15439/978-83-969601-1-5.
[24] The Anthropometric Survey of US Army Personnel (ANSUR 2) Female data files, 2012. https://www.openlab.psu.edu/ansur2/
[25] Hotzman, J., Gordon, C.C., Bradtmiller, B., Corner, B.D., Mucher, M., Kristensen, S., Paquette, S., and Blackwell, C.L. 2011, "Measurer's Handbook: US Army and Marine Corps Anthropometric Surveys, 2010-2011," Report No. NATICK/TR-11/017.
[26] RDocumentation. https://www.rdocumentation.org/packages/caret/versions/4.47/topics/train