

Supervised Learning Techniques for Body Mass Estimation in Bioarchaeology

Vlad-Sebastian Ionescu^(✉), Gabriela Czibula, and Mihai Teletin

Faculty of Mathematics and Computer Science, Babeş-Bolyai University,
1, M. Kogălniceanu Street, 400084 Cluj-Napoca, Romania
{ivlad,gabis}@cs.ubbcluj.ro, tmir1584@scs.ubbcluj.ro

Abstract. The task of estimating the body mass from human skeletal remains based on bone measurements is an important one in bioarchaeology and forensic anthropology. Most of the current literature deals with this problem through mathematical linear regression formulas applied to various bones. In order to improve the existing results, two supervised learning-based regression models are proposed, using *artificial neural networks* and *support vector machines*, which are useful for expressing good (usually nonlinear) mappings between skeletal measurements and body mass. Several experiments performed on an open source data set show that the proposed applications of machine learning-based algorithms lead to better results than the current state of the art. Thus, the proposed methods are useful for producing good body mass estimations from skeletal measurements.

Keywords: Bioarchaeology · Body mass estimation · Machine learning · Artificial neural networks · Support vector machines

1 Introduction

Body mass estimation from human skeletal remains represents a problem of major importance in paleontological and archaeological research, since it can provide useful information about past populations [21], such as their healthiness, different social aspects, the influence of environmental factors and others. Obtaining correct estimations for body mass from skeletons is also important in forensic death investigations regarding unknown skeletal remains [18].

In the bioarchaeological literature it is agreed that postcranial measurements have a direct relationship to the body size and produce the most accurate estimates [3]. Estimation of the body mass from human skeletons represents a challenge in forensic death investigations concerning unidentified remains [18]. A major problem in research related to body mass estimation is caused by a lack of research collections. The current methods for estimating body mass from the skeletons are: *mechanical*, *morphometric* [2] and a combination of biomechanical and morphometric methods [18]. The morphometric methods consist of directly reconstructing the body size from the skeletal elements, while the mechanical methods provide functional associations between skeletal elements and body mass [2].

In the machine learning literature, *artificial neural networks* [6,26] and *support vector machines* [12] are well known supervised learning models, with a wide range of applications. They are adaptive systems that are able to learn an input/output mapping (target function) from data. After the systems were trained by receiving some examples for the target function, they will be able to provide an approximation of the target output for new inputs.

In this paper, two machine learning-based regression models are used for estimating the body mass from human skeletons, using bone measurements: *artificial neural networks* and *support vector machines*. The experiments are carried out on a publicly available database. The obtained results are compared with the results of several body mass estimation regression formulas existing in the bioarchaeological literature. The results obtained by the application of machine learning models outperform the previous state of the art. Performed experiments indicate a superior performance of the proposed models relative to other similar approaches existing in the literature and highlight the efficiency of using machine learning-based regression techniques for the problem of human body mass estimation. To the best of the authors' knowledge, the proposed approaches are novel, since there are no existing machine learning approaches to learn, in a supervised manner, to estimate body mass from human skeletons.

The remainder of the paper is structured as follows. Section 2 presents the problem of archaeological body mass estimation and its relevance in bioarchaeology, as well as the fundamentals of the machine learning-based methods which will be used in this paper. Section 3 introduces the machine learning-based methods for estimating the body mass from human skeletons that were applied. Experiments on a publicly available data set are given in Sects. 4 and 5 contains an analysis of the obtained results and comparisons with related work from the literature. Section 6 provides the conclusions of the paper and indicates several future research directions.

2 Background

This section starts by briefly presenting the body mass estimation problem, emphasizing its relevance and importance within the archaeological research. Then, the fundamentals of *artificial neural networks* and *support vector machines* are presented.

2.1 Body Mass Estimation

Body mass estimation is a very important problem for modern archeology. It can provide certain knowledge about past populations, such as indicators for [21]: the past population's health, the effects of different environmental factors on past populations (e.g. subsistence strategy, climatic factors), social aspects etc. The ability to obtain accurate body mass estimations from skeletons is also essential in forensic death investigations concerning unidentified skeletal remains

[18]. Consequently, it is essential for bioarchaeologists to develop and use body mass estimation methods that are as accurate as possible.

However, designing an accurate method for solving this problem remains a great challenge, because there are many factors which should be taken into account [2]. Some decisions that are to be taken in this process [21]: which are the most relevant skeletal measurements to use, which is the appropriate statistical approach to apply, which should be the skeletal sample to use etc.

Most of the existing statistical methods for body mass estimation use linear regression formulas that usually consider one or a few bones measurements. These formulas are usually developed on particular data sets, and it is questionable whether or not they would perform well on previously unknown data.

Supervised machine learning-based models are therefore likely to be a good alternative to existing methods for body mass estimation, since they can be retrained on new data sets easily. Moreover, particular techniques to avoid the overfitting problem can be used in order to develop models that show good potential to generalize well on unseen data.

2.2 Artificial Neural Networks and Support Vector Machines

Artificial neural networks [15, 25] are machine learning models with a wide application area in domains like pattern recognition, speech recognition [27], prediction [14], system identification and control. In structural similarity with biological neural systems, artificial neural networks [17] consist of a set of interconnected computational units, also referred to as neurons. One motivation for the distributed representation is to capture the parallelism expressed by natural neural systems [17].

An *artificial neural network* (ANN) [1, 20] is an adaptive system that learns a function (an input/output map) from data. Adaptive means that the system parameters are repeatedly adjusted during the *training phase*. After the training phase, the Artificial Neural Network parameters are fixed and the system is used to solve the problem at hand (the testing phase). The Artificial Neural Network is built with a systematic step-by-step procedure to optimize a performance criterion or to follow some implicit internal constraint, which is commonly referred to as the learning rule.

In a supervised learning scenario, an input instance is presented to the neural network together with the corresponding target response [13]. These input-output pairs are often provided by an external supervisor. An error is represented by the difference between the desired response and the system's output. This error information is fed back to the network and the system parameters are adjusted in a systematic manner (the learning rule). The process is repeated until an acceptable performance is achieved.

Support Vector Machines were developed for classification by Cortes and Vapnik [7], but they can be adapted for regression quite easily, resulting in the so-called ε -support vector regression algorithm (*SVR*). The hyperparameter ε controls the level to which the algorithm is allowed to make mistakes.

Formula (1) describes the underlying SVR numerical optimization problem. The following notations are used: x_i is a training instance, y_i is its target output, b is a bias term, w is the weights vector being searched and C is a regularization hyperparameter. The variables ξ_i^- and ξ_i^+ are used to make the learning feasible by allowing for some degree of error [28].

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i^- + \xi_i^+) \\ & \text{subject to} && \begin{cases} y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i^- \\ (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \end{cases} \end{aligned} \quad (1)$$

3 Machine Learning-Based Body Mass Estimation

This section introduces the proposed machine learning-based approaches (*artificial neural networks* and *support vector machines*) for estimating the body mass of human skeletal remains, based on bone measurements.

Consider a data set of human skeletal remains denoted by $\mathcal{H} = \{h_1, h_2, \dots, h_n\}$ in which each instance h_i represents a human skeleton. Each skeleton is characterized by m features representing different bone measurements which are relevant for the body mass estimation problem. The measurements are numerical values and correspond to several significant bones in the body. Thus, an instance h_i may be viewed as an m -dimensional vector $h_i = (h_{i1}, h_{i2}, \dots, h_{im})$ where $h_{ij} \forall 1 \leq i \leq n$ represents the value of the j -th measurement applied to the i -th skeleton. For each human remain h_i , the *body mass* of the individual is known and is denoted by bm_i . No feature extraction or selection was performed. The features were used as they appear in the data set from the archaeological literature.

The ANN and SVM models are used as regressors for the *body mass* estimation problem - they provide an estimation e_i for the target *body mass* for each individual h_i from the data set \mathcal{H} .

Before applying the machine learning-based methods, the data set is *pre-processed*. First, the data is normalized using the *Min-Max* normalization method. ANN and SVM models are known to be sensitive to data normalization, requiring the values inputted to them to be of identical orders of magnitude in order to be able to perform well.

After normalization, a statistically-based feature selection is applied in order to determine how well the measurements are correlated with the target body mass output. The dependencies between the features and the target body mass are determined using the Pearson correlation coefficient [30]. A high Pearson correlation implies that the two variables are linearly related, which makes linear and non-linear learning models likely to perform well. A low Pearson correlation means that a linear model will likely not perform well, but does not give any information about non-linear models.

Also as a data preprocessing step, a self organizing feature map (SOM) [29] is used in order to obtain a two dimensional view of the data that will learned. The trained SOM is visualized using the U-Matrix method [11]. The U-Matrix values can be viewed as heights giving a U-Matrix landscape, with high places encoding instances that are dissimilar while the instances falling in the same valleys represent data that are similar. On the U-matrix, it is possible to observe outliers, i.e. input instances that are isolated on the map. The visually observed outliers are eliminated from the training data. This process is detailed in the experimental part of the paper.

After the data preprocessing step, as in a supervised learning scenario, the regression ANN and SVM models are trained during *training* and then *tested* in order to evaluate the performance of the obtained models. These steps are further detailed in their respective sections.

3.1 Training

The following sections present details about how the ANN and SVM models are built during the *training* step.

The ANN Model. The ANN model's capability to represent non-linear functions is used in order to solve the weight estimation problem. The requirement for the problem at hand is to design a model which uses the known measurements as input features, processes them in its hidden layers and gives the results corresponding to the estimation.

For the experiments, a *feedforward* neural network is built. It will be trained using the *backpropagation-momentum* learning technique [24].

The ANN is composed of three layers: *input*, *hidden* and *output*. The input layer is composed of n_i (equal to the dimensionality of the input space) neurons. A single hidden layer is used, the number n_h of hidden neurons being computed as $n_h = \lceil \sqrt{n_i \cdot n_o} \rceil$. Bias neurons are also used in the input and hidden layers. On the output layer, there is a single neuron corresponding to the *weight* output. All hidden and output units use the sigmoid activation function.

During *training phase*, the squared error between the network output value and the target value for the current training instance is computed. The error is propagated backwards through the network, followed by the gradient computation [17]. All the weights are updated in order to reduce the observed error for the current example.

In order to avoid the convergence to a local minimum, *stochastic gradient descent* is used during training, for building the network. The idea behind this variation [17] is to approximate the true gradient descent search by updating the weights incrementally, following the calculation of the error for each individual example.

The *momentum technique* is also added to the ANN model, in order to speed up the convergence and to avoid local minimums [23].

A validation subset consisting of $\lceil n/10 \rceil$ randomly chosen instances (n denotes the number instances from the input data set) is extracted from the training data set and used for optimizing the number of training iterations. Ten randomly initialized ANNs are trained in parallel and at the end of the training, the ANN that has the best performance on the validation set is kept. The obtained ANN is then used for testing.

The SVM Model. The SVM model consists of ε -SVR with multiple choices of kernels and a randomized hyperparameter search. A random search has been shown to outperform classical grid searches for this problem [4]. The random search consists of specifying either a probabilistic distribution for each hyperparameter or a discrete set of possible values. Then, for a given number of iterations, random values are selected for each hyperparameter, according to the given probability distributions or discrete sets of values. A model is constructed each iteration, and evaluated using 10-fold cross validation. The best model returned by the random search is then tested using Leave One Out cross validation, as described below.

3.2 Testing

For evaluating the performance of both the ANN and SVM models, a *leave-one out* cross-validation is used [31]. In the *leave-one out* (LOO) cross-validation on a data set with n instances, the learning model is trained on $n-1$ instances and then the obtained model is tested on the instance which was left out. This is repeated n times, i.e. for each instance from the data set.

During the cross-validation process, the *mean of absolute errors (MAE)* is computed using the Formula (2)

$$MAE = \frac{1}{n} \sum_{i=1}^n |bm_i - e_i| \quad (2)$$

In the above formula e_i is the estimated body mass for the i -th test instance (as provided by the ANN/SVM regressor), and bm_i represents the real body mass for the i -th test instance (known from the data set).

For the ANN, 20 LOO cross-validations will be performed, since the randomness of several steps (weights initialization, the selection of the validation set). As a preliminary step within a leave-one out cross-validation process, the validation subset (see Sect. 3.1) will be extracted. The remaining subset will be used for evaluating the model performance using LOO, as described above. The MAE values reported for the 20 runs are averaged and a statistical analysis on the obtained results is performed.

The SVM is tested using a single LOO cross-validation, since there is no randomness in building the SVM regressor, thus the obtained results are always the same. There is randomness in the initial shuffling of the data, but experiments have shown that this did not influence the results in any significant manner. The random search for the hyperparameters optimization is run for 200 iterations.

4 Experimental Evaluation

In this section, an experimental evaluation of the proposed machine learning models (described in Sect. 3) is provided on several case studies derived from an open source archaeological data set. Original implementations were used for the ANN and the *SOM* employed in the data preprocessing step. The *scikit-learn* machine learning library was used for the SVM implementation [19].

4.1 Data Set and Case Studies

The data set considered in the experiments is an archaeological data set publicly available at [10]. The database at [10] was developed through a research project [9] and is a skeletal database composed of forensic cases to represent the ethnic diversity and demographic structure of the United States population [10]. The database contains 1009 human skeletons and 86 bones measurements representing: postcranial measurements (length, diameter, breadth, circumference, and left and right, where appropriate) of the clavicle, scapula, humerus, radius, ulna, sacrum, innominate, femur, tibia, fibula, and calcaneus [10]. Only the instances for which the (*forensic* and/or *cadaver*) *body mass* was available were extracted from this database.

An analysis of the bioarchaeological literature revealed 10 measurements which are used for human body mass estimation and have a good correlation with the body size (stature and/or body mass) [2]: (i) *femoral head diameter* (ii) *iliac breadth*, *femur bicondrial length* (iii) *clavicle*, *humerus*, *radius*, *ulna*, *femur*, *tibia* and *fibula*. From these measurements, the *femural* measurements seem to produce the most accurate body mass estimations [2].

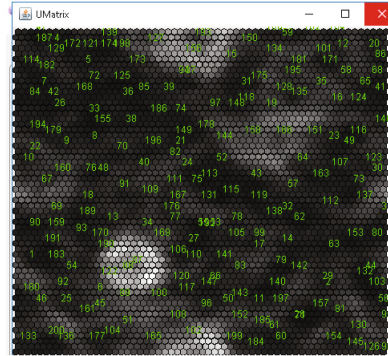
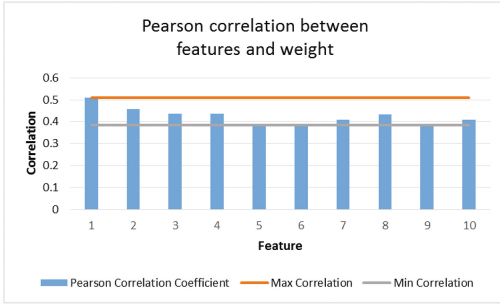
In case of two measurements for any of these bones (left and right), their mean was used. If only one measurement existed in the database, it was used as is. Instances containing missing bone measurements were not considered in any of the case studies.

The following case studies are considered, with the aim of analyzing the relevance of the previously mentioned measurements for the problem of body mass estimation.

- The **first** case study consists of 200 instances characterized by 3 measurements - (i) and (ii).
- The **second** case study consists of 146 instances characterized by 8 measurements - (i) and (iii).
- The **third** case study consists of 135 instances characterized by all 10 measurements - (i), (ii) and (iii).

4.2 Data Preprocessing

As mentioned in Sect. 3, before building the ANN and SVM models, the training data is preprocessed. After normalizing the data using the *Min-Max* normalization method, the Pearson correlation coefficients between the features



(a) Pearson correlation coefficients between the features and the target weights.

(b) The U-Matrix visualization.

Fig. 1. Pearson correlation and U-Matrix visualization.

(measurements) and the target weights are computed. Figure 1(a) illustrates the correlations between the 10 measurements and the target output on the third data set (case study).

Figure 1(a) shows that the first three features have the highest correlation with the body mass. The *femoral head diameter* has the maximum correlation with the body mass (0.5086), while the length of the *tibia* has a correlation of only 0.3858 with the body mass. Analysing the correlations from Fig. 1(a) it is expected that the best performance of the ANN and SVM models is on the first case study (using only the first three measurements).

In order to determine possible outliers within the training data, a self organizing map is trained on the first data set (consisting of 200 instances). The U-Matrix visualization of the trained SOM is depicted in Fig. 1(b). On the map one can see four small regions (with boundaries outlined in white). The instances from these regions may be viewed as possible outliers in the data set, since they are isolated from the other instances. This way, 8 instances can be visually identified as possible outliers. These instances are removed from the training data sets.

4.3 Results

In this section, an experimental evaluation of the proposed application of ANN and SVM machine learning regressors (described in Sect. 3) is provided on the case studies described in Sect. 4.1.

For the ANN parameters, a learning rate of 0.3 momentum values of 0.2 were used. These were chosen by using a classic grid search over several values lower than 1, and choosing the values which consistently provided the best results.

For the random search of SVM hyperparameters, the uniform probability distribution was used. The intervals or sets from which each hyperparameter is drawn are as follows:

- **kernel**: from the set $\{rbf, linear, sigmoid, poly\}$ (i.e. RBF, linear, sigmoid, polynomial kernels).
- ε : from $[0, 1)$.
- γ : from $[0, 1)$.
- b : from $[0, 1)$.
- **Polynomial degree** for the polynomial kernel: from the set $\{1, 2, 3, 4, 5\}$.
- C : from $[0, 1)$.

Note that not all hyperparameters apply to every kernel choice.

For the case studies considered for evaluation (Sect. 4.1), after preprocessing the training data set as indicated in Sect. 4.2, the ANN and SVM regressors are built through training (see Sect. 3.1).

For evaluating the performance of the trained ANN and SVM, 20 runs of Leave One Out (LOO) cross-validation were used for the ANN model and a single run for the SVM model, because the latter does not use any random numbers, so its results will always be the same.

ANN Results. The MAE values along with the minimum, maximum, mean value and standard deviation, obtained using the ANN, for each case study performed, are given in Table 1. In this table, the MAE values are given in kilograms.

Table 1. Results obtained using the ANN, considering all 20 LOO cross-validations.

Case study	Mean (kg)	Min (kg)	Max (kg)	Stdev (kg)
First case study	8.766	7.751	10.195	0.682
Second case study	9.593	8.301	11.259	0.77
Third case study	8.945	7.63	10.338	0.649

Table 1 indicates, as expected, that the best results were obtained for the first case study, when only the first 3 measurements are used for characterizing the skeletal remains. Figure 2 depicts the values for the MAE measure obtained during the 20 runs of the LOO cross-validation process applied on the first case study. The average of these values, as well as their standard deviations are also indicated. The small values for the standard deviation reveal a good performance of the proposed ANN model application.

SVM Results. The results obtained using the SVM are presented in Table 2. The best values used for the hyperparameters (including the used kernel function) are depicted in the last column of the table.

As shown in Table 2, the SVM obtained a performance similar to the ANN: the best performance on the first case study and the worst performance on the second case study.

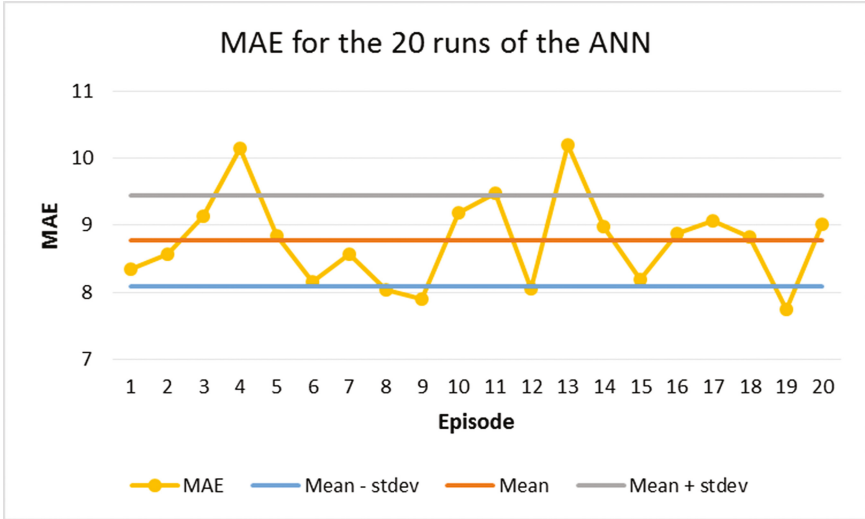


Fig. 2. MAEs for the ANN on the first case study.

Table 2. Results obtained using the SVM.

Case study	MAE (kg)	Example best hyperparameters
First case study	8.4897	kernel: <i>linear</i>
		ϵ : 0.027578797176389447
		C: 0.641027338498186
Second case study	9.4918	kernel: <i>linear</i>
		ϵ : 0.09576656837738184
		C: 0.24862712124238084
Third case study	8.6843	kernel: <i>polynomial</i>
		ϵ : 0.04798752786733451
		C: 0.1465703832031966
		degree: 2
		γ : 0.7681948068876795
b: 0.42975272058542213		

5 Discussion and Comparison to Related Work

In this section, an analysis is provided for the approaches introduced in Sect. 3 for body mass estimation from bone measurements. Then, a comparison with similar approaches from the literature is conducted.

As shown in the experimental part of the paper, both the ANN and SVM models have provided about the same performances for the body mass estimation problem. The SVM slightly outperformed the ANN, with at most 0.28 MAE.

The experimental values obtained for the (average) MAE considering all case studies are summarized in Table 3. The last column from Table 3 contains the MAE value (averaged over all LOO cross-validations). The best values, for each case study, are highlighted.

Table 3. MAE values obtained using the ANN and the SVM models on the considered case studies

Case study	Model	MAE (kg)
First case study	ANN	8.766
	SVM	8.4897
Second case study	ANN	9.593
	SVM	9.4918
Third case study	ANN	8.945
	SVM	8.6843

Table 3 shows that the best machine learning-based regressor for estimating the human body mass from skeletal remains is the SVM, when only 3 measurements (*femoral head diameter*, *iliac breadth* and *femur bicondrial length*) are used for the skeletal elements. This is to be expected, since these three measurements showed the highest correlation with the target body mass (Fig. 1(a)).

Analysing the results from Table 3 it can also be seen that the worst results were obtained, for both ANN and SVM, on the second case study. Therefore, the *iliac breadth* and *femur bicondrial length* are also important in estimating the body mass and the measurements for the *clavicle*, *humerus*, *radius*, *ulna*, *femur*, *tibia* and *fibula* do not improve the body mass estimation results.

It is worth mentioning that the outliers removal step performed during the *data preprocessing* step has increased the performance of the ML regressor. Table 4 illustrates the effect of removing the outliers from the training data of the ANN. A significant reduction for the MAE value was obtained for the third case study.

5.1 Comparison to Related Work

In the following, a brief review of the recent human body mass estimation literature is given, with the aim to compare the used ML regressors to the existing related work. As far as the authors are aware, there are no existing machine learning-based models (like the ANN and SVM models applied in this paper) for the problem of body mass estimation from skeletal remains.

A comparison between several body mass estimation methods was conducted by Auerbach and Ruff in [2] (2004). The authors proposed to test some existing methods on a great variety of subjects. They used skeletal remains of 1173 adult skeletons of different origins and body sizes, both males and females. Three

Table 4. Comparative MAE values - with and without outliers removal using the ANN.

Case study	MAE reduction (%)	Outliers removal	MAE (kg)
First case study	19.39	Yes	8.766
		No	10.874
Second case study	19.22	Yes	9.593
		No	11.875
Third case study	25.68	Yes	8.945
		No	12.036

femoral head-based regression formulas were tested and compared on the considered skeletal sample: Ruff et al. [22] (1991), McHenry [16] (1991) and Grine et al. [8] (1995). The authors concluded that for a very small body size range (Pygmoids), the formula of McHenry (1992) can provide a good body mass estimation. For very large body sizes, the formula of Grine et al. (1995) should be used, whereas for the other samples the formula of Ruff (1991), or the average of the three techniques would be the best approach.

Ruff et al. provided in [21] (2012) new body mass estimation equations that are generally applicable to European Holocene adult skeletal samples. Body mass estimation equations were based on femoral head breadth. 1145 skeletal specimens were obtained from European museum collections, from time periods ranging from Mesolithic to the 20th century [21]. On these data sets, the regression formulas introduced in [21] provided better results than the previous formulas from Ruff et al. [22] (1991), McHenry [16] (1991) and Grine et al. [8] (1995).

The data sets used in the previously mentioned papers are not publicly available, that is why an exact comparison of the approaches introduced in this paper to the previously mentioned approaches cannot be performed. Since the authors have not been able to find experiments in the literature related to body mass estimation using the same data set as in this paper, the following comparison to related work was conducted. The regression formulas introduced in Ruff et al. [8,16,21,22] were applied on the data sets used in this paper (all three case studies) and the obtained MAE values were compared with the ones provided by the ANN and SVM regression models. The results of the comparison are given in Table 5. In this table, 95% confidence intervals [5] were used for the obtained results. The comparison is graphically depicted, for all three case studies, in Fig. 3. In this figure, the first two dashed bars correspond to the ANN and SVM models used in this paper, while the other bars correspond to the above mentioned four approaches from the literature. For the ANN, the 95% confidence intervals of the average MAE are also illustrated.

In Table 5 and Fig. 3 it can be observed that the MAE values obtained by the ANN and SVM methods are smaller than those obtained using regression formulas from the literature. One can notice that for the ANN even the upper limit of the 95% confidence interval of the mean MAE is below the MAE from the

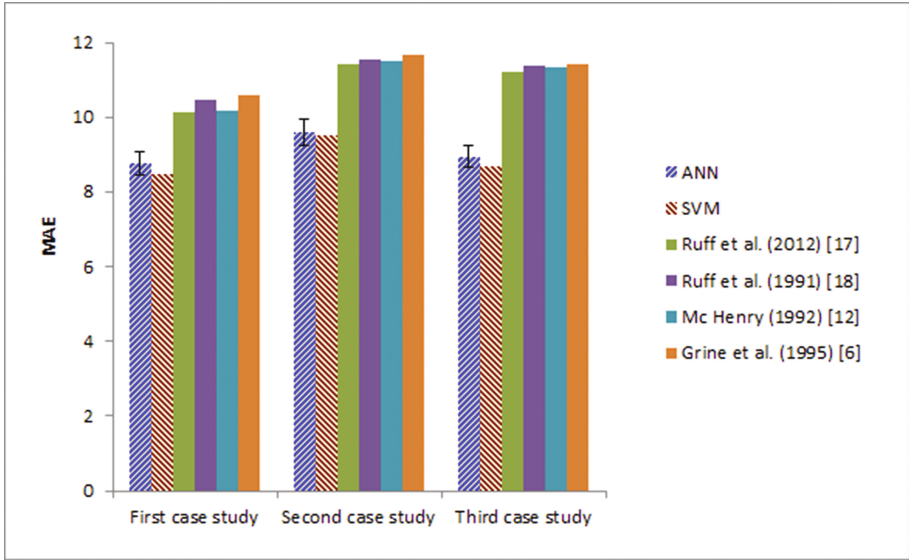


Fig. 3. Comparison to related work

literature. This is somehow predictable because the previously stated regression formulas only use one measurement, the *femoral head anterior-posterior breadth*, while the proposed approaches are using multiple measurements. This is the best measure of the performance of the machine learning approach, since the experiments are performed on the same data sets. Note that the ANN and SVM models provided better results, even if the evaluation of the machine learning based models was obtained using multiple cross-validation runs (in order to avoid overfitting), whereas the formulas from the literature were obtained using the entire data set. Another major advantage of the proposed approaches with respect to the literature is that estimations are made without knowing the *sex* characteristics, which is mainly used in the existing literature.

Table 5 shows that, considering both the approaches from the literature as well as the ANN and SVN, the best performance was obtained on the first case study.

It has to be mentioned that most of the researchers from the bioarchaeological fields develop regression formulas for body mass estimation based on a data set which is also used for testing the developed formulas, without using a testing set independent from the training data and without using any type of cross-validation. This may lead to overfitting, as for the regression formulae from the literature which provided good performances on the data they were trained on (about 4–5 MAE), but when applied on an unseen test data (our case studies) they provide large MAE values (see Sect. 4.3). It is likely that the methods from the body mass estimation literature would provide larger errors under the testing methodology used for the proposed machine learning applications.

Table 5. Comparison between the machine learning approaches and similar related work. 95% confidence intervals are used for the ANN results.

Approach	First case study MAE (kg)	Second case study MAE (kg)	Third case study MAE (kg)
ANN Implementation	8.766 ± 0.299	9.593 ± 0.338	8.9452 ± 0.284
SVM Implementation	8.4897	9.4918	8.6843
Ruff et al. (2012) [21]	10.115	11.43	11.202
Ruff et al. (1991) [22]	10.476	11.556	11.359
Mc Henry (1992) [16]	10.176	11.514	11.324
Grine et al. (1995) [8]	10.609	11.656	11.431

The main advantage of the machine learning approaches proposed in this paper over the simple mathematical ones is that the machine learning models can be retrained on new data, and they are able to generalize well if specific methods for avoiding overfitting are used, like the ones described in Sect. 3.1. Moreover, it can be computationally costly to develop mathematical formulas on new data sets, which will likely not generalize well anyway. Instead, a machine learning regressor (like ANN or SVM) can easily learn from new data sets, resulting in learning models that are able to perform well on unseen data having the same features as the training one.

6 Conclusions and Future Work

Two supervised learning regression models were applied for estimating the body mass from human skeletal remains based on bone measurements. The idea to apply machine learning for the body size estimation problem is a novel one. Several case studies were conducted on a publicly available data set from the bioarchaeological literature. The obtained results outperformed classical statistical methods for body size estimation, which makes the proposed approaches useful for the field of bioarchaeology.

Further work will be done in order to improve the data preprocessing step of learning, by automatically detecting and removing the outliers from the training data set. Since the archaeological data sets usually contain missing measurements, methods for dealing with these missing values will be investigated. Applying other machine learning-based regression models for the body mass estimation problem, like *radial basis function networks* and *k-nearest neighbors*, is also a direction of interest for the authors.

Acknowledgment. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-RU-TE-2014-4-0082.

References

1. Arora, P., Srivastava, S., Singhal, S.: Analysis of gait flow image and gait Gaussian image using extension neural network for gait recognition. *Int. J. Rough Sets Data Anal.* **3**(2), 45–64 (2016)
2. Auerbach, B., Ruff, C.: Human body mass estimation: a comparison of “morphometric” and “mechanical” methods. *Am. J. Phys. Anthropol.* **125**(4), 331–342 (2004)
3. Ruff, R.B.: Body mass prediction from skeletal frame size in elite athletes. *Am. J. Phys. Anthropol.* **113**(4), 507–517 (2000)
4. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012)
5. Brown, L.D., Cai, T.T., Dasgupta, A.: Interval estimation for a binomial proportion. *Statistical Science* **16**, 101–133 (2001)
6. Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A.S., Balas, V.E.: Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. *Neural Comput. Appl.* 1–12 (2016)
7. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
8. Grine, F., Jungers, W., Tobias, P., Pearson, O.: Fossil homo femur from berg aukas, northern namibia. *Am. J. Phys. Anthropol.* **26**, 67–78 (1995)
9. Jantz, R.J., Moore-Jansen, P.H.: Database for Forensic Anthropology in the United States, 1962–1991 (ICPSRversion) (1998). University of Tennessee, Dept. of Anthropology, Knoxville, TN [producer] (1998). Inter-university Consortium for Political and Social Research, Ann Arbor, MI (2000)
10. Jantz, R.J., Moore-Jansen, P.H.: Database for Forensic Anthropology in the United States, 1962–1991 (ICPSR). University of Tennessee, Department of Anthropology (2000). <http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/2581>
11. Kaski, S., Kohonen, T.: Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In: *Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pp. 498–507. World Scientific (1996)
12. Kriti Virmani, J., Dey, N., Kumar, V.: PCA-PNN and PCA-SVM based CAD systems for breast density classification. In: Hassanien, A.E., Grosan, C., Fahmy Tolba, M. (eds.) *Applications of Intelligent Optimization in Biology and Medicine*, pp. 159–180. Springer International Publishing, Cham (2016)
13. Kröse, B., van der Smagt, P.: An introduction to neural networks. The University of Amsterdam (1996)
14. Linstrom, K.R., Boye, A.J.: A neural network prediction model for a psychiatric application. In: *International Conference on Computational Intelligence and Multimedia Applications*, pp. 36–40 (2005)
15. Maji, P., Chatterjee, S., Chakraborty, S., Kausar, N., Samanta, S., Dey, N.: Effect of Euler number as a feature in gender recognition system from offline handwritten signature using neural networks. In: *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1869–1873, March 2015
16. McHenry, H.: Body size and proportions in early hominids. *Am. J. Phys. Anthropol.* **87**, 407–431 (1992)
17. Mitchell, T.M.: *Machine Learning*. McGraw-Hill Inc., New York (1997)

18. Moore, M.K.: Body mass estimation from the human skeleton. Ph.D. thesis, The University of Tennessee, Knoxville (2008)
19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
20. Rojas, R.: *Neural Networks: A Systematic Introduction*. Springer, Heidelberg (1996)
21. Ruff, C., Holt, B., Niskanen, M., Sladk, V., Berner, M., Garofalo, E., Garvin, H., Hora, M., Maijanen, H., Niinimki, S., Salo, K., Schuplerov, E., Tompkins, D.: Stature and body mass estimation from skeletal remains in the European holocene. *Am. J. Phys. Anthropol.* **182**(4), 601–617 (2012)
22. Ruff, C., Scott, W., Liu, A.C.: Articular and diaphyseal remodeling of the proximal femur with changes in body mass in adults. *Am. J. Phys. Anthropol.* **86**, 397–413 (1991)
23. Rumelhart, D.E., Hinton, G.E., Williams, R.: Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (eds.) *Parallel Data Processing*, vol. 1, pp. 318–362. The MIT Press, Cambridge (1986)
24. Russell, S., Norvig, P.: *Artificial Intelligence - A Modern Approach*. Prentice Hall International Series in Artificial Intelligence. Prentice Hall, Upper Saddle River (2003)
25. Samanta, S., Ahmed, S.S., Salem, M.A.M.M., Nath, S.S., Dey, N., Chowdhury, S.S.: Haralick features based automated glaucoma classification using back propagation neural network. In: Satapathy, C.S., Biswal, N.B., Udgata, K.S., Mandal, J. (eds.) *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, vol. 1, pp. 351–358. Springer International Publishing, Cham (2015)
26. Sirshendu, H., Tanmay, B., Nilanjan, D., Hassanien, A.E.: Encryption using speeded up robust features extraction technique and artificial neural network. In: *Innovations, and Applications, Image Feature Detectors Foundations* (2015)
27. Skowronski, M.D., Harris, J.G.: Automatic speech recognition using a predictive echo state network classifier. *Neural Netw.* **20**(3), 414–423 (2007)
28. Smola, A., Schlkopf, B.: A tutorial on support vector regression. *Stat. Comput.* **14**(3), 199–222 (2004)
29. Somervuo, P., Kohonen, T.: Self-organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.* **10**, 151–159 (1999)
30. Tufféry, S.: *Data Mining and Statistics for Decision Making*. Wiley, New York (2011)
31. Wahba, G., Lin, Y., Zhang, H.: GACV for support vector machines, or, another way to look at margin-like quantities. In: *Advances in Large Margin Classifiers*, pp. 297–309 (2000)