

Master Degree in Statistics for Data Science
Academic Year 2019-2020

Master Thesis

Classification models for landmarks configurations. Geometric morphometrics applied to acute malnutrition diagnosis.

Ricardo Hortelano Sánchez

Laura Medialdea

Ana Arribas Gil

Madrid, September 2020

AVOID PLAGIARISM

The University uses the **Turnitin Feedback Studio** program within the Aula Global for the delivery of student work. This program compares the originality of the work delivered by each student with millions of electronic resources and detects those parts of the text that are copied and pasted. Plagiarizing in a TFM is considered a **Serious Misconduct**, and may result in permanent expulsion from the University.



[Include this code in case you want your Master Thesis published in Open Access University Repository]

This work is licensed under Creative Commons **Attribution – Non Commercial – Non Derivatives**

SUMMARY

Geometric morphometric techniques are a collection of tools which allows visualization and quantification of shape changes among biological forms. 2D landmark-based configurations extracted from limbs and trunk images of children aged 6-59 months have been recently designed in order to study morphological variation related to ontogenetic and populational factors. Furthermore, such new templates registering children shape variation have been used to analyse acute malnutrition shape pattern, together with optimal nutritional condition's one. A linear discriminant analysis was performed to build a classification method for assessing whether target children could be malnourished or not (in press). This project is called SAM Photo Diagnosis App Project® and has been developed by Action Against Hunger since October 2015. All methods developed by SAM Photo project were programmed in a smartphone App with the aim of providing communities worldwide an offline easy to use screening tool for acute malnutrition. This Master's thesis reviews the current classification and methodological techniques used in the field, presents new ones and performs an exhaustive and quantitative analysis of the methodologies applied to children malnutrition detection from a body picture improving the state-of-the-art classification results up to 10% and detecting age and sex groups especially sensitive to misdiagnoses.

Keywords:

Geometric Morphometrics, Malnutrition Classification, SAM Photo Diagnosis app, Machine learning, Methodological Review

CONTENTS

1. INTRODUCTION	1
1.1. Objectives	2
1.2. Undernutrition and diagnosis	3
1.3. Geometric Morphometrics	4
1.4. Classification applied to Geometric Morphometrics	5
1.5. SAM Photo Diagnosis App project	6
2. METHODOLOGY	7
2.1. State-of-the-Art for landmark based classification methods	7
2.1.1. Landmark location definition	7
2.1.2. Multivariate analysis of landmarks information	8
2.1.3. Question Formulation	11
2.2. Proposed Methodology	11
2.2.1. Multivariate analysis of landmarks information	11
2.2.2. Visual Analysis of methods	16
2.3. Methodology Combination	16
2.4. Metrics	18
3. RESULTS	20
3.1. State-of-the-Art Results	20
3.2. Current Results	21
3.3. Comparison	22
3.3.1. Procrustes with/without Colinearity Trim	22
3.3.2. Procrustes with Allometric Reduction with/without Colinearity Trim . . .	23
3.3.3. Procrustes with PCA	25
3.3.4. Distances	25
3.3.5. Distances with Colinearity Trim	27
3.3.6. Distances with PCA	29
3.3.7. Procrustes with Distances with/without Colinearity Trim	30
3.3.8. Procrustes with Distances with PCA	31

4. CONCLUSIONS	36
4.1. Discussion	36
4.2. Future Work	37

LIST OF FIGURES

2.1	Example landmarks of the left arm of one of the children.	8
2.2	Superimposing that performs Procrustes allowing to compare visually two different objects, in this case on the wing of an insect. Source: wikipedia/Procrustes_analysis	9
2.3	Result of applying Procrustes to the dataset of study. In gray, procrustes coordinates \mathcal{P} of each of the children, in black, centroids $\bar{\mathcal{P}}$ of each coordinate.	9
2.4	Coordinates covariances before and after applying Procrustes. a) Covariance matrix Σ before Procrustes, b) Covariance matrix Σ after Procrustes, c) Covariance matrix Σ^* after Procrustes and Colinearity feature trim. In red colinears close to -1 values, in blue colinears close to +1 values and in white colinears close to 0 values.	12
2.5	Sample representation of the distances in matrix $D(\mathcal{L})$. The numerical value of each matrix element corresponds to the length of the associated black segment.	13
2.6	Comparison of different dimensional reduction techniques on the same dataset. PCA, UMAP and Variational Autoencoder. Source: Github lmccinnes/umap issue #113.	17
2.7	The graph that defines all the possible methods studied.	18
3.1	The four distances that contain the most information for all the methods. The distance between Landmark 15 and 16 is the one that contains the most information for the classification methods. The second most informative distance is the 13 and 14. It corresponds to the distance that is measured in the malnutrition diagnosis by MUAC.	22
3.2	Comparison of mean accuracy over all experiments between the data processing method and the classification method.	23
3.3	UMAP transformations of: a) Procrustes; b) Procrustes + Colinearity Trim; c) Procrustes + Allo. Shape Reduction; d) Procrustes + Allo. Shape Reduction + Colinearity Trim; e) Procrustes + PCA. Shape Reduction + Colinearity Trim. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.	26

3.4 UMAP transformations of: a) Distances; b) Distances + Colinearity Trim; c) Distances + PCA. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.	28
3.5 UMAP transformations of: a) Procrustes + Distances; b) Procrustes + Distances + Colinearity Trim; c) Procrustes + Distances + PCA. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.	32
4.1 Performance comparison of state of the art methods according to age group and sex.	
4.2 Performance comparison of thesis methods according to age group and sex.	

LIST OF TABLES

1.1	World Health Organization growth standards for malnutrition.	4
3.1	Best and worst two models for state-of-art methodology.	20
3.2	Top 2 models by experiment.	22
3.3	Best and worst two methods for Procrustes methodology.	24
3.4	Best and worst two methods for Procrustes + Colinearity Trim methodology.	25
3.5	Best and worst two methods for Procrustes + Allometric reduction methodology.	27
3.6	Best and worst two methods for Procrustes + Allometric reduction + Colinearity Trim methodology.	29
3.7	Best and worst two methods for Procrustes + PCA methodology.	30
3.8	Best and worst two methods for Distances methodology.	31
3.9	Best and worst two methods for Distances + Colinearity Trim methodology.	33
3.10	Best and worst two methods for Distances + PCA methodology.	34
3.11	Best and worst two methods for Procrustes + Distances methodology. . .	34
3.12	Best and worst two methods for Procrustes + Distances + Colinearity Trim methodology.	35
3.13	Best and worst two methods for Procrustes + Distances + PCA methodology.	35

1. INTRODUCTION

Malnutrition refers to a human condition caused by having a diet in which one or multiples nutrients are unbalanced with the normal quantities, causing health problems or even dead. Malnutrition can refer to excessive or deficient intakes of nutrients. An excessive intake could yield to overweight and obesity problems causing diet-related noncommunicable diseases (such as heart disease, stroke, diabetes and some cancers). Deficient or unbalanced intakes could yield to undernutrition health problems, classified as wasting, stunting, underweight, and micronutrient deficiencies by the World Health Organization (WHO).

Based on WHO definition¹:

"Wasting condition is defined as low weight-for-height. It often indicates recent and severe weight loss, although it can also persist for a long time. It usually occurs when a person has not had food of adequate quality and quantity and/or they have had frequent or prolonged illnesses. Wasting in children is associated with a higher risk of death if not treated properly. Stunting is defined as low height-for-age. It is the result of chronic or recurrent undernutrition, usually associated with poverty, poor maternal health and nutrition, frequent illness and/or inappropriate feeding and care in early life. Stunting prevents children from reaching their physical and cognitive potential. Underweight is defined as low weight-for-age. A child who is underweight may be stunted, wasted or both."

This condition is critical for children under five since the lack of these nutrients are essential for the body to correctly produce essential substances to the grow develop such as enzymes or hormones. Also, malnutrition increases health care costs, reduces productivity and slows economic growth. Poverty amplifies the risk of malnutrition so poor people are more likely to be affected by this condition. Women, children and adolescents are at the highest risk groups.

Based on WHO data² more than 60 millions of children worldwide are wasted or severely wasted, mainly in low- and middle-income countries. In these same countries, rates of obesity in children are rising. Around 45% of deaths among children under the age of 5 years are linked to undernutrition. Combating malnutrition in all its forms is one of the greatest global health challenges and multiples organizations worldwide exists to act against it.

Current diagnosis methodology is based on anthropometric measurements that are not always applicable on some context or situations. Action Against Hunger is currently developing a new project that aims to mitigate and reduce this problematic. The SAM Photo Diagnosis App project tries to prove that geometric morphometric techniques are

¹<https://www.who.int/health-topics/malnutrition>

²<https://www.who.int/news-room/fact-sheets/detail/malnutrition>

useful for such task. To do this, the project has images of children at different levels of nutrition and specific landmarks have been detected in the images that are used to classify future undiagnosed children.

In this context, wasting is also named acute malnutrition and for this project we work with four different states of nutrition: severe acute malnutrition (SAM), moderate acute malnutrition (MAM), risk of malnutrition (RIS) and normal nutritional status (NOR).

Geometric morphometric techniques are based on the study of the variation of biological form between organisms by identifying similar landmarks. A first stage of the SAM photo project has already been completed and yielded positive results to correctly diagnose two extreme conditions (NOR and SAM). On next stages of the project, the need to know the effectiveness of the model to identify intermediate stages of the disease (MAM and RIS) arises.

In order to discriminate between different populations, the SAM Photo diagnosis project currently uses state-of-the-art techniques in the field to pre-process the data and linear discriminant methods to classify as LDA. This thesis seeks to improve the methodology used in the SAM project through the analysis of currently used methods and the proposal of new methods. For this, a quantitative and qualitative analysis of the different possible combinations is carried out. The procedure followed in the thesis can be extended to other biology fields with few to none changes and the code in R can be freely used and reviewed on GitHub³.

1.1. Objectives

The objectives of the current thesis are the following:

- Carry out a bibliographic review about the techniques currently used to diagnose malnutrition using geometric morphometrics in order to understand their operation and limitations.
- Propose new data preparation techniques and evaluate their performance using a binary classification problem framework to find possible improvements to the methodology.
- Perform an exhaustive and quantitative analysis of the possible combinations of data preparation and binary classification techniques in order to find the best possible combination for the malnutrition diagnosis problem using landmark data extracted from left arm photos of children.
- Evaluate the performance of the classification algorithms to detect systematic advantages for the type of data studied.

³https://github.com/RicardoHS/geomorph_malnutrition_classification_models

- Evaluate the performance of current and new techniques, discriminating by age group and sex to detect groups especially vulnerable to being misdiagnosed.

1.2. Undernutrition and diagnosis

The exact definition of what is malnutrition and what types of malnutrition exist varies depending on the institution that defines the disease. The diagnosis can also vary depending on the method used and in fact its use varies depending on the country of origin of the patient [14]. For this thesis it is used the WHO growth standards (Table 1.1) and Action Against Hunger definitions for wasting state:

- **Normal nutritional status (NOR):** The child is in a correct state of nutrition, according to his age, size and sex.
- **Risk of Malnutrition (RIS):** The child is at risk of entering an unhealthy status. Steps should be taken to lower the risk and start monitoring.
- **Moderate Acute Malnutrition (MAM):** Also known as wasting, the child is in an initial state of malnutrition and should be treated to correct the condition and avoid major problems in the short and long term.
- **Severe Acute Malnutrition (SAM):** The worst case. It is defined by a very low weight for height by visible severe wasting, or by the presence of nutritional oedema. Treatment must be immediate to avoid the death of the child. This is responsible, directly or indirectly, for 35% of deaths among children under five.

The two methods used to diagnose malnutrition are MUAC (mid upper arm circumference) and WHZ (Weight-for-Height z-score). The first is obtained from the direct measurement of the child's arm. The second is obtained by measuring the child's weight and height and comparing it against predefined growth standards⁴. The WHZ for a child is the number of standard deviations of its standardised measure below or above the reference median value for the age, height and sex.

The main reasons for the WHO to use the -3 standard deviation classification in the WHZ scale to diagnose SAM are: 1) Children below this cut-off have a highly elevated risk of death compared to those who are above; 2) These children have a higher weight gain when receiving a therapeutic diet compared to other diets, which results in faster recovery; 3) In a well-nourished population there are virtually no children below -3 SD (<1%); 4) There are no known risks or negative effects associated with therapeutic feeding of these children applying recommended protocols and appropriate therapeutic foods.

Quoting "A Joint Statement by the World Health Organization and the United Nations Children's Fund":⁵

⁴https://www.who.int/childgrowth/standards/weight_for_length_height/en/

⁵https://apps.who.int/iris/bitstream/handle/10665/44129/9789241598163_eng.pdf

Diagnosis	WHZ	MUAC
NOR	(1, -1]	$135\text{mm} \leq X$
RIS	(-1,-2]	$125\text{mm} \leq X < 135\text{mm}$
MAM	(-2, -3]	$115\text{mm} \leq X < 125\text{mm}$
SAM	(-3, ∞)	$X < 115\text{mm}$

Table 1.1. WORLD HEALTH ORGANIZATION GROWTH STANDARDS FOR MALNUTRITION.

"Regarding MUAC, the WHO standards for mid-upper arm circumference-for-age show that in a well-nourished population there are very few children aged 6 to 60 months with a MUAC less than 115 mm. Children with a MUAC less than 115mm have a very high risk of death compared to those above it. Therefore, it is recommended to increase the cut-off point from 110 to 115 mm to define SAM with MUAC."

The treatment of severe malnutrition is based on the administration of Ready-to-use Therapeutic Food (RUTF). The RUTF is a ready-to-use paste that does not need to be mixed with water (which avoids accidental contamination with bacteria). The mixture is based on peanut butter mixed with dried skimmed milk and vitamins and minerals which can be consumed directly by the child and ensure a rapid weight gain. It can be stored for three to four months without refrigeration, even at tropical temperatures.

1.3. Geometric Morphometrics

In the field of study of shape, there has always been a historical tradition in terms of scientific techniques, mainly used in biology studies. The objective of the field is to correctly characterize associations, causes and effects regarding the study of form. Nowadays, the field is based on four principles defined by Bookstein [29] in 1991; landmark location definition, multivariate analysis of landmarks information, question formulation and statistical result analysis. It is commonly used to answer evolutionary questions. For instance, it's common to study the bones or organs of a living being and compare it against populations of different races or ages [28] using these four principles based methodology.

A fundamental change occurred in the 1980's when the compilation of coordinates (or landmarks) of key aspects of the object to study began [1]. This collection of Cartesian landmarks along with other approaches that provide a mathematical description of biological forms based on size and shape is called geometric morphometrics. The change allowed to use statistical multivariate analysis and more sophisticated visualizations that allowed the field to move forward along with the statistics development [23] and it continues like this until today.

In the more recent years, another change started with the inclusion of more

sophisticated statistical tools and machine learning methods [8]. This last change fostered the development of more sophisticated analysis and algorithms that allowed to further expand the boundaries of the field. Nevertheless this field of study is currently under development.

More recently, these new applied techniques have been used to develop automatic tools that discriminate between different types of malnutrition in children using only the information about their body shape [21]. In order to do so, the correctly collection of data corresponding to specific points on the anatomy of the organism is a key aspect. By using this technique, it's possible to characterize the morphological differences (directional and magnitude) of landmarks between populations and analyze the influences of key aspects that change the shape. For instance, some recent studies have analysed the influence of virus [2], degenerative disorders [17] or gene expressions [13] in these shape changes.

Geometric morphometrics is an advance compared to the traditional morphometric methods, in which landmarks were not analyzed but direct measurements were made on the body of study. According to some authors [12] [24] [26], traditional morphometrics failed to find differences in shape and to isolate variables that depended on it. However, and as will be seen later, a combination of modern morphometrics with some traditional aspects entails a substantial information improvement in some cases, at least in statistical and information theory [27] terms.

1.4. Classification applied to Geometric Morphometrics

The use of classification methods in geometric morphometrics problems has been dominated almost from its modern form by the use of Linear Discriminant Analysis (LDA). Due to the characteristics that LDA offers (mainly a good balance between accuracy and interpretability of the models) this method has been widely used since it is aligned with the four points described by Bookstein [29].

More recently, other algorithms, such as machine learning algorithms or deep learning algorithms, have been used in some modern research, some examples are [8], [30], [22], [7]. However, this methodology has not been widely adopted by researchers despite its good results [25]. Even more, a detailed reading of the use of these more modern algorithms reveals that their use is not the most appropriate, since some problems arise to the practitioners when choosing the algorithm, errors are made on the data preparing stage or even some researchers glorify the methods [31] [3] when there are numerous studies that reveal all the problems, hidden biases and limitations that they entail [15] [18] [5] [10] [6]. So although there is a change in trend regarding the use of these methods, their full potential is not being used yet. This thesis works in this direction by showing an analysis of the methodologies and performance of the different classification algorithms in order to show a guide that clarifies its use to practitioners in the case of malnutrition diagnosis. Although the methodology followed can be replicated

in other areas.

1.5. SAM Photo Diagnosis App project

The SAM Photo Diagnosis project [20] arises from the observed need to reduce the intra- and inter-observer errors that currently occur when diagnosing malnutrition. As the diagnosis by WHZ and MUAC depends in part on the skill of the technician making the measurements, there is a derivative error that depends entirely on the person making these measurements. This problem can lead to erroneous diagnoses in addition to the costs derived from training technicians. In addition, there are remote regions where it is not possible to transport the medical equipment necessary to perform these measurements.

To solve these problems, the project is based on measurement using a photo through an Android mobile. In this way the error derived from the measurement through a technician is reduced and the diagnosis is extended wherever there could be a mobile phone.

The choice of Android is very important since its use in LMIC (low and middle income countries) has grown in recent years and is expected to continue to do so. In addition, its use is intended for community health workers who, for the most part, already have one at their disposal.

In this way, the diagnosis would become a non-invasive and inexpensive method, since it only needs the investment of the cost of a mobile phone. The download of the application is free and the use of an active internet connection is not required to use it.

Thus, SAM Photo Diagnosis App project brings together everything previously presented and develops a prototype app to classify SAM, based on data collected from Senegalese and Spanish children. It is sponsored by Action Against Hunger and is currently in development. The project has already managed to find morphometric differences between NOR and SAM based on landmark, age and sex information and aims to find these same differences in four nutritional stages NOR, RIS, MAM, SAM. This work aims to contribute knowledge to SAM Photo Diagnosis App project regarding this last point, studying the behaviour of different classification methods for the four nutritional groups.

2. METHODOLOGY

The methodology used in this work in order to classify different types of malnutrition is based on the previous work of Medialdea et al. [21]. Her work has been the base where this thesis has started, using similar data and the landmark template previously defined. Later, other methodologies used in recent literature have been reviewed and progress has been made based on these, which led to a deep analysis of these methods to detect possible improvements or problems. It is important to note that each method and algorithm used has implications that can limit its combination with other methods. To define and clarify all the possible combinations studied, an acyclic directed graph has been defined and is presented in section 2.3.

Next we present the methods that are currently being used in the field, later we present new proposals for methods that lead to better results and finally we present the design carried out to combine and analyze these methods in a quantitative way.

2.1. State-of-the-Art for landmark based classification methods

As previously mentioned, Bookstein [29] states that the field is based on four concepts:

2.1.1. Landmark location definition

Landmark location refers to two key aspects. Data gathering methodology and antropometric measurements (landmark collection) based on a previously defined template. We define the set of raw landmarks for an individual $\mathcal{L} = L_1, L_2, \dots, L_p$ where L_i is a pair of Cartesian coordinates (x_i, y_i) . Based on this notation, we have $2p$ number of variables regarding the landmark data.

Data

The data provided by Action Against Hunger consist of morphometric data from the left arm of 563 children between 6–59 months of Spanish or Senegalese nationality. In particular, the data correspond to 554 Senegalese children (142 MAM, 129 NOR, 143 RIS, 140 SAM) and 9 Spanish children (9 SAM). The dataset contains 279 boys and 284 girls. The criteria used to diagnose malnutrition vary depending on the child. In some cases MUAC has been used, in others WHZ and in others both at the same time.

These data also contain information regarding the age of the children. This age is grouped into five different classes following the same criteria as in Medialdea et al. [21]: Group 0) 6–12 months; Group 1) 13–24 months; Group 2) 25–36 months; Group 3) 37–48 months; and Group 4) 49–59 months.

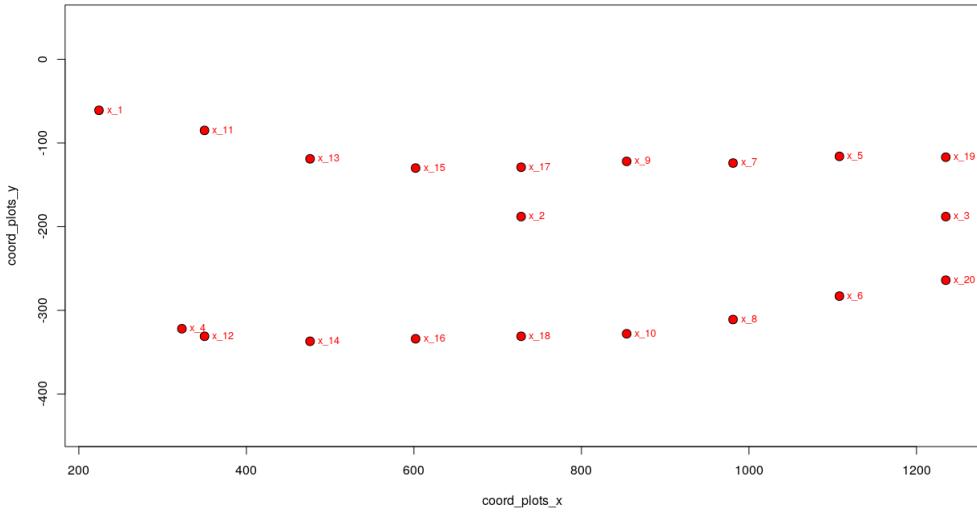


Fig. 2.1. Example landmarks of the left arm of one of the children.

Antropometric Measurements

Although in the Medialdea [21] study the landmarks collected refer to the entire body of the children, in the present work only those referring to the left arm have been used, thus limiting the scope. The choice to only use measurements of the left arm comes from previous work in which the best classification rates were obtained by only using these landmarks. In addition, the limitation in this aspect is aligned with the objectives of this thesis, by reducing the complexity in the data when studying the presented techniques. A total of $p = 20$ landmarks is used as can be seen in the figure 2.1.

2.1.2. Multivariate analysis of landmarks information

In order to convert the landmarks collected into something that can be used by statistical tools to obtain knowledge, some pre-procesing techniques are usually used. Next we review the techniques and later we present the new proposals.

Generalized Procrustes Analysis (GPA)

This technique [11] has been largely used in the field to compare shapes of different individuals and populations and it is one of the first techniques that researchers usually use (Figure 2.2). The technique estimates the best way to rotate, scale and reflect the landmarks \mathcal{L} of the whole sample in order to reduce the differences in shape between individuals and apply the transformation to the whole dataset. The new coordinates $\mathcal{P} = P_1, P_2, \dots, P_p$ are called procrustes coordinates and are a map of the original landmarks $\text{GPA} : \mathcal{L} \mapsto \mathcal{P}$. Each new pair of coordinates is related with the same procrustes centroid $\bar{\mathcal{P}} = \bar{P}_1, \bar{P}_2, \dots, \bar{P}_p$ where $\bar{P}_i = (\bar{x}_i, \bar{y}_i)$ are the centroids coordinates. Ordinary Procrustes

Analysis allows to perform this transformation over two individuals. The generalized form allow to do the same but with a undetermined amount of individuals at the same time.

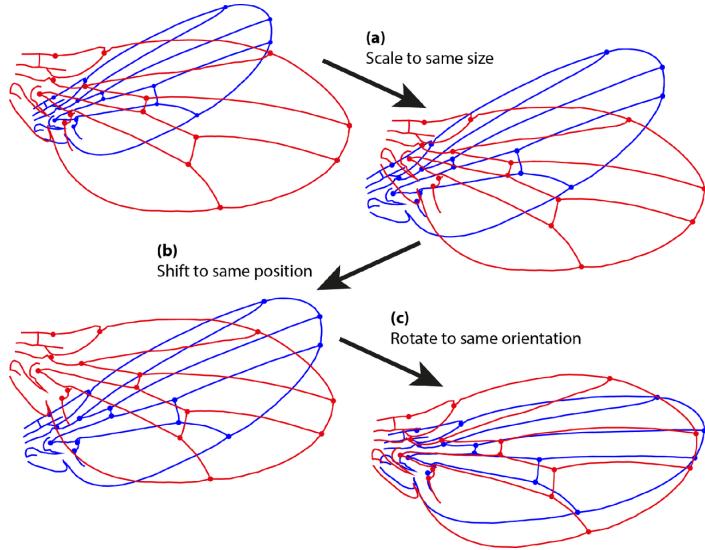


Fig. 2.2. Superimposing that performs Procrustes allowing to compare visually two different objects, in this case on the wing of an insect. Source: wikipedia/Procrustes_analysis

The technique is widely used since it allows to compare individual observations visually. The Procrustes centroid $\bar{\mathcal{P}}$ allows to compare populations on a simplified way and the new coordinates \mathcal{P} allow to perform better statistical studies by reducing some noise.

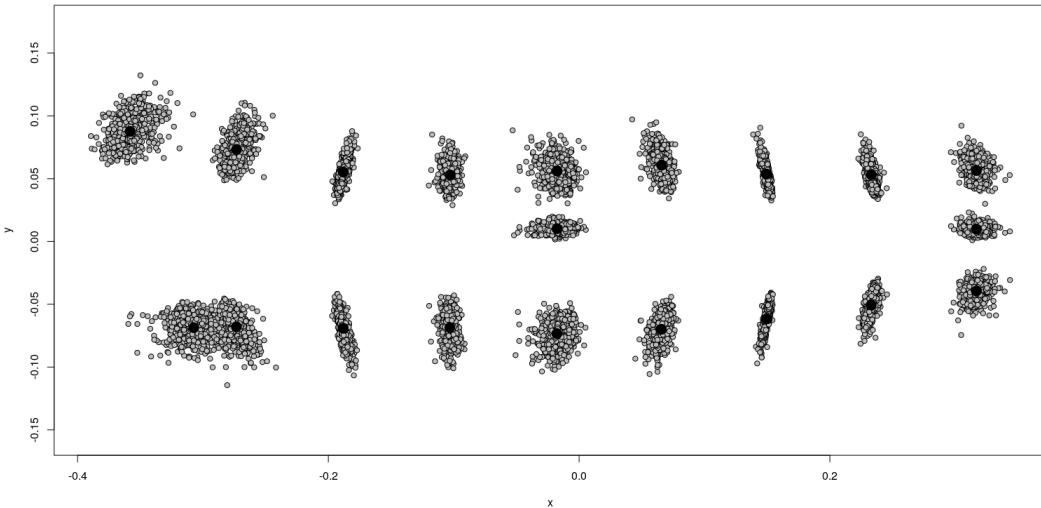


Fig. 2.3. Result of applying Procrustes to the dataset of study. In gray, procrustes coordinates \mathcal{P} of each of the children, in black, centroids $\bar{\mathcal{P}}$ of each coordinate.

Reduction of Allometric Effect

It requires of a previous Procrustes analysis. It's designed to reduce the landmark effects of size on shape. The technique allows to compare individuals regardless of the rate that shape might have with respect to size. In this particular case it is used to reduce the effect that age has on the body shape of the children. It consists on performing a linear regression between the log procrustes centroid sizes and the procrustes coordinates. The residuals of the linear regression are the new coordinates.

Principal Component Analysis

PCA is one of the most common and widely used statistical tools with a bunch of different and useful applications. It consists on performing a singular value decomposition from the data matrix $X = U\Sigma W^T$. The elements of this composition have several properties that allow PCA to be very versatile allowing the technique to be used for multiple purposes. In this context it is used to reduce dimensionality $T_L = XW_L$ maintaining only the principal components that explain the desired variance of the data. The L indicates the number of principal components retained. For this thesis, the threshold is set on a minimal proportion of explained variance of 95% regardless of the number of principal components L what this entails. It can be used before or after a Procrustes analysis but it is normal to do it afterwards since GPA is the technique that allows to reduce a lot of noise and bias (landmark drift) that may be in the data.

Linear Discriminant Analysis

The three previous methods are usually used partially or in their entirety before reaching the last step to analyze, classification. As discussed previously, it appears that the only method used by most studies in the field is LDA. LDA finds a linear combination of landmarks that allows to discriminate two or more different populations ($y = 0, y = 1$). It assumes homoscedasticity and a normal distribution for the conditional density functions $p(\vec{x}|y = 0)$ and $p(\vec{x}|y = 1)$, specifically for the two populations scenario it's called Fisher's linear discriminant:

$$S = \frac{(\vec{w} \cdot (\vec{\mu}_1 - \vec{\mu}_0))^2}{\vec{w}^T (\Sigma_0 + \Sigma_1) \vec{w}}, \quad \vec{w} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

where μ_0, μ_1 are the means of the two classes of observations and Σ_0, Σ_1 his covariances. Because it's assumed homoscedasticity $\Sigma_0 = \Sigma_1$.

LDA can be used to reduce dimensionality or as most of the time in this field, to classify. The use of LDA to classify has multiple limitations. Mainly its limitation in finding only linear relationships and the strong assumptions of normality and homoscedasticity. Although in this case, the only possible relationship between two

landmarks can be linear, the inclusion of a possible relationship with a third landmark is very limited. Later we will present other methods (not only classification algorithms) that allow us to solve this problem.

2.1.3. Question Formulation

According to Bookstein, the third principle of the field is based on not only the analysis of the possible relationships or correlations that the previous statistical tools can give us, but to give value to such analyzes it is necessary to convert them back to biology with questions of the style "What is the nature of the covariance between the reference data and the covariance factor?"

In this regard, listing the applications in this section does not make much sense, since it varies completely from one field to another in biology or paleontology. Furthermore, this is the main point at which researchers focus their work, since it is the main source of practical knowledge.

However, in the next section, we present some new proposals that can be used by researchers to approach studies from a new perspective.

2.2. Proposed Methodology

2.2.1. Multivariate analysis of landmarks information

The new methodological proposals are presented below with emphasis on their contribution on shortcomes that the usual techniques may exhibit.

Colinearity Feature Trim

This technique basically remove colinear features. Removing highly colinear features is crucial for some techniques. LDA is one of these techniques because it requires a matrix inverse step which is very numerically unstable when features are colinear. It is very important to note this problem as it is one of the main causes that previous research cannot be replicated or even that its insights are incorrect. The most used geometric morphometrics software like MorphoJ [16] do not warn of this problem when they perform LDA with highly correlated variables and it is a problem to highlight since the data of landmarks coordinates are usually highly correlated.

In some previous investigations PCA has been used to solve this same effect. Unfortunately, when using PCA the true meaning of the original landmarks is hidden, so analysis as ANOVA on LDA does not add much value. By using this technique, this problem is solved and since only landmarks are removed that do not provide new information, the quality of the analysis or future classification is not degraded.

The trim procedure is the following. Generate the covariance matrix of the data Σ with elements σ_{ij} . Loop over all rows i , if a row has an entry with a value higher than the threshold $\sigma_{vz} > \epsilon$, remove the row v and the corresponding column z from Σ . By doing this we can ensure that high redundant features are removed, leaving a smaller matrix Σ^* in the process. Albeit this could be seen as a good technique, using it with the raw points \mathcal{L} or Procrustes coordinates \mathcal{P} is not optimal at all. The covariance characterizes the linear relationship between two variables (e.g. $\text{Cov}(x_2, y_3)$), but in morphometric, each landmark requires two (x_i, y_i) or three variables (x_i, y_i, z_i) (one for each dimension), so although for instance the x coordinate of two landmark are highly correlated the y coordinate could not and vice versa.

For this work, a threshold $\epsilon = 0.80$ has been used in all cases, although it can and should be chosen with care if used for an anthropometric study. If not used with care, a significant loss of information is more than likely, do not forget that the technique basically eliminates highly redundant landmarks.

This technique is used in conjunction with generalized Procrustes analysis. From a statistical point of view, GPA reallocates the linear correlation between landmarks. Some landmarks reduce the linear correlation while others increase it, so, removing the highly colinear features would help the classification step (Figure 2.4).

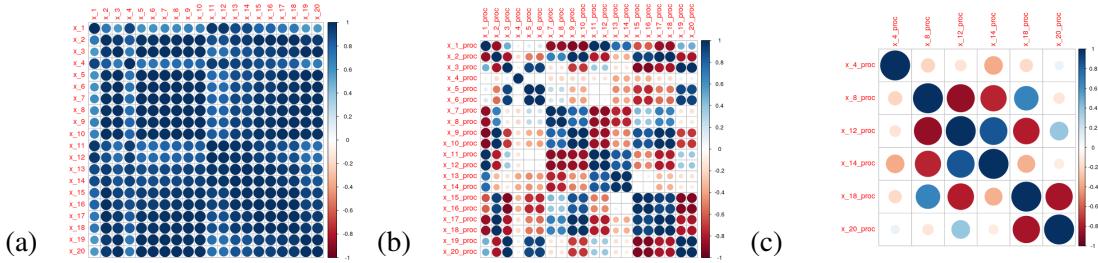


Fig. 2.4. Coordinates covariances before and after applying Procrustes. a) Covariance matrix Σ before Procrustes, b) Covariance matrix Σ after Procrustes, c) Covariance matrix Σ^* after Procrustes and Colinearity feature trim. In red colinears close to -1 values, in blue colinears close to +1 values and in white colinears close to 0 values.

Distance Matrix Computation

This new proposal is the one that mainly allows a substantial improvement in terms of classification with respect to the already established techniques. Instead of using the raw landmark coordinates \mathcal{L} or the procrustes coordinates \mathcal{P} , the euclidean distance $d(L_v, L_z) v, z \in \{1, \dots, p\}$ of each point with all the other points is computed. This generates a diagonal distance matrix D that contains the geometrical relationships between landmarks and also adds more variables, which is good for some algorithms. Thus, each distance contains the geometrical relationship between two points:

$$D(\mathcal{L}) = \begin{bmatrix} d(L_1, L_2) & d(L_1, L_2) & \dots & d(L_1, L_{p-1}) & d(L_1, L_p) \\ & d(L_2, L_3) & \dots & d(L_2, L_{p-1}) & d(L_2, L_p) \\ & & \ddots & \vdots & \vdots \\ & & & d(L_{p-2}, L_{p-1}) & d(L_{p-2}, L_p) \\ & & & & d(L_{p-1}, L_p) \end{bmatrix}$$

The application of this technique implies a substantial change with respect to the established methodology. Almost no references to the enormous information provided by dealing with distances instead of landmarks have been found in the reviewed literature. The only references found refer to these distances in visual terms and their use for classification has not been found.

The key to this technique is that it allows classification algorithms to focus directly on the linear relationships between points, rather than forcing the algorithms to find that individual relationship. The original dataset has more information, since each feature contains information from two original landmarks and also the size of the data grows. Based on the results obtained, it seems plausible that the relevant information to discriminate on different populations is in the information of the distances rather than in the coordinates alone.

This technique also opens up the possibility of improving the "Question formulation" concept since the most modern classification methods provide important information about which features are the most useful for them. So for instance, using this technique, it is almost immediate to find that the distance between landmark 1 and landmark 8 is the one that provides the most information.

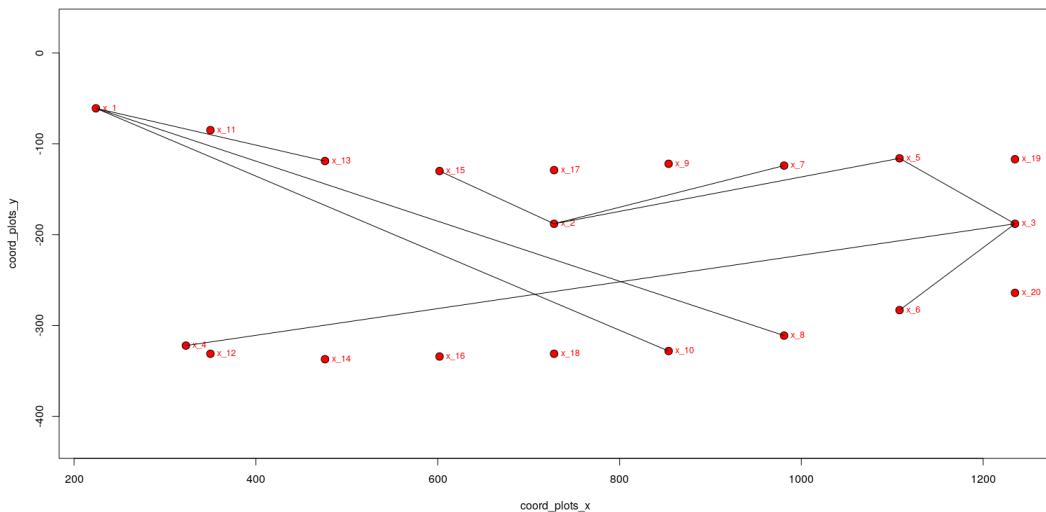


Fig. 2.5. Sample representation of the distances in matrix $D(\mathcal{L})$. The numerical value of each matrix element corresponds to the length of the associated black segment.

By counterpart, this technique has two added problems. On one hand, the generated

features grows on the order of $O(p^2)$. On the other hand, many of these new features are redundant, so a cleanup and removal step for these features is almost mandatory. Despite this, the gain far outweighs the new problems.

Finally, the use of this technique allows the inclusion of an infinity of non-parametric classification methods since the analysis of the importance of the features becomes much easier and more direct to analyze. It's likely that the inclusion of these methods has been hampered by the limitation they have to explain results. Many of these methods are considered "black box" since they only allow to be used to classify, but not to explain the reason for classifying. It is of little use for a biologist to know that a certain or a set of landmarks are the most important because this kind of information gives no knowledge about the interaction between landmarks. By using this technique, one can automatically know that the relationship between two landmarks is the most important, from a statistical point of view. In this way, a richer way of analyzing this type of relationship from the biological point of view opens up like the one referring the MUAC diagnosis.

Proposed classification algorithms to study

In this thesis, only three classification methods have been considered. Although more could have been added, the technical problems of use and the problem of performing "fine-tuning" of some hyperparameters that vary completely between algorithms means that the inclusion of more algorithms only adds complexity to the study. For this reason, only those that are probably the three most used classification algorithms in other fields have been included. The selection of the best classifiers has been made by performing Leave-One-Out Cross Validation and fine-tuning the hyperparameters by means of grid search in the case of neural network, for the other methods the default parameters have been used.

To evaluate the feature importance of each method, the 'varImp()' method of the R package 'caret' is used. Each model has its own way of assessing feature importance and is discussed below along with the values of each grid search.

- Support Vector Machine [9]

SVM is a widely used non-probabilistic classification method. The method builds a hyperplane that optimally separates two classes $(-1, 1)$. This optimal separation is defined as the maximum distance between the closest points of any class, called the functional margin, and generally speaking, the larger this functional margin, the lower the classifier error will be outside the training sample. The hyperplane is defined as a set of points \vec{x} that satisfies $\vec{w} \cdot \vec{x} - b = 0$ where \vec{w} is the normal vector to the hyperplane. To use the method on a probabilistic framework Platt scaling can be performed.

The classifier in its original form only builds linear discriminators, so in order to solve problems where the classes are not linearly separable a kernel function $k(x, y)$

can be used. This kernel function maps the values to larger dimensions. The radial basis kernel (also known as Gaussian kernel) has been used for this work. This choice of the kernel is due to its non-parametric regularization properties and proven performance on a bunch of different fields.

$$RBF(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

The variable importance by caret is performed using the area under the ROC curve.

- Random Forest Classifier [4]

This algorithm is based on the idea of ensemble learning by ensemble several decision trees on a better classifier. Decision trees are known to fit training data quite well but have very little generalization properties. Random Forest builds the classifier by taking several previously built decision trees with subsamples of the data and averaging the result. By doing this, the ability to adjust to the training data is not impaired and instead the ability to generalize when encountering new data is substantially improved. Random Forest has proven to be one of the best machine learning algorithms available for handling tabulated data, regardless of whether it is discrete or continuous.

There are multiple implementation of the algorithm, but the main idea can be described as: given a set of decision trees classifiers f_b trained following a bootstrap aggregating fashion policy, the final prediction of the random forest classifier f^* is

$$f^* = \frac{1}{B} \sum_{b=1}^B f_b(x)$$

For this thesis, a total of $B = 50$ decision trees have been used. The variable importance is computed using the differences between the prediction accuracy on the out-of-bag portion of the data and the prediction accuracy on the out-of-bag portion but permuting each predictor variable. These differences are averaged over all trees, and normalized by the standard error.

- Fully Connected Neural Network

In the last decade, artificial neural networks have gained enormous popularity, mostly due to advances in computer vision and natural language processing. However, its use in other fields of science resists. This is due to several factors, the main one being that they are difficult to perform hyperparameter tuning and their systematic superiority to other existing methods has not been demonstrated either. Due to this, it has been proposed in this thesis for study in its most generic form, that is, fully connected network where all neurons are connected to the next layer. This decision is based mainly on the limitations of having a small amount of data, limiting the use of other variants of the method, such as convolutional networks.

Of the three methods, this is the only one that requires an explicit hyperparameterization, which gives it a slight advantage over the other two methods from the beginning. For the selection of the network, grid search has been used with values of 2, 4 and 5 for the hidden layer and values of 0.1 and 0.01 for the decay regularizing parameter.

The method can be defined as:

$$f(x) = S \left(\sum_i w_i g_i(x) \right)$$

The S function is called the activation function, logistic in this case. The element w_i refers to the weight of the corresponding neuron and the function $g_i(x)$ refers to the preceding function of the network.

2.2.2. Visual Analysis of methods

For the statistical comparison when a problem is high dimensional (as in the case of the use of the distance matrix) we propose to use Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [19]. UMAP is a dimensionality reduction technique that is gaining popularity lately. The technique can be understood as similar to PCA in terms of dimensionality reduction except that the transformations it performs are non-linear, with which internal data structures are preserved and can be viewed in 2 dimensions. This technique allows to see limitations and visually compare the techniques studied here by being able to visually compare different representations by adding new data (such as sex or age) or by analyzing the different representations after using the studied methods.

Although the technique can give inconclusive results in some cases, in others it can help to give an idea of the limitations of some methods that transform the data. The main premise is that in the representation it offers, similar data are close, so if there is a well-defined structure, this representation will show small clusters as in figure 2.6.

The use of this method is not presented as a tool to gain biological knowledge but rather as a tool for statistical comparison of the methodologies studied and in this way to be able to visually analyze the transformations of the data that are carried out with each method.

2.3. Methodology Combination

It's important to note that all these techniques consist on coordinate transformation of the initial landmarks that yield to new informative coordinates/variables or coordinate/variable elimination that remove non-informative or redundant variables so

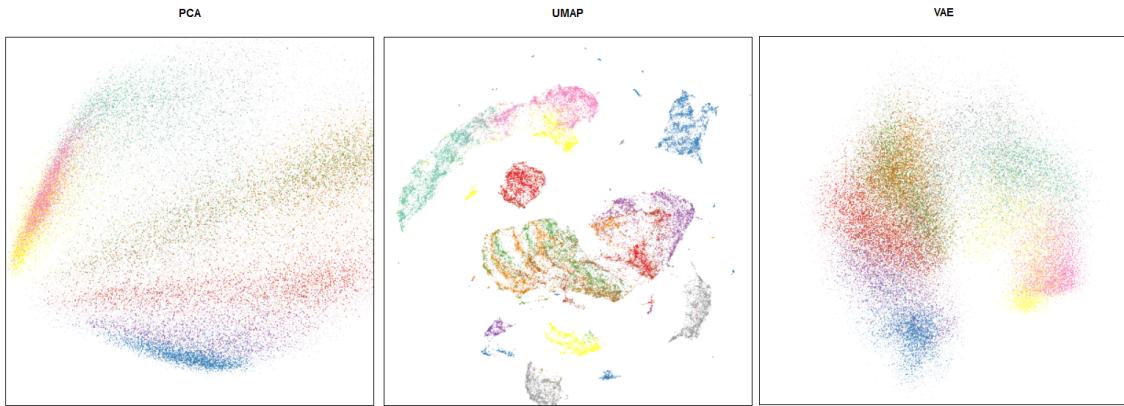


Fig. 2.6. Comparison of different dimensional reduction techniques on the same dataset. PCA, UMAP and Variational Autoencoder. Source: Github lmcmnnes/umap issue #113.

all these methods can be applied following different orders and the different combinations yields to best or worse characterizations of the groups.

In order to correctly define which methodologies and in what order they can be used, a directed graph without cycles has been defined. In this graph, each node represent the use of a previously defined possible method and each edge defines the order of application of one method to another. In this graph the starting point is always the raw landmark coordinate data and the last node is always some classification algorithm.

Based on this graph, all possible paths (possible types of geometric morphometrics analysis) have been defined and tested and quantified using different classification metrics widely used in the field.

Based on this graph, all possible methodologies combinations are:

- DATA → DISTANCES → CLASSIFICATION_ALGORITHM
- DATA → DISTANCES → COLINEARITY TRIM → CLASSIFICATION_ALGORITHM
- DATA → DISTANCES → PCA → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → DISTANCES → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → DISTANCES → COLINEARITY TRIM → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → DISTANCES → PCA → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → ALLOMETRIC REDUCTION → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → ALLOMETRIC REDUCTION → COLINEARITY TRIM → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → COLINEARITY TRIM → CLASSIFICATION_ALGORITHM
- DATA → PROCRUSTES → PCA → CLASSIFICATION_ALGORITHM

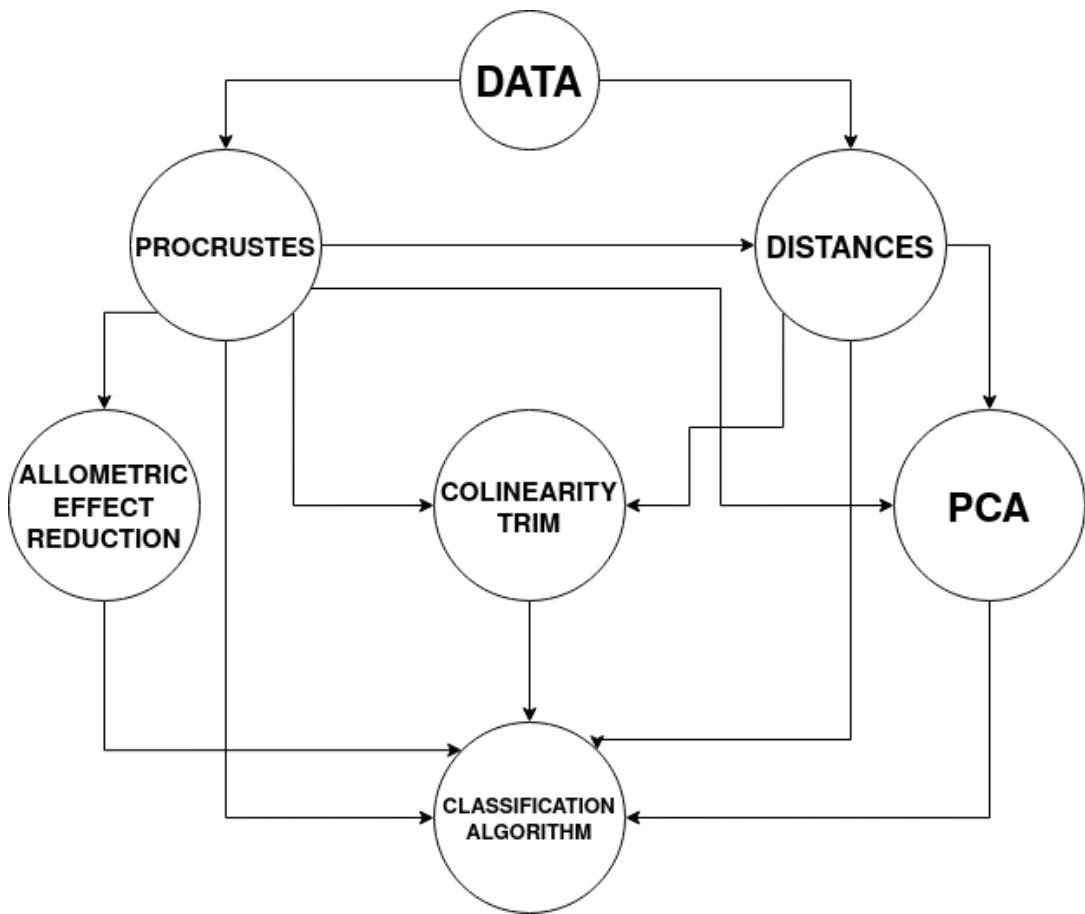


Fig. 2.7. The graph that defines all the possible methods studied.

2.4. Metrics

To measure the performance of each of the possible methodological combinations, different metrics have been used, each one focused on a particular aspect of the classification. All of these metrics have been defined on a binary-fashion excluding way, that is, for each element it is defined that it is to be positive or negative, but never at the same time. Following this definition, the following sets have been used:

- **True Positives (TP):** They are the positive elements in the dataset that have been classified as positive by the model.
- **True Negatives (TN):** They are the negative elements in the dataset that have been classified as negative by the model.
- **False Positives (FP):** They are the negative elements in the dataset that have been classified as positive by the model.
- **False Negatives (FN):** They are the positive elements in the dataset that have been classified as negative by the model.

The metrics used are:

- **Accuracy:** Used to measure the overall behavior of the model. It can be used in a sentence like "The model gets it right X% of the time".

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** Used to measure the behavior of the model with respect to producing false positives. It can be used in a sentence like "When the model predicts positive, it does it well X% of the time"

$$\frac{TP}{TP + FP}$$

- **Recall (also called sensitivity):** Used to measure the ability of the model to correctly classify all positive elements. It can be used in a sentence like "The model finds all the positive elements X% of the time"

$$\frac{TP}{TP + FN}$$

- **Specificity :** Used to measure the ability of the model to correctly classify all negative elements. It can be used in a sentence like "The model finds all the negative elements X% of the time"

$$\frac{TN}{TN + FP}$$

- **F1 Score:** Used to measure the precision and recall at the same time on a single metric.

$$\frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$$

3. RESULTS

This chapter presents the results of the study. First, the results using the state-of-the-art methods are shown in order to define a groundtruth to later show the results of each of the methods carried out with a detailed analysis of advantages and problems.

3.1. State-of-the-Art Results

Table 3.1 shows the results obtained following the current field methodology. In some cases like MAM-NOR (~71% accuracy) or NOR-SAM (~72% accuracy) the classification obtains high results and in all cases the use of allometric reduction worsens the results. This effect, as seen below, occurs in all the methods that use it.

The use of categorical data such as sex or age group does not necessarily improve the results of the methodology used and its use depends on the particular case of study.

In the annex to this document you can see figure 4.1. In it the performance of the best methods in table 3.1 depending on the age or sex of the child is shown.

Finally, in the next section you will find the results obtained using all the methods presented and how they obtain better results in all cases.

experiment	classification_method	method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	LDA	PROCRUSTES-LDA	sex+age_group	0.719557	0.666667	0.717105	0.767606	0.741497
		PROCRUSTES-LDA	age_group	0.715867	0.658915	0.712418	0.767606	0.738983
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex+age_group	0.461255	0.372093	0.487342	0.542254	0.513333
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex	0.457565	0.364341	0.484277	0.542254	0.511628
MAM_RIS	LDA	PROCRUSTES-LDA	age_group	0.603509	0.559441	0.593548	0.647887	0.619529
		PROCRUSTES-LDA	sex+age_group	0.592982	0.566434	0.586667	0.619718	0.602740
		PROCRUSTES-ALLOM_REDUCTION-LDA		0.459649	0.461538	0.457746	0.457746	0.457746
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex+age_group	0.459649	0.440559	0.459459	0.478873	0.468966
MAM_SAM	LDA	PROCRUSTES-LDA	sex+age_group	0.536082	0.550336	0.524823	0.521127	0.522968
		PROCRUSTES-LDA	age_group	0.515464	0.550336	0.503704	0.478873	0.490975
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex+age_group	0.491409	0.530201	0.477612	0.450704	0.463768
		PROCRUSTES-ALLOM_REDUCTION-LDA	age_group	0.487973	0.503356	0.475177	0.471831	0.473498
NOR_RIS	LDA	PROCRUSTES-LDA		0.555147	0.622378	0.534483	0.480620	0.506122
		PROCRUSTES-LDA	age_group	0.544118	0.594406	0.520661	0.488372	0.504000
		PROCRUSTES-ALLOM_REDUCTION-LDA	age_group	0.463235	0.531469	0.427350	0.387597	0.406504
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex+age_group	0.444853	0.496503	0.409836	0.387597	0.398406
NOR_SAM	LDA	PROCRUSTES-LDA	sex+age_group	0.726619	0.744966	0.705426	0.705426	0.705426
		PROCRUSTES-LDA	age_group	0.719424	0.738255	0.697674	0.697674	0.697674
		PROCRUSTES-ALLOM_REDUCTION-LDA	age_group	0.528777	0.557047	0.492308	0.496124	0.494208
		PROCRUSTES-ALLOM_REDUCTION-LDA	sex	0.525180	0.577181	0.487805	0.465116	0.476190
RIS_SAM	LDA	PROCRUSTES-LDA	sex+age_group	0.630137	0.617450	0.617450	0.643357	0.630137
		PROCRUSTES-LDA	age_group	0.623288	0.610738	0.610738	0.636364	0.623288
		PROCRUSTES-ALLOM_REDUCTION-LDA	age_group	0.482877	0.530201	0.469697	0.433566	0.450909
		PROCRUSTES-ALLOM_REDUCTION-LDA		0.469178	0.496644	0.456522	0.440559	0.448399

Table 3.1. BEST AND WORST TWO MODELS FOR STATE-OF-ART METHODOLOGY.

3.2. Current Results

The two best results of each experiment can be consulted in the table 3.2. From the results obtained we can observe a clear advantage in the experiments in which the distance matrix has been used instead of the procrustes coordinates. Only in the case of MAM-NOR does the use of procrustes only slightly improve other combinations without the use of distances. The following table shows the two best method combinations for the six types of experiments carried out: MAM-NOR, MAM-RIS, MAM-SAM, NOR-RIS, NOR-SAM, RIS-SAM. It should be noted that the inclusion of data referring to age or sex does not always improve the classification and in the case of NOR-RIS its addition only worsens the classification making the best model SVM without extra variables. A total of 792 combinations have been made, these include a model for each experiment type combined with each possible pre-procesing method combination combined with each of the three classification methods proposed combined with each of the four combinations of extra variables (only coordinates, sex, age-group, sex + age group).

In figure 3.2 the average performance of each method can be observed depending on the methodology used. As can be seen, random forest and SVM have very similar performances for similar methodologies and it is the neural network that takes advantage of other methods while decreasing the performance on the others. This difference is known in some areas and is sometimes used to do boosting or stacking with these different methods. It is worth noting the lower performance of the allometric reduction in all cases. A possible explanation for this is presented in the next section.

In a similar way to the previous section, in the annex you can see in figure 4.2 a comparison by age and sex of the best results obtained from table 3.2. In it, a systematic improvement can be observed in all the groups compared to those obtained following the methods of the state of the art. The most notable improvement is the one referring to NOR-SAM where in some cases the classification is close to 100% and almost always above 75% for other cases.

Finally, it is important to highlight one of the advantages of using distances between reference points instead of just coordinates. As these methods provide information on which distances provide more information, we can use the latter to draw new biological conclusions. As can be seen in figure 3.1, the four most important distances for the models are those corresponding to the circumference of the arm at different points. The most important of all is the distance between the 15 and 16 landmark and in fact it is the distance used in the malnutrition diagnosis by MUAC. These results are the same regardless of the classification algorithm used.

experiment	method	classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES	NNET	age_group	0.756458	0.720930	0.756757	0.788732	0.772414
	PROCRUSTES-DISTANCES	NNET	age_group	0.749077	0.728682	0.756944	0.767606	0.762238
MAM_RIS	PROCRUSTES-COLINEARITY_TRIM	NNET	age_group	0.638596	0.622378	0.632653	0.654930	0.643599
	DISTANCES	RF	age_group	0.635088	0.629371	0.631944	0.640845	0.636364
MAM_SAM	DISTANCES-PCA	NNET	age_group	0.621993	0.684564	0.626984	0.556338	0.589552
	DISTANCES	NNET	sex+age_group	0.621993	0.677852	0.625000	0.563380	0.592593
NOR_RIS	DISTANCES-COLINEARITY_TRIM	SVM		0.658088	0.762238	0.673077	0.542636	0.600858
	DISTANCES-PCA	NNET	sex	0.643382	0.720280	0.642857	0.558140	0.597510
NOR_SAM	DISTANCES-PCA	NNET	sex+age_group	0.787770	0.785235	0.761194	0.790698	0.775665
	DISTANCES	RF	sex	0.787770	0.805369	0.773438	0.767442	0.770428
RIS_SAM	DISTANCES-PCA	NNET	sex+age_group	0.695205	0.697987	0.687500	0.692308	0.689895
	DISTANCES	RF	sex	0.681507	0.664430	0.666667	0.699301	0.682594

Table 3.2. TOP 2 MODELS BY EXPERIMENT.

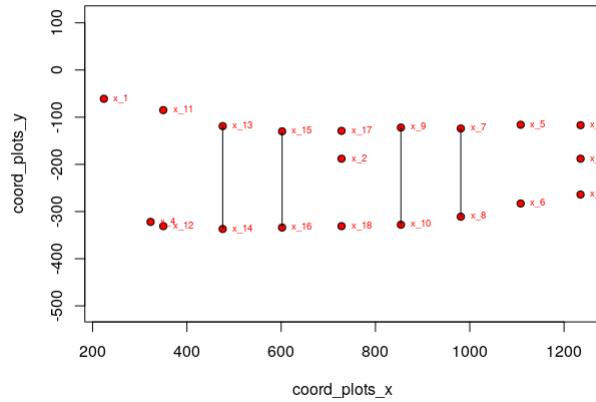


Fig. 3.1. The four distances that contain the most information for all the methods. The distance between Landmark 15 and 16 is the one that contains the most information for the classification methods. The second most informative distance is the 13 and 14. It corresponds to the distance that is measured in the malnutrition diagnosis by MUAC.

3.3. Comparison

The following section presents a detailed analysis for each methodological combination, the results obtained, and a transformation of the data using UMAP to give a possible visual explanation of why some methods work better than others.

3.3.1. Procrustes with/without Colinearity Trim

By the experiments carried out it can be observed that Procrustes is only the best option for MAM-NOR. This is quite noticeable since it is usually the procedure that is followed by default in most scientific works. The best method that works together with Procrustes is undoubtedly a neural network, being by far the best option with respect to random forest or SVM. It is also notable that the inclusion of age or sex data does not, as a rule,

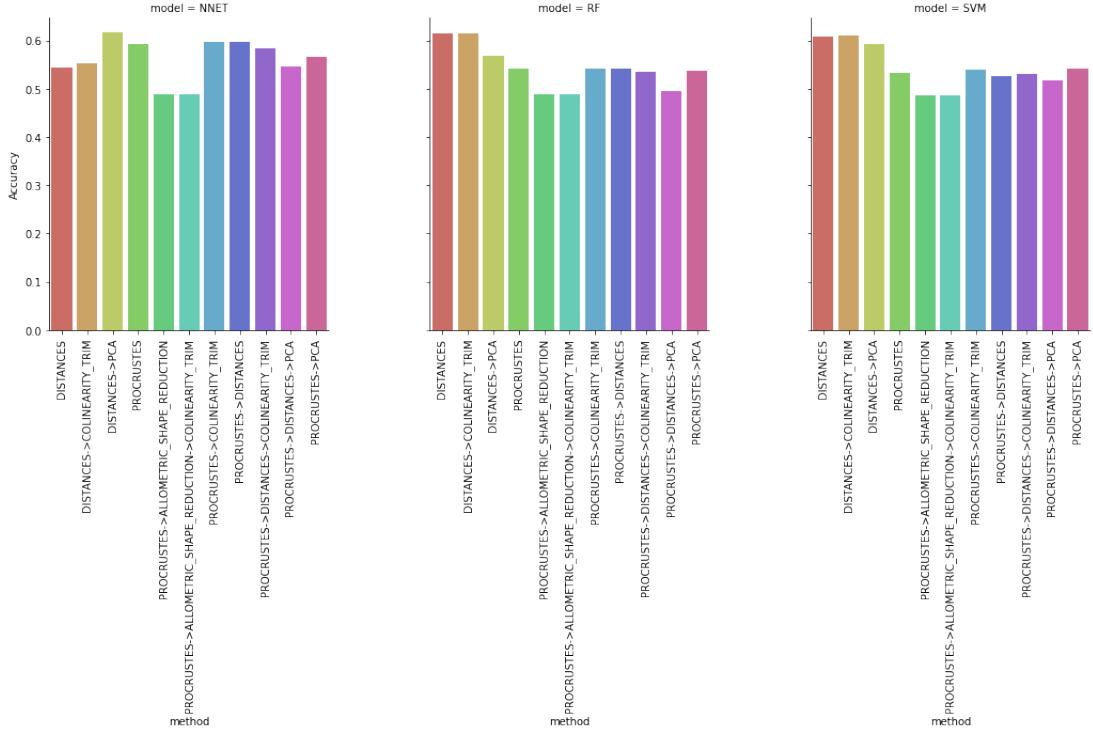


Fig. 3.2. Comparison of mean accuracy over all experiments between the data processing method and the classification method.

improve the classification results. The latter can be explained by looking at the figure 3.3a. It can be seen how the inclusion of new data does not affect the internal structure, making it quite clear that in this case, when using Procrustes, the inclusion of new data does not provide new information to the classifier, so the presence of these new data in the best classifiers becomes marginal.

If we observe the effect of removing highly correlated variables (Table 3.4) after using Procrustes, we see a clear regularizing effect on the classification. In most cases the best models are equal to or slightly worse than the models that only use Procrustes while the worst models see their performance increased in most cases. Punctually an improvement is observed with respect to only Procrustes in the best models of MAM-RIS (~2%), NOR-RIS (~1.5%) and NOR-SAM (~1%) while the worst methods also see improved their accuracy. If we observe figure 3.3b we can see that the internal structure of the data is modified with respect to only Procrustes but no deterioration is observed, with which we can conclude that using Colinearity Trim provides a regularizing effect that in most cases provides stability between models and can even lead to better rankings in some other cases.

3.3.2. Procrustes with Allometric Reduction with/without Colinearity Trim

From what is observable in the table 3.5 we can conclude that both the chosen method and the inclusion or not of data referring to age or sex do not have a determining effect

experiment	method	classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES	NNET	age_group	0.756458	0.720930	0.756757	0.788732	0.772414
		NNET	sex+age_group	0.723247	0.666667	0.718954	0.774648	0.745763
		SVM		0.630996	0.651163	0.659091	0.612676	0.635036
		SVM	sex	0.612546	0.651163	0.645669	0.577465	0.609665
		NNET	age_group	0.617544	0.622378	0.617021	0.612676	0.614841
		NNET	sex+age_group	0.617544	0.622378	0.617021	0.612676	0.614841
	PROCRUSTES	RF	sex+age_group	0.533333	0.531469	0.531469	0.535211	0.533333
		SVM	sex+age_group	0.522807	0.482517	0.519481	0.563380	0.540541
		NNET	sex	0.546392	0.557047	0.535211	0.535211	0.535211
		NNET		0.525773	0.590604	0.515873	0.457746	0.485075
		RF	age_group	0.470790	0.476510	0.458333	0.464789	0.461538
		RF	sex+age_group	0.457045	0.436242	0.447368	0.478873	0.462585
NOR_RIS	PROCRUSTES	NNET		0.584559	0.706294	0.580000	0.449612	0.506550
		NNET	sex	0.580882	0.699301	0.574257	0.449612	0.504348
		RF	age_group	0.507353	0.545455	0.480000	0.465116	0.472441
		SVM		0.507353	0.657343	0.473118	0.341085	0.396396
		NNET	age_group	0.744604	0.758389	0.723077	0.728682	0.725869
		NNET	sex+age_group	0.741007	0.765101	0.724409	0.713178	0.718750
NOR_SAM	PROCRUSTES	SVM	sex+age_group	0.661871	0.697987	0.640000	0.620155	0.629921
		RF	sex+age_group	0.658273	0.684564	0.632812	0.627907	0.630350
		NNET	sex	0.647260	0.691275	0.651515	0.601399	0.625455
		NNET		0.636986	0.704698	0.648000	0.566434	0.604478
		SVM	sex	0.554795	0.583893	0.547445	0.524476	0.535714
		SVM		0.551370	0.550336	0.541096	0.552448	0.546713

Table 3.3. BEST AND WORST TWO METHODS FOR PROCRUSTES METHODOLOGY.

on the classification. In fact, the differences in the accuracy levels between the best and worst methods in each experiment are very similar.

This is easily understandable, if one look at the figure 3.3c. In it, the internal structure is clearly defined in small clusters since allometric reduction is applied (this structure was not present when using only Procrustes) and it does not change when new data is included. In this way, it is clearly understandable why this methodology is the one that shows the worst results of all those studied. Once an individual has fallen within these clusters, the classification becomes stochastic in terms of the probabilities that the discriminator has been able to learn within that cluster. As is understandable, if a transformation of the data causes several individuals with different diagnoses to fall at the same point, there is no possible way to differentiate them except by a prior learned probability for that specific cluster.

In the same way, if we observe the results when adding collinearity trim to the coordinates after the allometric reduction of shape (Table 3.6) we can observe very similar results if not the same, that is, in this case, reducing or eliminating highly correlated variables does not have no effect. By looking at figure 3.3d we can clearly see why this similarity is due. The destruction of information caused by allometric reduction cannot be compensated for by eliminating redundant variables (which reduces possible existing noise), probably because allometric reduction has already eliminated all that noise while removing other useful information for classification. So we can finally

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-COLINEARITY_TRIM	NNET	age_group	0.741697	0.697674	0.740000	0.781690	0.760274	
		NNET	sex+age_group	0.723247	0.674419	0.721854	0.767606	0.744027	
		SVM		0.642066	0.705426	0.685950	0.584507	0.631179	
		SVM	sex	0.630996	0.744186	0.694444	0.528169	0.600000	
		NNET	age_group	0.638596	0.622378	0.632653	0.654930	0.643599	
		NNET		0.607018	0.594406	0.602740	0.619718	0.611111	
MAM_RIS	PROCRUSTES-COLINEARITY_TRIM	RF		0.540351	0.538462	0.538462	0.542254	0.540351	
		SVM	sex	0.526316	0.489510	0.522876	0.563380	0.542373	
		NNET	sex+age_group	0.546392	0.610738	0.539683	0.478873	0.507463	
		NNET	sex	0.546392	0.557047	0.535211	0.535211	0.535211	
		RF	sex+age_group	0.460481	0.456376	0.448980	0.464789	0.456747	
		RF	sex	0.450172	0.422819	0.441558	0.478873	0.459459	
NOR_RIS	PROCRUSTES-COLINEARITY_TRIM	NNET		0.595588	0.727273	0.597938	0.449612	0.513274	
		NNET	sex	0.591912	0.720280	0.591837	0.449612	0.511013	
		SVM		0.518382	0.615385	0.490741	0.410853	0.447257	
		RF	sex+age_group	0.511029	0.538462	0.484375	0.480620	0.482490	
		NNET	age_group	0.751799	0.778523	0.738095	0.720930	0.729412	
		NNET	sex+age_group	0.744604	0.765101	0.726562	0.720930	0.723735	
NOR_SAM	PROCRUSTES-COLINEARITY_TRIM	RF	sex	0.665468	0.684564	0.638462	0.643411	0.640927	
		RF	age_group	0.636691	0.637584	0.602941	0.635659	0.618868	
		NNET	age_group	0.643836	0.651007	0.636364	0.636364	0.636364	
		NNET		0.640411	0.704698	0.650794	0.573427	0.609665	
		SVM		0.568493	0.563758	0.557823	0.573427	0.565517	
		SVM	sex	0.561644	0.597315	0.555556	0.524476	0.539568	

Table 3.4. BEST AND WORST TWO METHODS FOR PROCRUSTES + COLINEARITY TRIM METHODOLOGY.

conclude that eliminating highly redundant variables after allometric reduction has no effect, since there are no variables to eliminate.

3.3.3. Procrustes with PCA

If we compare the use of Procrustes + PCA (Table 3.7) and only Procrustes (Table 3.3) we can see a systematic reduction of ~2% in the accuracy of all the best models when using PCA while in the worst models there is also a minor reduction. Looking at the figure 3.3e we clearly see the effect that PCA has on the data by adding the information about sex. PCA together with a balanced discrete binary feature, separates the data into two structures. The benefit or not of this fact is relative. While in this case it is clearly counterproductive, in the case of the use of distances it is just the opposite.

After all this we can affirm that in this case, the combined use of Procrustes and PCA to develop a classification model is counterproductive.

3.3.4. Distances

If we compare the results obtained when using the distance matrix (Table 3.8) with its closest method, that is, Procrustes (Table 3.3), we can observe a clear improvement in almost all cases, in particular, MAM-RIS (~2%), MAM-SAM (~5%), NOR-RIS (~3%), NOR-SAM (~4%) and RIS-SAM (~4%). Only in the case of procrustes MAM-NOR shows an improvement over the use of the distance matrix of ~6%. The reasons for this difference are not clear using the chosen methodology. If we compare the UMAP

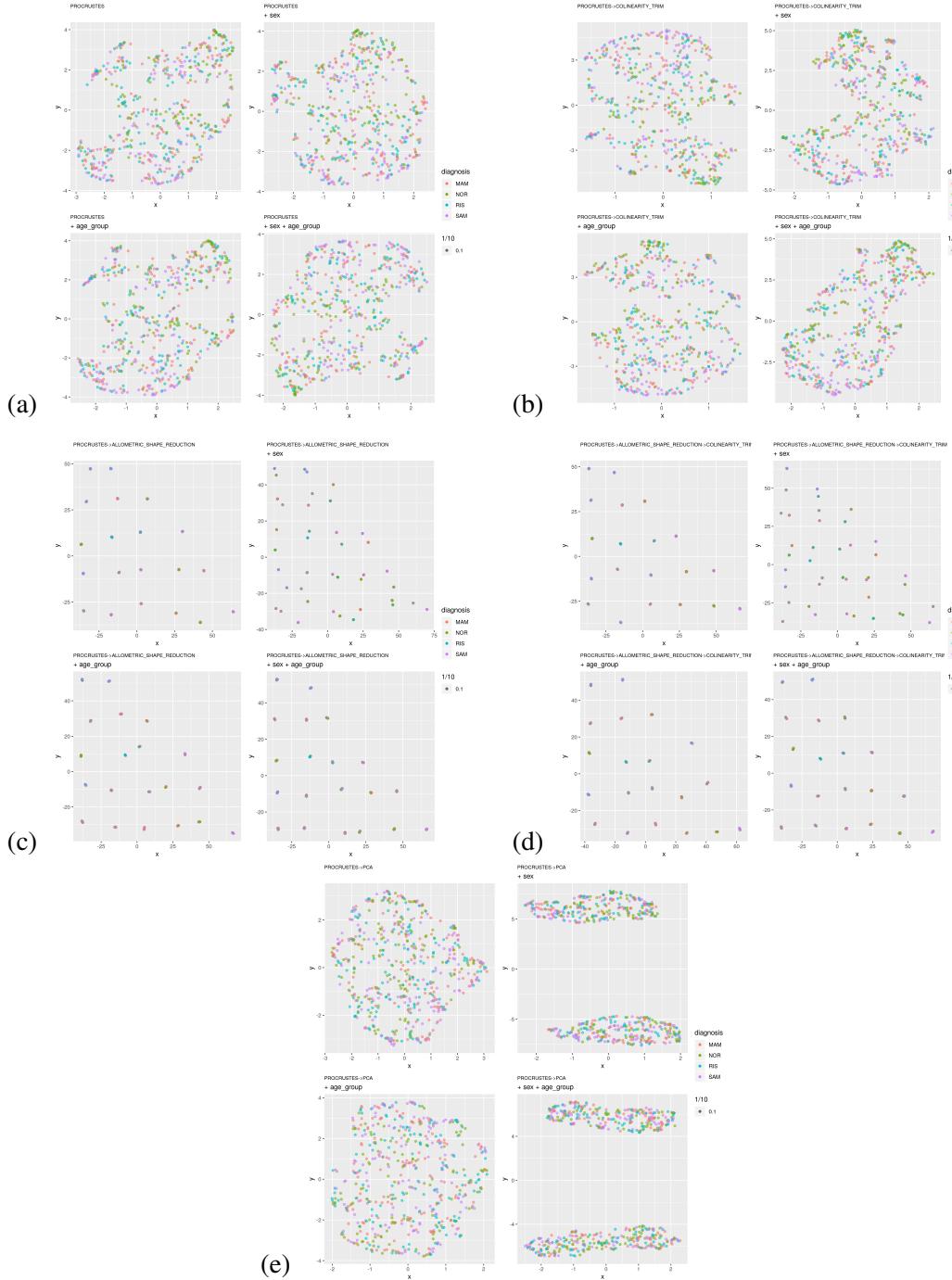


Fig. 3.3. UMAP transformations of: a) Procrustes; b) Procrustes + Colinearity Trim; c) Procrustes + Allo. Shape Reduction; d) Procrustes + Allo. Shape Reduction + Colinearity Trim; e) Procrustes + PCA. Shape Reduction + Colinearity Trim. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	RF	age_group	0.568266	0.503876	0.581699	0.626761	0.603390	
		RF		0.549815	0.465116	0.563291	0.626761	0.593333	
		NNET	sex	0.498155	0.410853	0.518987	0.577465	0.546667	
		NNET	sex+age_group	0.472325	0.372093	0.496894	0.563380	0.528053	
MAM_RIS	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	NNET	age_group	0.568421	0.587413	0.569343	0.549296	0.559140	
		RF	sex+age_group	0.526316	0.510490	0.523810	0.542254	0.532872	
		RF	age_group	0.484211	0.468531	0.482993	0.500000	0.491349	
		RF	sex	0.484211	0.419580	0.484472	0.549296	0.514851	
MAM_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	SVM	sex+age_group	0.525773	0.677852	0.520000	0.366197	0.429752	
		SVM		0.522337	0.684564	0.515464	0.352113	0.418410	
		RF		0.505155	0.536913	0.492647	0.471831	0.482014	
		NNET	sex+age_group	0.494845	0.543624	0.480916	0.443662	0.461538	
NOR_RIS	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	NNET	sex+age_group	0.573529	0.622378	0.553719	0.519380	0.536000	
		NNET	age_group	0.558824	0.615385	0.537815	0.496124	0.516129	
		RF	sex	0.474265	0.517483	0.443548	0.426357	0.434783	
		NNET	sex	0.470588	0.496503	0.441860	0.441860	0.441860	
NOR_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	RF	age_group	0.607914	0.630872	0.576923	0.581395	0.579151	
		RF		0.579137	0.617450	0.547619	0.534884	0.541176	
		SVM		0.539568	0.865772	0.512195	0.162791	0.247059	
		NNET	sex+age_group	0.532374	0.570470	0.496063	0.488372	0.492188	
RIS_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION	SVM	age_group	0.551370	0.778523	0.576923	0.314685	0.407240	
		SVM	sex+age_group	0.547945	0.785235	0.573333	0.300699	0.394495	
		RF	sex	0.523973	0.563758	0.514925	0.482517	0.498195	
		RF		0.520548	0.583893	0.511811	0.454545	0.481481	

Table 3.5. BEST AND WORST TWO METHODS FOR PROCRUSTES + ALLOMETRIC REDUCTION METHODOLOGY.

transformations (Distances Figure 3.4a, Procrustes Figure 3.3a) there is no significant difference. Both methodologies show similar behaviors when adding the data referring to sex or age, that is, the structure is affected but maintains the characteristics. In addition, both figures show a defined internal structure, and although they are geometrically different (indicating that the internal structure is not the same) both present similar characteristics since they present small agglomerations on some sides or a wide structure with distinguishable parts (indicating that a method can estimate a discriminant function of those parts). The fact that the classification by Procrustes is the best only for the particular case of MAM-NOR and does not stand out in any of the other cases does not help to explain the reasons.

Except for this detail, it is clear that the use of the matrix of distances between landmarks adds more information than the coordinates alone and this has a positive impact on the classification methods. In addition, the possibilities of directly studying the distances between landmarks and the existing relationships instead of the landmarks coordinates add an extra from the methodological point of view as can be seen in the example of figure 3.1. We can conclude that this method is generally preferable over the use of coordinates alone.

3.3.5. Distances with Colinearity Trim

When comparing the effect of eliminating distances with high correlation (Table 3.9) against the use of all distances (Table 3.8) we can see an improvement in the accuracy of the MAM-SAM (~2%) and NOR-RIS (~5%) models. In contrast, we see a worsening in MAM-RIS (~2%) and RIS-SAM (~1%). In the other experiments the results are similar.

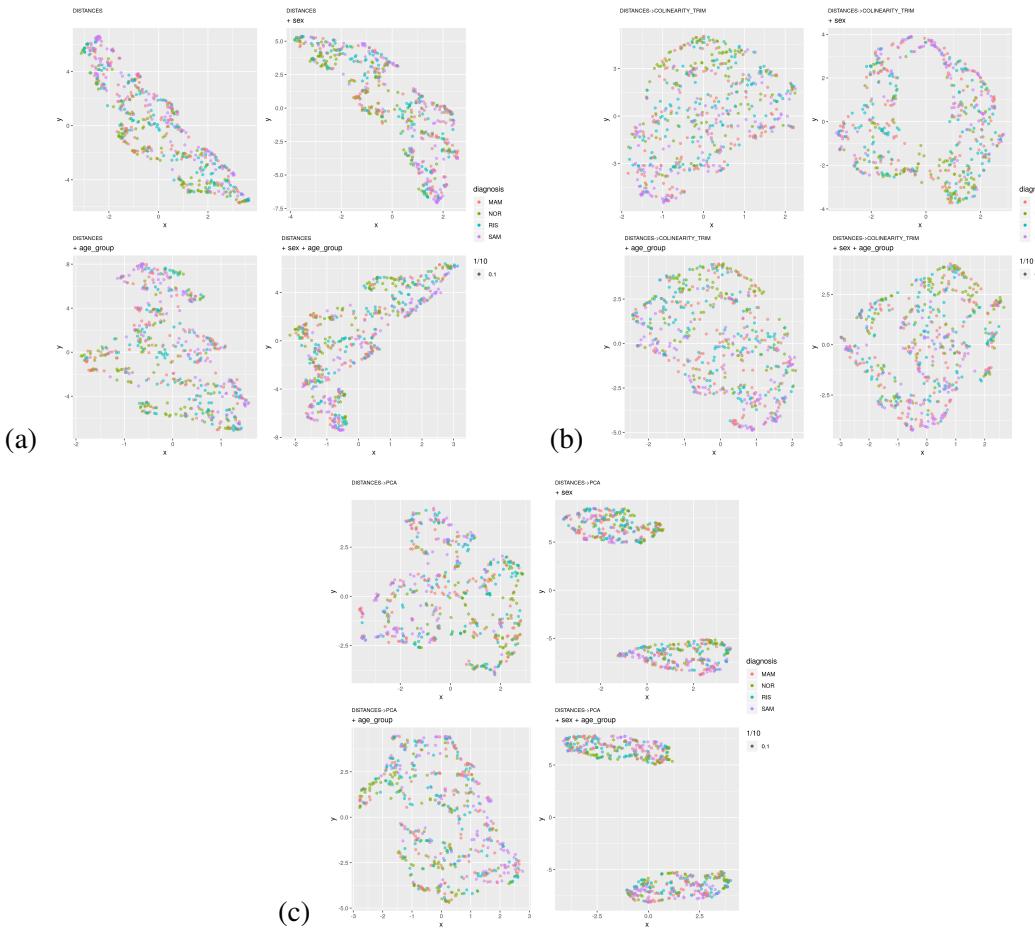


Fig. 3.4. UMAP transformations of: a) Distances; b) Distances + Colinearity Trim; c) Distances + PCA. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION-COLINEARITY_TRIM	RF	age_group	0.568266	0.503876	0.581699	0.626761	0.603390	
		RF		0.549815	0.465116	0.563291	0.626761	0.593333	
		NNET	sex	0.498155	0.410853	0.518987	0.577465	0.546667	
		NNET	sex+age_group	0.472325	0.372093	0.496894	0.563380	0.528053	
		NNET	age_group	0.568421	0.587413	0.569343	0.549296	0.559140	
		RF	sex+age_group	0.526316	0.510490	0.523810	0.542254	0.532872	
		RF	age_group	0.484211	0.468531	0.482993	0.500000	0.491349	
MAM_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION-COLINEARITY_TRIM	SVM	sex+age_group	0.525773	0.677852	0.520000	0.366197	0.429752	
		SVM		0.522337	0.684564	0.515464	0.352113	0.418410	
		RF		0.505155	0.536913	0.492647	0.471831	0.482014	
		NNET	sex+age_group	0.494845	0.543624	0.480916	0.443662	0.461538	
		NNET	sex+age_group	0.573529	0.622378	0.553719	0.519380	0.536000	
		NNET	age_group	0.558824	0.615385	0.537815	0.496124	0.516129	
		RF	sex	0.474265	0.517483	0.443548	0.426357	0.434783	
NOR_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION-COLINEARITY_TRIM	RF	sex	0.470588	0.496503	0.441860	0.441860	0.441860	
		RF	age_group	0.607914	0.630872	0.576923	0.581395	0.579151	
		RF		0.579137	0.617450	0.547619	0.534884	0.541176	
		SVM		0.539568	0.865772	0.512195	0.162791	0.247059	
		NNET	sex+age_group	0.532374	0.570470	0.496063	0.488372	0.492188	
		SVM	age_group	0.551370	0.778523	0.576923	0.314685	0.407240	
		SVM	sex+age_group	0.547945	0.785235	0.573333	0.300699	0.394495	
RIS_SAM	PROCRUSTES-ALLOMETRIC_SHAPE_REDUCTION-COLINEARITY_TRIM	RF	sex	0.523973	0.563758	0.514925	0.482517	0.498195	
		RF		0.520548	0.583893	0.511811	0.454545	0.481481	

Table 3.6. BEST AND WORST TWO METHODS FOR PROCRUSTES + ALLOMETRIC REDUCTION + COLINEARITY TRIM METHODOLOGY.

If we compare the UMAP transformations (Figure 3.5 a, b) we see a clear change in the representation of the data. However, this change is not always positive, which may indicate that by eliminating some highly correlated distances, certain important information has been eliminated with them in certain cases.

So, we can conclude that depending on the case, the elimination of highly correlated distances can be beneficial, and that adjusting the maximum correlation level to eliminate could reach an equilibrium where the use of this technique together with the distance matrix compensates the detriment of some specific cases.

3.3.6. Distances with PCA

After applying PCA to the distance matrix (Table 3.10) we can observe a systematic improvement against its non-use (Table 3.8) in all cases except in MAM-RIS. This fact is similar to the one founded when comparing Distances versus Procrustes, that is, there is a systematic improvement in all cases except one. The explanations that UMAP can provide (Figure 3.5c) are not conclusive. A behavior similar to that found in Procrustes + PCA (Figure 3.3e) is observed, but except for this fact, the characteristics are similar and the representation does not provide anything conclusive. A plausible answer that could explain this is the same one presented above. In the case of MAM-RIS the differences between populations must be marked by small differences in some specific features and by eliminating the main components of PCA that explain less variability, we eliminate within the key information for this case. However, it seem plausible that these PCAs contains in both ways information and noise, and by removing these PCAs, the noise that prevented a better classification in the other cases is also removed.

In this case, the benefits of using PCA together with the distance matrix method is not

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-PCA	NNET	sex+age_group	0.701107	0.666667	0.707483	0.732394	0.719723	
		NNET	age_group	0.693727	0.666667	0.703448	0.718310	0.710801	
		RF		0.623616	0.550388	0.628205	0.690141	0.657718	
		SVM		0.605166	0.480620	0.603550	0.718310	0.655949	
MAM_RIS	PROCRUSTES-PCA	NNET	sex+age_group	0.596491	0.587413	0.593103	0.605634	0.599303	
		SVM	age_group	0.578947	0.587413	0.578571	0.570423	0.574468	
		RF		0.529825	0.503497	0.526667	0.556338	0.541096	
		SVM	sex	0.522807	0.545455	0.522059	0.500000	0.510791	
MAM_SAM	PROCRUSTES-PCA	RF	age_group	0.529210	0.530201	0.517241	0.528169	0.522648	
		NNET		0.529210	0.604027	0.520325	0.450704	0.483019	
		SVM	sex+age_group	0.491409	0.738255	0.458333	0.232394	0.308411	
		SVM		0.487973	0.671141	0.461538	0.295775	0.360515	
NOR_RIS	PROCRUSTES-PCA	NNET	sex	0.562500	0.608392	0.540984	0.511628	0.525896	
		NNET	age_group	0.558824	0.559441	0.533333	0.558140	0.545455	
		RF		0.496324	0.538462	0.467742	0.449612	0.458498	
		RF	sex	0.496324	0.545455	0.467213	0.441860	0.454183	
NOR_SAM	PROCRUSTES-PCA	NNET	sex+age_group	0.726619	0.738255	0.702290	0.713178	0.707692	
		SVM	age_group	0.715827	0.758389	0.704918	0.666667	0.685259	
		SVM	sex	0.672662	0.731544	0.661017	0.604651	0.631579	
		NNET	sex	0.661871	0.718121	0.647059	0.596899	0.620968	
RIS_SAM	PROCRUSTES-PCA	NNET	sex+age_group	0.626712	0.664430	0.626866	0.587413	0.606498	
		NNET	age_group	0.602740	0.651007	0.603053	0.552448	0.576642	
		RF	sex+age_group	0.551370	0.530201	0.539474	0.573427	0.555932	
		RF	age_group	0.544521	0.523490	0.532895	0.566434	0.549153	

Table 3.7. BEST AND WORST TWO METHODS FOR PROCRUSTES
+ PCA METHODOLOGY.

clear. For a practitioner, care should be taken about the number of PCs chosen and by this, on how much variability retained. Therefore, the only recommendation that can be given is to carry out a detailed study for each particular problem, since a systematic use of this technique could lead to a detriment in the results, although its correct use, on the contrary, can provide a substantial benefit.

It should also be noted that the benefits of being able to directly analyze which distances between landmarks are the most important are eliminated, the detailed analysis of the linear combinations between distances that PCA give us can still provide new conclusions, although of course, if they exist these will be more complex and difficult to give meaning.

3.3.7. Procrustes with Distances with/without Colinearity Trim

In the case of calculating the distances between the Procrustes coordinates (Table 3.10) we can observe a systematic decrease (~3% on average) in all cases except in MAM-NOR, where the use of this method improves the classification at ~5%. The possible cause of this has already been discussed in the methodological analyzes in which this difference was also present. Given the possible explanation for the poor performance of this technique in general, we can observe in figure (3.5a) a form quite similar to that observed in only Procrustes (Figure 3.3a), which seems to indicate that the Procrustes transformation eliminates and/or normalizes the information that could be

experiment	method		extra_cols	Accuracy	Specificity	Precision	Recall	F1
		method						
MAM_NOR	DISTANCES	RF	age_group	0.693727	0.713178	0.721805	0.676056	0.698182
		RF		0.682657	0.697674	0.708955	0.669014	0.688406
		NNET	sex	0.627306	0.496124	0.619883	0.746479	0.677316
		NNET		0.579336	0.403101	0.576923	0.739437	0.648148
MAM_RIS	DISTANCES	RF	age_group	0.635088	0.629371	0.631944	0.640845	0.636364
		SVM	sex	0.617544	0.622378	0.617021	0.612676	0.614841
		NNET	age_group	0.543860	0.664336	0.555556	0.422535	0.480000
		NNET	sex+age_group	0.533333	0.629371	0.539130	0.436620	0.482490
MAM_SAM	DISTANCES	SVM	sex+age_group	0.597938	0.637584	0.593985	0.556338	0.574545
		SVM		0.591065	0.617450	0.583942	0.563380	0.573477
		NNET	sex	0.522337	0.577181	0.511628	0.464789	0.487085
		NNET	age_group	0.508591	0.671141	0.494845	0.338028	0.401674
NOR_RIS	DISTANCES	SVM	age_group	0.610294	0.762238	0.626374	0.441860	0.518182
		RF		0.610294	0.685315	0.601770	0.527132	0.561983
		NNET		0.569853	0.545455	0.542254	0.596899	0.568266
		NNET	sex+age_group	0.562500	0.762238	0.564103	0.341085	0.425121
NOR_SAM	DISTANCES	RF	sex	0.787770	0.805369	0.773438	0.767442	0.770428
		RF	sex+age_group	0.784173	0.805369	0.771654	0.759690	0.765625
		NNET	sex+age_group	0.654676	0.637584	0.617021	0.674419	0.644444
		NNET	age_group	0.651079	0.744966	0.648148	0.542636	0.590717
RIS_SAM	DISTANCES	RF	sex	0.681507	0.664430	0.666667	0.699301	0.682594
		RF	age_group	0.674658	0.664430	0.662162	0.685315	0.673540
		NNET	age_group	0.558219	0.671141	0.562500	0.440559	0.494118
		NNET	sex+age_group	0.558219	0.550336	0.547297	0.566434	0.556701

Table 3.8. BEST AND WORST TWO METHODS FOR DISTANCES METHODOLOGY.

extracted by computing the distance matrix. Furthermore, if we compare these two figures with the UMAP of only distances (Figure 3.4a) we can see a clear difference in the geometry, which could indicate (together with the observed results) that the internal structure obtained from the distances contains more information than the resulting from applying Procrustes alone or before calculating these distances.

If we look at the table 3.12 we can observe a behavior similar to that observed in section 3.3.5 in which the elimination of highly correlated variables reduced the performance of the methods in some cases. In this case, we can observe a reduction in accuracy in all cases. If we observe the UMAP representation, we see the same characteristics as those observed in the case of section 3.3.1. With which we can conclude that the combined use of Procrustes coordinates distances with an elimination of the correlated distances is counterproductive or that at least requires a specific adjustment in the pruning levels for each particular case.

3.3.8. Procrustes with Distances with PCA

As in previous cases, it seems that the use of PCA and the subsequent noise reduction by choosing a subset of PCs that explain less variability than the total is counterproductive

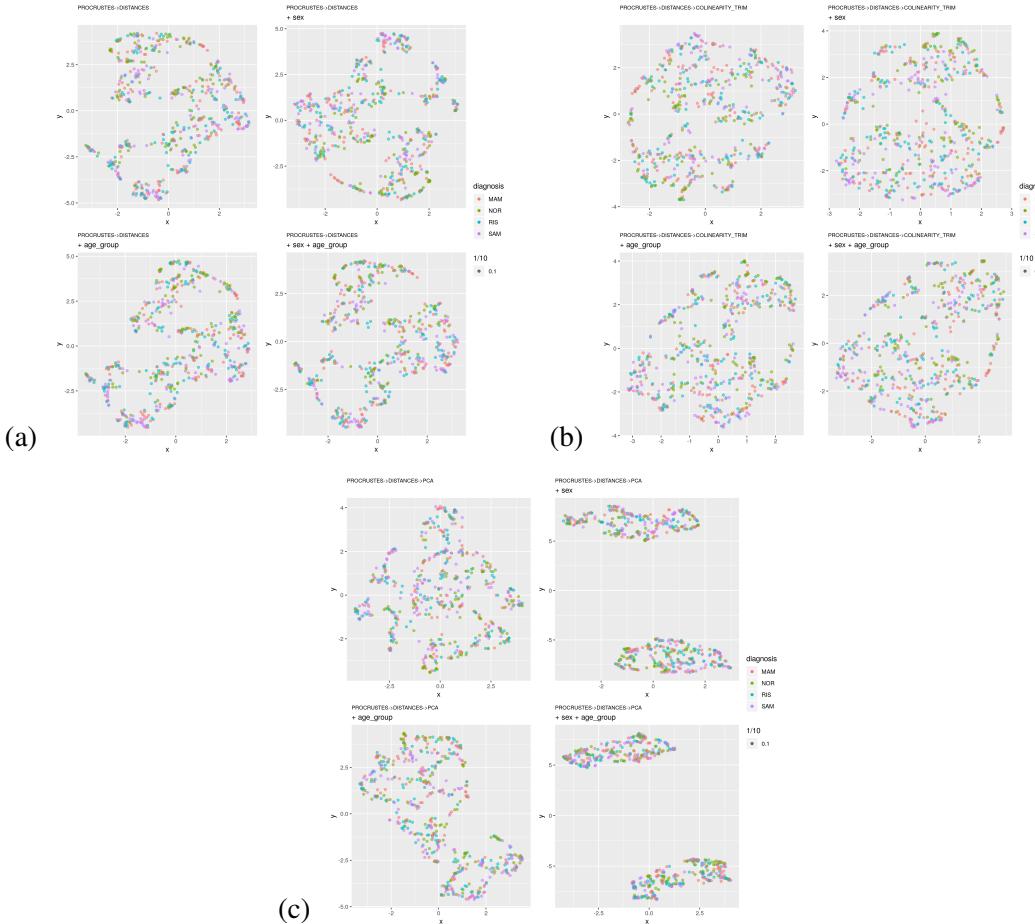


Fig. 3.5. UMAP transformations of: a) Procrustes + Distances; b) Procrustes + Distances + Colinearity Trim; c) Procrustes + Distances + PCA. The four subplots of each figure represent the UMAP by: upper-left) Just Landmark data; upper-right) Landmark + Sex data; lower-left) Landmark + Age Group data; lower-right) Landmark + Sex + Age Group data.

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	DISTANCES-COLINEARITY_TRIM	SVM	sex+age_group	0.697417	0.736434	0.734375	0.661972	0.696296	
		RF	sex+age_group	0.690037	0.689922	0.710145	0.690141	0.700000	
		NNET	age_group	0.623616	0.713178	0.675439	0.542254	0.601562	
		NNET	sex+age_group	0.616236	0.674419	0.655738	0.563380	0.606061	
MAM_RIS	DISTANCES-COLINEARITY_TRIM	RF	sex	0.614035	0.629371	0.615942	0.598592	0.607143	
		RF	age_group	0.592982	0.601399	0.592857	0.584507	0.588652	
		SVM	age_group	0.550877	0.601399	0.554688	0.500000	0.525926	
		SVM	sex+age_group	0.547368	0.594406	0.550388	0.500000	0.523985	
MAM_SAM	DISTANCES-COLINEARITY_TRIM	RF		0.618557	0.644295	0.613139	0.591549	0.602151	
		RF	sex+age_group	0.611684	0.651007	0.609023	0.570423	0.589091	
		NNET	sex	0.529210	0.610738	0.520661	0.443662	0.479087	
		NNET	sex+age_group	0.525773	0.651007	0.518519	0.394366	0.448000	
NOR_RIS	DISTANCES-COLINEARITY_TRIM	SVM		0.658088	0.762238	0.673077	0.542636	0.600858	
		SVM	age_group	0.625000	0.762238	0.642105	0.472868	0.544643	
		NNET		0.555147	0.580420	0.531250	0.527132	0.529183	
		NNET	age_group	0.547794	0.671329	0.530000	0.410853	0.462882	
NOR_SAM	DISTANCES-COLINEARITY_TRIM	RF		0.780576	0.791946	0.761538	0.767442	0.764479	
		RF	sex+age_group	0.780576	0.818792	0.778689	0.736434	0.756972	
		NNET		0.665468	0.738255	0.657895	0.581395	0.617284	
		NNET	age_group	0.647482	0.624161	0.608392	0.674419	0.639706	
RIS_SAM	DISTANCES-COLINEARITY_TRIM	RF		0.671233	0.624161	0.647799	0.720280	0.682119	
		SVM	sex	0.667808	0.697987	0.669118	0.636364	0.652330	
		NNET		0.595890	0.624161	0.591241	0.566434	0.578571	
		NNET	sex+age_group	0.571918	0.516779	0.555556	0.629371	0.590164	

Table 3.9. BEST AND WORST TWO METHODS FOR DISTANCES + COLINEARITY TRIM METHODOLOGY.

in the case of having calculated the distances of the Procrustes coordinates. The results are in table 3.13. From the previous analyzes, we can conclude that the use of Procrustes eliminates in the process some useful information for the classification while eliminating certain noise (the method is better in the case of MAM-NOR) so it seems like the method acts in a certain way as a regularizer. If we combine this method with the calculation of distances, it seems that the information available is even less, so if we reduce the variability explained using PCA, the information available in the data is seriously reduced. We can conclude that in this particular case, the PCA is not necessary or, in case of use, the amount of retained variability must be carefully chosen.

experiment	method	classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	DISTANCES-PCA	NNET	sex+age_group	0.712177	0.767442	0.758065	0.661972	0.706767
		NNET	age_group	0.708487	0.751938	0.748031	0.669014	0.706320
		RF		0.634686	0.620155	0.652482	0.647887	0.650177
		NNET	sex	0.634686	0.697674	0.677686	0.577465	0.623574
MAM_RIS	DISTANCES-PCA	SVM	age_group	0.596491	0.636364	0.603053	0.556338	0.578755
		RF	sex	0.578947	0.601399	0.580882	0.556338	0.568345
		SVM	sex	0.550877	0.580420	0.552239	0.521127	0.536232
		SVM	sex+age_group	0.505263	0.524476	0.503650	0.485915	0.494624
MAM_SAM	DISTANCES-PCA	NNET	age_group	0.621993	0.684564	0.626984	0.556338	0.589552
		NNET	sex+age_group	0.621993	0.677852	0.625000	0.563380	0.592593
		RF	sex+age_group	0.522337	0.543624	0.510791	0.500000	0.505338
		RF		0.518900	0.536913	0.507143	0.500000	0.503546
NOR_RIS	DISTANCES-PCA	NNET	sex	0.643382	0.720280	0.642857	0.558140	0.597510
		NNET	sex+age_group	0.636029	0.699301	0.629310	0.565891	0.595918
		RF		0.540441	0.552448	0.515152	0.527132	0.521073
		RF	sex	0.533088	0.552448	0.507692	0.511628	0.509653
NOR_SAM	DISTANCES-PCA	NNET	sex+age_group	0.787770	0.785235	0.761194	0.790698	0.775665
		NNET	age_group	0.780576	0.812081	0.774194	0.744186	0.758893
		RF	sex	0.712230	0.731544	0.689922	0.689922	0.689922
		RF		0.705036	0.704698	0.674074	0.705426	0.689394
RIS_SAM	DISTANCES-PCA	NNET	sex+age_group	0.695205	0.697987	0.687500	0.692308	0.689895
		NNET	sex	0.674658	0.664430	0.662162	0.685315	0.673540
		RF	sex	0.606164	0.617450	0.598592	0.594406	0.596491
		RF		0.575342	0.536913	0.560510	0.615385	0.586667

Table 3.10. BEST AND WORST TWO METHODS FOR DISTANCES + PCA METHODOLOGY.

experiment	method	classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-DISTANCES	NNET	age_group	0.749077	0.728682	0.756944	0.767606	0.762238
		NNET		0.730627	0.697674	0.734694	0.760563	0.747405
		RF	sex+age_group	0.619926	0.573643	0.630872	0.661972	0.646048
		SVM	sex	0.616236	0.620155	0.639706	0.612676	0.625899
MAM_RIS	PROCRUSTES-DISTANCES	NNET		0.607018	0.622378	0.608696	0.591549	0.600000
		NNET	age_group	0.600000	0.608392	0.600000	0.591549	0.595745
		SVM	age_group	0.519298	0.454545	0.515528	0.584507	0.547855
		SVM	sex+age_group	0.519298	0.454545	0.515528	0.584507	0.547855
MAM_SAM	PROCRUSTES-DISTANCES	NNET		0.546392	0.671141	0.546296	0.415493	0.472000
		NNET	sex	0.546392	0.610738	0.539683	0.478873	0.507463
		SVM	sex	0.487973	0.610738	0.467890	0.359155	0.406375
		RF		0.481100	0.483221	0.468966	0.478873	0.473868
NOR_RIS	PROCRUSTES-DISTANCES	NNET	sex+age_group	0.591912	0.629371	0.572581	0.550388	0.561265
		NNET		0.588235	0.622378	0.568000	0.550388	0.559055
		SVM	age_group	0.514706	0.727273	0.480000	0.279070	0.352941
		RF		0.514706	0.566434	0.487603	0.457364	0.472000
NOR_SAM	PROCRUSTES-DISTANCES	NNET	age_group	0.748201	0.765101	0.728682	0.728682	0.728682
		NNET	sex+age_group	0.733813	0.758389	0.716535	0.705426	0.710938
		SVM	sex	0.651079	0.630872	0.612676	0.674419	0.642066
		SVM		0.647482	0.624161	0.608392	0.674419	0.639706
RIS_SAM	PROCRUSTES-DISTANCES	NNET		0.647260	0.671141	0.644928	0.622378	0.633452
		NNET	sex+age_group	0.640411	0.677852	0.641791	0.601399	0.620939
		SVM		0.551370	0.516779	0.538462	0.587413	0.561873
		SVM	sex+age_group	0.537671	0.536913	0.527397	0.538462	0.532872

Table 3.11. BEST AND WORST TWO METHODS FOR PROCRUSTES + DISTANCES METHODOLOGY.

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET	age_group	0.730627	0.682171	0.728477	0.774648	0.750853	
		NNET	sex+age_group	0.712177	0.651163	0.707792	0.767606	0.736486	
		SVM	sex	0.619926	0.604651	0.638298	0.633803	0.636042	
		RF	age_group	0.616236	0.589147	0.631944	0.640845	0.636364	
MAM_RIS	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET	age_group	0.603509	0.594406	0.600000	0.612676	0.606272	
		NNET	sex+age_group	0.589474	0.573427	0.585034	0.605634	0.595156	
		RF		0.515789	0.503497	0.513699	0.528169	0.520833	
		RF	sex	0.505263	0.475524	0.503311	0.535211	0.518771	
MAM_SAM	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET	sex	0.549828	0.583893	0.540741	0.514085	0.527076	
		RF	age_group	0.529210	0.536913	0.517483	0.521127	0.519298	
		SVM	sex	0.498282	0.597315	0.482759	0.394366	0.434109	
		RF	sex	0.474227	0.496644	0.460432	0.450704	0.455516	
NOR_RIS	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET	sex	0.577206	0.657343	0.562500	0.488372	0.522822	
		NNET		0.566176	0.671329	0.552381	0.449612	0.495726	
		RF	sex+age_group	0.507353	0.545455	0.480000	0.465116	0.472441	
		SVM	sex	0.500000	0.706294	0.454545	0.271318	0.339806	
NOR_SAM	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET	age_group	0.751799	0.785235	0.741935	0.713178	0.727273	
		NNET	sex+age_group	0.730216	0.778523	0.725000	0.674419	0.698795	
		SVM	sex	0.672662	0.671141	0.639706	0.674419	0.656604	
		SVM	sex+age_group	0.669065	0.704698	0.648000	0.627907	0.637795	
RIS_SAM	PROCRUSTES-DISTANCES-COLINEARITY_TRIM	NNET		0.640411	0.697987	0.648438	0.580420	0.612546	
		NNET	age_group	0.626712	0.644295	0.621429	0.608392	0.614841	
		SVM	sex+age_group	0.554795	0.543624	0.543624	0.566434	0.554795	
		SVM	age_group	0.551370	0.577181	0.543478	0.524476	0.533808	

Table 3.12. BEST AND WORST TWO METHODS FOR
PROCRUSTES + DISTANCES + COLINEARITY TRIM
METHODOLOGY.

experiment	method		classification_method	extra_cols	Accuracy	Specificity	Precision	Recall	F1
MAM_NOR	PROCRUSTES-DISTANCES-PCA	NNET	age_group	0.671587	0.627907	0.677852	0.711268	0.694158	
		NNET	sex+age_group	0.645756	0.666667	0.674242	0.626761	0.649635	
		NNET		0.579336	0.627907	0.612903	0.535211	0.571429	
		SVM	sex	0.557196	0.620155	0.591667	0.500000	0.541985	
MAM_RIS	PROCRUSTES-DISTANCES-PCA	NNET	sex	0.554386	0.531469	0.550336	0.577465	0.563574	
		NNET	sex+age_group	0.543860	0.559441	0.543478	0.528169	0.535714	
		SVM	sex	0.480702	0.587413	0.473214	0.373239	0.417323	
		SVM	sex+age_group	0.473684	0.545455	0.467213	0.401408	0.431818	
MAM_SAM	PROCRUSTES-DISTANCES-PCA	NNET	sex	0.512027	0.476510	0.500000	0.549296	0.523490	
		NNET	sex+age_group	0.508591	0.530201	0.496403	0.485915	0.491103	
		RF		0.460481	0.463087	0.448276	0.457746	0.452962	
		RF	sex+age_group	0.457045	0.456376	0.445205	0.457746	0.451389	
NOR_RIS	PROCRUSTES-DISTANCES-PCA	NNET		0.577206	0.587413	0.553030	0.565891	0.559387	
		NNET	sex	0.577206	0.622378	0.557377	0.527132	0.541833	
		RF		0.470588	0.510490	0.440000	0.426357	0.433071	
		RF	sex	0.466912	0.545455	0.429825	0.379845	0.403292	
NOR_SAM	PROCRUSTES-DISTANCES-PCA	NNET	sex+age_group	0.690647	0.691275	0.659259	0.689922	0.674242	
		NNET	age_group	0.687050	0.684564	0.654412	0.689922	0.671698	
		RF	sex	0.607914	0.637584	0.578125	0.573643	0.575875	
		SVM	sex	0.593525	0.590604	0.557971	0.596899	0.576779	
RIS_SAM	PROCRUSTES-DISTANCES-PCA	NNET	age_group	0.595890	0.610738	0.588652	0.580420	0.584507	
		NNET	sex+age_group	0.592466	0.671141	0.598361	0.510490	0.550943	
		RF	sex	0.506849	0.530201	0.496403	0.482517	0.489362	
		RF		0.489726	0.489933	0.479452	0.489510	0.484429	

Table 3.13. BEST AND WORST TWO METHODS FOR
PROCRUSTES + DISTANCES + PCA METHODOLOGY.

4. CONCLUSIONS

4.1. Discussion

This thesis has reviewed the techniques currently used in geometric morphometrics, qualitatively and quantitatively analyzing their performance in the detection of malnutrition in children detecting a misuse of some of its widely used techniques, partly burdened by the software that is normally used.

In particular, it has been detected that the use of Procrustes on landmarks has a distributive effect of the existing correlations between the coordinates, causing that in the resulting procrustes coordinates there are highly correlated variables while others see this value reduced. In the same way, it has been shown that a situation like this entails problems when using classification methods sensitive to high correlations, such as LDA, where the matrix inversion leads to numerical instabilities. As LDA is the classification algorithm chosen by default in most scientific studies, it is possible that many of these have obtained unreliable results.

As a possible solution to this problem, a new technique has been proposed, called Colinearity Feature Trim, which involves the systematic elimination of highly correlated variables when they exceed a previously established threshold. The technique can be used before or after applying Procrustes, but due to its regularizing purpose, it makes more sense to use it after Procrustes. However, this proposal has not shown good results as it has been put into practice, except in some cases, probably because the threshold value has not been adjusted specifically for each case. However, it has shown certain regularizing capacities that make the use of the technique preferable in some cases.

In contrast, the use of allometric reduction has shown the worst results of the entire thesis. In no case the method has been included in any of the best results and on the contrary, it is present in almost all the worst results obtained.

The use of PCA, which is already widely used, has shown good results as it has been used as a dimensionality reducer by reducing the explained variance. However, its use is not systematically better in all cases, although this may be due to the fact that a personalized adjustment has not been made for each specific case. Despite the good results obtained from using the method, its use is not entirely recommended to be applied to classification problems. When making linear combinations of the original variables, the resulting variables are not of direct analysis if what you want is to draw conclusions from the modeling, a very important aspect in the fields where geometric morphometrics are used.

To solve this last problem, the use and calculation of a distance matrix between all the landmarks has been proposed. This technique has been shown to improve the state

of the art in four of the six types of experiments carried out and it's the second best technique (very close to the first one) in the other two experiments. In addition, this new way of approaching the classification problem has an added value by allowing to analyze relationships between landmarks directly. Due to this, it has been possible to demonstrate that the diagnosis of malnutrition using MUAC is made in the place of the arm that provides more information, and if this technique is extended to other parts of the body, it's likely that other areas could be detected where to perform this measurement, possibly with better results.

Regarding the proposed classification algorithms, namely Support Vector Machines, Random Forest and Fully Connected Neural Networks, a systematic superiority of neural networks has been shown. Despite this, and because the three algorithms are universal predictors (they can approximate any function) it's likely that the advantage obtained by neural networks is due to the fact that it has been the only algorithm in which a fine tuning has been carried out, since it is explicitly required. The results presented in this thesis seems to show that the choice of a classification algorithm is a totally dispensable aspect and the choice must be linked to ease of use, since, when performing fine tuning, all algorithms perform the work in a similar way. In this sense, I recommend to a practitioner the systematic choice of random forest as it is the simplest algorithm to use and has shown very close results to neural networks, which are much more difficult to use.

Finally, the combined analysis of all the proposed methodologies has allowed the detection of specific methods for each type of experiment (MAM-NOR, MAM-RIS, MAM-SAM, NOR-RIS, NOR-SAM, RIS-SAM) that improve in all the cases those used in the state of the art of the field. This improvement varies between 2% and 10% of accuracy depending on the case. Likewise, the improvement in the classification of groups that depend on age and sex has improved systematically in all cases, which has made it possible to show cases in which the classification can have a performance close to 100% as in males between 49-59 months in the case of MAM-NOR or in males between 37-48 months in the case of NOR-SAM.

As already mentioned in the introduction, the entire R code used both to carry out the study and to build the graphics is public and freely accessible for use and replication in https://github.com/RicardoHS/geomorph_malnutrition_classification_models. Care has been taken in programming structure so that the methodology followed can be replicated in other fields by making as few changes as possible.

4.2. Future Work

Finally, I want to record some possible future work paths, which either have not been included in this thesis or could be interesting to study.

The main one consists in the transformation of the binary classification problem into a multiclass one. This can be done either by using methods that allow multiclass classification or by grouping classifiers in a hierarchical structure.

Other secondary roads can be: the use of landmarks referring to the whole body and not only the left arm; the use of other modern classification methods that have proven their usefulness repeatedly (such as GradientBoostingTrees); the repetition of the study carried out but with the realization of fine tuning in the models and techniques that have shown good results and finally the construction of classifiers that have the objective of reducing the error in a specific age group or sex.

BIBLIOGRAPHY

- [1] Dean C. Adams, F. James Rohlf, and Dennis E. Slice. Geometric morphometrics: Ten years of progress following the ‘revolution’. *Italian Journal of Zoology*, 71(1):5–16, 2004.
- [2] Brent A. Berger, Vincent A. Ricigliano, Yoland Savriama, Aedric Lim, Veronica Thompson, and Dianella G. Howarth. Geometric morphometrics reveals shifts in flower shape symmetry and size following gene knockdown of cycloidea and anthocyanidin synthase. *BMC Plant Biology*, 17(1):205, Nov 2017.
- [3] James Bewes, Andrew Low, Antony Morphett, F. Donald Pate, and Maciej Henneberg. Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. *Journal of Forensic and Legal Medicine*, 62:40 – 43, 2019.
- [4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
- [5] Robert Challen, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova. Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3):231–237, 2019.
- [6] Irene Chen, Fredrik D Johansson, and David Sontag. Why is my classifier discriminatory? In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 3539–3550. Curran Associates, Inc., 2018.
- [7] Lloyd A. Courtenay, Rosa Huguet, Diego González-Aguilera, and José Yravedra. A hybrid geometric morphometric deep learning approach for cut and trampling mark classification. *Applied Sciences*, 10(1):150, Dec 2019.
- [8] Lloyd A. Courtenay, José Yravedra, Rosa Huguet, Julia Aramendi, Miguel Ángel Maté-González, Diego González-Aguilera, and Mari Carmen Arriaza. Combining machine learning algorithms and geometric morphometrics: A study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 522:28 – 39, 2019.
- [9] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. pages 155–161, 1997.
- [10] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58, 1992.
- [11] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, Mar 1975.
- [12] Hayley Green and Darren Curnoe. Sexual dimorphism in southeast asian crania: A geometric morphometric approach. *Homo : internationale Zeitschrift für die vergleichende Forschung am Menschen*, 60:517–34, 10 2009.

- [13] Rebecca M. Green, Weiguo Feng, Tzulip Phang, Jennifer L. Fish, Hong Li, Richard A. Spritz, Ralph S. Marcucio, Joan Hooper, Heather Jamniczky, Benedikt Hallgrímsson, and Trevor Williams. Tfap2a-dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in fgf8 gene dosage. *Disease Models & Mechanisms*, 8(1):31–43, 2015.
- [14] Emmanuel Grellety and Michael H. Golden. Weight-for-height and mid-upper-arm circumference should be used independently to diagnose acute malnutrition: policy implications. *BMC Nutrition*, 2(1):10, Feb 2016.
- [15] Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1):195, Oct 2019.
- [16] Chris Klingenberg. Morphoj: An integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11:353–357, 03 2011.
- [17] Giuseppina Laganà, Nicolò Venza, Valeria Paoloni, Fabio Bertoldo, Giovanni Ruvolo, and Paola Cozza. A 3d geometric morphometric analysis of the palatal morphology in marfan’s syndrome: A preliminary study. *Journal of Clinical and Diagnostic Research*, 12:ZC14–ZC17, 01 2018.
- [18] Agostina J. Larrazabal, Nicolás Nieto, Victoria Peterson, Diego H. Milone, and Enzo Ferrante. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences*, 117(23):12592–12594, 2020.
- [19] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.
- [20] Laura Medialdea, Iván Molina Allende, and Antonio Vargas Brizuela. Development of a sam photo diagnosis app. 57:66, March 2018.
- [21] Laura Medialdea, Cayetana Bazaco, Manuel Domingo D’Angelo del Campo, Carlos Sierra-Martínez, Rolando González-José, Antonio Vargas, and María Dolores Marrodán. Describing the children’s body shape by means of geometric morphometric techniques. *American Journal of Physical Anthropology*, 168(4):651–664, 2019.
- [22] Mathieu Quenu, Steven A Trewick, Fabrice Brescia, and Mary Morgan-Richards. Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia. *Journal of Molluscan Studies*, 86(1):35–41, 02 2020.
- [23] F. Rohlf and Leslie Marcus. A revolution in morphometrics. *Trends in ecology evolution*, 8:129–32, 04 1993.
- [24] A. Rosas and M. Bastir. Thin-plate spline analysis of allometry and sexual dimorphism in the human craniofacial complex. *Am. J. Phys. Anthropol.*, 117(3):236–245, Mar 2002.

- [25] Frédéric Santos, Pierre Guyomarc'h, and Jaroslav Bruzek. Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic Science International*, 245:204.e1 – 204.e8, 2014.
- [26] Katrin Schaefer, Philipp Mitteroecker, Philipp Gunz, Markus Bernhard, and Fred L. Bookstein. Craniofacial sexual dimorphism patterns and allometry among extant hominids. *Annals of Anatomy - Anatomischer Anzeiger*, 186(5):471 – 478, 2004.
- [27] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.
- [28] Ian D. Stephen, Vivian Hiew, Vinet Coetzee, Bernard P. Tiddeman, and David I. Perrett. Facial shape analysis identifies valid cues to aspects of physiological health in caucasian, asian, and african populations. *Frontiers in Psychology*, 8:1883, 2017.
- [29] D. Stoyan. Bookstein, f. l.: Morphometric tools for landmark data. geometry and biology. cambridge university press, cambridge/new york/port chester/melbourne/sydney 1991, xviii, 435 s., £ 50.00; \$ 89.95. *Biometrical Journal*, 35(4):512–512, 1993.
- [30] Bert. Van Bocxlaer and Roland. Schultheiß. Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology*, 36(3):497–515, 08 2010.
- [31] Wen Yang, Xiaoning Liu, Kegang Wang, Jiabei Hu, Guohua Geng, and Jun Feng. Sex determination of three-dimensional skull based on improved backpropagation neural network. *Computational and Mathematical Methods in Medicine*, 2019:1–8, 01 2019.

ANNEXES

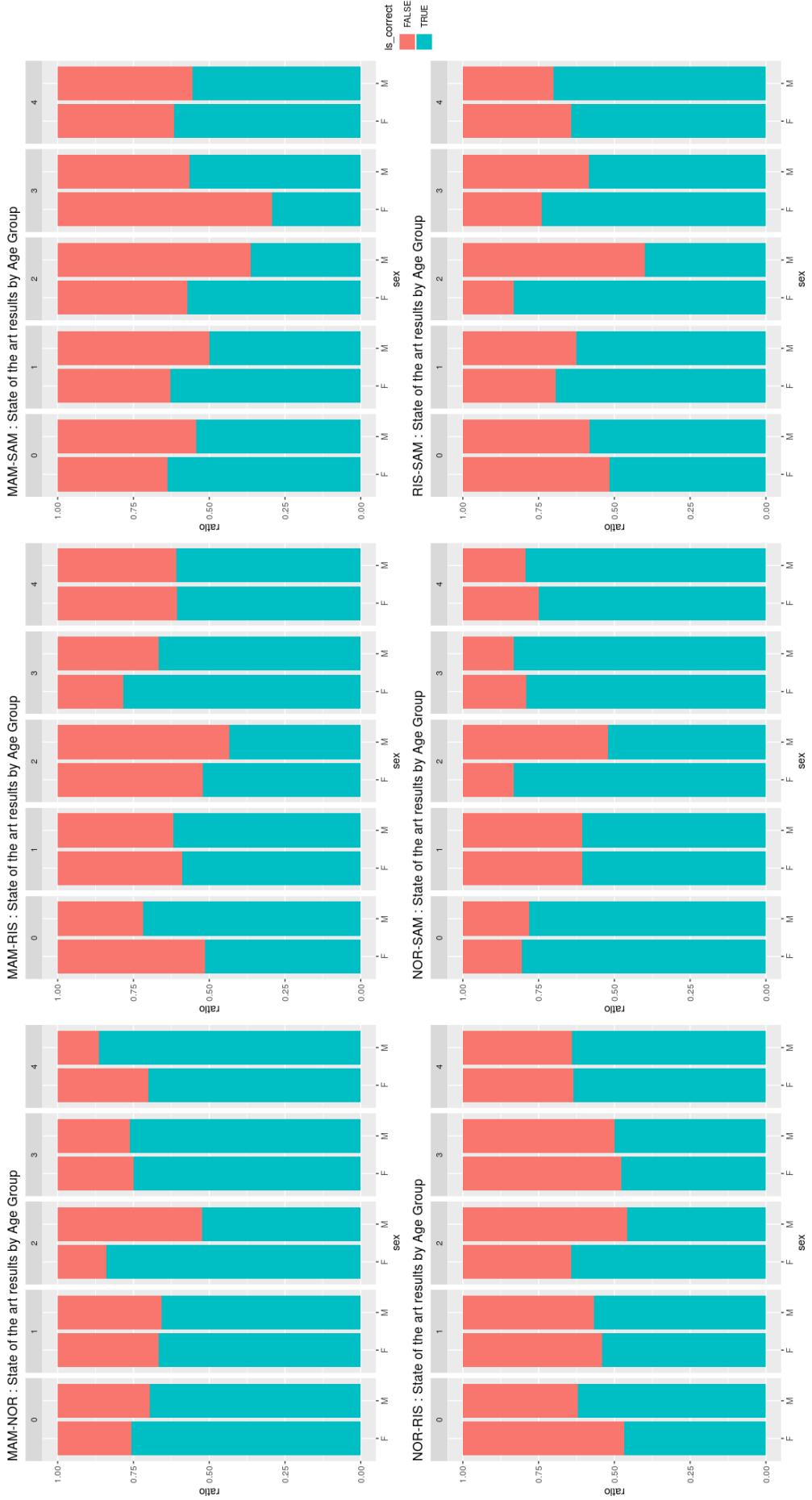


Fig. 4.1. Performance comparison of state of the art methods according to age group and sex.

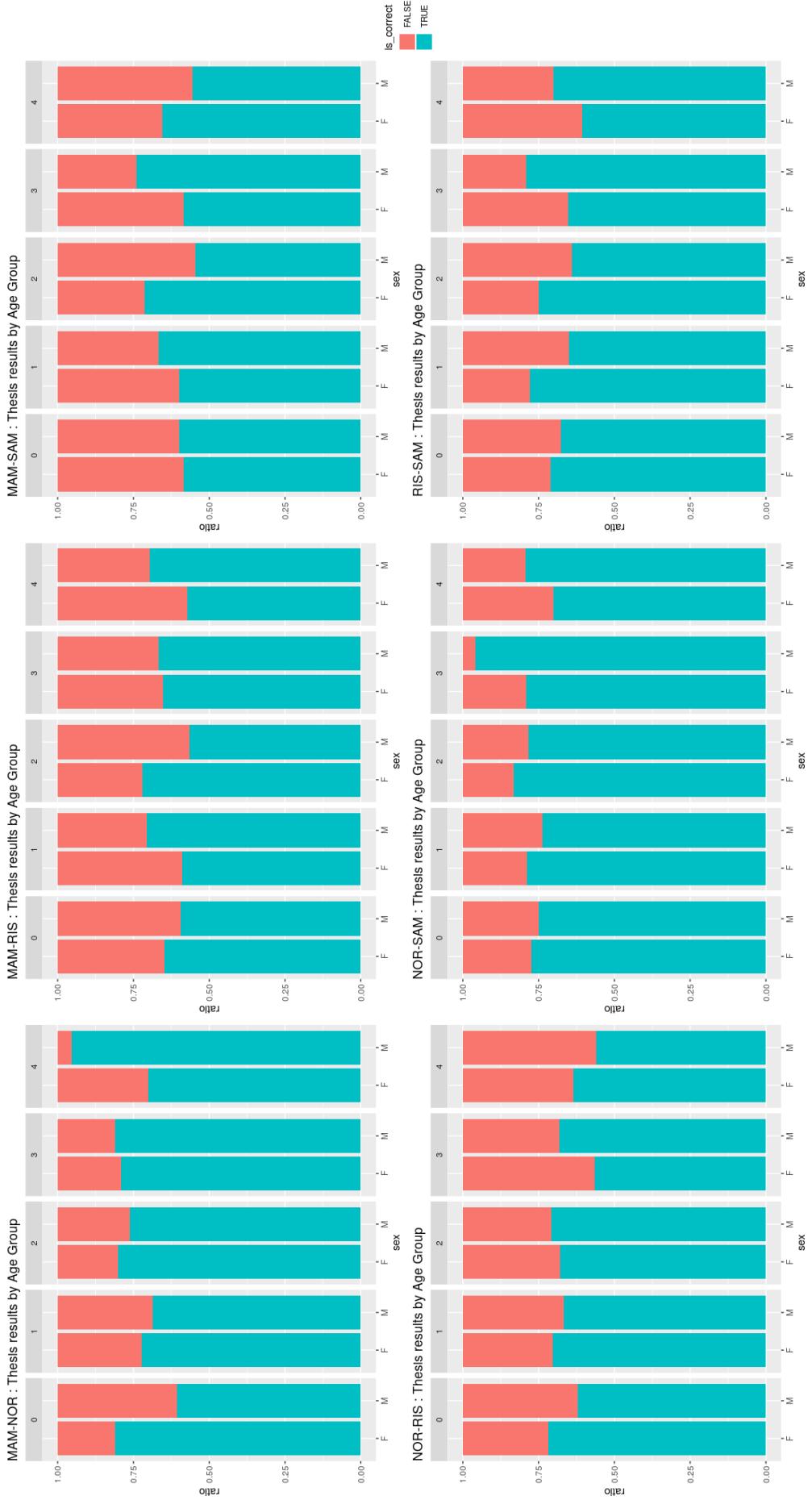


Fig. 4.2. Performance comparison of thesis methods according to age group and sex.