Nonlinear Machine Learning Model for Sex Prediction Using Lumbar Spine Geometry

Reza Rahimi[1], Allison Clouthier [1]
[1]School of Human Kinetics, University of Ottawa, Ottawa, Canada.
Email: rrahi080@uottawa.ca

Summary

We developed and verified a machine-learning algorithm to examine the geometric differences in the lumbar spine based on sex. The model was designed for binary classification, using a nonlinear decision boundary to capture complex relationships in the data. To prevent overfitting and achieve an optimal balance between bias and variance, robust cross-validation techniques were implemented. The suggested supervised learning strategy surpassed both statistical shape modelling and random labelling techniques, achieving higher mean accuracy, F1 score, precision, and recall. Additionally, this model allows for the development of a morphable model, facilitating in-depth investigations into the connections between spinal shape, sex, and performance.

Introduction

Understanding the differences in 3D morphology based on demographic factors is crucial, as these variations can affect joint function and dysfunction. Statistical shape modeling (SSM) has been widely employed to investigate 3D geometric variations, as demonstrated by Tang et al. [2] and Clouthier et al. [1], who highlighted its effectiveness in quantifying the relationship between individual characteristics, such as sex, and lumbar spine geometry. However, traditional SSM techniques, including Principal Component Analysis (PCA) and Partial Least Squares (PLS), are constrained by their reliance on assumptions of linear correlations and Gaussian data distributions. These limitations can result in inaccurate or unrealistic representations when applied to complex, nonlinear datasets. To overcome these constraints, we created a machine learning algorithm that identifies non-linear connections in the geometry of the lumbar spine. This leads to more precise classification based on gender.

Methods

A machine learning algorithm was created by analyzing computed tomography (CT) scans of 87 individuals [1]. These images were converted into three-dimensional models of the spinal column's vertebrae in 3D Slicer software. Subsequently, a Procrustes alignment technique was used to align the lumbar spines (L2-L5). The XGBoost model identified the most important features for binary classification based on sex. These selected landmarks (Fig.1) were used to reconstruct the meshes and facilitate classification. Due to the high-dimensional and nonlinear nature of the reconstructed meshes, a support vector machine (SVM) [3] with a nonlinear decision boundary was trained to classify the meshes based on sex. The model's performance was assessed through stratified k-fold cross-validation. The model's effectiveness was comprehensively evaluated using metrics like mean accuracy, F1 score, precision, and recall, while consistency was assessed through variance and standard deviation analysis. Overfitting was monitored via the learning curve from cross-validation.

Results and Discussion

The proposed model effectively captured data nonlinearity through a nonlinear decision boundary for binary classification, achieving a mean accuracy of 84.71%, an F1 score of 84.17%, a precision of 88.25%, and a recall of 81.39%. It exhibited low variance in accuracy, precision, recall, and F1 score, with values of 0.0078, 0.0100, 0.0131, and 0.0078, respectively, indicating reliability and consistency despite the small sample size. The learning curve (Fig. 1) shows high convergence of training and validation accuracies with a minimal gap, highlighting the model's strong generalization and reduced risk of overfitting. These results demonstrate the model's robustness in classifying high-dimensional, nonlinear lumbar spine data by sex.

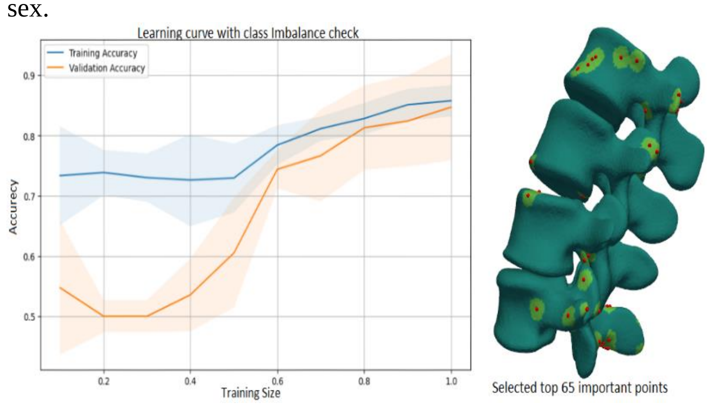

Figure 1: Learning curve Stratified K-fold cross validation and extracted top important points

Conclusions

The proposed model accurately classifies lumbar spine geometry by sex, leveraging a nonlinear decision boundary to capture complex variations in the data. It outperformed previous linear methods, highlighting the importance of capturing nonlinearity in shape analyses. Additionally, it enables a morphable framework for studying sex-related geometric differences and their links to demographics and function.

References

[1] Clouthier, A, et al. J Biomech, **146:** 111421, 2023.

[2] Tang, L, et al. J Biomech, **130**: 110821, 2022.

[3] Gareth, J, et al. (2023). An introduction to statistical learning with application in Python; Spring.