Predicting Spinal Condition using various biomechanical attributes of the pelvis and lumbar spine

 ABSTRACT

Purpose—To compare rates of spinal condition (normal, disk hernia, and spondylolisthesis) classification of machine learning algorithms trained on six biomechanical features of orthopedic patients.

Methods— Three-hundred sixty patients can be categorized into normal vertebral structure, individuals with a disk hernia, and individuals with spondylolisthesis. Six biomechanical attributes that are derived from the shape and orientation of the pelvis and lumbar spine were utilized in random forest and support vector machine models. Algorithms based from random forest models and support vector machine models were used to categorize individuals into one of the three spinal classifications. Five-fold cross validation was utilized to train and test the SVM and RF algorithm and subsequently produce corresponding CV errors and standard errors. The main goal was to obtain a predictive model with the best cross validation errors and cross validation standard errors.

Results— An SVM with a linear kernel performed the best on our dataset. AUC was the final measure used to select the best model.

Conclusion—The results of this study indicate that both the SVM and RF models have achieved a relatively satisfying level of precision in classifying spinal conditions. These machine learning algorithms show great potential to serve as an assistance for physicians in diagnosis.

INTRODUCTION

Machine learning applications in orthopedic medicine have become increasingly more prevalent in experimental research. While the application of machine learning must be assessed in structured frameworks, similar to extensive clinical trials, its exploratory benefits have shown to be helpful (Cabitza et. al, 2018). Specifically, previous research utilizing random forests in classifying orthopedic conditions allowed for a clearer clinical interpretability of result (Kotti et. al, 2017). Moreover, random forests models accompanied with the ability to utilize multiple decision trees, have been used for several applications in spine research (Galbusera et. al, 2019). Most closely related to this research is the use of decision trees to guide referral of a patient to a physiotherapist for lower back pain symptoms (Nijeweme-d'Hollosy et. al, 2016). Additionally, support vector machines (SVMs) can be powerful tools in orthopedic contexts. In spine research, SVMs have previously been used to accurately categorize grading of the degeneration of the spinal disks, in addition to scoliosis curve severity (Oktay & Akgul, 2013; Seoud et. al, 2010).

Spondylolisthesis is "a unilateral or bilateral defect (fracture or separation) in the vertebral pars interarticularis, usually in the lower lumbar vertebrae", while the symptoms include "pain that spreads across the patient's lumbar region and radiates into posterior legs" (D'Hemecourt, 2021). Due to lower back pain being a common complaint and symptom of various conditions, spondylolisthesis is often misdiagnosed by physicians (Syrmou et. al, 2010). Similarly, disk herniation presents as back pain, and is commonly misdiagnosed. Both conditions can be extremely painful and contribute to chronic back pain if mis-identified and treatment fails. Moreover, a study from Johns Hopkins Hospital reported that 40-80% of chronic pain patients are misdiagnosed, thus indicating the value in more research aiming to accurately diagnose conditions that result in chronic pain (Hendler, 2016).

The purpose of this study is to accurately categorize spinal osteoarthritis conditions into the categories: normal, disk herniation, and spondylolisthesis based on biomechanical attributes of the shape and orientation of the pelvis and lumbar spine. Additionally, this study seeks to compare the accuracy of categorizations between a support vector machine and a random forest model. An accurate machine learning algorithm could serve as an assistance tool for physicians to consult while considering a diagnosis, with the hopes of reducing misdiagnosis. The reduction of misdiagnoses could lead to more efficient treatment plans and the reduction of pain. Ultimately, assisting in the determination of treatment as surgical or non-surgical would be beneficial.

METHODS

These data were obtained from a selection of three-hundred sixty patients during the medical residence period of Dr. Henrique da Mota, in the Group of Applied Research in Orthopedics (GARO), at the Centre Medico-Chirurgical de Réadaptation des Massues in Lyon, France. The patients were volunteers that were asymptomatic for Disk Hernia and Spondylolisthesis. These three-hundred sixty patients can be categorized into normal vertebral structure and abnormal vertebral structure. The abnormal categorization can be further split into individuals who have a disk hernia and individuals with spondylolisthesis. In this paper, the goal was to create an algorithm to accurately predict individuals into one of the three categorizations for vertebral conditions. There were 100 patients categorized to have normal vertebral structures, 60 patients to show evidence of a disk hernia, and 150 patients with spondylolisthesis.

The predictors for our models include six biomechanical attributes that are derived from the shape and orientation of the pelvis and lumbar spine. The biomechanical attributes include pelvic incidence, pelvic tilt, pelvic radius, sacral angle, lumbar lordosis angle, and grade of spondylolisthesis.

Pelvic incidence is defined as the angle between a line perpendicular to the sacral plate at its midpoint and a line connecting this point to the femoral head axis (Boulay et. al., 2014). Pelvic incidence was demonstrated to be a key factor for regulation of spinal sagittal curves, thus making it of interest as a predictor for spinal conditions (Legaye et. al, 1998).

Pelvic tilt measures the degree to which the pelvis is rotated forward such that the back of the pelvis rises. Research has suggested that pelvic tilt is related to posture, which is an interest in the development of spinal conditions (Day et. al, 1984).

The lumbar lordosis angle is the angle formed between the long axis of the lumbar vertebrae and the sacrum. It is closely tied to posture, thus indicating its potential to predict spinal conditions.

Sacral slope is the angle between the horizontal and sacral plate, specifically shown to be a necessary spinal parameter in the study of sagittal balance and progression of spondylolisthesis (Drazin et. al, 2015).

Pelvic radius, the distance between the hip axis and the posterior superior corner of the S1 endplate, has been studied to be effective in the assessment of spinal sagittal alignment for degenerative spondylolisthesis. Thus, indicating that pelvic radius may be a useful predictor in the prediction algorithms (Zhou et. al, 2018).

The grade of spondylolisthesis is determined by measuring how much of a vertebral body has slipped forward over the body beneath it. This is of interest in categorization of spinal conditions, as a higher grade would indicate a more severe spinal condition.

A random forest model (RF) was compared with two support vector machine models (SVM), with the goal of selecting the model with the better performance. To select the model with the best performance, observations forming the training and testing sets were chosen randomly. The random selection of participants to form training and testing sets will assist in obtaining better error estimation for categorizing the individuals into normal, disk hernia, and spondylolisthesis conditions.

The random forest model utilized a grid search technique to select the best parameter values. The grid search considered 50, 250, and 500 trees in combination with 2, 3, and 6 predictors at the random forest splits. The best tuning parameters for a given training set were selected using the out-of-bag mean squared error in the training set. Five-fold cross validation was utilized to train and test the RF algorithm and subsequently produced corresponding CV errors and standard errors. The support vector machine models were trained with a linear kernel and a radial basis kernel, respectively. The grid of parameters that were tested included the epsilon parameter with values 0, 0.25, 0.50, 0.75, and 1.00, coupled with the cost parameter with values of 1 to 5. Five-fold cross validation was utilized to train and test the SVM algorithm and subsequently produced corresponding CV errors and standard errors. The main goal was to obtain a predictive model with the best cross validation errors and cross validation standard errors. All model tuning and testing was completed in R version 4.0.3 (R Core Team, 2020).

## RESULTS

<u>Summary Statistics</u>

Table 1

Characteristic	N	Hernia, N = 60[1]	Normal, N = 100[1]	Spondyl., N = 150[1]	p-value[2]
pelvic_incidence	310	48 (11)	52 (12)	72 (15)	<0.001
pelvic_tilt	310	17 (7)	13 (7)	21 (12)	<0.001
lumbar_lordosis_angle	310	35 (10)	44 (12)	64 (16)	<0.001
sacral_slope	310	30 (8)	39 (10)	51 (12)	<0.001
pelvic_radius	310	116 (9)	124 (9)	115 (16)	<0.001
degree_spondylolisthesis	310	2 (6)	2 (6)	52 (40)	<0.001

[1]Mean (SD)
[2]One-way ANOVA

Correlation Plots of each Variable

Figure 1
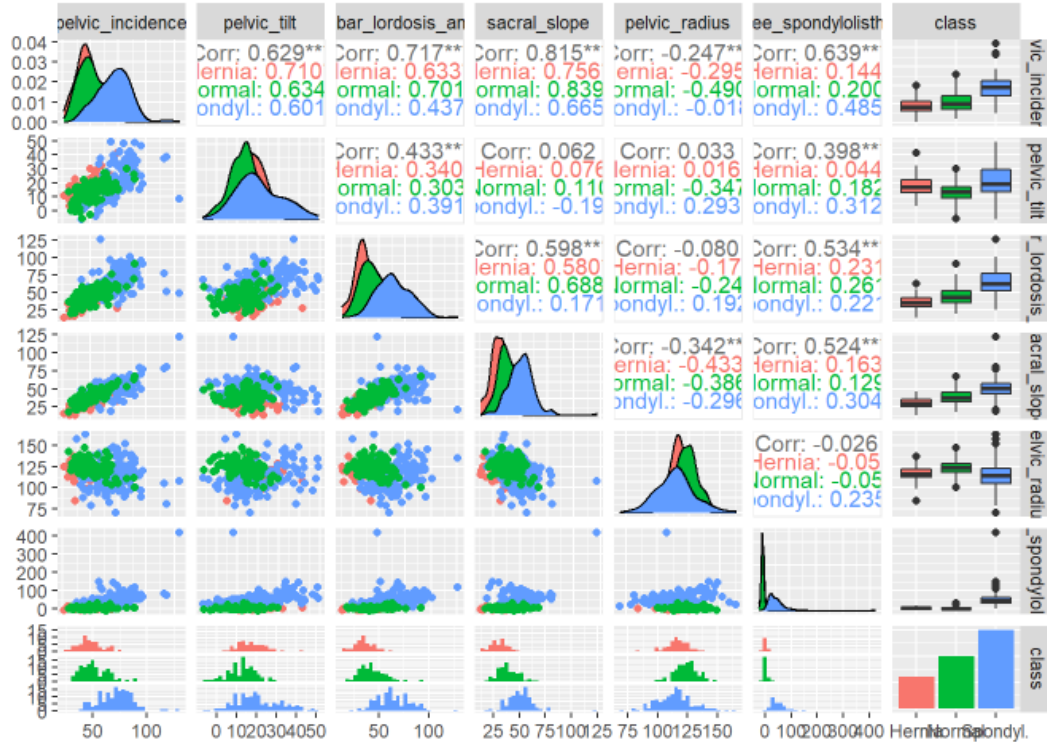
Per Class Error and Standard Error for each Algorithm

Table 2

	Hernia	Normal	Spondylolisthesis
SVM (Linear)	0.333(0.167)	0.170(0.057)	0.047(0.018)
SVM (RBF)	0.283(0.095)	0.180(0.110)	0.060(0.037)
Random Forest	0.433(0.199)	0.230(0.104)	0.040(0.043)

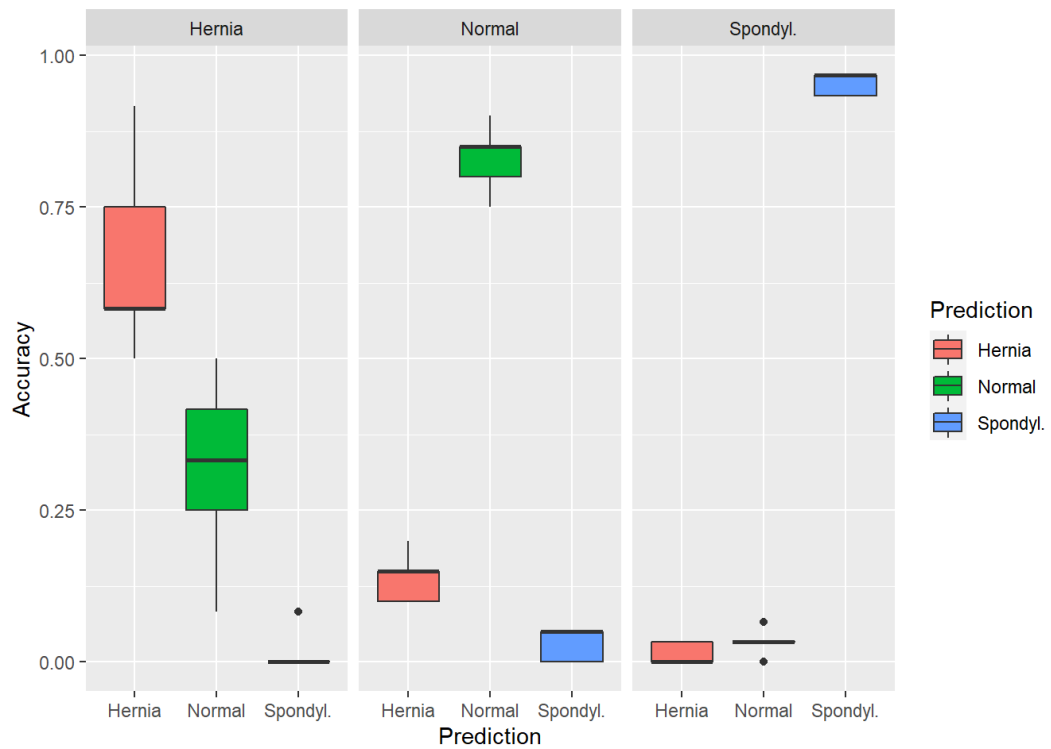(1) Error (Standard Error)

AUC for each Algorithm

Table 3

	AUC
SVM (Linear)	0.952
SVM (RBF)	0.944
Random Forest	0.923

Accuracy Boxplots of Correctly Classifications for each Algorithm

SVM (Linear)

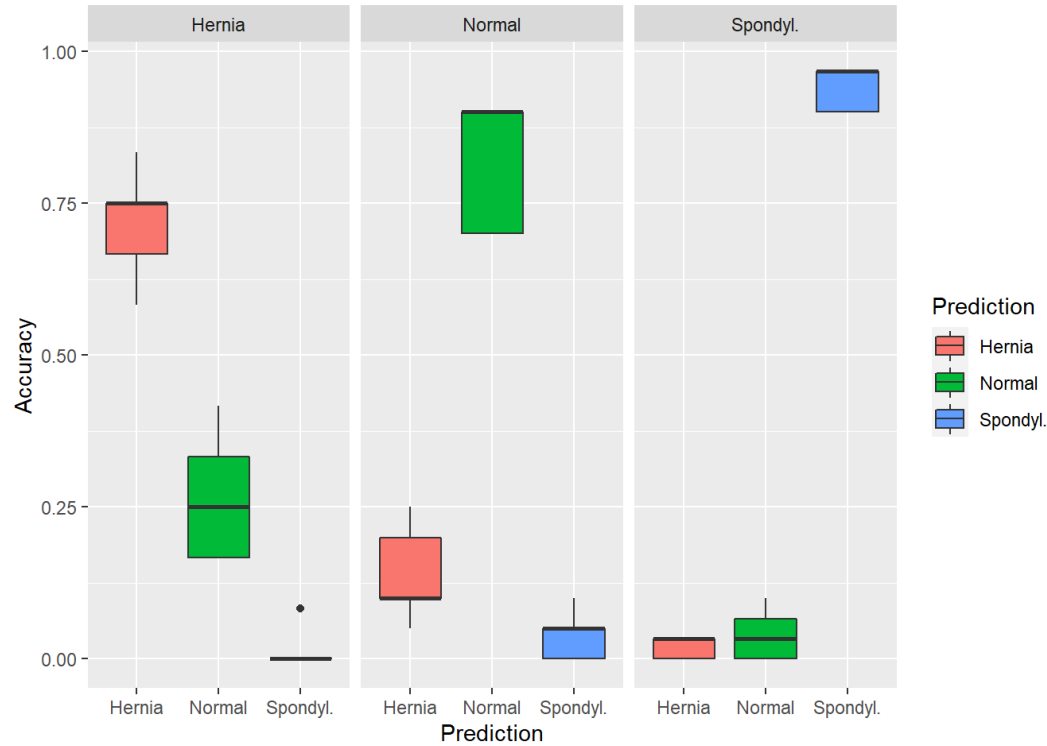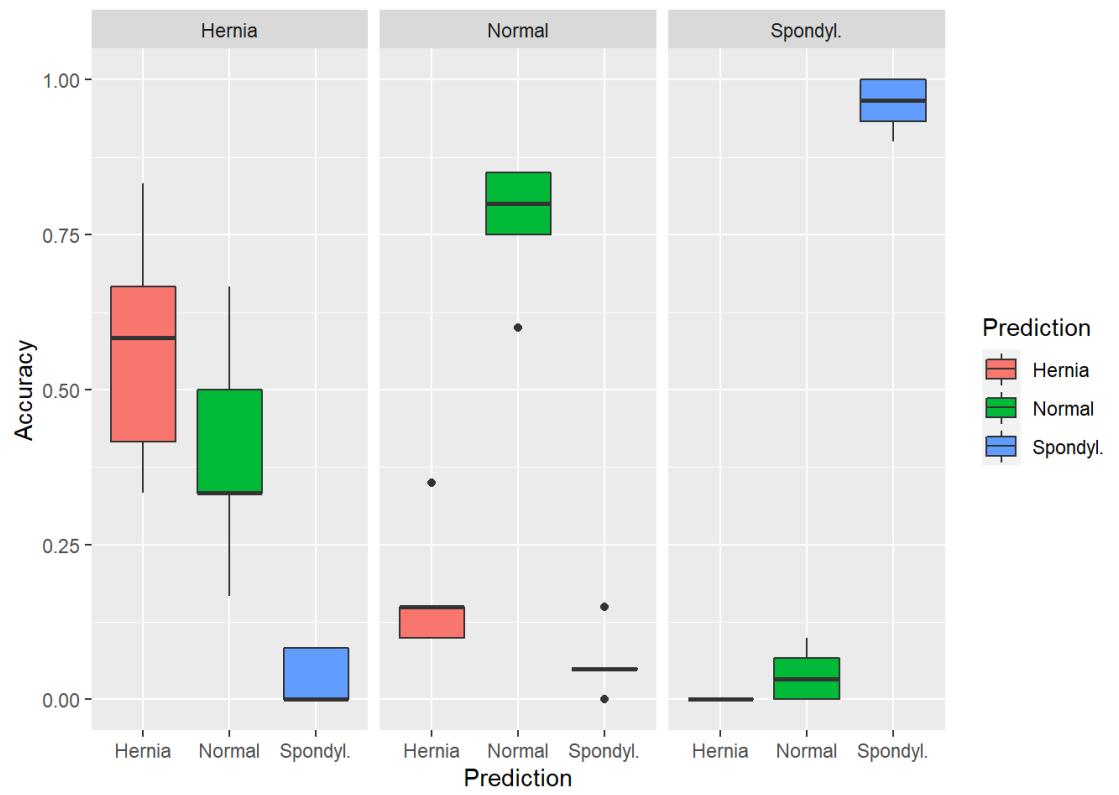**Figure 2**

SVM (RBF)

Figure 3

Random Forest

Figure 4

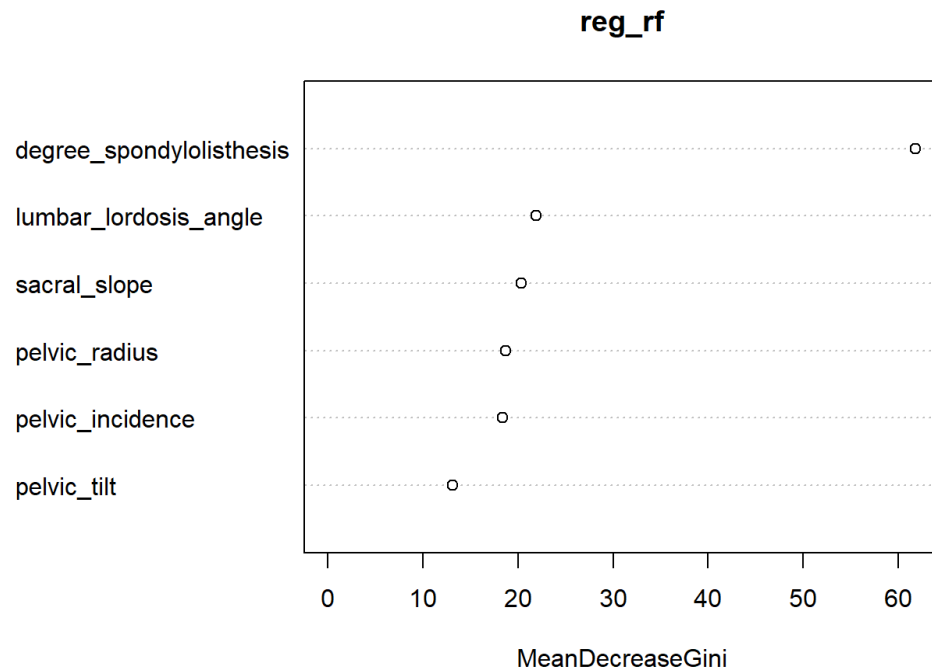Example of Variable Importance Plot from Random Forest

Figure 5

Table 1 shows summary statistics of the data including each features' (all numeric) mean and standard deviation organized by spinal condition class. Figure 1 displays correlation plots by each feature in the dataset (including the outcome class). Some of the features, based on plots alone, appear to be somewhat correlated. The max correlation between two features is 0.815. In addition, looking at the distribution plots by outcome class of each feature, some patterns appear. Most notably with the pelvic incidence, lumbar lordosis angle, sacral slope and the grade of spondylolisthesis. This potentially indicates that these features are important when predicting the outcome. Table 2 displays the average error and standard error of each outcome class by each algorithm. From this table alone, all errors and standard errors appear relatively similar from class to class. Regarding the largest outcome class, spondylolisthesis, Random Forest performs the best with the lowest average error, and linear SVM has the lowest standard error. In Table 3 the multiclass AUC is displayed for each algorithm. The linear SVM has the largest AUC of 0.952. Given sensitivity and specificity are not exactly available with a multiclass outcome, using AUC as our measure of best performance is acceptable. "Multiclass.roc" from the pROC R package was used to calculate AUC, and since ROC/AUC are typically calculated from binary classification, this method instead uses the Hand-Till method to calculate AUC. (Hand et al 2001) Hand-Till do this by averaging AUC over all pairs of classes. It certainly isn't ideal, but is likely the closest way we can calculate AUC, and find a sound measure of performance given the performance of all 3 algorithms are similar.

In Figure 2, 3, and 4 boxplots of accuracy by the actual class on top and predicted class on bottom appear. Boxplots are calculated using accuracies of each class per fold. Unfortunately, increasing fold count decreased performance, so there are only 5 observations per boxplot. Every figure has the highest accuracies for correctly classifying the outcome, as desired. Based solely off the plots, the least variance and highest accuracies appear from the linear SVM plot (figure 2). Though all algorithms appear to perform well. Lastly, in figure 5, there is an example of a variable importance plot calculated from one fold of the Random Forest model. This plot shows the mean decrease in the Gini importance for each variable in the model. This metric measures a variable's mean decrease in node impurity weighted by the proportion of samples that reach a given node in a tree in random forest. Essentially, this measures the importance of a variable across all trees in the forest. The higher the mean decrease in Gini importance, the more important the variable is to the RF model. In all folds of the RF model, the degree of spondylolisthesis had the highest mean decrease in Gini importance, and was the most important variable for the model by this metric.

There was some additional concern about using degree_spondlolisthesis as a feature. Due to being short on time and the time that SVM takes to run, we ran an additional Random Forest model without degree_spondylolisthesis. AUC decreased to 0.861, and the standard errors of the errors increased slightly (errors were wholly slightly larger). In addition, in 4/5 folds, lumbar_lordosis_angle had the largest MeanDecreaseGini, which indicates that this feature is also important. This short analysis validates the importance of degree_spondylolisthesis as a feature. In its absence though, the model could still perform relatively well.

DISCUSSION

This study tests and compares the performance of different machine learning algorithms on the prediction of spinal condition classification. Our results indicate that all three algorithms that were tested in the study (SVM-linear, SVM-RBF, and RF), although they differ in precision on different spinal conditions, show very promising performance.

All three algorithms performed the best in terms of prediction accuracy on the spondylolisthesis condition, second on the normal condition, and the lowest on the disk hernia condition. The high level of accuracy on spondylolisthesis is one of the promising results applying machine learning algorithms on spinal condition classification, as it is very commonly misdiagnosed by physicians (Syrmou et. al, 2010). Although the accuracy level of disk hernia is the lowest among the three conditions, we were still able to get an above 60% accuracy across all three algorithms. One potential explanation for the low prediction accuracy on disk hernia could be that it can occur in any part of the spine and people can have very different symptoms varying from pain to weakness to numbness, which makes it very difficult to diagnose as it shares some common symptoms with other spinal conditions. Overall, the three algorithms tested in our study showed promising reliability in prediction precision across the three spinal conditions.

However, this study is not without limitations. One of the limitations is that among the six predictors we used, some of them show a high level of collinearity (eg., pelvic incidence and

sacral slope, see Figure 1). This could have the potential to reduce the precision of our estimation and classification, which could weaken the statistical power of our models. Some potential treatments include, for example, testing the significance level of collinearity, and/or using a subset of predictors instead of all six to clear out potential collinearity.

Another limitation of the study is that one of the predictors, the grade of spondylolisthesis, appeared to have a notably higher mean decrease in the Gini importance compared to the other predictors for the RF model (see Figure 5). It might have been a predictor that dominated the classification process and has the potential to bias our estimation.

Overall, our study showed some promising results of the application of machine learning algorithms on the prediction of spinal condition classification. While in practice these conditions could be easily misdiagnosed by physicians, we believe that these machine learning algorithms, whilst not perfect, have the potential to serve as a reliable additional assistance tool for physicians to consult while considering a diagnosis.

CONCLUSION

In conclusion, the results of this study indicate that both the SVM and RF models have achieved a relatively satisfying level of precision in classifying spinal conditions. These machine learning algorithms show great potential to serve as an assistance for physicians in diagnosis. Based on our current study, we recommend that future studies develop more condition specific classification algorithms (as opposed to a one-for-all-spinal-conditions classification), and be more cautious on predictions selection.

REFERENCES

Boulay C , Bollini G, Legaye J, Tardieu C, Prat-Pradal D, Chabrol B, Jouve J, Duval-Beaupère G, Pélissier J, (2014). Pelvic Incidence: A Predictive Factor for Three-Dimensional Acetabular Orientation—A Preliminary Study, *Anatomy Research International*, vol. 2014, Article ID 594650, 9 pages https://doi.org/10.1155/2014/594650

Cabitza F, Locoro A, Banfi G. (2018, Jun). Machine Learning in Orthopedics: A Literature Review. *Front Bioeng Biotechnol*. 6:75. doi:10.3389/fbioe.2018.00075

Day J, Smidt G, Lehmann T, (1984, April). Effect of Pelvic Tilt on Standing Posture, *Physical Therapy*, Volume 64, Issue 4, Pages 510–516, https://doi.org/10.1093/ptj/64.4.510

D'Hemecourt P, Micheli L. (2021) Spondylolysis and spondylolisthesis in child and adolescent athletes: Clinical presentation, imaging, and diagnosis. In: UpToDate, Post, TW (Ed).

Drazin D, Hussain M, Harris J, Hao J, Phillips M, Kim T, Johnson J, & Bucklen B, (2015). The role of sacral slope in lumbosacral fusion: a biomechanical study, *Journal of Neurosurgery: Spine SPI*, *23*(6), 754-762. Retrieved Apr 30, 2021, from https://thejns.org/spine/view/journals/j-neurosurg-spine/23/6/article-p754.xml

Galbusera F, Casaroli G, Bassani T. (2019, Mar). Artificial intelligence and machine learning in spine research. *JOR Spine*. 2019;2(1):e1044. doi:10.1002/jsp2.1044

Hand, D.J., Till, R.J. (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45, 171–186. https://doi.org/10.1023/A:1010920819831

Hendler N, (2016). Why Chronic Pain Patients are Misdiagnosed 40 to 80% of the Time?. Journal on Recent Advances in Pain. 2. 94-98. 10.5005/jp-journals-10046-0051.

Kotti M, Duffell L, Faisal A, McGregor A, (2017). Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med. Eng. Phys.* 43, 19–29. 10.1016/j.medengphy.2017.02.004

Legaye J, Duval-Beaupère G, Hecquet J, Marty C. (1998). Pelvic incidence: a fundamental pelvic parameter for three-dimensional regulation of spinal sagittal curves. Eur Spine J ;7(2):99-103. doi: 10.1007/s005860050038. PMID: 9629932; PMCID: PMC3611230.

Nijeweme-d'Hollosy W, van Velsen L, Soer R, Hermens H. (2016) Design of a web-based clinical decision support system for guiding patients with low back pain to the best next step in primary healthcare *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016)*. Cham: Springer International Publishing

Oktay A, Akgul Y. (2013). Simultaneous localization of lumbar vertebrae and intervertebral discs with SVM-based MRF. *IEEE Trans Biomed Eng*;60:2375-2383

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna,

Seoud L, Adankon M, Labelle H, Dansereau J, Cheriet F. (2010) Prediction of scoliosis curve type based on the analysis of trunk surface topography *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE); 10.1109/ISBI.2010.5490322.

Syrmou E, Tsitsopoulos PP, Marinopoulos D, Tsonidis C, Anagnostopoulos I, Tsitsopoulos PD. (2010) Spondylolysis: a review and reappraisal. *Hippokratia*.;14(1):17-21.

UCI Machine Learning. (2017). Biomechanical Features of Orthopedic Patients. Retrieved April 04, 2021. From https://www.kaggle.com/uciml/biomechanical-features-of-orthopedic-patients

Zhou Z, Hou C, Li D, Si J, (2018). The pelvic radius technique in the assessment of spinopelvic sagittal alignment of degenerative spondylolisthesis and lumbar spinal stenosis, Journal of Orthopaedic Science, Volume 23, Issue 6, Pages 902-907, ISSN 0949-2658, https://doi.org/10.1016/j.jos.2018.06.021.