Using Machine Learning to Diagnose Orthopedic Patients

Will McCullough

2023-09-01

Introduction

Machine learning is a powerful set of tools which can be used across a wide range of fields, including economics, public public policy, marketing, and particularly healthcare. Discovering insights where theory and research alone would not predict, machine learning is beneficial for uncovering the symptoms or issues leading to the correct diagnosis of a patient's health issue. One particularly complex issue is diagnosing orthopedic patients with back pain, who may be suffering from any number of similarly presenting issues, requiring different treatment methods.

This research attempts to diagnose, or classify, orthopedic patients by their biomechanical features, or physical measurements, as either having a healthy or unhealthy spine, followed by diagnosing the particular injury the patient is suffering from either a herniated disc, Spondylolisthesis, or is healthy.

The biomechanical predictors being assessed in this research are:

- Pelvic Incidence: The angle of the lower vertebrae in relation to the base of the hips, a fixed measurement in degrees. (1)
- Pelvic Tilt: The degree which an individual's hips are tilted forward, or sometimes backward relative to having hips perpendicular to the floor.
- Lumbar Lordosis Angle: The angle made by completing a triangle between the first and the fifth lower vertebrae, considered normal within a range of 20-45 degrees generally. (2)
- Sacral Slope: A measure related to pelvic incidence, the sacral slope is between the sacral end plate and a horizontal line. (3)
- Pelvic Radius: the distance from the hip axis to the upper rear corner of the sacral endplate. (4)
- Degree of Spondylolisthesis: The degree to which the lower vertebrae is out of alignment with the tailbone.

Physiology tells us that any deviation in these metrics will likely spell trouble for a patient's back, ultimately leading to pain and mobility issues. Our first model assesses the specific metrics which most influence someone to have an abnormal lower back and require intervention. The second model further assesses the combination of abnormal metrics which would classify someone as having a herniated disc, spondylolisthesis, or having no issues. This research is beneficial to the medical field by enhancing care providers' ability to accurately asses patients and correctly diagnose them, allowing patients to begin a treatment plan and recover sooner.

Methods and Analysis

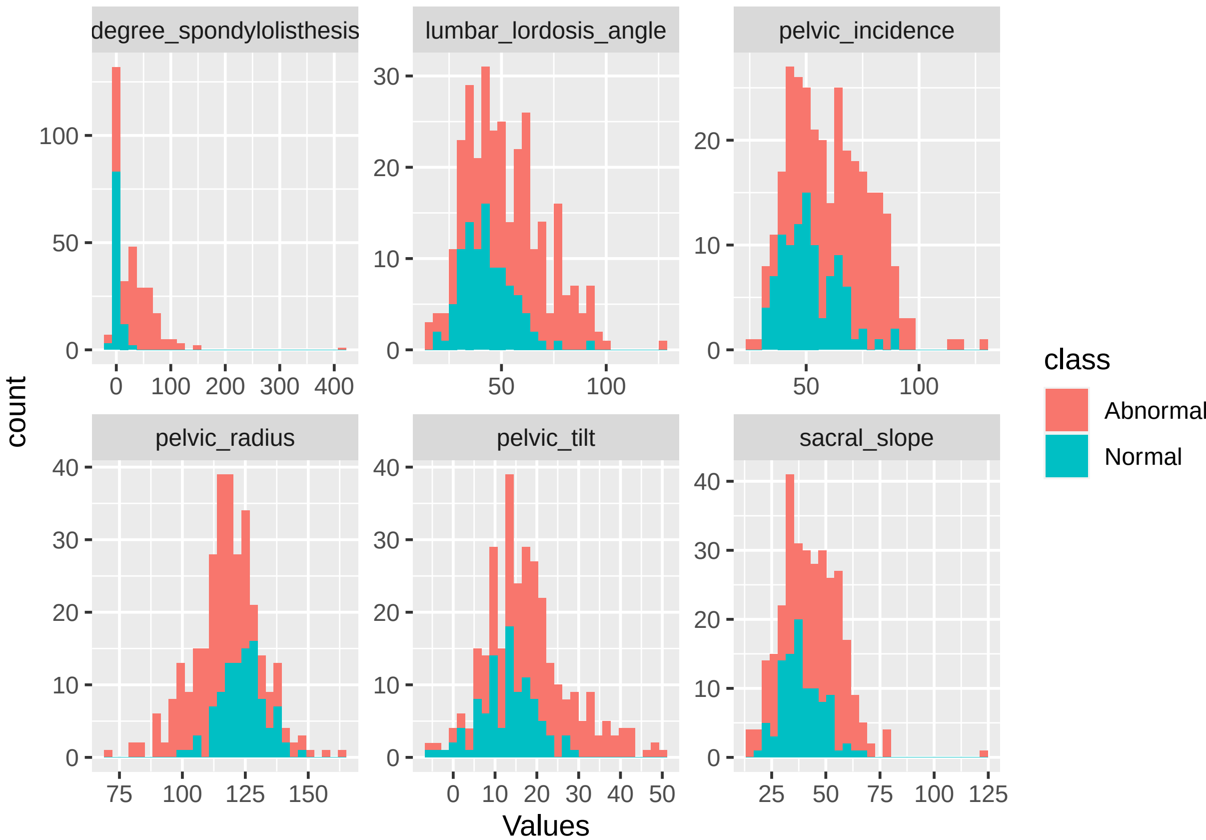
Assessing first the binary classification model is loading, cleaning, and summarizing the data set. Data for this project comes from the University of California Irvine data repository and is mostly clean. There are 310 total observations, split into 210 abnormal and 100 normal observations. Code loading the dataset and summary statistics for all six descriptor variables are listed below.

```r
if(!require(readr)) install.packages("readr", repos = "http://cran.us.r-project.org")
if(!require(tidyverse)) install.packages("tidyverse", repos = "http://cran.us.r-project.org")
if(!require(caret)) install.packages("caret", repos = "http://cran.us.r-project.org")
if(!require(data.table)) install.packages("data.table", repos = "http://cran.us.r-project.org")
if(!require(Matrix)) install.packages("Matrix", repos = "http://cran.us.r-project.org")
if(!require(ggplot2)) install.packages("ggplot2", repos = "http://cran.us.r-project.org")
if(!require(knitr)) install.packages("knitr", repos = "http://cran.us.r-project.org")

library(readr)
library(tidyverse)
library(caret)
library(data.table)
library(ggplot2)
library(knitr)

# load binary data set
ortho_dataset <- read_csv("archive/column_2C_weka.csv")

ortho_dataset <- ortho_dataset %>% mutate(class = as.factor(class)) %>% rename(pelvic_tilt = 'pelvic_til
summary(ortho_dataset)
```

```
##  pelvic_incidence  pelvic_tilt     lumbar_lordosis_angle  sacral_slope
##  Min.   : 26.15   Min.   :-6.555   Min.   : 14.00        Min.   : 13.37
##  1st Qu.: 46.43   1st Qu.:10.667   1st Qu.: 37.00        1st Qu.: 33.35
##  Median : 58.69   Median :16.358   Median : 49.56        Median : 42.40
##  Mean   : 60.50   Mean   :17.543   Mean   : 51.93        Mean   : 42.95
##  3rd Qu.: 72.88   3rd Qu.:22.120   3rd Qu.: 63.00        3rd Qu.: 52.70
##  Max.   :129.83   Max.   :49.432   Max.   :125.74        Max.   :121.43
##  pelvic_radius    degree_spondylolisthesis      class
##  Min.   : 70.08   Min.   :-11.058          Abnormal:210
##  1st Qu.:110.71   1st Qu.:  1.604          Normal  :100
##  Median :118.27   Median : 11.768
##  Mean   :117.92   Mean   : 26.297
##  3rd Qu.:125.47   3rd Qu.: 41.287
##  Max.   :163.07   Max.   :418.543
```

To get a better understanding of how variable values are distributed, we assess a histogram plot, colored by classification. Shown here, we see that the variables fall into three buckets. Pelvic Radius and Pelvic Tilt appear uniformly distributed, Lumbar Lordosis Angle and Pelvic Incidence have bimodal distributions, while Sacral Slope and Degree of Spondylolisthesis have a right skewed distribution.

```r
# plots of predictors by classification
plot_dat <- ortho_dataset %>% pivot_longer(!class, names_to = "Predictor", values_to = "Values" )
summary_plots <- plot_dat %>% ggplot(aes(Values, fill = class)) +
  geom_histogram() +
  facet_wrap(~Predictor, scales = "free")

summary_plots
```



Comparing the distributions by classification, it is clear that all apparent outliers belong to the Abnormal group, while the normal grouping generally exhibits lower variance. This is particularly pronounced in the distributions for Degree of Spondylolisthesis and Pelvic Tilt. These dynamics are further explored in the following table comparing the means and standard deviations of each variable.

```
#overall means
ortho_means <- ortho_dataset %>% select(-class) %>% sapply(, FUN = mean)
# mean and sd for predictors by classification
# normal mean and sd first
normal_means <- ortho_dataset %>% filter(class== "Normal") %>% select(-class) %>% sapply(, FUN=mean)
normal_sds <- ortho_dataset %>% filter(class== "Normal") %>% select(-class) %>% sapply(, FUN=sd)

# abnormal
abnormal_means <- ortho_dataset %>% filter(class== "Abnormal") %>% select(-class) %>% sapply(, FUN=mean
abnormal_sds <- ortho_dataset %>% filter(class== "Abnormal") %>% select(-class) %>% sapply(, FUN=sd)

class_summary <- data.frame(Normal_Mean = normal_means,
                            Normal_Sd = normal_sds,
                            Abnormal_Mean = abnormal_means,
                            Abnormal_Sd = abnormal_sds)

class_summary
```

```
##                          Normal_Mean Normal_Sd Abnormal_Mean Abnormal_Sd
## pelvic_incidence           51.685244 12.368161      64.69256    17.66213
## pelvic_tilt                12.821414  6.778503      19.79111    10.51587
## lumbar_lordosis_angle      43.542605 12.361388      55.92537    19.66947
## sacral_slope               38.863830  9.624004      44.90145    14.51556
```

```
## pelvic_radius              123.890834  9.014246     115.07771    14.09060
## degree_spondylolisthesis     2.186572  6.307483      37.77771    40.69674
```

The next step is to explore the accuracy of a few different machine learning algorithms in correctly classifying the data. For the binary classification we fit a General Linear Model, a Linear Discriminant Analysis, a Naive Bayes model, a K-Nearest Neighbors model, a Generalized Additive Model using Loess, and a Random forest Model. These models were chosen for this project because of their demonstrated ability to quickly and efficiently generate accurate predictions. The results of these models, using default tunings, are in the below table.

```
set.seed(33)
test_index <- createDataPartition(y = ortho_dataset$class , times = 1, p = 0.25, list = FALSE)
test <- ortho_dataset[test_index,]
train <- ortho_dataset[-test_index,]

# train models to determine which models to further evaluate
set.seed(21)
models <- c("glm", "lda", "naive_bayes", "knn", "gamLoess", "rf")
fits <- lapply(models, function(model){
  print(model)
  train(class ~ ., method = model, data = train)
})
```

```
## [1] "glm"
## [1] "lda"
## [1] "naive_bayes"
## [1] "knn"
## [1] "gamLoess"
## [1] "rf"
```

```
names(fits) <- models
preds <- sapply(fits, function(f){ predict(f, newdata = test)}) %>% as.data.frame()

acc <- data.frame(glm_acc = mean(test$class==preds$glm),
                  lda_acc = mean(test$class==preds$lda),
                  bayes_acc = mean(test$class==preds$naive_bayes),
                  knn_acc = mean(test$class==preds$knn),
                  gamloess_acc = mean(test$class==preds$gamLoess),
                  rf_acc = mean(test$class==preds$rf))
acc
```

```
##     glm_acc   lda_acc bayes_acc   knn_acc gamloess_acc    rf_acc
## 1 0.8205128 0.7948718 0.7051282 0.8333333    0.7948718 0.7820513
```

```
# test voting model with algos >80% accuracy
set.seed(22)

models2 <- c("glm", "knn")
vote_fits <- lapply(models2, function(model){
  print(model)
  train(class ~ ., method = model, data = train)
})
```

```
## [1] "glm"
## [1] "knn"
```

```r
names(fits) <- models2
vote_preds <- sapply(fits, function(f){ predict(f, newdata = test)}) %>% as.data.frame()

votes <- rowMeans(vote_preds == "Abnormal")

voting_preds <- ifelse(votes >= 0.5, "Abnormal", "Normal") %>% as.factor()
mean(test$class == voting_preds)
```

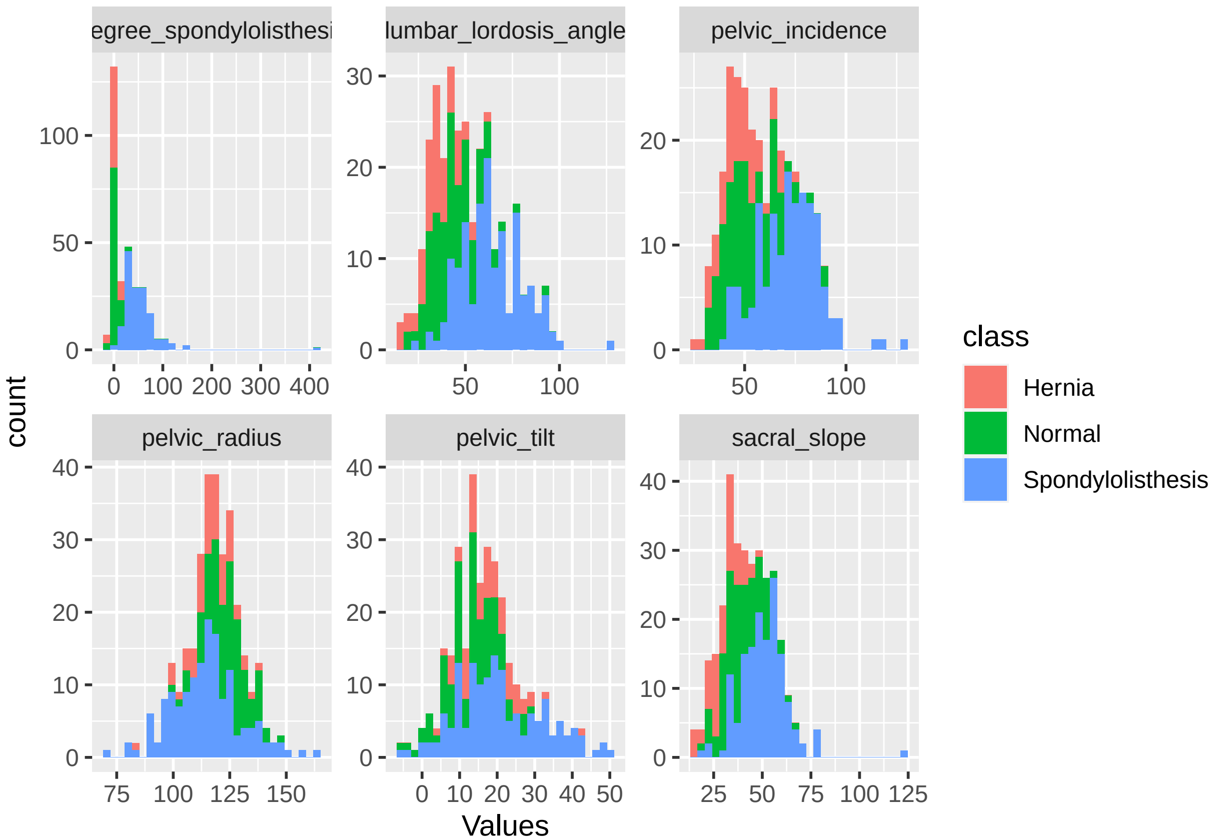
```
## [1] 0.8076923
```

Some of these models performed better than others, with K-nearest Neighbors and the generalized linear model being the most accurate models. A voting model is then created using the predictions of these two top performing models with an accuracy of 0.8077.

With this initial analysis of the binary classification task, we move on to the three-way classification problem, diagnosing patients with either a disc hernia, spondylolisthesis, or healthy.

The below plots show the distribution of the six predictor variables colored by diagnosis. These values are the same as the binary classification dataset, and the distribution characteristics are the same.

```r
# load 3-way classification dataset
multi_class_data <- read_csv("archive/column_3C_weka.csv")
multi_class_data <- multi_class_data %>% mutate(class = as.factor(class))

# plots of predictors by classification
plot_dat_multi <- multi_class_data %>% pivot_longer(!class, names_to = "Predictor", values_to = "Values
summary_plots_multiclass <- plot_dat_multi %>% ggplot(aes(Values, fill = class)) +
  geom_histogram() +
  facet_wrap(~Predictor, scales = "free")

summary_plots_multiclass
```



New details emerge, in the degree of spondylolisthesis, lumbar lordosis angle, and the sacral slope variables. It is clear that the while the herniated disc and healthy patients record similar readings under these metrics, patients with spondylolisthesis appear to have higher readings in all three metrics. This is verified by the table below showing the means and standard deviations of the variables by diagnosis.

```
# mean and sd for predictors by classification
normal_means <- multi_class_data %>% filter(class== "Normal") %>% select(-class) %>% sapply(, FUN=mean)
normal_sds <- multi_class_data %>% filter(class== "Normal") %>% select(-class) %>% sapply(, FUN=sd)

# hernia
herni_means <- multi_class_data %>% filter(class== "Hernia") %>% select(-class) %>% sapply(, FUN=mean)
herni_sds <- multi_class_data %>% filter(class== "Hernia") %>% select(-class) %>% sapply(, FUN=sd)

#spondylolisthesis
spondy_means <- multi_class_data %>% filter(class== "Spondylolisthesis") %>% select(-class) %>% sapply(
spondy_sds <- multi_class_data %>% filter(class== "Spondylolisthesis") %>% select(-class) %>% sapply(, F

class_summary <- data.frame(Normal_Mean = normal_means,
                            Normal_Sd = normal_sds,
                            Hernia_Mean = herni_means,
                            Hernia_Sd = herni_sds,
                            Spondylolisthesis_Mean = spondy_means,
                            Spondylolisthesis_Sd = spondy_sds)
class_summary
```

```
##                            Normal_Mean Normal_Sd Hernia_Mean Hernia_Sd
## pelvic_incidence             51.685244 12.368161   47.638407 10.697131
## pelvic_tilt                  12.821414  6.778503   17.398795  7.016708
```

```
## lumbar_lordosis_angle      43.542605 12.361388   35.463524  9.767795
## sacral_slope               38.863830  9.624004   30.239612  7.555388
## pelvic_radius             123.890834  9.014246  116.474968  9.355720
## degree_spondylolisthesis    2.186572  6.307483    2.480251  5.531177
##                          Spondylolisthesis_Mean Spondylolisthesis_Sd
## pelvic_incidence                       71.51422             15.10934
## pelvic_tilt                            20.74804             11.50617
## lumbar_lordosis_angle                  64.11011             16.39707
## sacral_slope                           50.76619             12.31881
## pelvic_radius                         114.51881             15.57999
## degree_spondylolisthesis               51.89669             40.10803
```

Next, we fit five machine learning algorithms to the training data including a Linear Discriminant Analysis, Naive Bases, K Nearest Neighbors, a Generalized Additive Model using Loess, and a Random Forest model. Given the three-way classification task, linear models are not applicable to this task. The accuracy results of these models are shown below.

```
# create test and train partitions
set.seed(33)
test_index_multi <- createDataPartition(y = multi_class_data$class , times = 1, p = 0.25, list = FALSE)
test_multi <- multi_class_data[test_index_multi,]
train_multi <- multi_class_data[-test_index_multi,]
summary(train_multi)
```

```
##  pelvic_incidence  pelvic_tilt     lumbar_lordosis_angle  sacral_slope
##  Min.   : 26.15   Min.   :-6.555   Min.   : 14.00        Min.   :13.37
##  1st Qu.: 47.57   1st Qu.:11.864   1st Qu.: 36.67        1st Qu.:33.11
##  Median : 59.38   Median :16.135   Median : 50.27        Median :42.35
##  Mean   : 60.24   Mean   :17.702   Mean   : 51.64        Mean   :42.54
##  3rd Qu.: 71.44   3rd Qu.:22.309   3rd Qu.: 63.00        3rd Qu.:51.89
##  Max.   :118.14   Max.   :49.432   Max.   :125.74        Max.   :79.70
##  pelvic_radius    degree_spondylolisthesis               class
##  Min.   : 70.08   Min.   :-11.058          Hernia           : 45
##  1st Qu.:110.68   1st Qu.:  1.506          Normal           : 75
##  Median :117.35   Median : 11.768          Spondylolisthesis:112
##  Mean   :117.22   Mean   : 25.015
##  3rd Qu.:125.05   3rd Qu.: 41.016
##  Max.   :157.85   Max.   :148.754
```

```
# test performance of a few models

model_list <- c("lda", "naive_bayes", "knn", "gamLoess", "rf")
set.seed(23)
multi_fits <- lapply(model_list, function(model){
  print(model)
  train(class ~ ., method = model, data = train_multi)
})
```

```
## [1] "lda"
## [1] "naive_bayes"
## [1] "knn"
## [1] "gamLoess"
## [1] "rf"
```

```
names(multi_fits) <- model_list
preds_multi <- sapply(multi_fits, function(f){ predict(f, newdata = test_multi)}) %>% as.data.frame()
multi_accuracy <- data.frame(lda_acc = mean(test_multi$class == preds_multi$lda),
```

```r
                              bayes_acc = mean(test_multi$class == preds_multi$naive_bayes),
                              knn_acc = mean(test_multi$class == preds_multi$knn),
                              gam_acc = mean(test_multi$class == preds_multi$gamLoess),
                              rf_acc = mean(test_multi$class == preds_multi$rf)
                              )

multi_accuracy
```

```
##     lda_acc bayes_acc   knn_acc   gam_acc    rf_acc
## 1 0.8076923 0.8333333 0.8333333 0.3846154 0.8717949
```

The accuracy of these models are quite different to those on the binary classification task. The Random Forest, KNN, and naive Bayes presented the greatest accuracy while the generalized additive model under performed. Fitting a second voting model using the top three predictors yields an accuracy of 0.8333.

```r
# knn model, bayes and random forest voting model on multiple classification

vote_model_list <- c("naive_bayes", "knn", "rf")
set.seed(23)
multi_vote_fits <- lapply(vote_model_list, function(model){
  print(model)
  train(class ~ ., method = model, data = train_multi)
})
```

```
## [1] "naive_bayes"
## [1] "knn"
## [1] "rf"
```

```r
names(multi_vote_fits) <- vote_model_list

preds_votes_multi <- sapply(multi_vote_fits, function(f){ predict(f, newdata = test_multi)}) %>% as.data

votes <- preds_votes_multi %>% mutate(norm_vote = ifelse(rowSums(preds_votes_multi=="Normal") >= 2, "No
                                                         ifelse(rowSums(preds_votes_multi=="Hernia") >=
                                                                ifelse(rowSums(preds_votes_multi=="Spond
mean(votes$norm_vote == test_multi$class)
```

```
## [1] 0.8333333
```

Results

The final results of the multiple models on the binary and tertiary classifications tasks are presented below. Overall, the models exhibited higher accuracy on the tertiary classification task compared to the binary classification, with the highest accuracy achieved being on the Random Forest model, which had an accuracy of 0.8718. This is compared to the highest accuracy on the binary classification, the K-nearest neighbors model which achieved 0.8333 accuracy. Only the Generalized additive model was less accurate on the tertiary classification than the binary, reducing its accuracy from 0.7949 to 0.3846154, which made it only slightly improved compared to guessing.

Binary Task		Tertiary Task	
Model	Accuracy	Model	Accuracy
LDA	0.7948718	LDA	0.8076923
Bayes	0.7051282	Bayes	0.8333333
KNN	0.8333333	KNN	0.8333333
GAM	0.7948718	GAM	0.3846154

Binary Task		Tertiary Task	
RF	0.7820513	RF	0.8717949
Voting	0.8076923	Voting	0.8333333
GLM	0.8205128		

The voting models developed did not improve accuracy over the top performing model in either task, with an accuracy of 0.8077 on the binary classification and 0.8333 on the tertiary classification, presenting reductions of 0.0256 and 0.0385 respectively.

Conclusion

This research was tasked with using machine learning to diagnose patients in across two metrics: normal or abnormal, and normal, having a herniated disc, or having spondylolisthesis. The Binary and Tertiary classification tasks posed unique challenges for these machine learning algorithms. The six dependent variables used in this analysis all proved to be significant predictors of patient classification.

Overall, the K-nearest neighbors model demonstrated the highest accuracy of 0.8333 out of the selected models on the binary classification task, while the random forest was most accurate for the tertiary classification with an accuracy of 0.8718. In both tasks it is clear that the voting model did not improve upon these scores, and actually lowered the accuracy to lowest denominator of the included models. This demonstrates how voting models are only as accurate as their least accurate component.

The finding that no single model was most accurate for both the binary and tertiary classification tasks is also insightful into the importance of model selection for supervised machine learning analysis.

It is also interesting that the tertiary classification generally saw higher accuracy relative to the binary classification. This is likely due to the observations in each group being more similar to each other across the six variables in the tertiary classification, while the abnormal group in the binary dataset saw higher variation across the six variables. Given that two diseases included in the Abnormal grouping in the binary task would present differently across the variables and present a challenge for supervised classification tasks such as Random Forest or Naive Bayesian regression.

This research also presents opportunities for further exploration, including using Random Forest for other medical diagnoses purposes. While this research focused on orthopedic medicine, additional research can extend to diagnosing any number of diseases with similarly presenting symptoms.

Sources

Dataset: https://www.kaggle.com/datasets/uciml/biomechanical-features-of-orthopedic-patients

Physiological Terms:

1. https://www.sciencedirect.com/topics/nursing-and-health-professions/pelvic-incidence
2. https://pubmed.ncbi.nlm.nih.gov/1354697/
3. https://www.sciencedirect.com/topics/nursing-and-health-professions/sacral-slope
4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877554/#