A ROBUST LOGISTIC REGRESSION FRAMEWORK FOR ORTHOPAEDIC DIAGNOSIS WITH VIF FEATURE SELECTION AND SMOTE BALANCING.

AM5510- COURSE PROJECT

BIOMEDICAL SIGNALS AND SYSTEMS

PROJECT REPORT

Submitted by

KAILAASH R.M. (AM25S005)

M.S BIOMEDICAL ENGINEERING

DEPARTMENT OF APPLIED MECHANICS AND BIOMEDICAL ENGINEERING

NOVEMBER, 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This project focuses on building a simple and effective machine learning model, primarily focused on **Logistic Regression** to classify orthopaedic conditions (Abnormal cases – Disc Herina, Spondylolisthesis versus Normal) using basic biomechanical measurements. It includes exploring the dataset through visual plots and statistical tests to understand how each feature behaves and how strongly they contribute to the condition. Since some features were highly correlated, VIF-based feature selection was used to pick the most meaningful parameters for both **binary and multi-class classification**.

Different machine learning models such as Logistic Regression, SVM, and Bayesian approaches were trained and compared. SMOTE was applied to handle the class imbalance, which helped the model perform more fairly across all classes. Among all the models tested, **Logistic Regression with the selected features** gave the most balanced and reliable results.

The final model was saved and **deployed using Streamlit**, allowing users to enter biomechanical values and get predictions directly through a simple web interface. The overall workflow remained clean, without data leakage, and showed how proper feature selection and class balancing can improve performance even with a small medical dataset. This project shows a practical way to build an interpretable and lightweight ML tool for supporting orthopaedic assessments.

CHAPTER 1

INTRODUCTION

1.1 ABOUT THE TASK:

The given dataset is organized into two files for two related classification tasks. It is a Biomedical Orthopedic dataset consists of six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine. In one dataset, the subjects are classified based on three spinal conditions that is “Normal”, “Disc Herina”, “Spondylolisthesis”, whereas in another dataset, the latter two are combined into one class called “Abnormal”, turning into a binary classification problem.

The task is to build simple yet efficient model (especially logistic regression) that can be able to learn the parameter distributions and classify the patient condition accordingly.

1.2 INTRODUCTION:

According to a multicentre MRI-based Indian study, the prevalence of lumbar disc herniation in asymptomatic individuals was approximately around 30%. Spinal Disorders like Disc Herina and Spondylolisthesis are common orthopaedic conditions, affecting the alignment and stability of the spine. These conditions can be diagnosed by inspecting the various biomechanical features, that talks about the shape and orientation of the pelvis and lumbar spine. The given dataset consists of six features namely **‘Pelvic Incidence’, ‘Pelvic Tilt’, ‘Pelvic Radius’, ‘Sacral Slope’, ‘Lumbar Lordosis Angle’** and **‘Grade of Spondylolisthesis’.** These features are obtained from the X-ray Images of 310 patients.

The main objective of this work is to find the most important and relevant features for classification and develop classification models specially build a logistic regression model, finding its optimal parameters, which can automatically distinguish between the spinal conditions. Two different tasks are binary classification (Normal Vs Abnormal) and multi-class classification (Normal vs Disc Herina vs Spondylolisthesis).

1.3 SPINAL CONDITIONS:

Disc Herina and Spondylolisthesis are among the most common causes of chronic low back pain and mobility issues. Hence, their early diagnosis of these conditions is necessary to avoid future complications.

Disc Herina is the condition where the soft inner material of an intervertebral disc comes out through its outer layer. Normally inside the vertebrae, there is a cushion like disc which can absorb shock. In **disc Herina,** the outer layer of this disc called (annulus fibrosus) start rupturing, causing the protrusion of the soft part (nucleus pulposus) and it presses the nearby nerves leading to back pain, leg pain and numbness. There will not be any slipping or position-change in the vertebral bone. This causes changes in the pelvic angles such as **tilt, incidence and slope.**

Spondylolisthesis is a more serious condition, where forward slipping of one vertebra over the another happens, leading to spinal instability. Due to the vertebral slip, bending of spine, forward tilting of pelvis happens. In order to compensate them, the pelvic angles increase. This condition leads to severe pain and posture change. Increase in the degree of spondylolisthesis values and other pelvic measurements might indicate this condition.

1.4 BIOMECHANICAL ATTRIBUTES:

The dataset consists of six biomechanical attributes that are obtained from the shape and orientation of the pelvis and lumbar spine. They are depicted in the figure 1.1

- Pelvic Incidence
- Pelvic tilt
- Lumbar Lordosis Angle
- Sacral Slope
- Pelvic Radius
- Grade of Spondylolisthesis

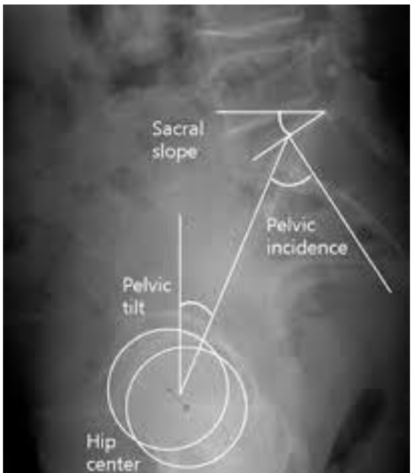

Figure 1.1 Pelvic Biomechanical Parameters

1.4.1 PELVIC INCIDENCE:

Pelvic incidence is the measure of angle between a line perpendicular to the sacral plate at its mid-point and a line connecting this point to the axis of femoral heads. It is an anatomical parameter, where higher value indicates increased Lumbar Lordosis.

1.4.2 PELVIC TILT:

It is the measure of angle between vertical and a line connecting the midpoint of the sacral plate to femoral heads. It changes with the posture of the subject. Higher pelvic tilt might indicate that pelvis is rotated backward to compensate for the spinal deformity.

1.4.3 SACRAL SLOPE:

It is defined as the **angle between the sacral plate and a horizontal line**. This angle indicates how much the sacrum (the base of the spine) tilts forward. A **higher sacral slope** means the sacrum is more inclined, usually associated with a **greater lumbar lordosis**. It can be expressed as the sum of Sacral Slope and Pelvic Tilt.

1.4.4 LUMBAR LORDOSIS:

Lumbar lordosis is the inward (ventral) curvature of the lumbar spine formed by the wedging of lumbar vertebral bodies and the intervertebral disks. It represents the **degree of inward curvature of the lumbar spine**. It is measured as the **angle** formed between the **superior endplate of the first lumbar vertebra (L1)** and the **superior endplate of the first sacral vertebra (S1).** This angle indicates how much the lower back curves toward the front of the body

1.4.5 PELVIC RADIUS:

It is a biomechanical parameter that represents the distance between the midpoint of the sacral endplate and the axis of the femoral heads. It is a geometric measure, which describes the size and shape of the pelvis in relation to the spine. It helps in assessing the Spino-pelvic alignment.

1.4.6 DEGREE OF SPONDYLOLISTHESIS:

It is the measure of tilt or slippage of angle between superior endplate of Sacrum and the inferior endplate of the last lumbar vertebra. It indicates the extent of forward slippage of one vertebra over the one below it.

CHAPTER 2

DATA ANALYSIS AND CLASSIFICATION

2.1 REQUIREMENTS:

Google Colab (python) was used to perform data analysis and train the model. The following libraries were required for all the tasks.

- Pandas: For performing computations on Data frame.
- NumPy: Numerical Operations
- Matplotlib: For plots and analysis
- Seaborn: Data visualizations
- Scikit-learn: Preprocessing, Data splitting, Model evaluation (Confusion Matrix)
- Stats model: Variance Inflation Factor
- Imblearn: SMOTE.

2.2 DATASET DISTRIBUTION:

The first dataset (column_3C_weka.csv) consists data of 310 patients where the class distribution is as follows:

- Disc Herina = 60 patients
- Normal = 100 patients
- Spondylolisthesis = 150 patients.

It is a multiclass-classification problem.

In the second dataset, (column_2C_weka.csv) Disc Herina and Spondylolisthesis are combined into one class as Abnormal. The distribution is as follows.

- Normal = 100
- Disc Herina = 210

It is a binary classification problem. Each dataset consists of 6 features.

2.3 EXPLORATORY DATA ANALYSIS:

At first, for both datasets, it was checked for any missing values and it is found that there are no missing values.

- Histogram of each feature is obtained to study the spread of the given parameters and how much overlap among the classes.
- Boxplot and pairwise (scatter plots) are obtained to visualize the data distribution and outliers etc.,
- The statistical parameters such as mean, standard deviation, minimum and maximum values are obtained for each feature to get the range and summary of the dataset.

2.4 FEATURE SELECTION:

- Simple feature selection tests (**t-test** for binary classification and **ANOVA** – for multi-class problem) were performed to estimate the statistical
- Correlation matrix of the features is obtained to study how strongly the features are related.
- VIF (Variance Inflation Factor) is obtained, which tells how much a feature is similar to other features in the dataset. If VIF is high, it indicates two features are very similar, giving almost the same information and one of them is necessary.
- It is a statistical measure used in regression analysis to assess how much the variance of an estimated regression coefficient is increased because of collinearity among independent variables. High VIF indicates that feature can make the regression model unreliable.

Automated VIF based Feature Selection:

- VIF library is imported from Stats model package.
- The threshold was fixed as 15.
- VIF values were computed for all features.
- Feature with highest VIF exceeding the threshold was automatically removed.
- The process is iteratively repeated until all the features have VIF within the threshold.

2.5 DATA BALANCING:

- From the dataset size, it is observed that in the first dataset, the number of Disc Herina patients data is low (65) compared to the Normal (100) and Spondylolisthesis (150). Similarly, in another dataset, Normal class is a minor class with only 100 datapoints. Though the major aim of machine learning in healthcare datasets is to improve the accurate classification of abnormal cases, to improve the recall of both the classes and build a stable model, SMOTE (Synthetic Minority Oversampling Technique) was performed on the minority class to increase its data points.

- The classification by the models were performed with and without SMOTE and their results were compared.

2.5.1 SMOTE:

SMOTE generates new synthetic samples for the minority class by interpolating between existing samples and their nearest neighbours in feature space. This approach increases minority class representation without simple duplication. It results in balanced and more generalizable dataset for classification.

For each minority class sample x_i, one of its k-nearest neighbors $x_{nn}$is selected, and a new synthetic point is created using linear interpolation. The basic equation of SMOTE is written in equation 2.1.

$$\boldsymbol{x}_{\text{new}} = \boldsymbol{x}_i + \boldsymbol{\delta} \times (\boldsymbol{x}_{nn} - \boldsymbol{x}_i) \quad \textbf{(2.1)}$$

where $\delta$is a random number in the range $[0,1]$. This process is repeated until the minority class reaches the desired size

2.6 LOGISTIC REGRESSION:

Logistic Regression is a supervised machine learning algorithm used for classification problem. It uses a sigmoid or SoftMax functions to convert inputs into a probability value between 0 and 1. It can be Binomial or Multinomial or ordinal.

Assumptions:

- Each data point is assumed to be independent of the others, means there is no correlation between input samples.
- It assumes a linear relationship between the independent variables and the log odds of the dependent variable which means the predictors affect the log odds in a linear way.
- The dataset should not contain extreme outliers.

2.6.1 Multi-Class Logistic Regression:

A multiclass logistic regression model was implemented manually using the SoftMax function to classify patients into *Normal*, *Disk Hernia*, or *Spondylolisthesis* categories. The model computes a linear combination of the input features, as shown in equation 2.2

$$\boldsymbol{z = X\theta + \theta^{bias}} \quad \textbf{(2.2)}$$

Where,

- X= feature matrix
- θ= weight parameters to be learned.
- $\theta^{bias} = \ bias\ term$ (intercept- initialized as column of ones)

The **SoftMax function** then converts these linear outputs into class probabilities using the equation 2.3

$$\boldsymbol{P(y = j \mid X) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}} \ \textbf{...(2.3)}$$

- where $K$is the total number of classes.(here K= 3)

The model parameters $\theta$are optimized by minimizing the **negative log-likelihood loss** (cross-entropy) using equation 2.4

$$\boldsymbol{L(\theta) = -\sum_{i\cdot} \sum_{j} y_{ij} \log\left(P(y = j \mid X_i)\right)} \textbf{...(2.4)}$$

Where:

- i indexes the samples (i = 1,2, … , n)
- j indexes the classes (j = 1,2, … , C)
- y_{ij} is the one-hot encoded label for sample i (1 if sample ibelongs to class j, 0 otherwise)
- $P(y = j \mid X_i; \theta)$is the predicted probability that sample i belongs to class j, given model parameters θ
- Θ are the parameters of the model (weights in logistic regression, neural networks, etc.

- The loss is **high** when the model assigns low probability to the true class.
- The loss is **minimized** when the model assigns probability close to 1 to the correct class.
- Minimizing $L(\theta)$is equivalent to **maximizing the likelihood** of the correct labels.

2.6.2 Binary Classification Logistic Regression:

The binary classification logistic regression model uses sigmoid function, defined by equation 2.5.

$$\boldsymbol{\hat{y}_i = \sigma(z_i) = \frac{1}{1+e^{-z_i}}, z_i = \theta^T X_i + \theta^{bias}} \quad \textbf{(2.5)}$$

For binary classification the loss function is simplified to

Taking the **negative log** to make it easier to minimize

$$\boldsymbol{L(\theta) = -\,log\,\mathcal{L}(\theta)} \qquad \textbf{(2.6)}$$

$$\boldsymbol{L(\theta) = -\sum_{i=1}^{n}[y_i \log\, \hat{y}_i + (1-y_i)\log\,(1-\hat{y}_i)]} \quad \textbf{(2.7)}$$

Optimization of weights:

The best weights or the parameters are estimated through the BFGS (Broydenn-Fletcher-Goldfarb-Shanno) optimization algorithm It is a quasi-Newton method. It approximates the inverse Hessian without computing it explicitly. It uses gradient information from previous steps to update the approximation, explained by equation 2.8

$$\boldsymbol{\theta_{k+1} = \theta_k - \alpha_k H_k^{-1} \nabla L(\theta_k)} \quad \textbf{(2.8)}$$

$$\boldsymbol{H = \frac{\partial^2 L}{\partial \theta^2}} \textbf{\textit{is the Hessian matrix.}} \ \textbf{(2.9)}$$

- Initialize the parameter with a guess for theta θ as zero or epsilon and inverse Hessian approximation H_0^{-1} as the identity matrix I.
- Compute the gradient of the loss at current θ_k.
- Find the search direction using the inverse Hessian approximation, using equation 2.10

$$\boldsymbol{p_k = -H_k^{-1} g_k} \ \textbf{...(2.10)}$$

- BFGS adjusts the step using curvature information, so the step is smarter and more direct.

- Update the parameters and gradient at new point: $\boldsymbol{g_{k+1}} = \boldsymbol{\nabla L(\theta_{k+1})}$**... (2.11)**
- Update the inverse Hessian Approximation using the following equations 2.12 and 2.13.

$$\boldsymbol{s_k = \theta_{k+1} - \theta_k, y_k = g_{k+1} - g_k} \ \textbf{... (2.12)}$$

$$\boldsymbol{H_{k+1}^{-1} = (I - \frac{s_k y_k^T}{y_k^T s_k}) H_k^{-1} (I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k}} \ \textbf{...(2.13)}$$

- Iterate the process until convergence

2.7 ML PIPELINE:

- **Data Preparation:** Select the relevant features as x and target classes as y. Encode the target (class) using Label Encoder.
- **Outlier Removal:** The outliers are removed using IQR (below 25^{th} percentile and above 75^{th} percentile).
- **Train Test Split:** Split the data into 80% training and 20% test data, with stratification to preserve distribution.
- **Class Balancing:** Apply SMOTE to balance the classes.
- **Bias:** Define a column of ones to X for intercept term.
- **Activation function:** Define SoftMax (Multi-class) or the Sigmoid function (Dual-class), that converts logits to probabilities.
- **Loss Function:** Loss function is defined as negative log-likelihood (cross entropy).
- **Initialize Parameters:** Initialize θas zeros.
- **Optimization:** Find the optimum parameters through BFGS Algorithm.
- **Extract Optimized parameters:** Reshape and extract the optimizer result.
- **Prediction:** Compute the probabilities and predicted class is argmax of probabilities.
- **Evaluation:** Evaluate the performance of the model using various metrices.
- **Model save:** The best trained logistic regression model is saved using pickle library.

Apart from this logistic regression model from scikit-learn library is also used and the results are compared. Bayesian probability-based classification and Support Vector Machine model was also performed. The bayes theorem for the estimation of theta is given by the following equation 2.14

$$\boldsymbol{p(\theta \mid \mathbf{data}) = \frac{p(\mathbf{data}|\theta)\, p(\theta)}{p(\mathbf{data})}} \ \textbf{(2.14)}$$

2.8 MODEL EVALUATION:

To evaluate the performance of the classification models, standard metrics such as Accuracy, Precision, Recall, and F1-score were used. These metrics are derived from the confusion matrix consisting of:

- TP: True Positives
- TN: True Negatives
- FP: False Positives
- FN: False Negatives

Accuracy: Measures the overall correctness of the model defined by equation 2.15

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \textbf{... (2.15)}$$

Precision: Measures how many predicted positives are actually positive, calculated by equation 2.16

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad \textbf{(2.16)}$$

Recall (Sensitivity): Measures how many actual positives are correctly identified, defined by equation 2.17.

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad \textbf{(2.17)}$$

F1-score: Harmonic mean of Precision and Recall, useful when classes are imbalanced.

$$\textbf{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \textbf{(2.18)}$$

2.9 DEPLOYMENT USING STREAMLIT:

The logistic regression model with the best accuracy, built using **scikit-learn**, has been deployed as a web application using **Streamlit**. The trained model is first saved using **pickle**, and then a Python program with the necessary code and interface is created to make the model accessible through Streamlit. All the files and programs are uploaded to a **GitHub repository**, making it easy to share. With this setup, anyone can enter the required feature values on the web app and get instant predictions from the model.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 EXPLORATORY DATA ANALYSIS:

The given datasets were analysed by extracting the statistical parameters such as mean, standard deviation as well as by means of various graphical plots such as histogram, box plots, scatter plots and correlation maps, in order to study the relationship between the features. It was performed for both the datasets and the results are discussed as follows.

3.1.1 HISTOGRAM:

The following figure 3.1 shows the histogram distribution of all the 6 features in the binary classification problem with the hue as classes. It shows the individual feature distribution and range difference between two classes.

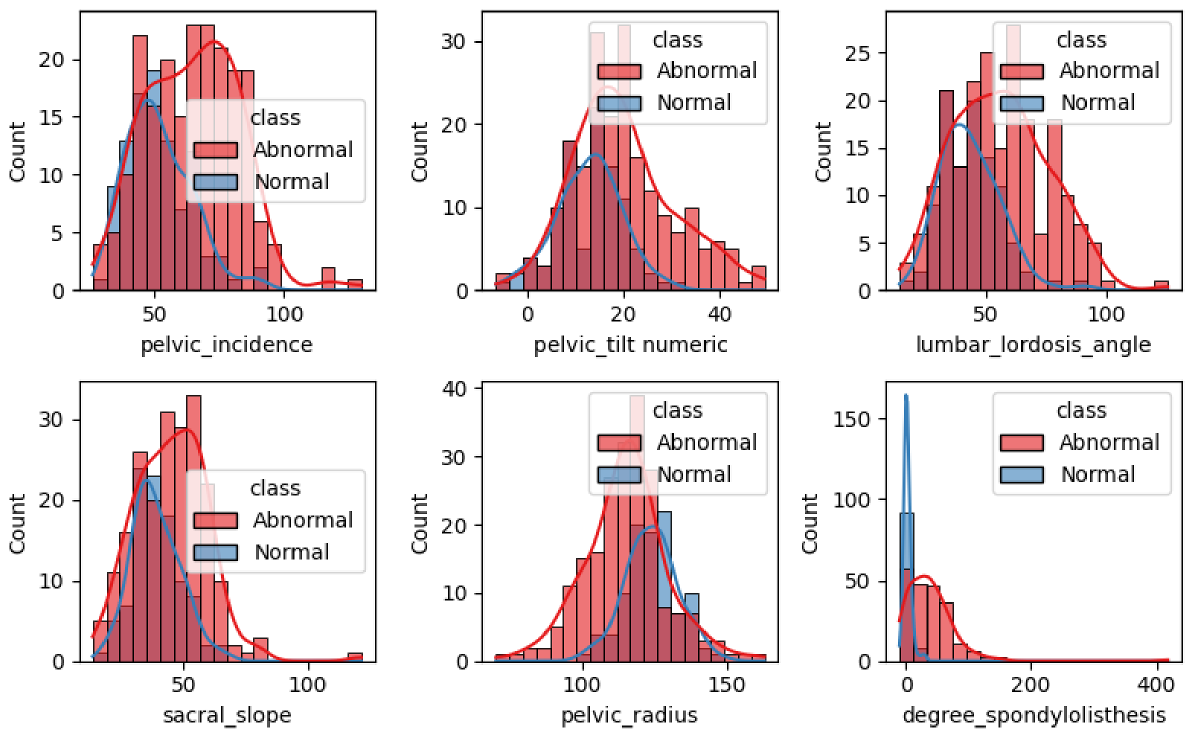

Figure 3.1 Histogram Distribution of Features - Binary Classification

From the plots, we can infer that, for abnormal class, all features except pelvic radius, extends up to a maximum range indicating higher mean value. Degree of Spondylolisthesis shows a clear distinction between two classes. Also, it is noted that there is overlap between normal and abnormal classes in the mid rages.

The following figure 3.2 depicts the histogram distribution of the features for the multi-class classification problem.

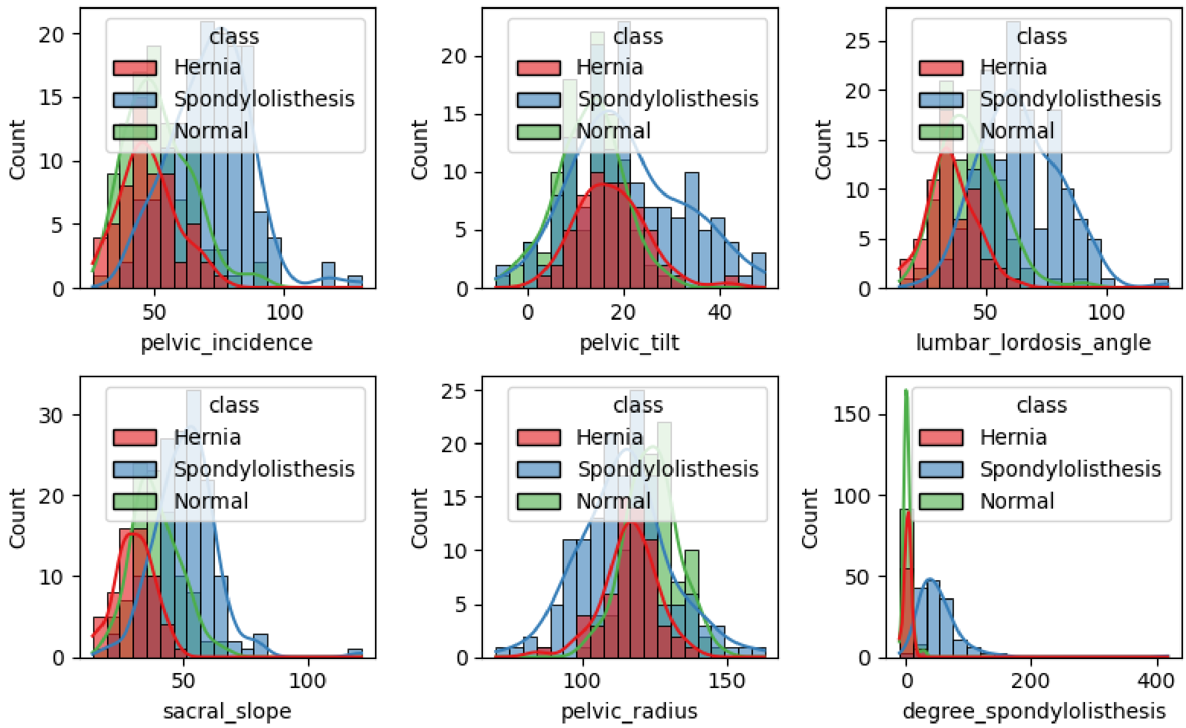

Fig 3.2 Histogram of features - tri class classification

From the figure, it is inferred that Spondylolisthesis cases show noticeably higher values in pelvic incidence, lumbar lordosis angle, sacral slope and especially degree of spondylolisthesis. Normal subjects cluster at lower values across most parameters. Herina cases show moderate feature values, lying between normal and spondylolisthesis ranges.

3.1.2 BOX PLOT:

The following figures 3.3 and 3.4 shows the box plots of the features of binary classification dataset and multi-class classification dataset respectively.

From the figure , it is visible that abnormal cases show higher values in all cases except pelvic radius. The huge difference in the degree of spondylolisthesis confirms the conclusions of the histogram plots.

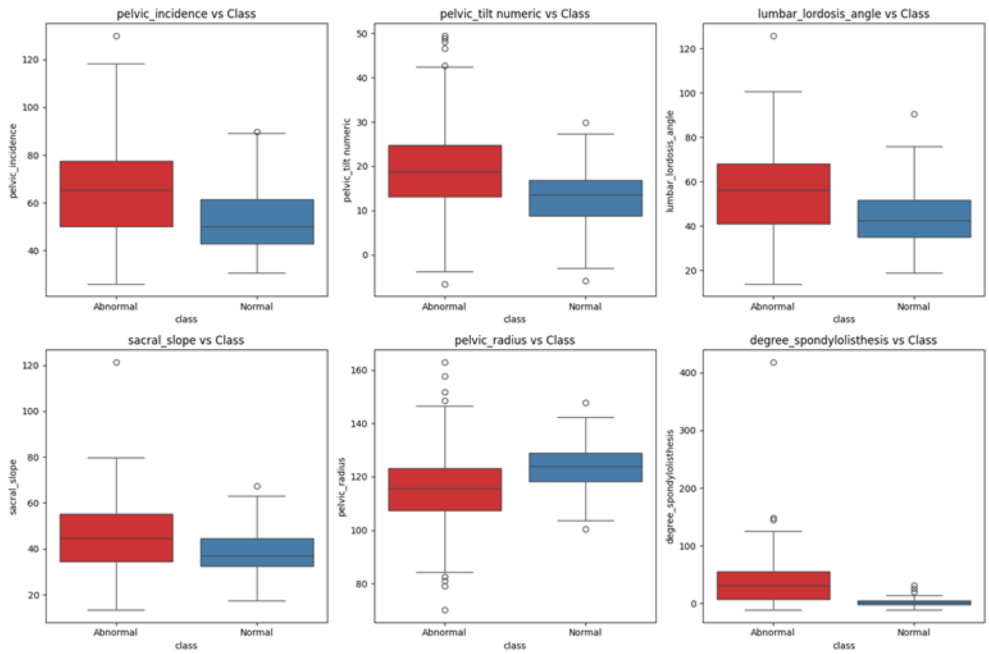

Figure 3.3 - Box Plot - Binary Classification

From the figure 3.4, **Spondylolisthesis** cases show the highest values across most features, especially **degree of spondylolisthesis**. **Hernia** cases have lower values in pelvic and spinal angles.

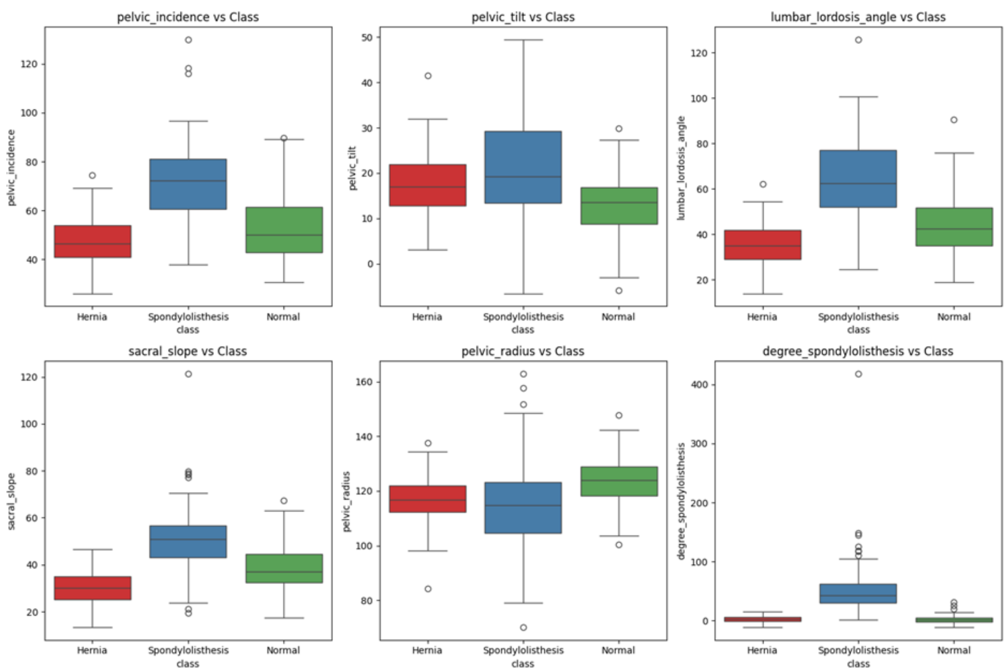

Figure 3.4 Box-Plot - Multi-class classification

3.1.3 PAIR PLOT:

The following figures 3.5 and 3.6 shows the pair -plot of the feature pairs of binary and multi-class dataset respectively. This help to visualize the correlation between two features for both the classes and how much they are distinct.

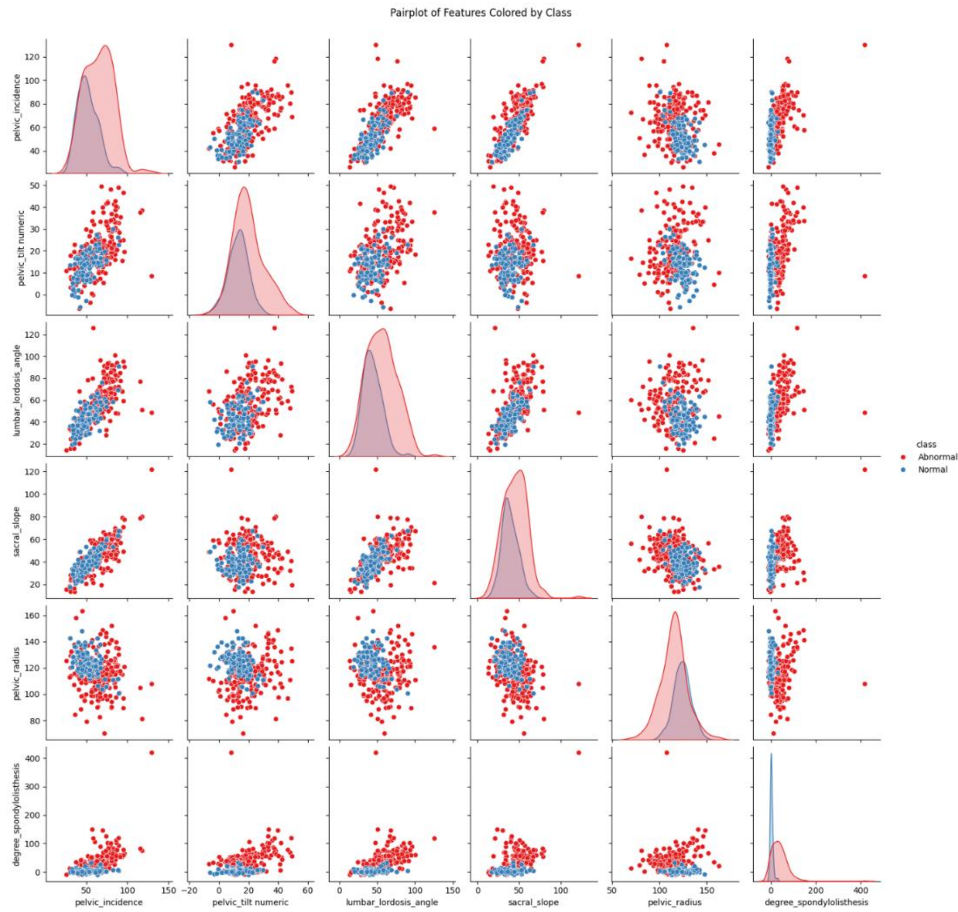

Figure 3.5 Pair plot - Binary Classification

From the figure , it is observed that, there exists a moderate overlap but notable separation between 'Normal' and 'Abnormal' classes, especially for the feature 'Degree of Spondylolisthesis". This also indicates positive correlation between sum feature such as Lumbar Lordosis angle and sacral slope.

The pair plot results for the multi-class dataset are shown in the figure . Spondylolisthesis is separable from the other classes in several feature pairs. Hernia overlaps heavily with Normal, suggesting harder discrimination. Some features, like Pelvic Incidence and Lumbar Lordosis Angle, demonstrate moderate correlation.

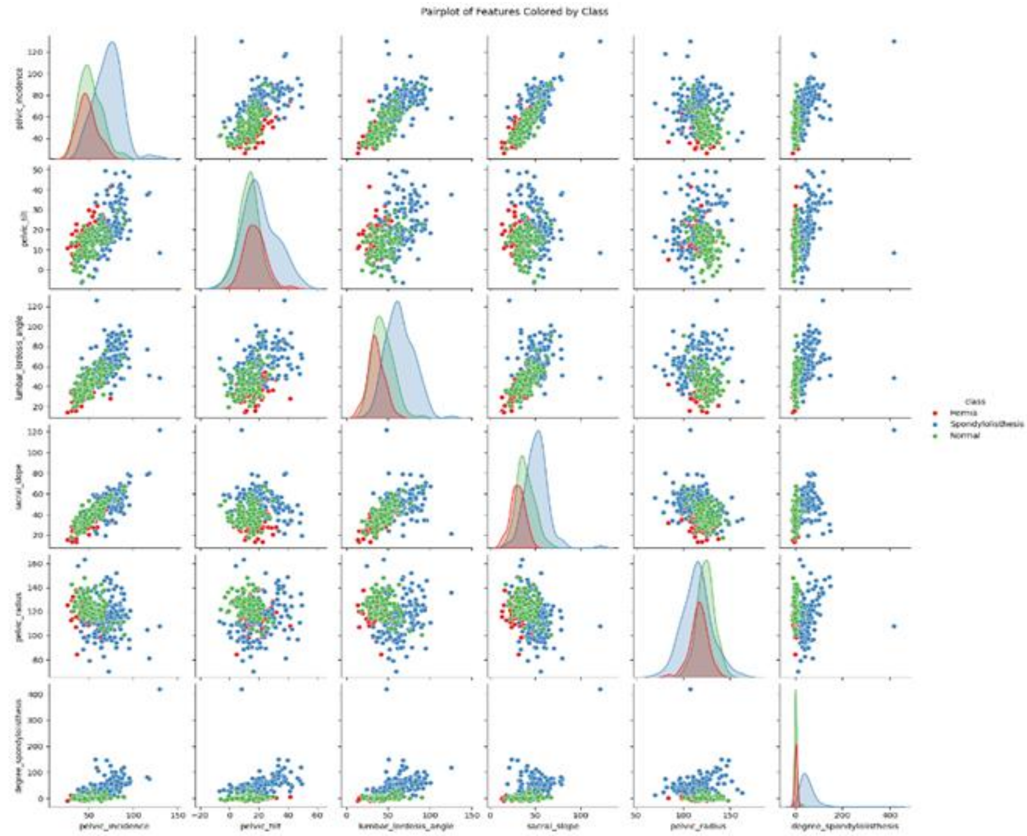

Figure 3.6 Pair plot- Multi-class Classification

3.1.4 FEATURE CORRELATION:

Feature Correlation Matrix was obtained , which gives the correlation value between all the feature pairs separately. It is shown in figure 3.6. It indicates there are strong positive relationships among key spinal parameters. Especially, pelvic incidence shows strong positive correlation with all features except pelvic radius. Positive correlation exists between lumbar lordosis angle and sacral slope. he presence of high correlations implies multicollinearity and supports using dimensionality reduction or regularization in predictive models to avoid redundancy and overfitting. Pelvic Radius shows inverse relationship (negative correlation) with other features.

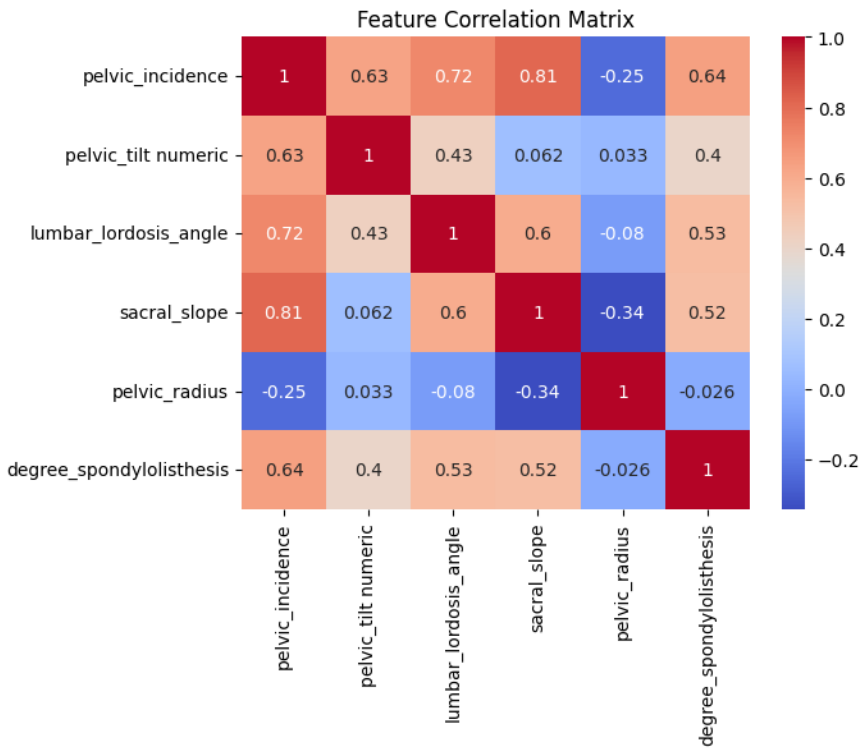

Figure 3.7 Feature Correlation Matrix

3.2 FEATURE SELECTION:

From the exploratory data analysis, it is observed that there exists positive correlation between the features, indicating the need to remove reductant features and select relevant features not only to improve the classification accuracy but also reduce computational cost. Statistical tests like t-test, ANOVA and Variance Inflation Factor are employed. The t-test for binary classification task and f-one-way test for tri-class classification task showed that p-value is 0 for all classes, indicating all features are important. The results are shown in the following figure 3.8.

```
pelvic_incidence: p-value =0.000
pelvic_tilt numeric: p-value =0.000
lumbar_lordosis_angle: p-value =0.000
sacral_slope: p-value =0.000
pelvic_radius: p-value =0.000
degree_spondylolisthesis: p-value =0.000
```

Figure **Error! No text of specified style in document.**3.8 - p-values of Features

The VIF test with threshold 10 was carried out to automatically select the features that have low VIF. In the binary classification task, the selected features are Pelvic tilt, Sacral slope and Degree Spondylolisthesis with VIF values 3.5, 3.7 and 1.9 respectively. The results are shown in the **figure 3.9.** For the tri-class classification, the threshold was set as 15, pelvic radius was also selected as a feature along with the above said three features. The result is shown in the **figure 3.10**

```
Dropping 'pelvic_incidence' with VIF=inf
Dropping 'lumbar_lordosis_angle' with VIF=18.89
Dropping 'pelvic_radius' with VIF=11.81

Final feature set with acceptable VIFs:
                   Feature       VIF
0      pelvic_tilt numeric  3.564297
1              sacral_slope  3.701406
2  degree_spondylolisthesis  1.945764
```

Figure 3.9 VIF Feature Selection-Binary Classification

```
Dropping 'pelvic_incidence' with VIF=inf
Dropping 'lumbar_lordosis_angle' with VIF=18.89

Final feature set with acceptable VIFs:
                   Feature        VIF
0              pelvic_tilt   4.791892
1             sacral_slope  10.982582
2            pelvic_radius  11.807684
3  degree_spondylolisthesis   2.308863
```

Figure 3.10 VIF Feature Selection-Multi-class classification

3.3 CLASSIFICATION

After the feature selection, the classification was carried out using various models under various conditions. The primary focus was on logistic regression, and it was performed with both selected features and all the features to test the improvement in classification accuracy due to feature selection. SMOTE was also employed to balance the class imbalance of the minority class, and classification metrices were evaluated. Support Vector Machine and Bayesian based Classification was also performed for the binary classification task.

3.3.1 BINARY CLASSIFICATION:

The classification models were employed for the Normal Vs Abnormal classification task. The results are discussed with following sections.

3.3.1.1 LOGISTIC REGRESSION:

Logistic Regression model for the binary classification task is build using sigmoid function, log loss function and BFGS optimization was built. It was tested for selected features and all features with and without SMOTE. Classification was also performed with outlier removal.

3.3.1.2 LOGISTIC REGRESSION WITH FEATURE SELCTION:

The Logistic Regression with **VIF selected three features** for the binary classification task, showed the highest accuracy **of 90.3% and ROC-AUC is 0.944.** The number of parameters is 4 and the optimized parameters values are [-0.27366603 -0.0808185 0.05949098 -0.14437984]. The results of logistic regression model with VIF selected feature is summarized in the following table 3.1 and the confusion matrix is shown in the figure 3.11.

Table 3.1: Classification Metrices – Logistic Regression with Feature Selection (Binary Classification)

CLASS	PRECISION	RECALL	F1-SCORE	SUPPORT
0-Abnormal	0.91	0.95	0.93	42
1-Normal	0.89	0.80	0.84	20

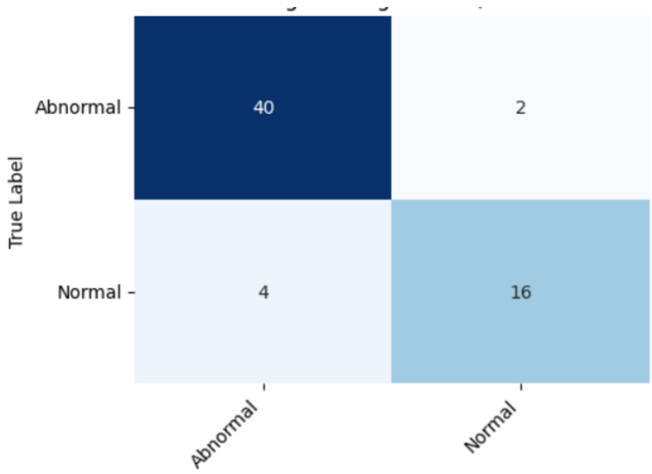

Figure 3.11 Confusion Matrix- Logistic Regression with Feature Selection

From the results it is inferred that, logistic regression with feature selection showed a higher recall for the abnormal class (0.95) but slightly lower for the Normal class (0.80). It is due to relatively lower test data points for the Normal class (minority class). This seems to be a good working for a medical dataset, not missing most of the abnormal cases.

3.3.1.3 LOGISTIC REGRESSION WITHOUT FEATURE SELECTION:

Logistic regression model was then trained with all the features with 100 iterations, and the test accuracy was reduced to 83.87%, due to the high positive correlation existing between the features. The results are shown in the following figure 3.12.

```
••• Test Accuracy: 0.8387096774193549
    Confusion Matrix:
     [[37  5]
     [ 5 15]]
    ROC-AUC: 0.9500000000000001
    Classification Report:
                  precision    recall  f1-score   support

               0       0.88      0.88      0.88        42
               1       0.75      0.75      0.75        20

        accuracy                           0.84        62
       macro avg       0.82      0.82      0.82        62
    weighted avg       0.84      0.84      0.84        62
```

Figure 3.12 Classification Metrics- Logistic Regression without Feature Selection

The optimized parameters are as follows:

Coefficients (θ1...θn): [[-0.07838704 0.07488117 -0.16335794 -0.00350585 0.10616331 0.03121346]]

Intercept (θ0): [-14.60129335]

Total parameters: 7

3.3.1.4 LOGISTIC REGRESSION WITH SMOTE AND OUTLIER REMOVAL:

To improve the classification accuracy of the minority class (Normal), the dataset is balanced by applying SMOTE to the Normal class (80 to 168), and then logistic regression is trained on the balanced dataset. The overall test accuracy reduced to 87.09% but the recall of the minority class increased to 95% and the AUC became 0.953. The results of SMOTE balancing and logistic regression is shown in the following figure 3.13 . The features used were **sacral slope, pelvic tilt, pelvic radius and degree of spondylolisthesis.** It was found that adding pelvic radius improved the accuracy, indicating it as distinctive feature for the normal class.

```
Original training set class distribution:
 0    168
1     80
Name: count, dtype: int64
Resampled training set class distribution:
 0    168
1    168
Name: count, dtype: int64
Accuracy: 0.8709677419354839
Confusion Matrix:
 [[35  7]
 [ 1 19]]
ROC-AUC: 0.9535714285714286

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.83      0.90        42
           1       0.73      0.95      0.83        20

    accuracy                           0.87        62
   macro avg       0.85      0.89      0.86        62
weighted avg       0.89      0.87      0.87        62
```

Figure 3.13 – Classification metrics- Logistic Regression with SMOTE

From the results, it is inferred that though SMOTE improved the classification accuracy for Normal class, recall for abnormal class reduced indicating balanced performance. But when

SMOTE is performed after Outlier removal using IQR, the accuracy drastically dropped to 78.6%. logistic Regression with Outlier Removal achieved 97% 'Abnormal' recall but had a substantially lower 'Normal' recall of 60% and overall accuracy of 84.5%.

3.3.1.5 SVM AND BAYESIAN CLASSIFICATION:

SVM model showed a classification accuracy of 87.1 %, with reduced recall for abnormal class (83%). But the ROC is highest of all that is 0.957, indicating more robustness and versatile. Bayesian probability-based model showed similar results that of logistic regression with feature selection. The overall comparison of the model's performance based on the features used is summarized in the following table 3.2

Table 3.2 Comparison of Model Performances- Binary Classification

S.NO	Model Name	Features Used	Accuracy	ROC-AUC	Precision (Class 0)	Recall (Class 0)	Precision (Class 1)	Recall (Class 1)
1.	Logistic Regression (VIF-selected features)	['pelvic tilt numeric', 'sacral slope', 'degree spondylolisthesis']	0.903	0.944	0.91	0.95	0.89	0.8
2.	Custom Logistic Regression	All features	0.838	0.95	0.88	0.88	0.75	0.75
3.	Logistic Regression (with Outlier Removal)	['pelvic tilt numeric', 'sacral slope', 'degree spondylolisthesis']	0.845	0.914	0.82	0.97	0.92	0.6
4.	Logistic Regression (with SMOTE)	['sacral slope', 'pelvic tilt numeric', 'pelvic radius', 'degree spondylolisthesis']	0.871	0.954	0.97	0.83	0.73	0.95
5.	Logistic Regression (with Outlier Removal and SMOTE)	['pelvic tilt numeric', 'sacral slope', 'degree spondylolisthesis', 'pelvic radius']	0.786	0.911	0.83	0.83	0.7	0.7
6.	SVM (with SMOTE)	['sacral slope', 'pelvic tilt numeric', 'pelvic radius', 'degree spondylolisthesis']	0.871	0.957	0.97	0.83	0.73	0.95
7.	Bayesian Logistic Regression	['pelvic tilt numeric', 'sacral slope', 'degree spondylolisthesis']	0.903	0.944	0.91	0.95	0.89	0.8

3.3.2 MULTI-CLASS CLASSIFICATION:

Several models primitively logistic regression model was also trained for the multi-class (3 classes) classification problem of the same dataset and the performances were evaluated life aforementioned binary classification task.

3.3.2.1 LOGISTIC REGRESSION WITH FEATURE SELECTION:

Logistic regression model was trained with the VIF based selected features and the classification accuracy was found to be 85%, slightly reduced compared to that of binary classification task. This was mainly due to major overlap between the Normal and the Disc Herina classes. Spondylolisthesis showed higher recall of 93% and f1 score 97%, indicating its clear distinction from other classes. The ROC-AUC was 0.95. The total number of parameters was found to be 15, and the loss converged in 33 iterations. The results are depicted in the figure 3.14 and the confusion matrix is shown in the figure 3.15.

```
...          Current function value: 113.719377
             Iterations: 33
             Function evaluations: 1372
             Gradient evaluations: 85
    Optimized Parameters (θ):
     [[ 16.6679182  -10.18798709  -6.48002292]
     [  0.02372817  -0.07604972   0.03782849]
     [ -0.17653658   0.04801941   0.05443974]
     [ -0.09526288  -0.10275772   0.19646362]
     [ -0.33278372  -0.16026256  -0.24840301]]
    Number of Parameters: 15

    Test Accuracy: 0.8548387096774194
    Confusion Matrix:
     [[ 9  3  0]
     [ 4 16  0]
     [ 1  1 28]]
    ROC-AUC (OVR): 0.9534325396825397
    Classification Report:
                        precision    recall  f1-score   support

               Hernia       0.64      0.75      0.69        12
               Normal       0.80      0.80      0.80        20
    Spondylolisthesis       1.00      0.93      0.97        30

             accuracy                           0.85        62
            macro avg       0.81      0.83      0.82        62
         weighted avg       0.87      0.85      0.86        62
```

Figure 3.14 Classification Results - Logistic Regression with Feature Selection (Multi-Class)

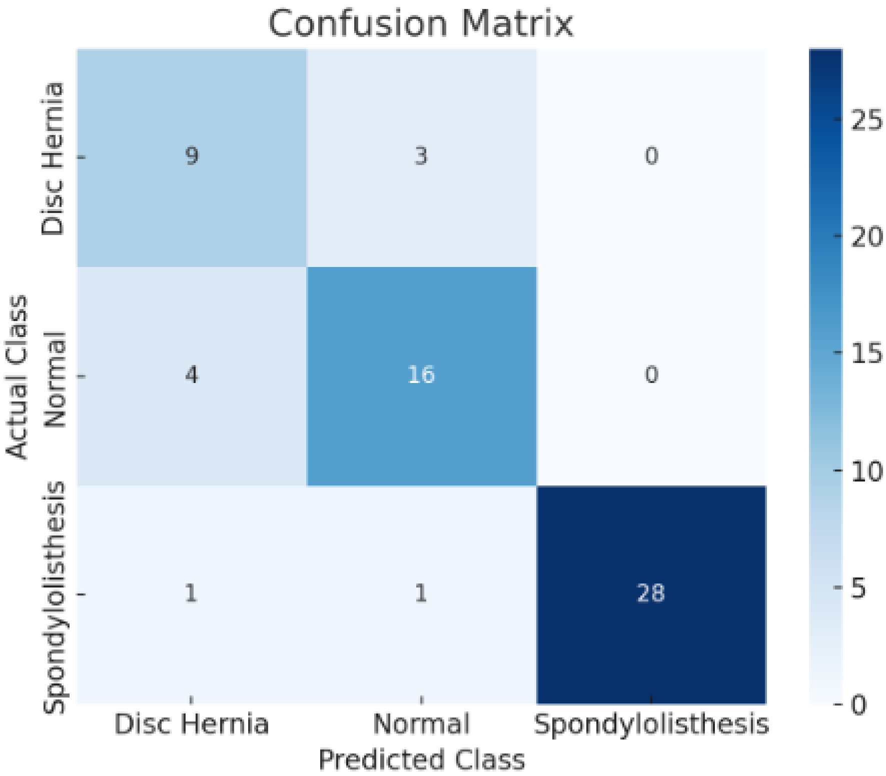

Figure 3.15 - Confusion Matrix- Logistic Regression with Feature Selection (Multi-Class)

3.3.2.2 LOGISTIC REGRESSION WITHOUT FEATURE SELECTION:

Logistic Regression Model was trained with all the six features. The accuracy was reduced to 82.2% and ROC-AUC was 0.94. The number of iterations of convergence was 42 and the number of parameters was 21. The results are shown in the figure 3.16

3.3.2.3 LOGISTIC REGRESSION WITH SMOTE:

Logistic Regression with SMOTE was performed with the selected features. It showed the highest accuracy of all with 87.09% and the ROC-AUC is 0.964. There was an overall improvement in the recall for all the cases. All the training dataset was balanced to 105 data points. It showed the 100% precision for Spondylolisthesis class. The results of logistic regression with SMOTE is shown in the figure 3.17 . The results indicated that in multi-class classification, class balancing played a vital role in both increasing the classification accuracy of all the classes and robustness.

```
          Current function value: 114.079255
          Iterations: 42
          Function evaluations: 1837
          Gradient evaluations: 83
Optimized Parameters (θ):
 [[ 1.68226043e+01 -1.10830450e+01 -5.73950502e+00]
 [ 6.46498698e-02 -8.11635422e-02 -5.74807041e-03]
 [-1.45911671e-01  4.72519897e-02 -1.89208486e-02]
 [-9.25209684e-02 -1.08044769e-01  1.99884684e-01]
 [-4.69075312e-01 -2.88054766e-01 -3.86275384e-01]
 [-1.19942220e-01 -7.41354074e-02 -4.00565340e-02]
 [-4.91902443e-02 -6.30477343e-02 -3.80360063e-02]]
Number of Parameters: 21

Test Accuracy: 0.8225806451612904
Confusion Matrix:
 [[ 9  3  0]
 [ 5 14  1]
 [ 1  1 28]]
ROC-AUC (OVR): 0.9498809523809525
Classification Report:
                    precision    recall  f1-score   support

           Hernia       0.60      0.75      0.67        12
           Normal       0.78      0.70      0.74        20
Spondylolisthesis       0.97      0.93      0.95        30
```

Figure 3.16 Classification Metrics- Logistic Regression without Feature Selection (Multi-Class)

```
Original training set class distribution:
 2    105
1     70
0     42
Name: count, dtype: int64
Resampled training set class distribution:
 2    105
1    105
0    105
Name: count, dtype: int64
Accuracy: 0.8709677419354839
Confusion Matrix:
 [[15  3  0]
 [ 7 23  0]
 [ 1  1 43]]
ROC-AUC: 0.9644824202082267

Classification Report:
              precision    recall  f1-score   support

           0       0.65      0.83      0.73        18
           1       0.85      0.77      0.81        30
           2       1.00      0.96      0.98        45

    accuracy                           0.87        93
   macro avg       0.83      0.85      0.84        93
weighted avg       0.88      0.87      0.87        93
```

Figure 3.17 - Classification Metrics- Logistic Regression with SMOTE (Multi-Class)

3.3.2.4 SVM AND GRADIENT BOOST:

The model trained using **SMOTE** for class balancing and a **Support Vector Machine (SVM)** with a linear kernel performed very well, achieving an **accuracy of 86.02%** and a **ROC-AUC score of 0.959**. By applying SMOTE, the imbalance between classes was reduced, which helped improve the **recall for all classes**, especially for the minority ones. This shows that the model can correctly identify both normal and abnormal spinal conditions more effectively. Overall, the linear SVM model proved to be accurate, consistent, and well-suited for this classification task. The results are showed in the figure 3.18. Gradient Boost model showed very least accuracy of 77.4%.

```
Original training set class distribution:
 2    105
1     70
0     42
Name: count, dtype: int64
Resampled training set class distribution:
 2    105
1    105
0    105
Name: count, dtype: int64
Accuracy: 0.8602150537634409
Confusion Matrix:
 [[15  3  0]
 [ 8 22  0]
 [ 1  1 43]]
ROC-AUC (OVR): 0.9590784832451499

Classification Report:
                    precision    recall  f1-score   support

          Hernia       0.62      0.83      0.71        18
          Normal       0.85      0.73      0.79        30
Spondylolisthesis       1.00      0.96      0.98        45

        accuracy                           0.86        93
       macro avg       0.82      0.84      0.83        93
    weighted avg       0.88      0.86      0.86        93
```

Figure 3.18 Classification Metrics SVM with Smote (Multi-Class)

3.3.2.5 MODELS PERFORMANCE SUMMARY:

The comparison of performances of all the models is summarized in the following table 3.3. VIF selected features are **Pelvic radius, Pelvic tilt, Sacral Slope and Degree of Spondylolisthesis.**

Table 3.3 Comparison of Model Performances – Multi-class Classification

S.NO	Model	Features Used	Accuracy	ROC-AUC (OVR)	Macro Avg Recall
1.	Logistic Regression	VIF selected features	0.854	0.9534	0.81
2.	Logistic Regression (without feature selection)	All features	0.8225	0.9498	0.78
3.	Logistic Regression with SMOTE	VIF selected features	0.8702	0.964	0.83
4.	SVM with SMOTE	VIF selected features	0.8602	0.9590	0.82
5.	Gradient Boosting (with SMOTE)	VIF selected features.	0.7741	0.9215	0.73

3.4 MODEL DEPLOYMENT USING STREAMLIT:

The logistic regression model trained using the VIF-selected features was saved as a pickle file and deployed using Streamlit. A Streamlit application was developed to allow users to input feature values through an interactive interface. A **Predict** button was included to generate outputs using the trained model. The program files, along with the required resources, were uploaded to a GitHub repository, and the web application was successfully launched. The following figure 3.19 shows the interface of the Streamlit application.

GitHub Repository Link: https://github.com/KAILAASH06/SPINAL-DISEASE-CLASSIFICATION

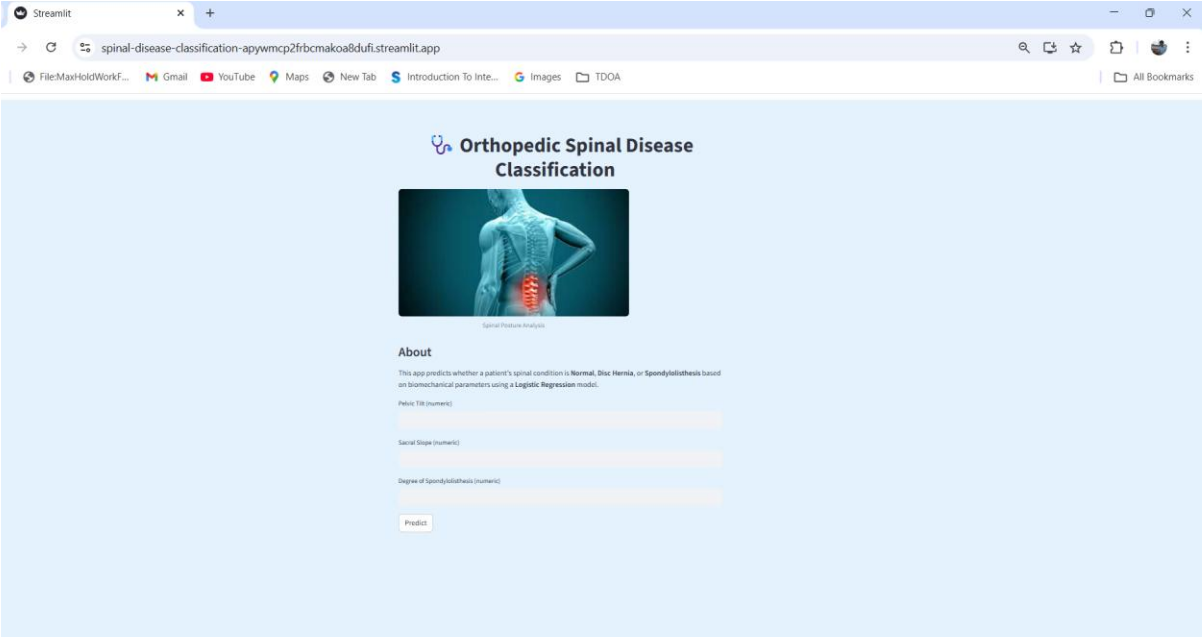

Figure 3.19 Interface of the Streamlit app

CHAPTER 4

CONCLUSION

4.1 CONCLUSION:

Several classification models including logistic regression, SVM, and a Bayesian probability-based approach were trained on the orthopaedic dataset with the biomechanical parameters. Data exploratory analysis was performed using various plots such as box plots, pair plots, and histograms. The t-test and ANOVA showed that every feature had a p-value of 0.0, indicating that each feature is significant. The correlation matrix showed that there exists more positive correlation between certain features, stressing the need for removal of redundant features. The Variance Inflation Factor test produced 3 significant features (Pelvic Tilt, Sacral Slope, Degree of Spondylolisthesis) for the binary classification task and 4 significant features (Pelvic Radius along with the above three features) for the multi-class classification task.

The Logistic Regression model trained with the selected features showed the highest accuracy of 90% for the binary classification task. But the classification accuracy for the minority class (Normal) was lower, around 80%. Then, class imbalance was handled using the SMOTE algorithm and the logistic regression model was retrained with the selected features. This showed an improvement in the classification of the Normal class (95%), with a trade-off in the reduction of classification accuracy for the Abnormal class (83%). The SVM model also showed a good accuracy of 87% for the binary classification task.

In the multi-class classification task, the logistic regression model trained on the SMOTE-balanced dataset using the VIF-selected features showed the highest accuracy of 87.02%. The Gradient Boost model showed the lowest accuracy of 77.4%. The best model was saved as a pickle file and deployed using Streamlit, and the Streamlit web application was launched using GitHub.

It is concluded that logistic regression outperformed all the models with 4 parameters for the binary classification task and 15 parameters for the tri-classification task. The SMOTE-based class balancing approach increased the classification efficiency for the minority class (Normal) but reduced the classification accuracy for the Abnormal class, yet high recall for the Abnormal class is desired for biomedical problems. Feature selection using VIF proved to be more successful in increasing the classification accuracy and reducing the computational cost. Data leakage tests also confirmed that there was no data leakage. Future improvements can be made in terms of generalizing the model for new datasets and making the model more robust.

REFERENCES

[1] **Dua, D., & Graff, C. (2019).** *UCI Machine Learning Repository: Orthopaedic Patients Dataset*. University of California, Irvine.
(Dataset source)

[2] **Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002).** *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321–357. (Class imbalance handling)

[3] **Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013).** *Applied Logistic Regression* (3rd ed.). Wiley. (Main model used)

[4] **Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005).** *Applied Linear Statistical Models*. McGraw-Hill/Irwin. (VIF and multicollinearity theory)

[5] **Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011).** *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
(ML implementation library)