# Lumber Spine Degenerative Classification using Deep Learning Models


## Abstract

This project focuses on the development and evaluation of deep learning models to classify lumbar spine degenerative conditions from medical images. The dataset used for this task is the RSNA 2024 Lumbar Spine Degenerative Classification dataset, which includes labelled medical images depicting various stages of lumbar spine degeneration. The objective is to build robust models capable of accurately identifying different degenerative conditions based on the provided images. To achieve this, multiple deep learning models are trained and evaluated using the EfficientNetV2 architecture, which has shown promising results in image classification tasks due to its efficiency and accuracy. The models undergo a rigorous training process, where their performance is assessed based on several metrics, including training accuracy, validation accuracy, and loss functions. Additionally, early stopping is implemented to prevent overfitting and ensure the models generalize well on unseen data. The report provides a detailed description of the dataset, the architecture used, and the evaluation techniques applied during the project. It also discusses the results obtained from various models and compares their performance. The insights derived from these evaluations can help in understanding the strengths and limitations of different models, offering valuable guidance for future research and potential

clinical applications in diagnosing lumbar spine degenerative conditions.


## Introduction

Degenerative diseases of the lumbar spine, such as neural foraminal narrowing, subarticular stenosis, and canal stenosis, are common conditions that can result in chronic pain, impaired mobility, and diminished quality of life. Early and accurate diagnosis through medical imaging plays a crucial role in facilitating timely and effective interventions. This project focuses on leveraging deep learning, specifically convolutional neural networks (CNNs), for the automated classification of lumbar spine degenerative conditions from MRI images.

### Objective

The primary objective of this project is to develop and evaluate deep learning models designed to classify lumbar spine degenerative conditions using MRI scans. By utilizing advanced AI techniques, this project aims to assist in the early detection of these conditions, providing valuable insights for clinical decision-making.

### Approach

To achieve the classification task, the project employs the EfficientNetV2 architecture, which is known for its efficiency and superior performance in image classification tasks. The model utilizes transfer learning, incorporating pre-trained weights from large-scale datasets to fine-tune the model on the relatively smaller lumbar spine MRI dataset. This approach helps mitigate the challenges posed by limited data and enhances the model's generalization ability, making it more robust for medical image analysis.

## Dataset

The dataset utilized in this project is the RSNA 2024 Lumbar Spine Degenerative Classification dataset, which consists of a comprehensive collection of MRI image sequences that provide valuable insights into various lumbar spine degenerative conditions. These MRI images offer detailed views of the spine from different perspectives, enabling accurate classification of conditions that can lead to chronic pain and mobility issues. The images come from various angles and have been labeled according to the specific lumbar spine condition they represent. This labeling is crucial for supervised learning, as it allows the models to be trained with clear targets for classification.

The primary MRI sequences in the dataset include:

**Sagittal T1**: This sequence captures detailed information regarding neural foraminal narrowing, a condition where the nerve roots become compressed, potentially leading to pain and numbness.

**Axial T2**: This sequence is essential for evaluating subarticular stenosis, a condition in which the spinal canal narrows and can cause nerve compression, often resulting in pain or weakness.

**Sagittal T2**: This sequence plays a key role in diagnosing canal stenosis, a condition that involves narrowing of the spinal canal, leading to possible nerve damage and symptoms like pain, tingling, and numbness.

Each MRI image in the dataset is accompanied by annotations that indicate the severity or presence of the corresponding degenerative condition. These annotations are important for training the deep learning models to differentiate between varying degrees of severity and conditions accurately.

To prepare the dataset for efficient use in deep learning models, the images undergo a pre-processing step, where they are resized and normalized. Resizing ensures that the images fit the input requirements of the models, while normalization standardizes the pixel values to facilitate better model convergence and performance.

### DataSplitting

To train and evaluate the models effectively, the dataset is split into distinct training and validation sets. This ensures that the models are trained on one portion of the data while evaluated on another, providing an unbiased estimate of their performance. Special attention is given to ensuring that each class is adequately represented in both the training and validation sets to prevent class imbalance issues, which could otherwise affect the model’s ability to generalize.

## Methodology

### ModelArchitecture:

For this project, the EfficientNetV2 model was selected due to its remarkable efficiency and superior performance in image classification tasks. EfficientNetV2 is an advanced iteration of the EfficientNet family, offering improvements in both accuracy and computational efficiency. This model is known for its ability to scale effectively, balancing depth, width, and resolution in the architecture, all while maintaining optimized model parameters. EfficientNetV2 uses a technique called compound scaling, which simultaneously adjusts depth, width, and resolution in a manner that maximizes performance without excessively increasing the computational cost. This makes it particularly well-suited for medical image analysis tasks, where accuracy and efficiency are both crucial.

**TransferLearning**
The model was initially pre-trained on the **ImageNet** dataset, a large-scale image classification benchmark. Transfer learning allows the EfficientNetV2 model to leverage features it learned from ImageNet, such as basic patterns, edges, textures, and shapes, which are common across various image domains. By applying this pre-trained knowledge, the model requires less data and time to train on the specific lumbar spine degenerative condition dataset. Moreover, this pre-training helps the model to generalize better and achieve higher performance when fine-tuned for the task at hand. The pre-trained model provides a solid foundation, making it much more efficient to adapt to medical image classification compared to training a model from scratch.

**Fine-Tuning**
After the initial training with pre-trained weights, the model undergoes fine-tuning to tailor it to the specific task of lumbar spine degenerative condition classification. Fine-tuning involves adjusting the last few layers of the neural network to specialize in recognizing patterns and features unique to the lumbar spine MRI sequences. This process allows the model to become more attuned to the specific nuances of the dataset, ensuring improved classification performance for conditions such as neural foraminal narrowing, subarticular stenosis, and canal stenosis.

**Model for MRI Sequences**
The EfficientNetV2 architecture was adapted and implemented separately for each of the three main MRI sequences in the dataset, each targeting a specific degenerative condition of the lumbar spine:

**Sagittal T1 Model**: This model is fine-tuned to focus on detecting **neural foraminal narrowing**, a condition where the nerve roots become compressed, causing pain or numbness. The model is optimized to identify subtle changes in the neural foramina that signal this condition.

**Axial T2 Model**: This model is specialized in identifying **subarticular stenosis**, a narrowing of the spinal canal at the nerve root level. It focuses on detecting patterns in the axial T2 sequences that indicate the presence of stenosis, which can lead to nerve compression and associated symptoms.

**Sagittal T2/STIR Model**: This model is tailored to classify **canal stenosis**, a condition where the spinal canal narrows, potentially compressing the spinal cord and nerves. The model is trained to recognize the distinct features of canal stenosis seen in sagittal T2/STIR MRI images, aiding in the accurate diagnosis of this condition.

Each of these specialized models is trained and evaluated separately to optimize performance on the specific conditions they target, ensuring high accuracy in classifying lumbar spine degenerative diseases.

**DataPreprocessing**
Data preprocessing is an essential step to ensure the MRI images are suitable for deep learning models. The key preprocessing steps applied to the RSNA 2024 Lumbar Spine Degenerative Classification dataset are:

**Resizing:**
All images were resized to a consistent dimension of **224x224 pixels** to meet the input size requirements of the EfficientNetV2 model. This step standardizes the images, enabling the model to process them efficiently and focus on learning the relevant features without being affected by varying image sizes.

**Normalization:**
To improve model training and

convergence, the images were normalized using the **mean and standard deviation values** from the ImageNet dataset. This standardization ensures that all pixel values are scaled similarly, helping the network converge faster and learn more effectively. By normalizing the images, it ensures that no individual feature dominates the learning process, resulting in a more stable and efficient model training.

These preprocessing steps—resizing and normalization—prepare the dataset to be fed into the EfficientNetV2 model, optimizing the training process and improving the model's ability to classify lumbar spine degenerative conditions accurately.

**TrainingSetup**
To optimize the training of the deep learning models for lumbar spine degenerative condition classification, several key components were selected.

**LossFunction:**
The **cross-entropy loss function** was chosen as it is effective for multi-class classification tasks. It helps the model minimize the error between predicted probabilities and true labels, driving it to make more accurate predictions.

**Optimizer:**
The **Adam optimizer** was used for its adaptive learning rate, which helps speed up convergence and maintain stability. It combines the benefits of momentum and RMSprop, making it ideal for deep learning tasks.

**Learning Rate Scheduler:**
A **step-based learning rate scheduler** was employed to reduce the learning rate by a factor of **0.1 every 2 epochs**. This approach fine-tunes the model towards the end of training, improving both convergence and accuracy.

**EarlyStopping:**
To prevent overfitting, **early stopping** was applied with a patience of **3 epochs**. If validation accuracy did not improve for 3 consecutive epochs, training was stopped early, saving resources and avoiding overfitting.

These strategies—cross-entropy loss, Adam optimizer, learning rate scheduler, and early stopping—ensured efficient and effective training, improving the model's performance while preventing overfitting.

**EvaluationMetrics**
The performance of the models was thoroughly evaluated using several important metrics that provide insights into their learning progress, accuracy, and generalization ability.

**Training and Validation Loss:**
The **training and validation loss** metrics track the model's learning process by calculating the difference between predicted values and actual labels. A decreasing loss indicates that the model is learning effectively and making better predictions over time. Training loss reflects how well the model fits the training data, while validation loss measures how well it generalizes to unseen data. Monitoring both losses is crucial for ensuring that the model is not overfitting or underfitting to the data.

**Training and Validation Accuracy:**
**Training accuracy** reflects the percentage of correct predictions on the training dataset, while **validation accuracy** measures the model's performance on unseen validation data. High accuracy on both training and validation sets suggests that the model is effectively learning to classify the lumbar spine degenerative conditions and generalizing well to new, unseen data. These accuracy metrics are essential for assessing how well the model

distinguishes between different degenerative conditions in the MRI images.

**EarlyStopping:**
**Early stopping** was implemented to prevent overfitting by monitoring the validation accuracy during training. If the validation accuracy did not improve for a predefined number of epochs, training was halted early. This helps prevent the model from learning noise or irrelevant patterns in the training data, which can reduce its ability to generalize to new data. Early stopping ensures that the model achieves optimal performance without wasting computational resources on unnecessary epochs.

These evaluation metrics—training and validation loss, accuracy, and early stopping—were essential for monitoring the model's performance and ensuring it achieved high accuracy while avoiding overfitting. Together, they helped ensure that the model would be robust and reliable when classifying lumbar spine degenerative conditions.

## Results

**4.1TrainingResults**
The models were trained for a maximum of 5 epochs, with early stopping implemented to prevent overfitting and optimize the training process. The results for each model are as follows:

**Model1:SagittalT1**
For the Sagittal T1 model, the training accuracy ranged from **77% to 78%**, while the best validation accuracy achieved was **78.24%**. Early stopping was triggered at **epoch 4**, indicating that the model's performance on the validation set had plateaued, preventing unnecessary training beyond this point.

**Model2:AxialT2**
The Axial T2 model showed a training accuracy between **70% and 71%**, with the best validation accuracy reaching **71.43%**. Notably, early stopping was **not triggered** for this model, meaning it continued training for the full 5 epochs without observing a plateau in validation accuracy.

**Model3:SagittalT2/STIR**
The Sagittal T2/STIR model achieved a training accuracy ranging from **87% to 88%**, with the best validation accuracy recorded at **87.13%**. Similar to Model 1, early stopping was triggered at **epoch 4**, preventing further training once the model's performance had stabilized.

These results demonstrate the varying levels of performance across the three models, with the Sagittal T2/STIR model showing the highest training and validation accuracy.

**Loss Curves**

The loss curves for each model provided valuable insights into the training process and the model's ability to generalize to unseen data.

**SagittalT1:**
The Sagittal T1 model exhibited **rapid convergence**, with the training accuracy quickly improving and validation accuracy plateauing early on. This pattern suggests that the model quickly learned the main features in the data but may have stopped improving once it captured the most salient patterns. The early plateau in validation accuracy could also indicate that the model was overfitting to the training data, as it struggled to improve its performance on unseen data after a certain point.

**AxialT2:**
The Axial T2 model showed a continuous increase in training accuracy, which was a positive indicator of the model's ability to learn the features in the training data. However, **validation accuracy remained**

**stable**, suggesting that the model was not generalizing well to the validation set. This discrepancy points to potential **overfitting**, where the model performs well on training data but fails to generalize to new, unseen examples. The stable validation accuracy despite increasing training accuracy suggests that the model may have memorized the training data instead of learning the underlying features.

**SagittalT2/STIR:**
The Sagittal T2/STIR model demonstrated **high training accuracy**, but its validation accuracy fluctuated throughout the training process. These fluctuations suggest that the model might be overfitting to the training data, as it showed a tendency to perform well on the training set but struggled with consistency on the validation set. The erratic validation accuracy could also imply that better regularization techniques or adjustments to the model's architecture were needed to improve its generalization ability.

**Discussion**

**Challenges**
During training, two main challenges affected model performance:

**Overfitting:**
Many models performed well on the training data but failed to generalize effectively on the validation data. To reduce overfitting, methods such as **data augmentation**, **dropout layers**, and **L2 regularization** could be incorporated. These would help the model generalize better and avoid memorizing the training data.

**ClassImbalance:**
Certain conditions were underrepresented in the dataset, which led to poor generalization for those conditions. Techniques like **resampling** or **class weighting** could help balance the classes and improve the model's performance on underrepresented conditions.

**5.2ModelPerformance**
The performance of each model revealed areas for improvement.

**SagittalT2/STIR:**
This model achieved the highest training accuracy, but validation accuracy improvements were limited, suggesting overfitting. Regularization techniques could help improve generalization to unseen data.

**SagittalT1 and AxialT2:**
Both models showed more **stable but suboptimal validation accuracy**, indicating that there is room for improvement in model tuning and data augmentation to boost their performance.

In conclusion, while the models performed reasonably well, addressing overfitting and class imbalance through regularization and augmentation will be important for improving generalization and overall performance.

**Conclusion**

This project successfully applied deep learning models, specifically **EfficientNetV2**, to classify lumbar spine degenerative conditions from MRI images. The models demonstrated strong **training accuracy**, indicating their ability to learn relevant features from the data. However, the models struggled to achieve significant improvements in **validation accuracy**, which often suggested **overfitting** to the training data.

The results highlight the potential of deep learning in the field of medical image analysis, particularly for classifying complex conditions such as those affecting the lumbar spine. Despite the challenges faced, the project underscores the effectiveness of **EfficientNetV2** in image classification tasks. However, to enhance the models' ability to generalize and prevent overfitting, further work is needed. Techniques such as better **regularization**, **data augmentation**, and addressing **class imbalance** should be explored in future work to improve model performance and robustness, ultimately enabling more reliable and accurate automated diagnosis in clinical settings.

### Future Work

Future work on this project will focus on improving model performance and addressing the challenges identified during training. Key directions include:

**AdvancedArchitectures:**
Exploring more advanced deep learning architectures, such as **transformers** or **hybrid models** that combine convolutional layers with **attention mechanisms**. These models have shown promise in improving performance, particularly in capturing long-range dependencies and enhancing feature extraction from complex data like medical images.

**DataAugmentation:**
Implementing additional **data augmentation strategies**, such as **rotation**, **flipping**, and **cropping**, will help increase the model's ability to generalize to unseen data. These techniques will allow the model to learn more robust features by simulating different variations of the MRI images, thereby improving its performance in real-world scenarios.

**ClassImbalanceSolutions:**
Investigating and addressing **class imbalance** using techniques such as **SMOTE (Synthetic Minority Over-sampling Technique)** or **class weighting** during training. These methods will ensure that the model is not biased towards the majority class, thus improving its ability to accurately classify underrepresented conditions.

## References:


1. V. M. Ravindra, S. S. Senglaub, A. Rattani, M. C. Dewan, R. Härtl, E. Bisson, et al., "Degenerative lumbar spine disease: Estimating global incidence and worldwide volume", *Global Spine J.*, vol. 8, no. 8, pp. 784-794, Dec. 2018.
2. Lumbar Disease Classification Using an Involutional Neural Based VGG Nets(INVGG) https://ieeexplore.ieee.org/document/10440340
3. Glassman SD, Bridwell K, Dimar JR, Horton W, Berven S, Schwab F. The impact of positive sagittal balance in adult spinal deformity. Spine (Phila Pa 1976). 2005;30(18):2024–9.
4. Lafage V, Schwab F, Patel A, Hawkinson N, Farcy JP. Pelvic tilt and truncal inclina- tion: two key radiographic parameters in the setting of adults with spinal deformity. Spine (Phila Pa 1976). 2009;34(17):E599–606.
5. Schwab FJ, Blondel B, Bess S, Hostin R, Shaffrey CI, Smith JS, et al. Radiographical spinopelvic parameters and disability in the setting of adult spinal deformity: a prospective multicenter analysis. Spine (Phila Pa 1976). 2013;38(13):E803–12.
6. Ruchi et al., "Lumbar Spine Disease Detection: Enhanced CNN Model with Improved Classification Accuracy," in IEEE Access, vol. 11, pp. 141889-141901, 2023, doi: 10.1109/ACCESS.2023.3342064.
7. B. H. Cho et al., "Automated measurement of lumbar lordosis on radiographs using machine learning and computer vision," Global spine

journal, vol. 10, no. 5, pp. 611–618, 2020.
8. O. Khan, J. H. Badhiwala, G. Grasso, and M. G. Fehlings, "Use of machine learning and artificial intelligence to drive personalized medicine approaches for spine care," World neurosurgery, vol. 140, pp. 512–518, 2020.
9. Grading scale of Lumbar Degenerative disease in full spine X-Ray https://www.scielo.br/j/coluna/a/cqTP8ypdGQrDDPBxFWmYzVm/
10. Tyler Richards, Jason Talbott, Robyn Ball, Errol Colak, Adam Flanders, Felipe Kitamura, John Mongan, Luciano Prevedello, Maryam Vazirabad.. (2024). RSNA 2024 Lumbar Spine Degenerative Classification. Kaggle. https://kaggle.com/competitions/rsna-2024-lumbar-spine-degenerative-classification.