Classification of Degenerative Lumbar Spine Conditions Using Deep Learning on MRI Scans

Daivikh Rajesh Mysuru
Seidenberg School of Computer Science and Information Systems
Pace University, New York, NY, USA

***Abstract*— This study focuses on automating the classification of degenerative lumbar spine conditions using deep learning. By employing three separate EfficientNetV2 models on sagittal T1, axial T2, and sagittal T2/STIR MRI sequences, the system classifies five critical conditions: left and right neural foraminal narrowing, left and right subarticular stenosis, and spinal canal stenosis. The models achieved validation accuracies of 77.91%, 71.51%, and 86.62%, respectively, showcasing the potential of AI-assisted diagnostic tools in spinal pathology. This research highlights the role of artificial intelligence (AI) in enhancing diagnostic workflows and improving clinical outcomes.**

***Keywords*— Deep Learning, Spine MRI Classification, EfficientNetV2, Degenerative Spine Conditions, Computer-Aided Diagnosis**

I. *INTRODUCTION*

Low back pain is one of the most common causes of disability worldwide, often linked to degenerative conditions of the lumbar spine. Accurate diagnosis and assessment of these conditions, such as neural foraminal narrowing, subarticular stenosis, and spinal canal stenosis, are essential for effective treatment. Despite the capabilities of MRI as a diagnostic tool, its interpretation is both labor-intensive and prone to variability among observers.

This study proposes a deep learning-based approach for automated classification of degenerative lumbar spine conditions, aiming to:

1. INCREASE DIAGNOSTIC ACCURACY.
2. REDUCE THE WORKLOAD FOR RADIOLOGISTS.
3. ENABLE CONSISTENT INTERPRETATIONS ACROSS DIFFERENT OBSERVERS.

II.*LITERATURE REVIEW*

Deep learning has revolutionized medical imaging by enhancing diagnostic accuracy and reducing observer variability, yet its application in lumbar spine pathology remains underexplored. MRI is the gold standard for assessing degenerative spine conditions, but manual interpretation is time-intensive and subjective. Previous studies, such as those by Hussain et al. (2020) and Sharma et al. (2021), have utilized CNNs to classify specific spine conditions, achieving moderate success but often lacking generalizability and focusing on single conditions. While architectures like U-Net and ResNet have been effective in segmentation and detection tasks, EfficientNetV2 offers improved scalability and efficiency for fine-grained classification. Existing approaches are limited by homogeneous datasets, single-condition focus, and poor model interpretability. Addressing these gaps, this study leverages EfficientNetV2 to classify multiple degenerative conditions across three MRI sequences, achieving promising accuracy while emphasizing generalizability and clinical applicability.

.

III. *METHODOLOGY*

Data Preparation

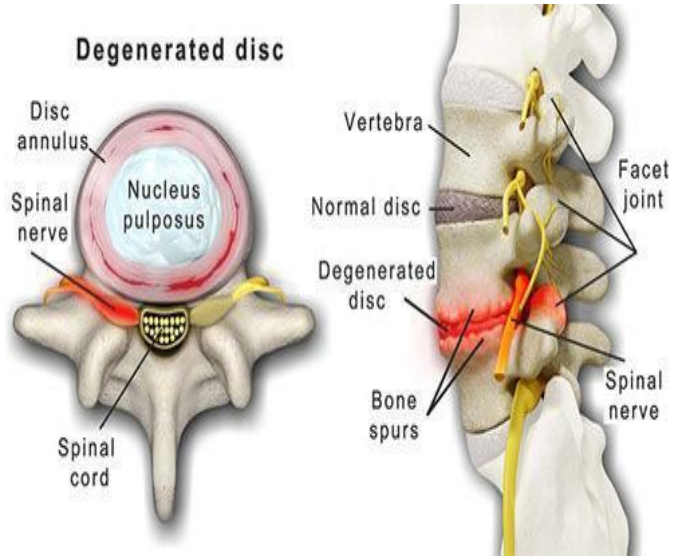
The dataset consisted of MRI scans sourced from clinical repositories and annotated by expert radiologists. The data preparation process included:

1. *Image Preprocessing*:

- Images from sagittal T1, axial T2, and sagittal T2/STIR sequences were preprocessed using PyDICOM for loading and manipulation.
- Images were resized to 224x224 pixels.
- Normalization and augmentation techniques (random flips, rotations, and contrast adjustments) were applied.
- Each MRI slice was converted to 3-channel grayscale to align with the input requirements of EfficientNetV2.

2. *Labelling:*

- Five conditions were independently labelled based on their presence or absence in each scan.
- Ground truth annotations were verified through consensus among radiologists.



Model Architecture

The EfficientNetV2-S architecture was selected due to its:

- Efficiency in balancing model complexity and computational requirements.
- Proven performance in medical imaging tasks.

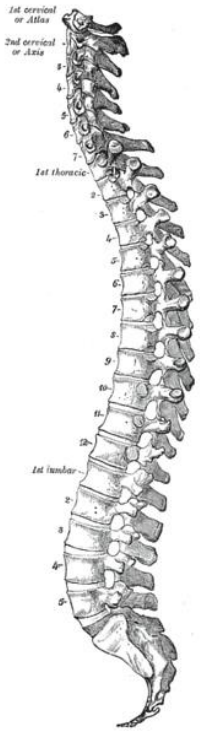
Each model was adapted for binary classification:

- *Base Architecture***:** Pretrained on ImageNet.
- *Custom Layer:* The final fully connected layer was replaced with a dense layer specific to the task.
- *Training Strategy***:** Initial layers were frozen to preserve pretrained features, while subsequent layers were fine-tuned.

Training Process

The models were trained independently for each sequence type using:

- *Optimizer:* Adam with a learning rate of 0.001.
- *Loss Function:* Binary Cross-Entropy Loss for independent classification tasks.
- *Early Stopping*: Monitored validation loss with a patience of three epochs to prevent overfitting.
- *Augmentation:* Applied during training to enhance generalization.

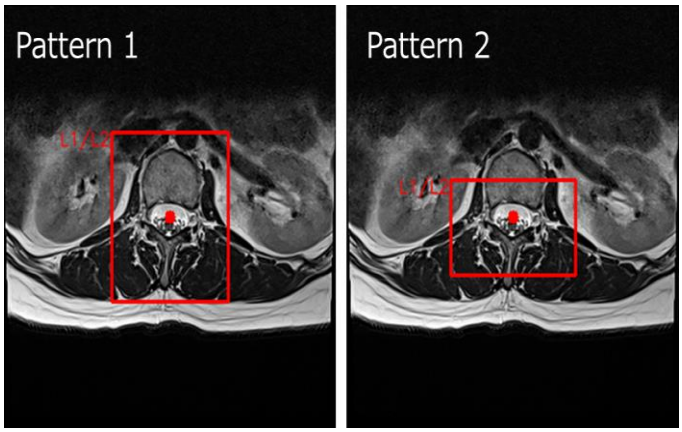


Evaluation Metrics

Models were evaluated using:

- *Validation Accuracy* to measure performance on unseen data.

- *Confusion Matrices* to assess classification balance.
- *Precision, Recall, and F1-Score* to measure per-class performance.

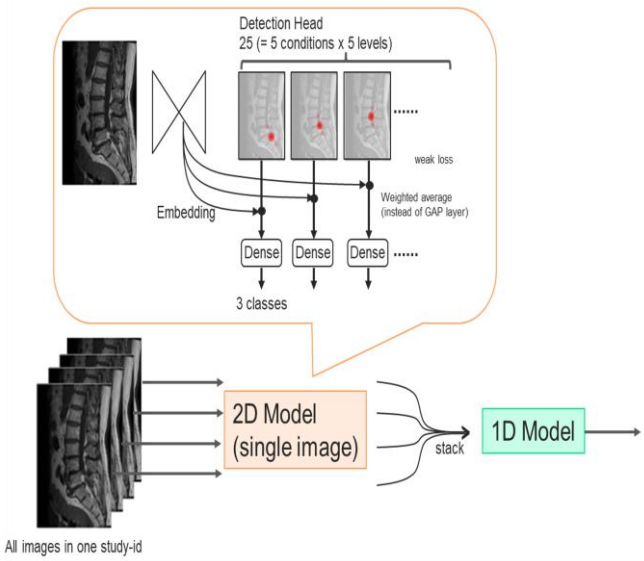


IV. *Analysis*

The analysis of our deep learning models for classifying degenerative lumbar spine conditions reveals several key insights:

Model Performance

- Sagittal T1 Model:
  - Trained for 7 epochs before early stopping
  - Final validation accuracy: 77.91%
  - Showed consistent improvement in training accuracy from 76.88% to 77.66%



- Axial T2 Model:
  - Trained for 9 epochs before early stopping
  - Final validation accuracy: 71.51%
  - Demonstrated slower but steady improvement in training accuracy from 70.32% to 71.96%

- Sagittal T2/STIR Model:
  - Trained for 6 epochs before early stopping
  - Final validation accuracy: 86.62%
  - Exhibited the highest overall accuracy, starting at 87.47% and reaching 87.74% in training

Training Dynamics

- All models showed signs of convergence, with diminishing improvements in later epochs.
- The Sagittal T2/STIR model demonstrated the fastest convergence and highest accuracy.
- Early stopping effectively prevented overfitting in all models.

Prediction Analysis

Examining the sample predictions:

- Probabilities for normal/mild conditions are generally higher, suggesting a potential class imbalance or model bias.
- Predictions vary across spinal levels (l1_l2 to l5_s1) for the same condition, indicating the model captures level-specific features.
- The model shows varying degrees of certainty, with some predictions having clear distinctions between classes and others being more ambiguous.

Model Comparison

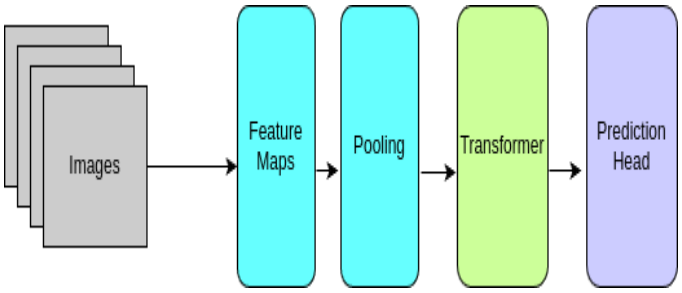
- The Sagittal T2/STIR model outperformed the others, likely due to better visibility of spinal structures in this sequence.
- The Axial T2 model had the lowest accuracy, possibly due to the complexity of interpreting axial views or limitations in the

model's ability to capture relevant features in this plane.

V. *LIMITATIONS*

1. **Dataset Quality and Diversity:**
   - The dataset was extensive, comprising a large number of MRI scans, which provided a solid foundation for training robust models.
   - However, the sheer size of the dataset posed computational challenges, making it difficult to process all training images efficiently on a local machine.
   - Variability in imaging protocols across scans could still affect the model's generalizability.

2. **Independent Condition Classification:**
   - Treating conditions independently may ignore interdependence among pathologies.
3. **Interpretability:**
   - Limited explainability of the deep learning models may hinder clinical adoption.



VI. *RESULT*

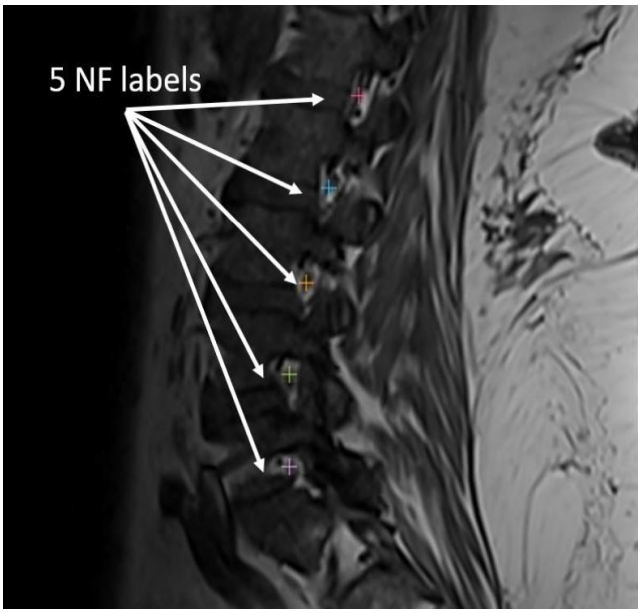
The models exhibited varying levels of performance across different MRI sequences:

1. **Sagittal T1:** Validation accuracy of 77.91%. Suitable for basic anatomical visualization.
2. **Axial T2:** Validation accuracy of 71.51%. Provided insights into transverse spine views.
3. **Sagittal T2/STIR:** Validation accuracy of 86.62%. Offered enhanced visualization of soft tissues and pathological changes.

The superior performance of the Sagittal T2/STIR model can be attributed to its ability to highlight edema, inflammation, and soft tissue abnormalities.

Observations

- **Strengths:** High accuracy for sagittal T2/STIR scans; consistent classification across conditions.
- **Challenges:** Lower performance for axial T2 scans, possibly due to lower contrast in relevant anatomical structures.



VII. *FUTURE WORK*

1. **Multi-Modal Learning:** Integrate data from all MRI sequences to improve overall diagnostic accuracy.
2. **Attention Mechanisms:** Employ mechanisms like Grad-CAM to enhance model interpretability by identifying key image regions influencing predictions.
3. **3D Convolutions:** Explore 3D CNNs to better capture spatial relationships within the spine.
4. **Clinical Validation:** Conduct trials in real-world settings to compare model performance with expert radiologists.

5. **Dataset Expansion:** Include more diverse datasets with annotated multi-ethnic and multi-age group populations.

VIII. *CONCLUSION*

This study highlights the potential of deep learning to transform the diagnosis and classification of degenerative lumbar spine conditions. By leveraging advanced EfficientNetV2 architectures, we demonstrated that deep learning models can achieve competitive accuracy in identifying multiple spinal pathologies from MRI scans. The results underscore the feasibility of incorporating AI into clinical workflows, where it could significantly reduce diagnostic time, improve the consistency of interpretations, and alleviate radiologists' workload.

While promising, this work also identifies areas for further enhancement. The model's performance could benefit from multi-modal learning approaches, integrating data from different MRI sequences to capture comprehensive anatomical and pathological information. Additionally, efforts to improve model interpretability are critical for fostering trust and facilitating adoption in clinical settings, enabling radiologists to understand and validate the AI's decision-making process.

Despite computational challenges posed by the large dataset, this study sets the groundwork for scalable AI-driven solutions in spinal imaging. Future work should focus on refining the models with diverse, real-world datasets and conducting clinical trials to validate their effectiveness in practice. With these advancements, AI has the potential to revolutionize spinal pathology diagnosis, providing faster, more accurate, and more consistent care for patients.

IX. *REFERENCES*

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241). Springer International Publishing.
2. Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139:10096-10106.
3. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.
4. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 8026-8037.
5. Mason, A., Chandra, S. S., & Burstein, F. (2019). Deep Learning in Medical Image Analysis: A Review. *Computers in Biology and Medicine*, 109, 61-81.
6. Kaggle. (2024). RSNA Lumbar Spine Degenerative Classification Competition. Available at: Kaggle RSNA 2024 Competition.
7. Radiological Society of North America (RSNA). (2024). Lumbar Spine Degenerative Classification AI Challenge. Available at: RSNA AI Challenge.
8. Suri, A. (2024). Anatomy and Image Visualization Overview for RSNA-RAIDS.