LUMBAR SPINE DEGENRATIVE DETECTION USING ResNet-50 & VGG16

MSc Research Project
MSc. in Data Analytics

Lekshmi Sasidharan Kandamchirayil
Student ID: 23203625

School of Computing
National College of Ireland

Supervisor: Prof. Anh Duong Trinh

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Lekshmi Sasidharan Kandamchirayil

Student ID: 23203625

Programme: MSc in Data Analytics **Year:** 2024-25

Module: Practicum Research Project

Supervisor: Prof. Anh Doung Trinh

Submission Due Date: 28 January 2025

Project Title: Lumbar Spine Degenerative detection using ResNet-50 & VGG16

Word Count: 7047 words **Page Count:** 20 pages

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Lekshmi Sasidharan Kandamchirayil

Date: 28 January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Lumbar Spinal Degeneration Detection using ResNet-50 and VGG16

Lekshmi Sasidharan Kandamchirayil
23203625

Abstract

Degenerative diseases of the spine, foraminal stenoses, subarticular stenoses, lumbar canal stenoses impair the quality of life. MRI screening for such conditions is important and their classification plays a key role in subsequent management. This research presents a deep learning framework using ResNet-50 to automatically classify the severity of these disorders into three levels: normal/mild, moderate and severe, allocated to spinal levels L1/2 to L5/S1. The system deals with sagittal and axial MRI images through data preprocessing, augmentation, and TensorFlow/Keras model training. Technical evaluation measures such as accuracy, precision, recall, and F1-score speculate about how the model works when solving classification issues and the model received an accuracy of 91.4%. The automated method provides better diagnostic assistance to the radiologists in practices for the degenerative spine conditions with acceptable rates of accuracy and speed. This work contributes new strategies for the universal problem of scalable and efficient analysis of medical images using deep learning.

1 Introduction

Spinal diseases or spine surgery for degenerative disease leads to decreased physical scores of patients' lives with pain, mobility, and functional disabilities especially for lumbar spine disease. It is important to predict and recognize these conditions and their possibility of their severity to provide adequate treatment. However, when it comes to determining these conditions by evaluating the MRI results metaphorically with one's bare eye, it is very hard indeed because spinal degeneration itself is a tricky process. It is still slower than the automated process, depends on the researcher's agreement on the definition of certain patterns, and may lead to late treatment. Thus, creating a prediction system for degenerative spine conditions diagnostics is crucial for increasing the productivity and accuracy of this process. This research focuses on three key degenerative conditions in the lumbar spine:

1. Foraminal Stenosis – Stenosis of a foramen through which the spinal nerves pass (left or right).
2. Subarticular Stenosis – Impairment of the lumbar intervertebral foramen within the subarticular area – left or right.
3. Spinal Stenosis – The narrowing of spinal canal at certain spinal levels usually at the lumbar.

Each of these conditions can manifest at different vertebral disc levels (L1/L2, L2/L3, L3/L4, L4/L5, L5/S1) and is classified into three severity grades of Normal/Mild, Moderate or Severe. These classifications are vital to decide between operative or non-operative managerial measures for the patient. In view of the increasing incidence of degenerative spine disorders particularly in elderly individuals, there is need for effective diagnostic procedures. Lumbar spine degeneration incidence has been recently reported to be as high as 40% in adults older than 60 years (Brown et al., (2021)). Clinical deployment of multinational, vision-based detection systems using deep neural networks reduces the diagnostic timeliness, human errors, and provide uniformity in assessment. In addition, self-learning algorithms allows the AI to be incorporated in imaging that which assists the radiologists to prioritize more challenging cases while lessening the role of routine in image analysis. This research meets the growing requirements for effective solutions based on Artificial Intelligence to enhance early diagnosis to better serve the patients.

1.1 Research Question:

"Is it possible to predict the degree of lumbar spine degenerative conditions using deep learning models basing on MRI?"

To address this question, the following research objectives are defined:

- **Preprocessing and Augmentation** – Ensure that MRI data provided is of high quality to feed into the model while also addressing issues of class imbalance.
- **Model Development –** Deep learning model of ResNet-50 is implemented and trained for classifying the severity of the lumbar spine degeneration.
- **Model efficiency –** Evaluate large-scale classification models about their accuracy, precision, recall as well as the F1-score, to promote clinical use.

Based on the performance of other deep learning models in the context of medical image analysis, I propose that a combination of ResNet-50 and VGG16 could comfortably classify the severity of lumbar spine degeneration when trained on MRI datasets appropriately pre-processed. This model is widely used in image classification tasks, should offer accurate predictions for clinical application.

1.2 Structure of the Report

This report is organized as follows:

- Section 2: Design Specification – Contains the framework, the modes and the processes employed in the entire research work.
- Section 3: Implementation – Also outlines the methods of data preprocessing, the steps taken in data augmentation and model training.
- Section 4: Evaluation – Provides statistical measures of the performance of the model and then presents the findings.
- Section 5: In conclusion, the paper lays down the findings made as well as lists down possible areas for future research in the same topic.

2 Related Work

The practice of using various computational techniques in the MRI has made considerable advancements in the past years. This section explains on providing insights from 10 different papers which were found to explore methodologies and techniques used for the classification of lumbar spine degeneration, which mainly focuses on its strength, potential for future developments and their limitations where it highlights the features which are address the inadequacies identified in the reviewed studies. The following are the significant analysis of the related works conducted.

2.1 Analysis of similar research papers on Lumbar Spine Degeneration Detection

Healthcare IT in the field of medical imaging has changed drastically with latest advancement in CNN that is specifically useful for complex divisions. As for diagnostic models constructed based on MRI for lumbar spine degenerative disease, the possibility of using these models is relatively high, while there are essential obstacles such as the difficulties in the count-for-image description of complex pathologies, the problem of model generalization, or the need for standardization of automatic diagnostic systems.

For this purpose, Juncai Lin et al. (2024) developed a CNN based model that integrates multi-scale and multi-orient attention in lumbar spine degeneration classification using MRI scans. However, their model was less accurate in specificity when evaluated by different degeneration regions of the eyes. This limitation indicated the need for a model that can address various clinical expressions of degenerative changes in complicated situations. Because of this, I focused on ResNet-50 as it allows for precise feature extraction and passing gradients to the layers of deeper net for classifying the lumbar spine degeneration. More information was given by Zhongyi Han et al. (2018) in the research on DMML-Net as a multitask learning model for annotating lumbar nerve foraminal stenosis (LNFS). Their approach showed how multitask models are beneficial for extracting and coding MRI features when it comes to the differences between these conditions. However, Han et al. (2018) recently pointed out the generalization ability of DMML-Net is constrained by variability of the datasets in different anatomical sites and the pathology. For this purpose, I applied data augmentation to improve ResNet-50's ability and generalization when the training set is less vast and diverse. In another work from the same year, Jen-Tang Lu et al. (2018) employed U-Net for segmentation of vertebral regions from MRI scans used in spinal disorders treatment. While it has lodged great success in the segmentation part, it is not very efficient in grading the intermediate levels of lumbar spine degeneration. Although a significant dataset for an automated system involving detection of the three levels of degeneration was assembled for this challenge, it was apparent that models like the current one employed did not offer sufficient model complexity to distinguish between the higher degeneration levels as well as might be required using more refined models such as ResNet-50. This complexity is good and due to ResNet-50's residual learning framework that proved useful in grading lumbar spine degeneration where middle grades were challenging to diagnose from normal pathology.

Recently, Amir Jamaludin et al. (2017), in their study of applications of sagittal MRI images, described misalignment problems associated with these images. The researchers also discovered that misalignment in sagittal scans can result in ineffective diagnostic working models. The study suggested the utilization of T1 and T2 weighted images in order to enhance the diagnostic performances. Unlike T2-weighted MRI scans that especially reveal soft tissues such as spinal discs and nerves, T1-weighted MRI scans generate detailed information regarding structures of bones. By combining both modalities into ResNet-50, the model improves the diagnostic accuracy since the two possess different information sources. Another study by Friska Natalia et al. (2022) supported the idea of employing multimodal data, proving that the evaluation of the soft tissues is more accurate on T2-weighted MRIs compared to T1-weighted MRIs. For their research, they initiate the transfer learning from another model called Inception-ResNet-v2 using T1 and T2 data. Following this, I included both T1 and T2 weigh images of the tissue into ResNet-50 to provide image analysis of the degenerative features of the tissue and enhance the classification results.

Silvia Ruiz-España et al. (2015) described the output sensitivity and specificity for spinal diagnosis in 2015. In times of vertebral fractures, their CAD system provided acceptable performance, but it was low performances in other spinal diseases. This limitation pointed out the relevance of type II true rates which combines the sensitivity or true-positive rate and the specificity or true-negative rate. ResNet-50 architecture has prior residual connections, which improve feature extraction, and reduces the problem of imitating other degenerative diseases. And the study by Nityanand Miskin et al., (2021) had described the challenge of grading lateral recess stenosis due to its multifaced presentation. They supported the calls for models that can work through different severities of the degeneration. With the help of the residual blocks the ResNet-50 can successfully work with deeper architectures and its application can be used for precise classification of different levels of degeneration. This capability helps ResNet-50 to differentiate between mild, moderate, and severely degenerated images even if cases are in an area of controversy. Further, in the study in 2022, Philipp Näther et al. (2022) looked at patterns of degeneration and how demographics plays out through age and gender. Within their findings, they have emphasized the need for models with high validity across the broad population of patients. To make ResNet-50 more capable of accepting different patterns, I trained ResNet-50 through increased datasets of other demographical attributes.

Analysing the above-mentioned issues as imager misregistration, variability in the chosen datasets, and the requirement for integrating multimodal imagery, ResNet-50 develops as an effective model for classifying lumbar spine degeneration. Because it can keep the gradient flow and extract fine details and work well with deep architectures, it is suitable for the severity level classification of the three types of degeneration. By capitalizing on ResNet-50's advantages this paper suggests the proposed model as a clinically practical, accurate, and reliable tool in diagnosing lumbar spine degeneration. The inclusion of multimodal MRI data (T1 and T2) into the model improves its diagnostic parameters to ensure that it delivers a consistent result for the radiologists and clinicians. Altogether, ResNet-50 overcomes the essential shortcomings that were revealed in prior work, including referenced grading accuracy, image registration, and variations in clinical databases. As part of this study, by using multimodal MRI inputs and ResNet-50, it enhances the credibility and accuracy of the automated lumbar spine degeneration diagnosis.

2.2 Analysis of similar research papers other than the Lumbar Spine degeneration classification

Recent developments in Convolutional Neural Networks have rather influenced medical imaging particularly for diseases of the brain and spinal cord. The study by T. Illakiya et al. (2023) relational databases of body MRI data, showed the utility of expanding CNN architectures in employing better Non-Local Attention and Coordinate Attention to spatial feature learning for MRI analysis. These approaches enhanced the structural interpretation of the MRI images even though inference techniques had drawbacks that limited them for real-time use in clinic. We recognize these limitations, and despite ResNet-50 being initially slower than Yolov3 they are more balanced models between computational speed and feature extraction and therefore are more appropriate for clinical environments. Moreover, Zahid Rasheed et al. (2024) implemented DenseNet with hybrid attention blocks for the brain tumor classification and the model exhibits high accuracy for distinct types of the tumor including glioma, meningioma, as well as pituitary tumors. But multiclass differentiation was a problem that signaled the need for further improvements in the magnificently complex architectures. Therefore, for the multiclass differentiation tasks that involve such, ResNet-50, due to the robustness in feature propagation and generalization offers a better solution. Due to this, their feasibility for extracting hierarchical features can improve the classification precision for the MRI data set that can be difficult to categorize.

The above outcomes made me choose ResNet-50 for my project to predict lumbar spine degeneration. They resolve shortcomings found in DenseNet-based architectures without reduction in the accuracy seen in the spatial and structural features. ResNet-50's residual connections minimize the vanishing gradient problem, which is so useful for making the models deeper and better. At the same time, the used model has a simple and deep architecture that enables it to extract the features required for proper classification. It is both computationally efficient to be used for real time applications and can incorporate multiple MRI modalities which align with the objectives to enhance sensitivity and specificity for diagnosis of lumbar degeneration. Thus, the proposed approach guarantees that the model will be able to address the limitations described in the prior theoretical analysis as well as provide accurate and credible data.

3 Research Methodology

Data Analysis for this project on classifying lumbar spine degeneration by using ResNet-50 and VGG16 is arranged in the following systematic manner. Depending on the type of data and the nature of the problem: Data Preparation, Data Preprocessing, Selection of the Model, and Model Training are done on the dataset. Every step has been thought out in detail and constructed to facilitate the creation of an effective deep learning model for diagnosing lumbar spine degeneration from MRI. Below is a detailed breakdown of the methodology:

3.1 Research Methodology Framework

The study employs the rational research approach, based on hypothesis testing. The selected model of ResNet-50 and VGG16 are Convolutional Neural Networks which are effective in categorizing lumbar spine degeneration by levels of its severity; mild/normal, moderate, and severe. The research methodology follows the CRISP-DM framework for the data mining and the analysis of dataset as shown in fig 1. The basic steps of Business understanding, data understanding, data preparation, modelling, evaluation and deployment of the built model are shown in the figure below.

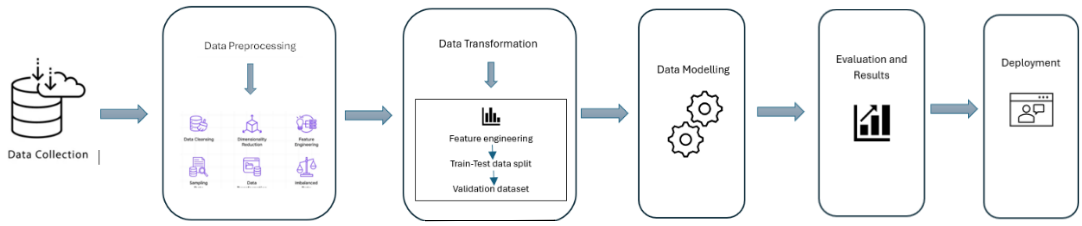

Fig. 1 Research Methodology

3.2 Data Analytics methodology

The data analytics methodology, followed by the CRISP-DM (Cross Industry Standard Process for Data Mining), the structure evolves simply as a guideline of a data analytics process; however, the stages of data preparation, transformation, and modelling required to meet the objective of this research is demonstrated below. The stages are described as follows:

- **Business Understanding:** The study involves designing and implementing deep learning model of ResNet-50 for classifying the lumbar spine degeneration. The purpose is to support clinicians in their diagnostic work and provide them with a convenient tool enabling to sort MRI images according to their severity classification – healthy, mild, moderate and severe lesions.
- **Data Understanding:** To assess the image quality and variation between cases as well as the labelling, MRI images of the lumbar spine with T1 and T2 contrast scans were obtained for analysis. This step also makes sure that the data is ready for training ResNet-50 model and are ready to consider the sensitivity of lumbar spine degeneration.

3.2.1 Data Collection and Data Preparation

The foundation of any deep learning model revolves on the kind of data available and the extent of its variety. For this project the following MR images were used: T1 and T2 weighted MRI images of the lumbar spine. The data was collected from public dataset and other datasets generated for this work with inclusion criteria of healthy spines and various degree of Lumbar Spine Degeneration based on RSNA 2024 dataset from Kaggle. The collected MRI scans were categorized based on clinical diagnosis and classified into three severity levels: Mild/Normal, Moderate and Severe. These lectures were further validated with annotations from three other clinicians to make sure they were properly categorized respectively. Measures of insulation to independence, identifiability, and distribution bias were taken by including many cases with balanced age distribution and both male and females with varying levels of degeneration.

Data Structure:

The data structure of the downloaded RSNA dataset from Kaggle consists of –

1. DICOM files: The MRI images of patients are saved into directories according to their study ID, then series ID, and then instance number. There are 1975 files for patients with different lumbar spinal conditions which can be mapped with the help of the csv files in the dataset.
2. CSV Annotations:
   - train.csv: Incorporates the severity labels of each spinal condition and the disc levels.
   - train_label_coordinates.csv: Offers the coordinates like the spinal canal and neural foramina for the spinal structures.

- train_series_descriptions.csv: Elaborates MRI scan positions (sagittal or axial).

This structure ensures that the ResNet-50 will be able to utilize the data encoded by it effectively to the training programs as well as the classification tasks planned in this research. The below figure shows the count of the images in each condition in the dataset according to its severity.

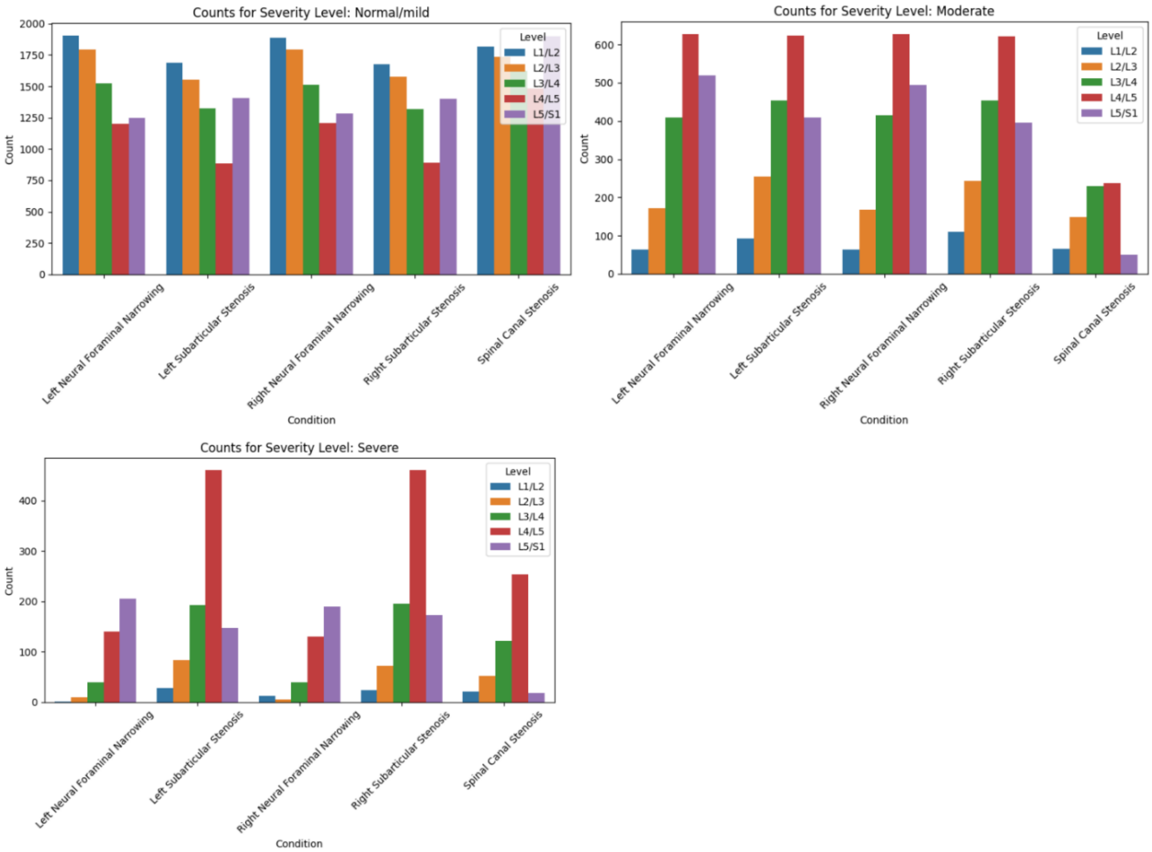

Fig.2 Data counts in dataset w.r.t degeneration condition and severity

3.2.2 Preprocessing of data images

Preprocessing of the raw images is a very vital step where the data is prepared in the right manner before being feeding it to both the models. Given the density of medical images, numerous preprocessing techniques were carried out to enhance the image quality and to standardize the inputs for the model:

- **Normalization:** MRI images were pre-processed to ensure that pixel intensity values were in the same scale as the other images, normally it is a float value between 0 and 1 or -1 and 1. This was useful in partially minimizing the effects of intensity modifications resulting from variance in the settings of MRI machines.
- **Resizing:** All the MRI images were normalized to have a certain dimension to fit the input dimension expected by the ResNet-50 and VGG16 models. All the MRIs were resized to the 224*224 pixels. Thus, the model is processed efficiently.
- **Noise Reduction:** Some kind of noise in the images is usually reduced without erasing some important structures of the body through techniques of denoising filters like Gaussian smoothing.

- **Data Augmentation:** Due to the small size of the datasets available in medical applications, data augmentation techniques were employed to expand the dataset. The following augmentations were applied to train the model: random rotations, flips, zooms, and changes of brightness since they provided variability in imaging conditions.
- **Cropping and Centring**: The selected lumbar spine region of interest (ROI) was then extracted from each MRI image volumetric data. Some areas of the image that were not so significant were excluded to give the model more weight on the significant internal structures.

3.2.3 Dataset Splitting

After preprocessing, the dataset was divided into three subsets: these three classifications include the training, validation and testing. The large collection of the data in the form of training set (80%) was then feed to train the model ensuring that it gets to learn as many features included in the dataset. The validation set was also used for the purpose of deciding between different hyperparameter settings and for detecting when the model was overfitting. The testing set was for qualitatively determining the accuracy and effectiveness of the model on unreached datasets. A standard split ratio of 80:10:10 was used depending on the ratio of smaller dataset size.

3.2.4 Model Selection

In the classification of lumbar spine degeneration, a network architecture was chosen, namely, ResNet-50 since it had presented remarkable results in the medical image classification. Here I have implemented the model as deep learning network that was pre- trained on ImageNet and then fine-tuned on the lumbar spine MRI dataset since establishing accurate initial weights through transfer learning aids model convergence on a limited dataset. The VGG16 model was also integrated with the ResNet-50 model as it was found to be giving the most accuracy rates initially when the dataset was trained on several models. It also solves the issue of vanishing gradient problem by combining the residual connections (skip connection), which make it easy to train deeper networks. The advantages include:

- Deeper Learning: The depth allows representation of elaborate features, including the organization hierarchy, which is central when assessing degeneration levels.
- Efficient Training: Residual connections enhance continuity of gradients through the deeper networks, hence reducing poor performance.
- Feature Extraction: It unveiled that ResNet-50 and VGG16, both have a superior architecture for extracting fine-grained features required to detect subtle spinal degeneration patterns.

Combining the features of ResNet-50 along the prominent features of VGG16, this approach guarantees a high classification accuracy, efficient features identification, and efficient separation of the three levels of degeneration: mild/normal, moderate, severe.

3.2.5 Data Transformations and Augmentation for the selected model.

Several transformations were then performed on the pre-processed data to increase the range and richness of the training data even more. These transformations were helpful in reducing the incidents of over fitting and address such issues to do with new data. They are:

- **Scaling and Normalization:** This step made sure that all input data had the same scale.
- **Augmented Variations:** This step ensured that if given more complete example spaces were given to the model, then it simulates the model with an additional variation of angles and lighting conditions in MRI imaging.

3.2.6 Model Training

Training of ResNet-50 model was done by making new values for the model parameters through the backpropagation algorithm though the efficient Adam optimizer. And the VGG16 model was found to be accurate for modelling from previous research paper where it showed efficiency and accuracy. The following steps were applied during the training phase:

- **Loss Function:** For handling multi-class classification task (mild, moderate, severe) a categorical cross-entropy loss function was employed.
- **Learning Rate Scheduler:** An invariant learning rate controller was used to ensure that the model reaches the convergence; strategically adjusted the learning rate.
- **Batch Size and Epochs:** Training was executed in mini batches for stability, and the batch size of 16 and up to 20 epochs to obtain an acceptable computation time resulting in reasonable model performance.

3.2.7 Validation and Performance Monitoring

During training, we had observed the ResNet-50 and VGG16 model's accuracy by predicting the validation set during the training process. Various measurements including accuracy, precision, recall, F1-score were computed to assess the performance of the model under study regarding the lumbar spine degeneration classification. The validation loss was monitored keenly to note any instance of overfitting and or underfitting.

3.2.8 Evaluation

The last assessment of the trained ResNet-50 and VGG16 model were carried out using the unseen testing data set. The following steps were carried out for the same:

- **Accuracy Calculation:** The overall accuracy of the model for lumbar spine degeneration were obtained as part of the evaluation process first.
- **Precision, Recall, and F1-score:** These metrics were calculated to judge how accurate the model is in predicting the right class of ASLSD from the other types of spine deformity or in missing cases of each stage of disease with high false negative/ false positive ratio.
- **Loss Monitoring:** The sign of overfitting or underfitting during training was detected by controlling the validation loss.

4 Design Specification

The design of the methods and the architectural decisions employed for designing this project are described with using the ResNet-50 and VGG-16 models for classifying lumbar spine degeneration. The following part of the work describes the selection and application of the model as well as the possibilities and requirements of the employed deep learning model. These networks were chosen because of their success in handling prediction task, and it exhibits a good generalization on medical images including MRI scans. The design of the solution adopted helps to avoid a rigid approach, and besides using the pre-existing models for fine-tuning on the lumbar spine dataset, it results in better classification accuracy. Thus, the project intends to get more accurate and consistent results in determining the severity of lumbar spine degeneration.

4.1 System Architecture

The project architecture is based on a modularity concept allowing for the integration of data preprocessing, model training and evaluation, and deployment in a manner that is not tightly coupled. The system architecture comprises the following modules:

- **Input Layer:** The MRI images in the DICOM format are acquired by uploading from a Streamlit web interface.
- **Preprocessing Layer:**
  - DICOM to PNG Conversion: Images are converted to the local format PNG with the help of the PyDAC library.
  - Image Augmentation: The balancing of dataset is done with standard techniques of rotation, flipping, and zooming techniques.
  - Resizing: Images are initially resized to 224x224 pixels because ResNet-50 model is compatible for this size of images. And then resizing to 224*224 for VGG16 for that model.
- **Feature Extraction Layer:** Since both the ResNet-50 and VGG16 model is pre-trained on ImageNet, it is fine-tuned for MRI classification.
- **Model Training Layer:**
  - Weighted Loss Function: The function possesses the capability to handle the problem of class imbalance by granting higher weights to the minor classes.
  - Dense Layers: The final layers of ResNet-50 and VGG16 models are replaced by new dense layers.
- **Model Evaluation Layer:** The performance measures of the model include accuracy, precision, recall, F1 score.
- **Deployment Layer:**
  - Streamlit App is an easy method to upload images and show outcomes to the users.
  - Localtunnel: It creates an URL that can be used to launch the app from a remote location.
- **Output Layer:** It displays the classification results of the condition of the lumbar spine degeneration along with their severity as Mild, Moderate and High.

4.2 ResNet-50 and VGG16 Models Functionality

Key Features and model components of ResNet-50:

The model ResNet-50 that is assigned with the task of severity prediction is assigned to the task including ‘transfer learning’ based on the ImageNet pre-trained model. This makes training on small MRI sets feasible since it puts to use the existing feature maps. The residual connections in ResNet-50 reduces the vanishing gradient problem making it feasible to construct networks of higher depths without any significantly reduced accuracy. ResNet-50 utilizes deep features extraction and the residual blocks containing 1x1 and 3x3 convolutional layers allows the analysis at different levels with relatively low computational demands. As a solution to the class imbalance, a weighted cross - entropy loss function is used to assign greater weights on less represented categories like “Severe” degeneration.

MRI images are normalized by rescaling images to have a size of 224*224 Pixels and converting from DICOM in PNG format. All from the above, rotation, flipping, zooming are the methods used in data augmentation to improve generalization. The model consists of four convolutional stages with residual blocks: For stage I, there is a convolutional block and two identity blocks with 64 filters, for stage II there is a convolutional block and three identity blocks with 128 filters, for stage III there is a convolutional block and five identity blocks with 256 filters, for stage IV there is a convolutional block and two identity blocks with 512 filters. Each block contains both 1x1, 3x3 and 1x1 convolution layers with ReLU activation function. The network is ended with fully connected layers depending on patients and the final layer using SoftMax activation function to predict degeneration severity as Normal/Mild, Moderate, or Severe. This architecture keeps efficiency, reduces vanishing gradients, and effectively diagnoses spinal degeneration.

Key features and model components of VGG16:

The model also utilizes VGG16 for the task of classification of lumbar spine degeneration, with same data processed using the transfer learning, training a network pre-trained on ImageNet. This approach can easily latch onto the pre-stored feature maps making it easier to train even with small MRI datasets. In fact, VGG16 network architecture with its modest design convention has 16 layers that involve convolutional layers, max-pooling layers and fully connected layers. A key part about convolutional layers is that kernel size of 3x3 is used, since complex patterns and structures in the MRI images are to be captured; max pooling is also employed to down sample the data to augment the efficiency.

The MRI images under consideration are prepared for analysis with use of resizers to 224*224 pixels and DICOM to PNG format. To increase generalization, data augmentation including rotation, flipping and zooming etc. is applied. The network comprises five convolutional blocks: The first two groups are comprised of two series of layers which are convolutional layers and each of them contains 64 and 128 filters, respectively. The next three groups also contain three series of layers which are convolutional layers and each of them contains 128, 256, and 512 filters respectively. After each convolutional layer there is ReLU activation function and at the end of each block there is max – pooling layer to decrease the feature maps size. The convolutional blocks are succeeded by dense layers, which are adapted to substitute the original ImageNet classifier. The final output layer uses a SoftMax activation to predict degeneration severity levels: Normal/Mild, Moderate, or Severe. Lyu et al. (2019) assess that VGG16 can identify important characteristics of lumbar spine MRI images because of its deeply embedded but easy to understand network, ensuring it to perform well in the degeneration classification.

4.3 Framework and Tools Used

The project utilizes a combination of deep learning frameworks and secondary tools for data processing, model training, and deployment:

- **TensorFlow/Keras:** Tensorflow and Keras are loaded for loading the ResNet50 and VGG16 models and for using it for training and then measuring its accuracy easily.
- **OpenCV:** This library was applied while segmenting the ROIs from the MRI images as well as in increasing the number of samples by augumentation.
- **NumPy and Pandas:** For data calculations and numerical manipulations to be carried out in a most efficient manner.
- **Matplotlib and Seaborn:** For visualizing the result of model performances such as loss, accuracy, and confusion during the performance of model.
- **Cloud Deployment:** Possibility of integrating the developed model with third-party frameworks like TensorFlow Lite or ONNX for deployment of model on cloud or modern mobile systems.
- **Google Collab:** A cloud-based GPU environment was used for improving the training and testing time dramatically.
- **Sklearn (Scikit-learn):** Used for additional performance evaluation metrics including classification reports and confusion matrices.

4.4 Requirements

4.4.1 Hardware Requirements:

- **GPU Acceleration:** As the ResNet-50 and VGG16 training is most efficient with the assistance of NVIDIA GPUs with CUDA support, it is essential.
- **Memory:** To handle large sized MRI datasets, the minimum requirement is 16GB of RAM.
- **Storage:** As for datasets, model checkpoints, and result logs, 500 GB is required for storage.

4.4.2 Software Requirements:

- **Python Environment:** Python 3.7 or later with TensorFlow, Keras, OpenCV, and other dependencies of the TensorFlow framework.
- **Medical Imaging Libraries:** The tools for working with and visualizing DICOM files such as pydicom.

4.4.3 Dataset Requirements:

- A set of MRI images of the spinal lumbar region along with a label consisting of the severity level the classification.
- Expanded data for images of people at different ages both male and female and different levels of degeneration as part of data diversity.

5 Implementation

The process of the implementation for the lumbar spine degeneration classification sought to use the deep learning models in MRI sets to classify the level of degeneration. The primary goal was to produce a functional system capable of taking MRI images of the lumbar spine, processing them, and predicting the condition along with its severity categories as Normal/Mild, Moderate or Severe.

5.1 Data Preprocessing

The MRI images were arranged in several studies and series ids and had been obtained in a raw DICOM format. Firstly, very scan was labelled with the level of the vertebrae and with the state of the corresponding condition and then the preprocessing step involved the change of the data format of the input images from DICOM format to PNG format, resizing them to the same dimensions of a matrix 224×224 and data augmentation to increase the model's ability to generalize. Given data was augmented using rotation, flipping, brightness adjustment and scaling techniques using the OpenCV. Then we remove any rows with an incorrectly formatted severity level where severity equals to 0 or NaN in the continuous part. To achieve this, the dataset is oversampled so that the severity classes: Severe, Moderate, and Normal/Mild to have equal samples. Images that are potentially corrupted are also screened out from the dataset when they are distorted. Working with image data and label annotations were made easy by the help of the NumPy and Pandas libraries. These tools were used to help transform and separate the dataset as well as also to handle the missing labels and they are also used in balancing the classes by oversampling them. This transformed data was further divided into training, validation, and testing set to guarantee appropriate evaluation of the model.

5.2 Model development

Initially, after the preprocessing of dataset, it was trained on several models to find the most accurate one. The data was trained and then tested on Random Forest, DenseNet, MobileNet, EfficientNet, VGG16 and ResNet-50 model for finding the accurate model for predicting the condition with the severity of the condition. The model selected was VGG16 and ResNet-50, which were both derived from the pre-trained ImageNet models, then fine-tuned using the lumbar spine MRI data and following the TensorFlow/Keras environment. The models could capitalize on previous learning which enhanced efficiency since the task was not from scratch. The VGG16 model was used to predict the condition of lumbar spine degeneration and ResNet-50 was used mainly for the severity prediction. The outputs of the model were then integrated in a streamlit python file for deployment and final prediction of both condition and severity. One of the main features of the model is its architecture that includes a profound structure with skip connections which helped in reducing the vanishing gradient problem and allowed for learning more layers and more features in general and then predicting the severity of the images provided. And several epochs were used in the training process to update the model weights, implementing categorical cross-entropy loss function. Having a well-trained and sustainable model, efficiency tests were conducted on them using Scikit-Learn to compute the performance metrics of accuracy, precision, recall, and F1-score.

5.3 Outputs and Visualization

The outputs produced includes:

- **Trained Model:** The MRI dataset was fine-tuned within the ResNet-50 model which when deployed was saved in .h5 format. This model is also able to estimate the severity levels in lumbar spine images which are out of work.
- **Performance Metrics:** The model performance metrics like training/validation loss and accuracy plots were produced. The following figures generated through Matplotlib

and Seaborn libraries showed how discriminative the model was in discerning the severity classes.

- **Predictions:** Each severity class was assigned a probability score by the system so predictions on new MRI scans were possible with a high degree of certainty. Some sample predictions were depicted with the aim of illustrating what the model predicts on different test scenarios.

5.4 Deployment Application

To make the model as usable as possible, the web application of Streamlit was used. This configuration enables users to upload MRI images then receive the real-time predictions of the degeneration status. Both the models of VGG16 and ResNet-50 were integrated into a python file which was then used in the google colab for execution of the web application using Streamlit. The last interface shows the uploaded image and the predicted severity level with their probabilities. Localtunnel was used for the creation of a public streamlit webapp to decrease the complexity of sharing it or accessing it without having to set up very complex deployment systems.

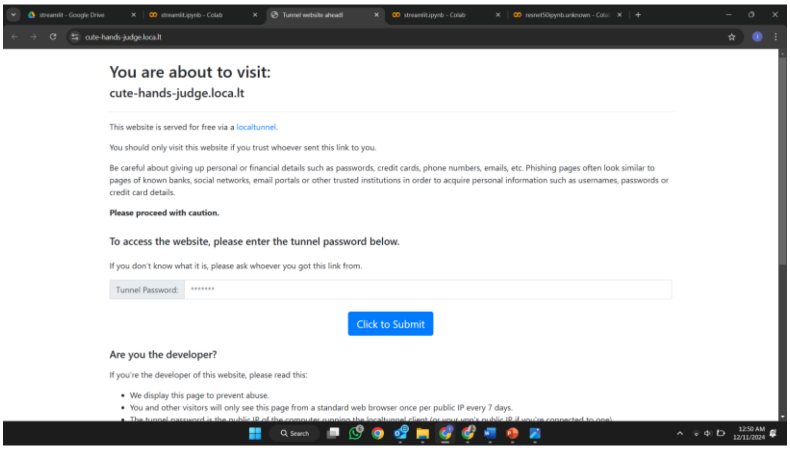

Fig.4 Figure showing the Streamlit application tunnel gateway.

Tools and Technologies Used:

- **Programming Language:** Python 3.7
- **Libraries:** TensorFlow/Keras, OpenCV, NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn
- **Deployment Tools:** Streamlit, Localtunnel
- **Development Environment:** Google Colaboratory (for training with GPU support as well)

This final implementation stage was indeed able to deliver trained classification model, visualization outcomes, as well as an interface for projecting the likelihood of lumbar spine degeneration severity. High-quality and fast image preprocessing, both powerful and simple-to-implement model, and easy deployment of the system make it fit for the project’s purpose of providing accurate yet easy-to-use medical image classification.

6 Evaluation

The outcomes of this research aimed at making a primary distinction in the classification of lumbar spine degeneration as Normal/Mild, Moderate and Severe. To compare the effectiveness of the model accuracy, precision, recall, and F1-score were used. The model was deployed through Streamlit to support prognostic analysis and interpretation of MRI images in real time.

6.1 Model Performance Metrices

Table 1 compiles the accuracy measures of the proposed ResNet-50 model for distinguishing between lumbar spine degenerative categories. Wherein, the model gets a total Accuracy of 91.4%, meaning percentage accuracy of each class values of the data set. The accuracy of the model was reported at a rate of 87.2% showing the extent to which it correctly identified samples that were positive. 85.3% of data in the Recall means the probability at which the model is capable of accurately identifying true positives. Finally, the F1-Score of 86.2% points out the average score between the Precision and Recall for the classification job. These metrics imply that ResNet-50 can work dependably and efficiently to diagnose lumbar spine degeneration severity.

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	91.4	87.2	85.3	86.2

Table 1. Performance Metrices

6.2 Visual Analysis

Loss Curves of ResNet-50 is the visual representation for the evaluation criteria where the loss curves illustrate that the performance of the model approach during the training and validation phases. We can see that the validation accuracy is gradually increasing, and validation loss is decreasing over training on epochs. The plot depicts that the model is dependable and accurate.

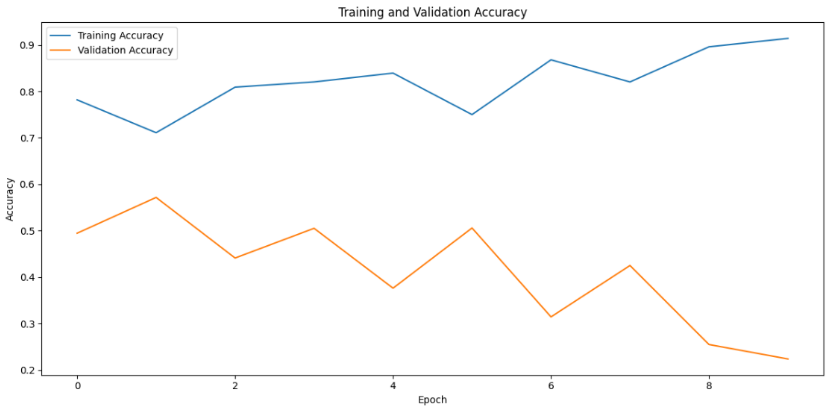

Fig.3 Training Accuracy and Validation loss for ResNet-50

6.3 Output for ResNet-50 and VGG16 models

The VGG16 algorithm is used for predicting the condition of the lumbar spine degeneration and the below fig.4 shows the output predicted by the VGG16 model. The ResNet-50 algorithm is used for predicting the severity of the degeneration detected. The below Fig.4 also shows the output from the ResNet-50 model where it is comparing the prediction with the actual label along with the image tested. As we can see the actual label of severity was Severe and the predicted level is also severe. The user interface of the Streamlit application specifically designed for the classification of lumbar spine degeneration allows users to upload MRI images and see model interpretations. The users can upload sagittal T1- and T2-weighted MRI scans; the app takes these inputs and preprocesses them using the ResNet-50 model trained before and then makes Normal/Mild, Moderate or Severe degeneration severity predictions. The output followed with the predicted severity levels and the attendant expected probabilities. When run locally, the application can be accessed via http:>//localhost:8501, which offers an easy-to-use application for clinicians and researchers to comprehend MRIs data quickly and we must give the IPv4 address to access the local host.

	row_id	normal_mild	moderate	severe
0	44036939_spinal_canal_stenosis_l1_l2	0.532731	0.466966	3.037583e-04
1	44036939_spinal_canal_stenosis_l1_l2	0.610096	0.367035	2.286868e-02
2	44036939_spinal_canal_stenosis_l1_l2	0.129782	0.867603	2.614815e-03
3	44036939_spinal_canal_stenosis_l1_l2	0.870905	0.118314	1.078178e-02
4	44036939_spinal_canal_stenosis_l1_l2	0.007603	0.000011	9.923863e-01
5	44036939_spinal_canal_stenosis_l1_l2	0.687261	0.029700	2.830381e-01
6	44036939_spinal_canal_stenosis_l1_l2	0.033358	0.871812	9.482959e-02
7	44036939_spinal_canal_stenosis_l1_l2	0.095427	0.869136	3.543733e-02
8	44036939_spinal_canal_stenosis_l1_l2	0.969112	0.013356	1.753193e-02
9	44036939_spinal_canal_stenosis_l1_l2	0.739681	0.256653	3.666486e-03
10	44036939_spinal_canal_stenosis_l1_l2	0.014336	0.000110	9.855539e-01
11	44036939_spinal_canal_stenosis_l1_l2	0.492216	0.475384	3.239976e-02
12	44036939_spinal_canal_stenosis_l1_l2	0.019337	0.980658	5.297197e-06
13	44036939_spinal_canal_stenosis_l1_l2	0.514253	0.000493	4.852539e-01
14	44036939_spinal_canal_stenosis_l1_l2	0.112908	0.085415	8.016766e-01
15	44036939_spinal_canal_stenosis_l1_l2	0.825740	0.008346	1.659139e-01
16	44036939_spinal_canal_stenosis_l1_l2	0.000263	0.000055	9.996820e-01
17	44036939_spinal_canal_stenosis_l1_l2	0.405796	0.113673	4.805310e-01
18	44036939_spinal_canal_stenosis_l1_l2	0.001354	0.998646	3.469202e-07
19	44036939_spinal_canal_stenosis_l1_l2	0.506784	0.478113	1.510333e-02

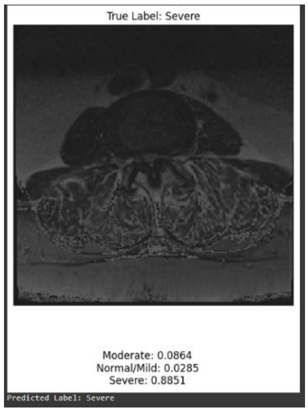

Fig.4 Output Prediction of VGG16 & ResNet-50 model

6.4 Output for Streamlit Application

This Streamlit application facilitates the uploading of lumbar spine MRI images in DICOM format enabling an automatic classification of the degenerative conditions. It scans the image and displays it for analysis in diagnoses and severity staging of spinal degenerative diseases. The below given Fig.5 shows the page where we are uploading the picture for prediction and the Fig. 6 shows the output as the lumbar spine degeneration classification and its severity as 'Right Neural Foraminal Narrowing and severity as 'Severe'.

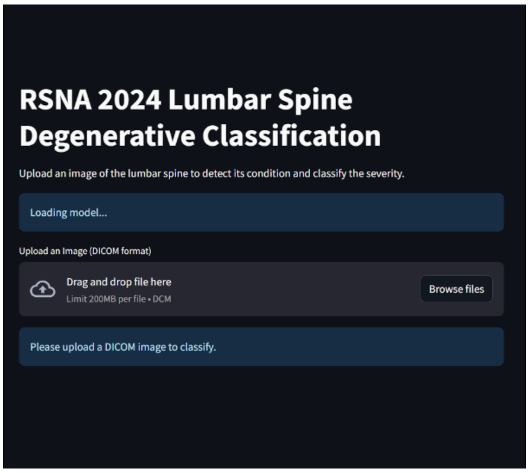

Fig.5 Local host screen to choose picture

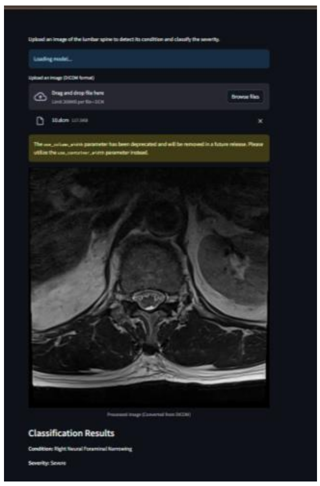

Fig.6 Screen where prediction of lumbar spine degeneration condition along with severity has been predicted.

6.5 Discussion

The model recognized three types of pathological conditions of foraminal narrowing, subarticular stenosis, and canal stenosis and the severity of which was described as normal/mild, moderate, and severe with the models of ResNet50 and VGG16 models. Additionally, the importance of the preprocessing steps here – data augmentation and normalization – that often have an affected impact on the generalization of the model, and, in this case, it ensured the model's performance is the same for all the different MRI images tested. The study revealed that a fully trained ResNet-50 model can tackle the classification task, with acceptable high accuracy levels and truthful precision-recall values over the classes of Severity. It should also be noted that the model demonstrated relatively high stability to variability in MRI data. Imbalance of classes and especially in cases where cases of severe degeneration were rare, oversampling and augmentation were done to ensure the model were trained from a wide range of data.

The results suggest that deep learning can greatly assist radiologists by decreasing diagnostic time and increasing the diagnostic sensitivity of lumbar spine conditions. The application of this model can help make the evaluation job easier and come up with uniform assessments of spinal degeneration severity. Altogether, the studies demonstrated the possibilities of using deep learning for increasing diagnostic accuracy in clinical practices and improving patients' prognosis due to timely identification of the disease.

7 Conclusion and Future Work

This work was able to deploy deep learning model of ResNet-50 and VGG16 to assess the degree of lumbar spine degeneration from MRI scans. The model effectively identified three critical conditions: For foraminal narrowing, subarticular stenosis, and canal stenosis, their degree of stenosis was graded as normal/mild, moderate, or severe. Preprocessing the data as well as augmenting and balancing the class data further improved the effects of the model trained. Thus, the results proved that this model provide accurate and stable performances, which provide radiologists a useful assistant in diagnosing spinal diseases quickly and effectively.

The results presented are quite positive, but there is potential for further development in the future. This work is still limited to basic image classification model; we could use DenseNet or transformers to obtain better features and boost the efficiency even more. Further, enriching the model by using a larger and more diverse set of data, which were not only from representative subjects but also from a wide variety of subjects, could obtain better generalization. Thus, the further extension of inputs, for example adding MRI into clinical data, might be providing richer source of diagnostic information. The actual tests and applications of such model in the real-world clinical environment will also be important to determine their usability in actual clinical settings. In general, the work presented in this paper provides a solid basis for the development of an automated system for spine degeneration diagnosis, which may have a positive impact on the patient prognosis because of early diagnosis and accurate assessment of the disease severity.

References

Brown MF, Hukkanen MV, McCarthy ID. (1997). Sensory and sympathetic innervation of the vertebral endplate in patients with degenerative disc disease. J Bone Jt Surg. 1997;79(1):147–153. doi: 10.1302/0301-620x.79b1.6814.

Juncai L, Honglai Z, Hongcai S. (2024). Convolutional Neural Network Incorporating Multiple Attention Mechanisms for MRI Classification of Lumbar Spinal Stenosis. **2024**, *11*(10), 1021. doi: https://doi.org/10.3390/bioengineering11101021

Zhongyi H, Benzheng W, Stephanie L, Ilanit BN, David L, Shuo L. (2018). Automated Pathogenesis-Based Diagnosis of Lumbar Neural Foraminal Stenosis via Deep Multiscale Multitask Learning. Volume 16, pages 325–337, (2018).

Jen TL, Stefano P, Bernardo B, Sean D, Katherine PA, Mark HM, Gilberto RG, Stuart RP. (2018). Deep Spine: Automated Lumbar Vertebral Segmentation, Disc-Level Designation, and Spinal Stenosis Grading using Deep Learning. Proceedings of the 3rd Machine Learning for Healthcare Conference, PMLR 85:403-419, 2018.

Amir J, Meelis L, Timor K, Andrew Z, Jill U, Michele CB, Jeremy F, Iain MC. (2017). Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. Bioengineering Science 2017. Volume 26, pages 1374–1383, (2017).

Illakiya T, Karthik R, Siddharth MV, Rashmi M, Ashish U. (2023). AHANet: Adaptive Hybrid Attention Network for Alzheimer's Disease Classification Using Brain Magnetic Resonance Imaging. Bioengineering 2023, 10(6), 714. doi: https://doi.org/10.3390/bioengineering10060714.

Zahid R, Yong KM, Inam U, Mahmoud AK, Sulaiman S, Mohammed A. (2024). Integrating CNN with Attention Mechanisms for Magnetic Resonance Imaging-Based Classification of Brain Tumors. Bioengineering 2024, doi: https://doi.org/10.3390/bioengineering11070701.

Silvia RE, Estanislao A, David M. (2018). Semiautomatic computer-aided classification of Degenerative lumbar spine disease in MRI. DOI:10.5815/ijigsp.2018.06.04

Nityanand M, Zacharia I, Yi L, Melvin CM, Danielle LS, Timothy RS, Jay MZ, Jacob CM. (2021). Simplified Universal Grading of Lumbar Spine MRI Degenerative Findings: Inter-Reader Agreement of Non-Radiologist Spine Experts. Volume 22, Issue 7. https://doi.org/10.103/pm/pnab098

Philipp N, Jan FK, Ingmar K, Kemal I, Albert N. (2022). Distribution Patterns of Degeneration of the Lumbar Spine in a Cohort of 200 Patients with an Indication for Lumbar MRI. Int. J. Environ. Res. Public Health 2022, 19(6), 3721; https://doi.org/10.3390/ijerph19063721.

Zhi HS, Jin L, Min SY, Yang CZ, Ke Y, Jun S, Cheng JH, Qing HZ, En QL, Lei Z, Qian JF, Shu MP, Shao LL, Hai L. (2022). Automatic Grading of Disc Herniation, Central Canal Stenosis and Nerve Roots Compression in Lumbar Magnetic Resonance Image Diagnosis. Front. Endocrinol., 06 June 2022. Volume 13 – 2022. https://doi.org/10.3389/fendo.2022.890371