National Institute of Technology Delhi

EEP405 Summer Internship-II

Integration of Convolutional Neural Network and Kolmogorov Arnold Network for Detection of Lumbar Spine Degeneration

By Aditya Chauhan
Summer Research Intern IIT Patna
Supervisor - Dr. Rishav Singh (CSE Dept IIT Patna)

Overview

01 Introduction

02 Problem statement

03 Objective

04 Methadology

05 Result & Discussion

06 Summary

Introduction

- Global Impact of Low Back Pain (LBP): LBP is the leading cause of disability worldwide, affecting over 619 million people, with rates rising as populations age.
- Degenerative Spine Conditions: Common causes include spondylosis, spinal stenosis, and foraminal narrowing, which often lead to pain, restricted mobility, and reduced quality of life.
- Role of MRI in Diagnosis: MRI provides detailed imaging essential for diagnosing and assessing spine degeneration but requires skilled interpretation, making it time-intensive and prone to variability.
- AI in Medical Imaging: AI, particularly deep learning models like Convolutional Neural Networks (CNNs), can aid radiologists by automating the detection and classification of spine degeneration, enhancing diagnostic accuracy and efficiency.
- Kolmogorov-Arnold Network (KAN) Integration: KANs, used alongside CNNs, help capture complex spatial relationships in MRI images, offering improved interpretability and accuracy for detecting degenerative conditions.
- Clinical Importance: AI-assisted models like CNN-KAN can standardize diagnoses, reduce diagnostic variability, and provide faster and more reliable assessments, benefiting radiologists and patients.

Brief about Lumbar Spine

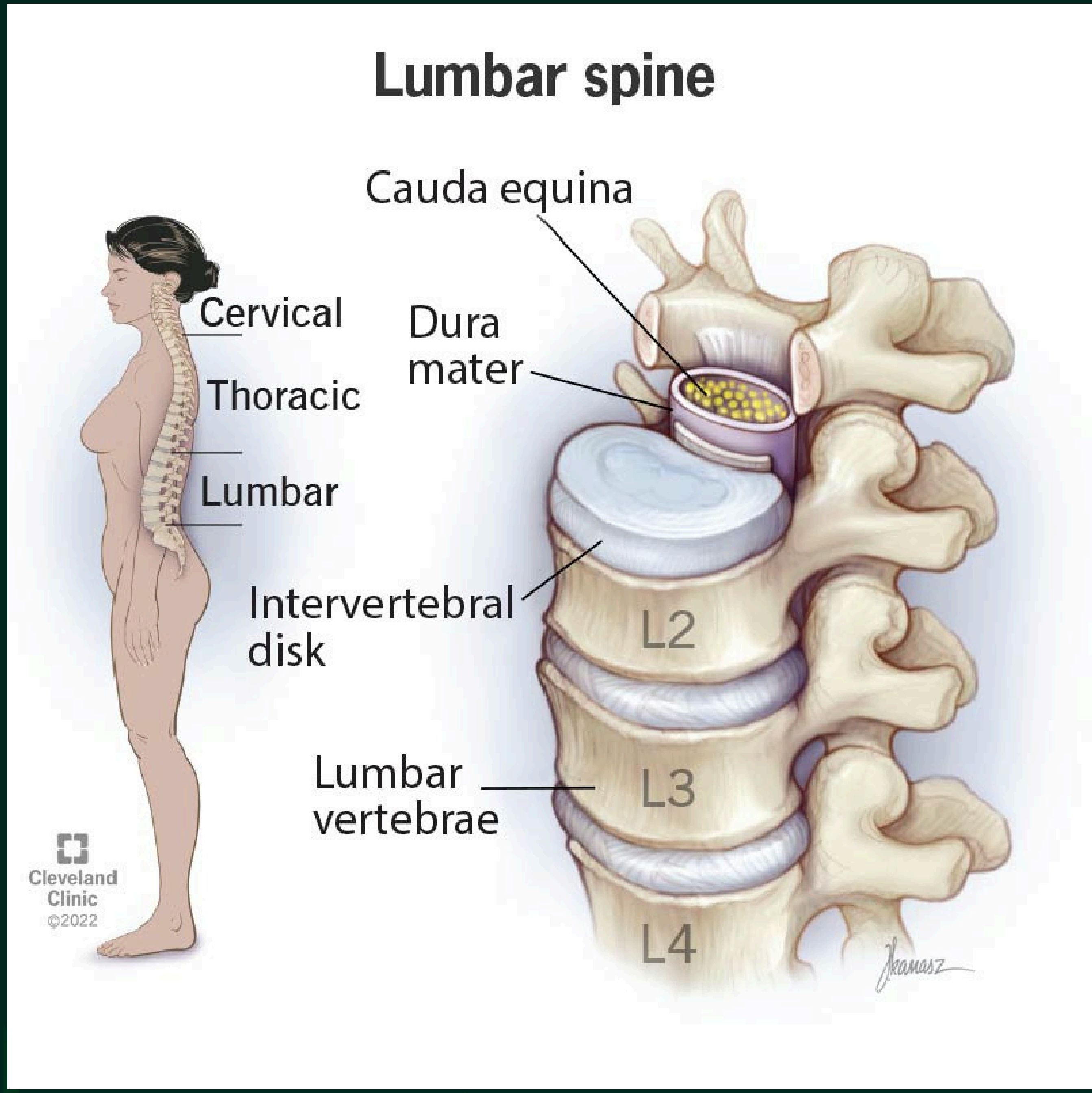
Location: The lumbar spine consists of five vertebrae
located in the lower back

Function:

- Supports the body's weight
- Surrounds and protects the spinal cord
- Enables a wide range of body motions

Common Conditions:

- Lower back pain
- Arthritis
- Degenerative bone and disk disease
- Spinal stenosis



Problem Statement

Classification of five lumbar spine degenerative conditions:

- Left Neural Foraminal Narrowing
- Right Neural Foraminal Narrowing
- Left Subarticular Stenosis
- Right Subarticular Stenosis
- Spinal Canal Stenosis

Dataset Used

- The dataset we used was release by Radiological Society of North America for the purpose of the Kaggle competition names Lumbar Spine Degenerative Classification
- It comprise of MRI images of the spine, focusing on the lumbar region. The images are in the form of standard DICOM files
- The MRI images are identified by a study_id and a series_id. A study_id may consist of one or more series_id
- According to information in the meta data files, each series_id corresponds to a certain kind of lumbar degeneration
- In total, this dataset has 147000 DICOM files and 5 CSV files storing meta data about the former, the dataset is of 35.3 GB data
- For the purposes of training, we have used DICOM files, and trained on both KAN and MLP-based networks. In both the cases, the input was processed through data augmentations and convolutional neural networks

About KAN

The **Kolmogorov-Arnold Theorem** states that any continuous multivariate function can be expressed as a sum of continuous univariate functions and addition operations. Formally, for any such function $f(x_1, x_2, \ldots, x_n)$, there exists a set of continuous univariate functions φ_q and ψ_{pq}, where:

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=0}^{2n} \phi_q \left(\sum_{p=1}^{n} \psi_{pq}(x_p) \right)$$

Comparison with MLPs

- Universal Approximation Theorem is the theoretical foundation for MLPs, while the Kolmogorov-Arnold Representation Theorem is the basis for KANs

- According to the Universal Approximation Theorem, a feedforward neural network with a single hidden layer containing a finite number of neurons can approximate any continuous function, given sufficient neurons and proper activation functions

- MLPs utilize activations functions such the tanh, sigmoid, ReLU, etc. to introduce non-linearity in the model. KANs, on the other hand, learn the required activations during the training process

- The nodes in MLPs are the points where the weighted sum is added to a bias, followed by application of activation function. On the other hand, nodes in KANs are the sites for simple addition

- The weights in KANs can be thought of as learnable activation functions, whereas they can be perceived as learnable parameters in the case of MLPs. This significant difference is the reason why nodes in KANs are only needed to sum up activation values along incoming edges

- Both KANs and MLPs use backpropagation for updating the parameters

Literature Review

Limitations of MRI Analysis Alone:

- MRI interpretation is time-intensive and relies heavily on radiologist expertise
- Subjectivity and variability in diagnosis due to differing levels of expertise

Challenges with Standard Machine Learning Models:

- Conventional machine learning models struggle with the high-dimensional nature of MRI data
- Limited interpretability and high computational demands

Drawbacks of Traditional CNNs:

- Difficulty capturing complex, non-linear relationships within MRI images
- Susceptible to overfitting without large datasets and intensive computational resources

Research Gap

- Limited Integration of Advanced Architectures: Most studies use traditional CNNs for medical imaging, with minimal exploration of hybrid models like CNN-KAN, which could improve interpretability and performance in complex data like MRI.

- Underutilization of Multivariate Decomposition: Limited research has explored the use of Kolmogorov-Arnold Networks (KAN) in medical imaging, despite their potential to model complex spatial relationships within MRI data efficiently.

- Lack of Focus on MRI Data Variability: Existing studies often overlook variability in MRI scans caused by factors like scanner type, image resolution, and patient-specific anatomical differences, which may affect model accuracy.

- Insufficient Analysis of Non-Linear Relationships: Few studies address the intricate, non-linear relationships in spine degeneration features, such as disc degeneration, vertebral misalignment, and nerve compression, which are crucial for accurate diagnosis.

Objective

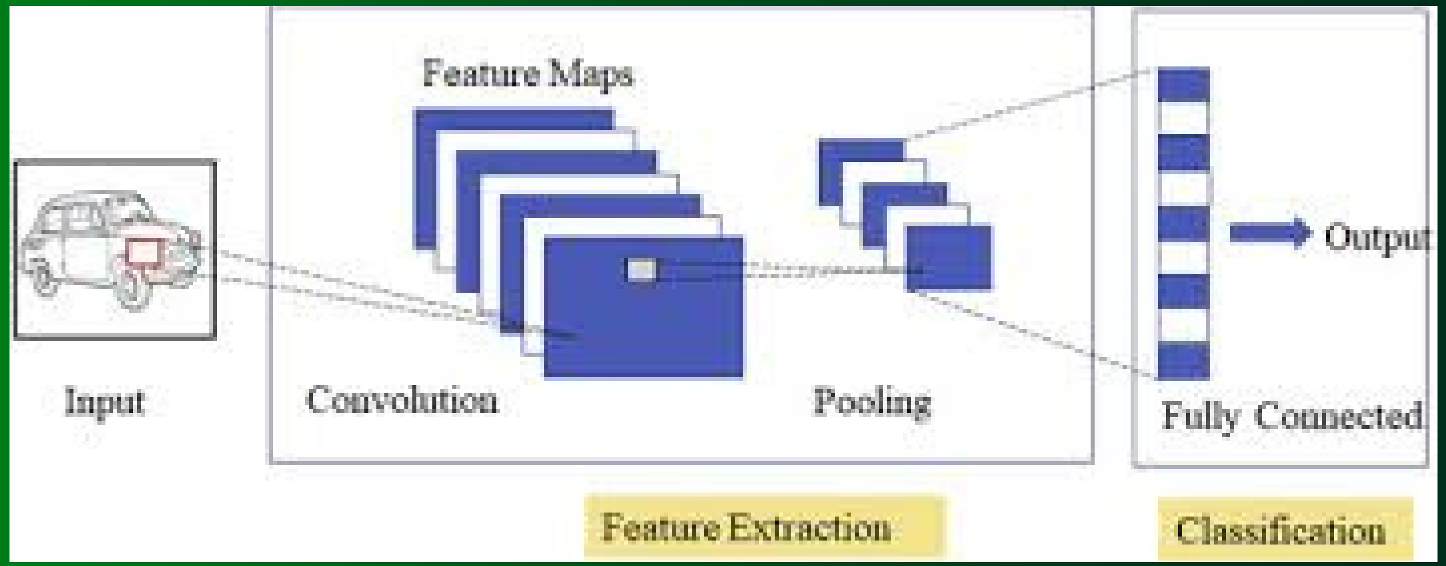
- Develop a Feature Extraction Method for Spine Degeneration Detection: Design a method for extracting relevant features from lumbar spine MRI data to accurately detect and classify degenerative conditions.

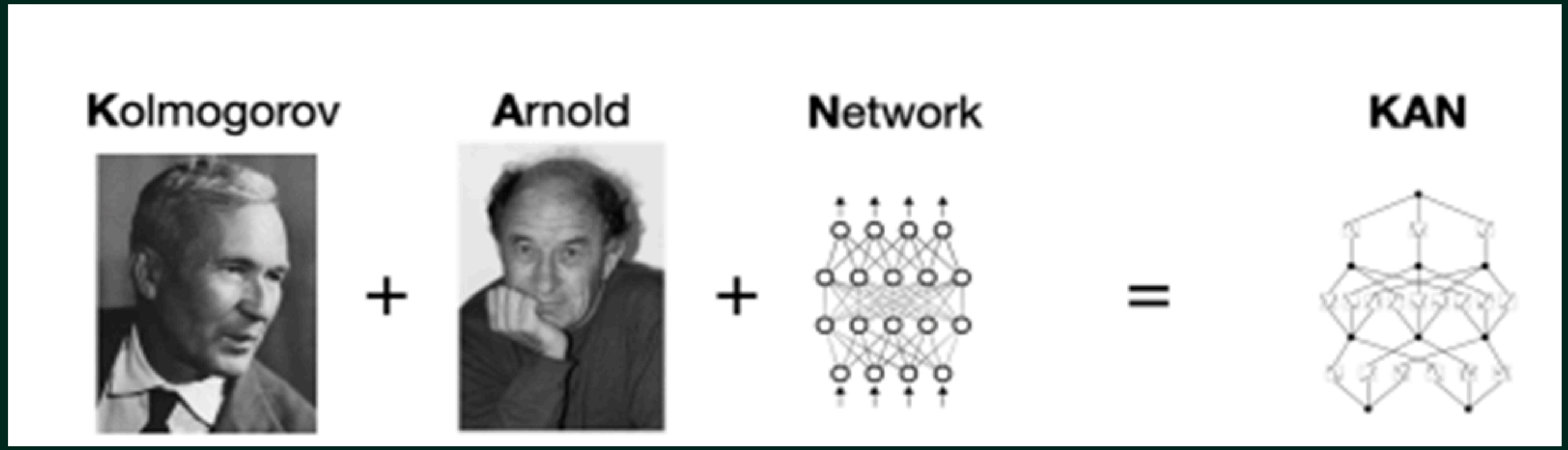
- Evaluate Machine Learning Algorithms for Spine Condition Classification: Assess the performance of various machine learning models, particularly CNN and Kolmogorov-Arnold Networks, in classifying degenerative spine conditions based on MRI data.

- Investigate the Influence of Clinical Factors on Degenerative Condition Detection: Explore the impact of factors such as disc degeneration, vertebral alignment, and spinal canal narrowing on the accuracy of degenerative spine condition classification.

Proposed Model

- Our experiment was aimed at training a convolutional neural network based on KAN instead of MLP
- The data augmentations we used are Random Brightness Contrast, Motion Blur, Median Blur, Gaussian Blur, Gaussian Noise, Optical Distortion, Grid Distortion, Elastic Transform, Shift Scale Rotate, Resize, Coarse Dropout, and Normalization





In our model, CNN layers capture and process detailed anatomical features from MRI images, such as vertebral alignment and disc shape, which are essential for identifying degenerative conditions

By integrating KAN with CNN, the model can interpret complex interactions between MRI features, like the relationship between disc degeneration and nerve compression, enhancing classification accuracy and interpretability.

Methodology

Dataset and Preprocessing Pipeline

Image Resizing → **Channel stacking** → **Data Augmentation** → **Normalization**

Image Resizing

All MRI scans were resized to a consistent dimension of 512x512 pixels to maintain uniformity across the dataset

Channel stacking

MRI scans from different sequences (e.g., T1-weighted, T2- weighted, and STIR) were stacked along the channel dimension, resulting in a 30- channel input for the model

Data Augmentation

To prevent overfitting and improve generalization, data augmentation techniques were applied during training. This included random brightness and contrast adjustments, motion blur, elastic transformations, and coarse dropout

Normalization

Each image was normalized to have a mean of 0.5 and a standard deviation of 0.5. This step ensures that the input data is on a similar scale, facilitating faster convergence during training

Model Architecture

KANConv2DLayer → **DenseNet Backbone** → **Global Pooling and Classification**

KANConv2DLayer

The model begins with a KANConv2DLayer, which applies spline based convolutions to extract fine-grained features from the

DenseNet Backbone

After the KAN layer, the model incorporates multiple layers from the DenseNet201 architecture, known for its dense connectivity pattern and efficient feature extraction

Global Pooling and Classification

The final layers of the model include a global average pooling layer, which reduces the feature maps to a single vector, followed by a fully connected layer that outputs the classification scores for each of the 75 labels corresponding to different spinal conditions

Training Process

Loss Function → **Optimizer** → **Learning Rate Scheduler** → **Mixed Precision Training**

Loss Function

We utilized the Cross-Entropy Loss function, which is commonly used for multi-class classification tasks

Optimizer

The AdamW optimizer was chosen for its ability to handle large-scale datasets and high-dimensional parameter spaces efficiently. AdamW combines the benefits of Adam (adaptive learning rates) and L2 regularization (weight decay)

Learning Rate Scheduler

A cosine learning rate scheduler with warmup was employed to adjust the learning rate dynamically throughout the training process

Mixed Precision Training

To improve training efficiency, we utilized Automatic Mixed Precision (AMP), which allows for training with both 16-bit and 32-bit floating-point numbers, reducing memory consumption and speeding up training on modern GPUs

Findings

- Training on KANs was slower than that of MLPs
- The overhead on GPU was significantly higher in case of KAN, leading to slower training and frequent exhaustion of allotted memory on cloud notebooks
- Over training periods of 10 epochs and 20 epochs on a sample of the data, MLPs performed better than KANs, though not by a large margin, but with lesser training time
- The batch size for the training data loader was reduced to a fixed value of four for KANs, as any larger batch was leading to failure of training due to exhaustion of cloud memory on Kaggle Jupyter environment

Result & Discussion

Training – VANILLA CNN (CNN + MLP)

Training – (KAN+CNN)

Visualizing Training and Validation Loss

- The visualization confirms the efficacy of the training procedure, demonstrating that the KAN effectively reduced loss over time, a sign of learning and adaptation

Epoch	Training Loss	Validation Loss	Best Weighted Log Loss Update
1	0.863042	0.749139	Updated from 1.200000 to 0.749139
2	0.735288	0.745552	Updated from 0.749139 to 0.745552
3	0.740586	0.753168	-
4	0.739538	0.746479	-
5	0.739034	0.737936	Updated from 0.745552 to 0.737936
6	0.734762	0.746083	-
7	0.737025	0.739064	-
8	0.728840	0.737367	Updated from 0.737936 to 0.737367
9	0.727065	0.732812	Updated from 0.737367 to 0.732812
10	0.728662	0.725210	Updated from 0.732812 to 0.725210
11	0.724374	0.728540	-
12	0.722049	0.725220	-
13	0.720861	0.719615	Updated from 0.725210 to 0.719615
14	0.720920	0.718220	Updated from 0.719615 to 0.718220
15	0.716651	0.712672	Updated from 0.718220 to 0.712672
16	0.712382	0.717087	-
17	0.713007	0.711341	Updated from 0.712672 to 0.711341
18	0.708734	0.708354	Updated from 0.711341 to 0.708354
19	0.704407	0.706738	Updated from 0.708354 to 0.706738
20	0.703651	0.699605	Updated from 0.706738 to 0.699605
21	0.700520	0.697399	Updated from 0.699605 to 0.697399

<u>ROC Curve</u> :

- The image displays ROC curves for classes 0, 1, and 2, along with a micro-average ROC curve. Class 0 has the highest performance (AUC=0.95), while classes 1 and 2 have lower AUC values (0.78 and 0.28, respectively).

- The micro-average ROC curve has an AUC of 0.61, indicating some interference between classes.

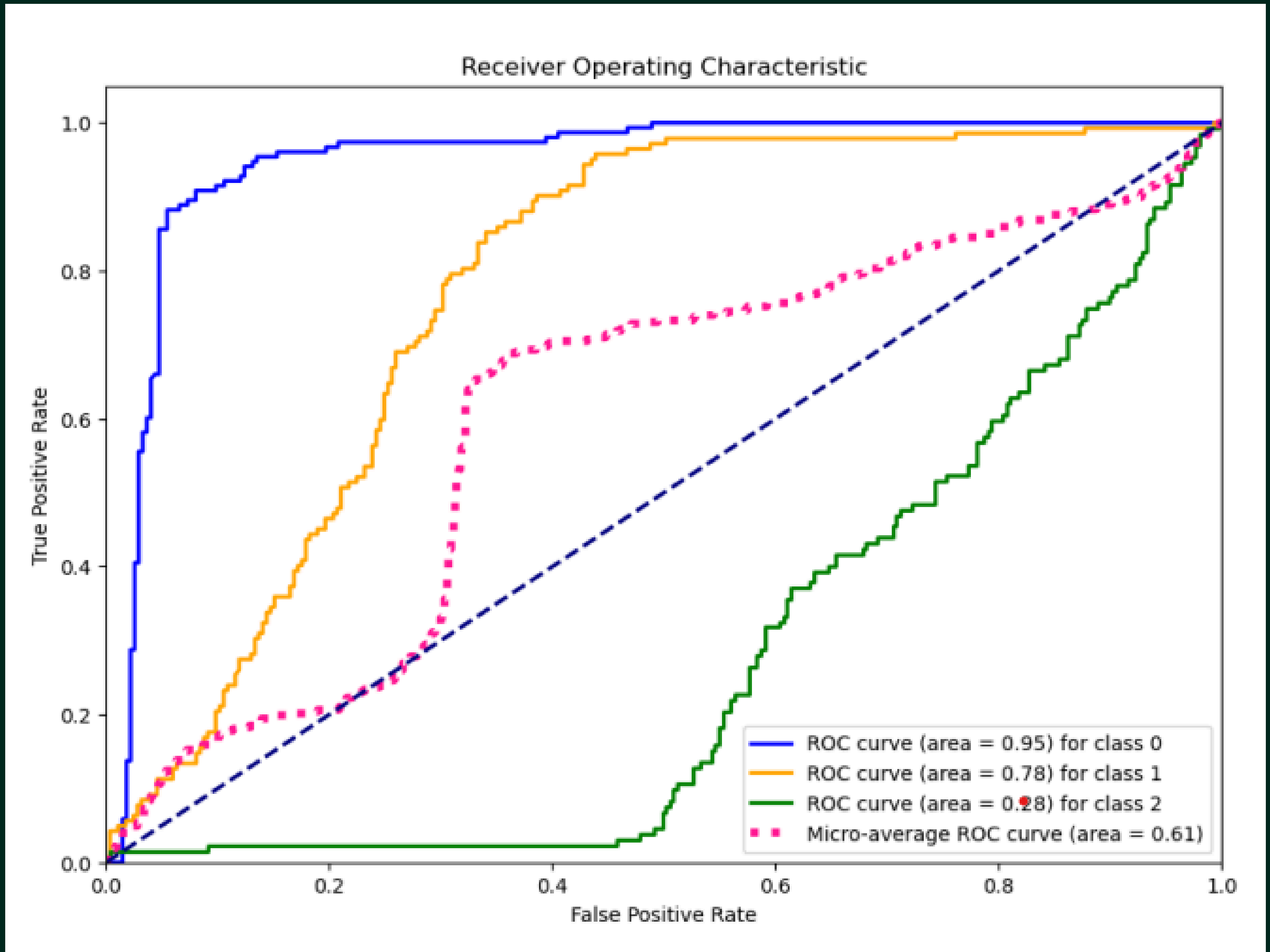


Fig :ROC curve and AUC values for different classes

Model Evaluation training curve :

- The Proposed Model (KAN With CNN) has the least No of fluctuations in Fig The Improvements in the proposed model are mainly due to the addition of the to the ReduceLROnplateau technique, making the model optimised with less no of batch size.

Fig Model losses and Model accuracy

Summary

- Potential of Kolmogorov-Arnold Network (KAN): Demonstrates promise in modeling complex relationships in medical imaging, particularly for classifying degenerative spine conditions

- Challenges of KAN: Faces limitations, including complex architecture, extended training times, interpretability issues, generalization challenges, and higher computational demands.

- Future Directions: Essential to address these challenges while leveraging KAN's strengths to develop more effective, reliable models for medical diagnostics

Thank You