Deep Learning for Lumbar Spine Diagnosis

Contributor
Ayush Tripathi

Contributor
Arjun Ashok

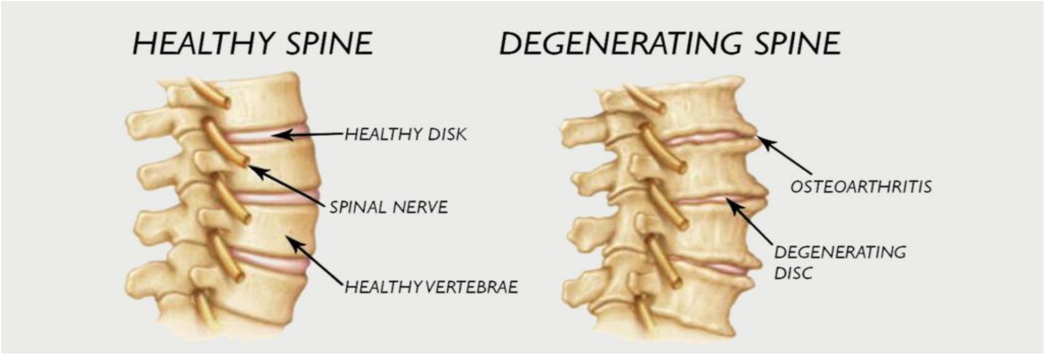
Contributor
Zhian Li

Outline

Background

Problem:

- Degenerative spine conditions adversely affect people's quality of life.

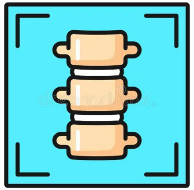


- Detecting these conditions is crucial for determining therapeutic plans for patients.

Background

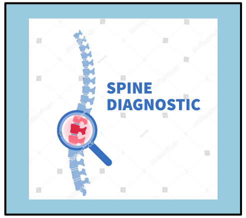
Motivation:

- Modern computer vision (CV) models demonstrate high accuracy in image classification
- These models have the potential to assist in repetitive diagnostic tasks, such as assessing spinal conditions, providing a supporting opinion for diagnosticians



Input: Medical images (spinal CT/MRI)

Processing: CV model analyzing the image

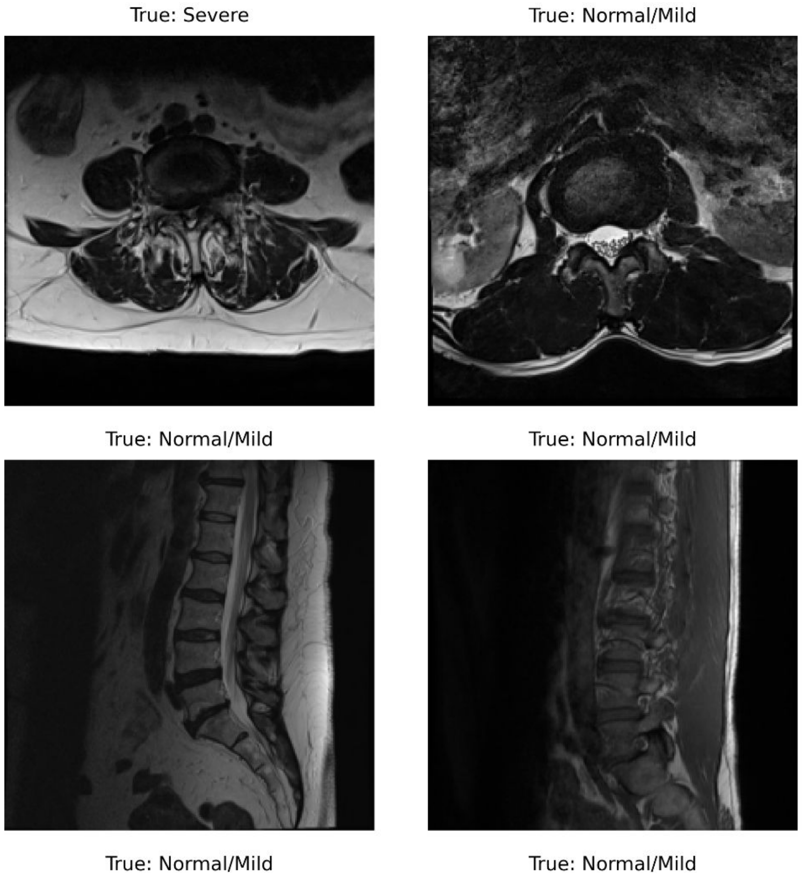


Output: severity assessment

Background

Prior Work:

- Mostly stems from RSNA Kaggle Competition
- Previous models achieved limited success
  - Weighted Log-Loss: 0.36
- Related studies focus on core design of individual model architectures



Unprocessed Samples of the Dataset

Dataset

- Source
  - Kaggle Competition
  - Anatomy & Image Visualization Overview–RSNA RAIDS

- Transforms Used
  - Resize to 224 x 224
  - Normalize Pixel brightness mean to 0.5, standard deviation to 0.5

Model Architectures

(1) Convolutional Neural Network (CNN)

- 3 blocks (Conv, ReLU, MaxPool)
- 1 classification head, 3 linear layers (ReLU, SoftMax)

(2) Modified CNN (MCNN)

- 3 blocks (Conv, GeLU, LayerNorm MaxPool)
- 1 classification head, 3 linear layers (ReLU, SoftMax)

(3) Residual Neural Network (ResNet)

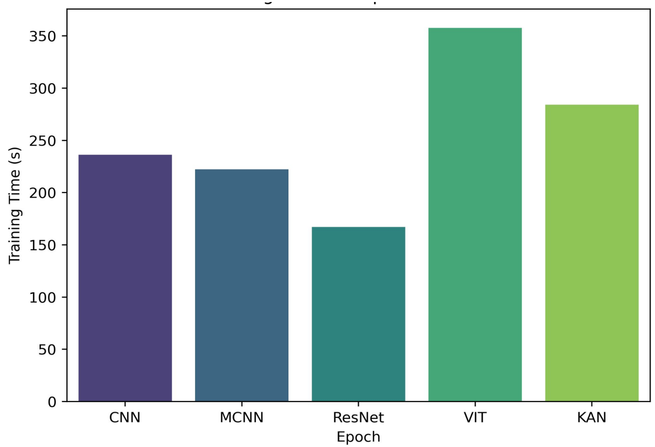
- ResNet-18 (18 layers deep)
- Not pre-trained for fair comparison

(4) Vision Transformer

- Patch Embedding (grid + projection via convolution)
- Positional Encoding (learn spatial relationships)
- Encoder Layers (multi-headed attention, feed-forward)
- Classifier Head (CLS token)

(5) Kolmogorov-Arnold Network

- 3 blocks (Conv, GeLU, LayerNorm, MaxPool)
- 1 linear layer to project from convolutional layer
- 2 Spline-linear layers (KAN) for classification head



Training Time per Epoch

TABLE I
PERFORMANCE ACROSS MODEL ARCHITECTURES

Architecture	Accuracy	Weighted Log Loss	Inference Time (ms)
CNN	**90.2%**	0.348	5.692
MCNN	89.5%	0.359	5.105
ResNet	89.5%	0.442	**4.179**
CKAN	**90.2%**	**0.342**	5.034
VIT	88.4%	0.395	8.596

Cumulative Performance Table

Results

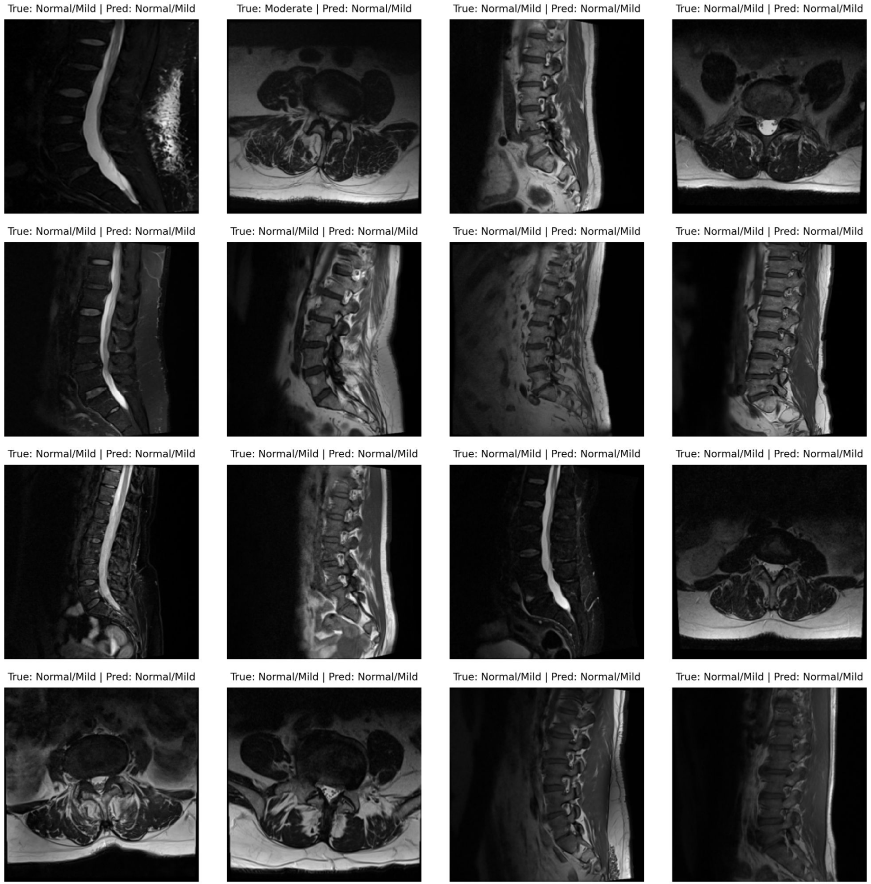
Computational Efficiency

- Training times similar across models, except Vision Transformer (~50% longer/epoch).
- Higher computational cost due to larger architecture.

Performance

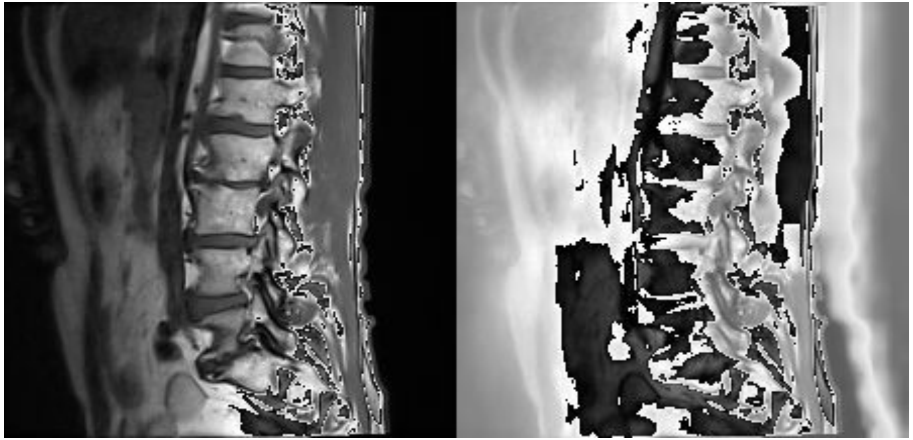
- All models achieved similar accuracy
- CKAN and MCNN excelled; VIT underperformed compared to ResNet despite lower WLL.
- ResNet was confidently wrong, while VIT showed uncertainty in incorrect predictions.
  - Their larger architectures struggled with the smaller dataset; pre-training may improve results.
- All models meet real-time deployment criteria, exceeding MRI framerate reqs.(≤40ms/frame).

Results

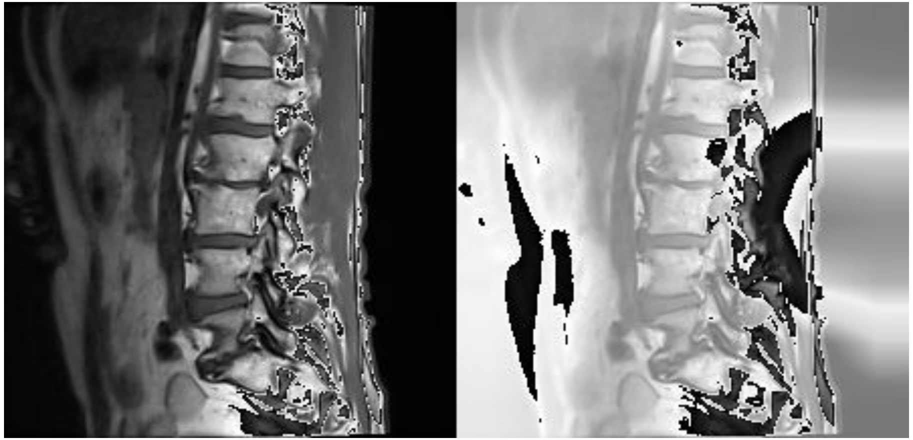


Processed Samples – Predicted With CKAN

CNN Interpretation



ResNet Interpretation

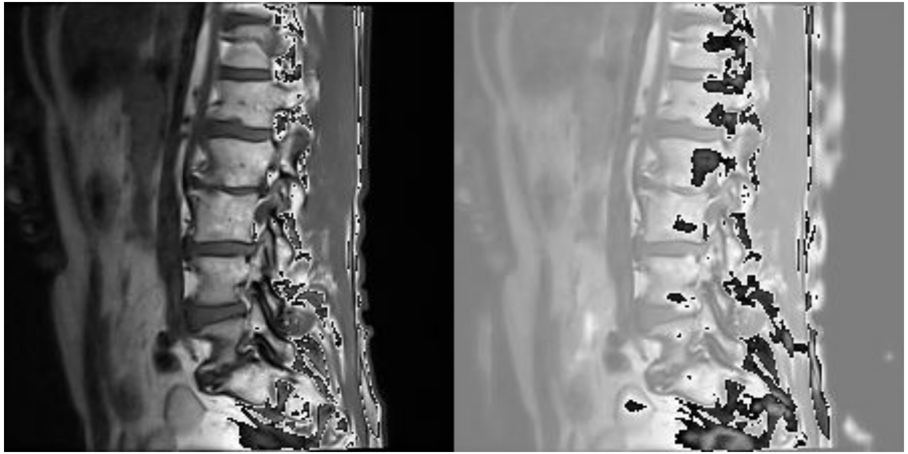


Results

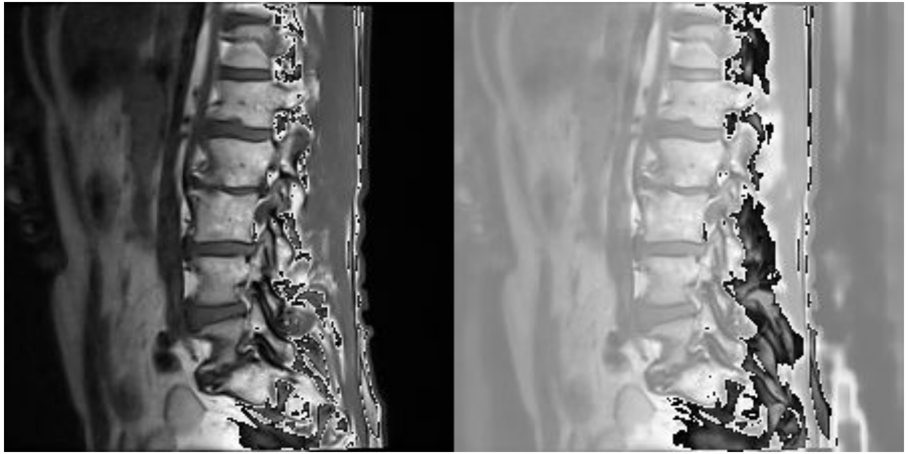
Model Interpretation

- Used Grad-Cam to visualize high-activation areas as heatmaps.
- CNN showed broader activations, suggesting less specific learning.
- ResNet misfocused on irrelevant regions, needing more training or data.
- CKAN focused on spinal fluid build-up, indicating strong localization.
- M-CNN seems similar to this, however CKAN has broader activations.

CKAN Interpretation



M-CNN Interpretation



Results

Model Interpretation

- Used Grad-Cam to visualize high-activation areas as heatmaps.
- CNN showed broader activations, suggesting less specific learning.
- ResNet misfocused on irrelevant regions, needing more training or data.
- CKAN focused on spinal fluid build-up, indicating strong localization.
- M-CNN seems similar to this, however CKAN does better at finding areas of interest.

Conclusions

- Lessons
  - KAN demonstrates real–world efficacy for replacing MLPs
  - Larger models (ResNet, VIT) require more data, favoring pre–training or smaller architectures for limited datasets
  - Smaller CNN–based models offer better interpretability and faster inference, crucial for medical applications
- Future work should explore pre–training on related/unrelated datasets and pseudo–tasks for performance boosts
- Threats to validity primarily from the imbalance dataset, though we weight our loss to compensate