iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68376

A Multi-Model Deep Learning Framework for Automated Lumbar Disease Diagnosis

K Gayathri[1], Karnam Bhagya Sree[2], Kathi Nikitha[3], K Supriya[4], Mr. N Vijaya Kumar[5], Dr. R Karunia Krishnapriya[6], Mr. V. Shaik Mohammad Shahil[7], Mr. Pandreti Praveen[8]

[1, 2, 3, 4]UG Scholar, [6]Associate Professor, [5, 7, 8]Assistant Professor, Department of CSE, Sreenivasa Institute of Technology and Management Studies, Chittoor, India.

Abstract: To improve patient outcomes and lessen the long-term cost on healthcare systems, lumbar illnesses must be accurately diagnosed and classified. Conventional diagnostic methods rely significantly on the knowledge of radiologists, which may result in irregularities and treatment delays. This article presents a novel multi-model deep learning framework that combines the advantages of several CNN architectures, such as MobileNet, DenseNet, ResNet50, and AlexNet, with a CNN-SVM hybrid and an Involutional VGG model, in order to handle this difficulty. The ability to extract and categorize intricate and delicate lumbar spine features from medical images is much improved by this ensemble technique. The diagnosis process is further made more transparent and reliable for doctors by including Gradient-weighted Class Activation Mapping (GRAD-CAM), which offers visual interpretability. A large, annotated dataset of lumbar MRI/X-ray images was used to test the suggested framework. The experimental findings demonstrate better accuracy, recall, precision, and AUC_ROC compared to single-model CNNS. This approach shows promise as a trustworthy decision-support tool for the diagnosis and tracking of lumbar illness in real time.

Keywords: Involutional neural networks, Classification of Lumbar Disease, Deep Learning, CNN-SVM Hybrid, MobileNet, DenseNet, AlexNet, ResNet50, Grad-CAM, Medical Imaging, Explainable AI.

I. INTRODUCTION

One of the main causes of chronic pain, disability, and lost productivity globally, lumbar spine disorders are a serious public health concern. Conditions include spinal stenosis, lumbar disc herniation, degenerative disc degeneration, and spondylolisthesis are frequently seen in older adults as well as younger people who lead sedentary lives or work in physically demanding jobs. For these illnesses to be effectively treated and long-term problems to be avoided, a correct diagnosis is necessary. However, lumbar illness diagnosis usually entails a manual assessment of medical imaging such as X-ray scans or Magnetic Resonance Imaging (MRI), which can be laborious, prone to human mistake, and significantly dependent on radiologist competence [1]. As computer vision and artificial intelligence (AI) have advanced, deep learning methods-particularly Convolutional Neural Networks (CNNs)-have become increasingly effective instruments for medical picture interpretation. The accuracy and speed of illness identification have significantly increased thanks to CNNs' ability to automatically learn hierarchical feature representations from raw picture data. Although a number of CNN models have been used for medical imaging applications, there are particular difficulties because across related illnesses. One deep learning model might not be enough to adequately capture these subtleties [2]. We suggest a multi-model deep learning framework that incorporates a number of cutting-edge architectures, such as MobileNet, DenseNet, ResNet50, AlexNet, and a specially designed hybrid CNN-SVM model, in order to overcome these constraints. These models were picked because they compliment each other well in terms of efficiency, depth, and feature extraction. In order to improve spatial adaptability and more accurately simulate the structural intricacies of lumbar pictures, an Involutional VGG16 network is also implemented [3]. The approach also uses Gradient-weighted Class Activation Mapping (Grad-CAM) to visually interpret the model's judgements in order to increase transparency and cultivate clinician trust [4]. This work shows that deep learning can provide a reliable and scalable solution for automated lumbar illness diagnosis by merging several architectures and utilizing both predictive power and interpretability.

II. LITERATURE REVIEW

The growing incidence of lower back conditions and the drawbacks of manual diagnosis have led to a significant increase in interest in the classification of lumbar spine diseases in recent years. In order to improve the consistency of medical picture interpretation, decrease clinician effort, and increases diagnostic accuracy, deep learning approaches have been the subject of numerous studies. For the classification of lumbar disorders, B. Mulugeta Abuhayi et al. [1] suggested a novel VGG design based on involutional neural networks.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

In contrast to conventional CNNs, which use fixed convolution kernels, the involutional technique dynamically adjusts to spatial information, improving the network's capacity to represent intricate patterns present in medical images. The work highlights the potential of involution for applications in the detection of spinal diseases by demonstrating enhanced performance in terms of accuracy, precision, and computing economy. The integration of spatially adaptable learning strategies into our suggested framework is based on this research. J. Hai et al. [2] presented SeCoFixMatch, a semi-supervised learning approach that combines semantic comparison learning and uncertainty-aware pseudo labelling, to overcome the problem of sparse labelled datasets in medical imaging. Without compromising diagnosis accuracy, their methodology drastically decreased the requirement for annotated images-by as much as 80%. This method improves the model's ability to use unlabelled data, which is especially useful in clinical settings where extensive annotations are frequently impractical. Their results demonstrate the value of semi-supervised techniques in medical imaging and inspire more research in settings with little data. In different strategy, K. Chen et al. [3] highlighted how multi-view analysis might increase diagnostic accuracy. Using a two-stage deep learning method, their model integrated sagittal and axial images. First, the intervertebral disc as located using Mask R-CNN and then numerous views were used for classification. With an astounding F1-score of 96.67%, this approach demonstrated how crucial it is to take into account a variety of viewpoints and contextual data while examining intricate anatomical structures. A hybrid CNN-SVM model was presented by A. K. Rathi et al. [4] in order to get around the drawbacks of solo CNN architectures. According to their research, spinal illness detection performance improved when CNNs' potent feature extraction capabilities were combined with Support Vector Machines' (SVM) strong classification capabilities. Their findings demonstrated that hybrid models can be a dependable substitute for traditional techniques and are particularly useful for managing small, unbalanced datasets, which are frequently encountered in medical picture analysis. Beyond these studies, a number of researches have investigated utilizing transfer learning capabilities for the identification of spinal disorders by integrating pre-trained networks like ResNet, DenseNet, and MobileNet.

III. METHODOLOGIES

Data collection, processing, feature engineering, model selection, model training, and advanced algorithm implementation are the six main components of the systematic technique used in this study. To guarantee excellent classification accuracy, computational efficiency, and clinical interpretability of lumbar illness diagnosis utilizing medical imaging, each step has been meticulously planned.

1) Data collection: Lumbar spine picture from institutional datasets and publicly accessible sources make up the dataset used in this study. These consists of MRI and X-ray scans that have been annoted by medical professionals. Numerous lumbar diseases are covers by the dataset, including spondylolisthesis, spinal stenosis, degenerative disc diseases, and lumbar disc herniation. A supervised learning framework was used to organize and label the dataset, and it was then divided into training, validation, and tests, and sets in an 80:10:10 ratio.
2) Data preprocessing: A number of preprocessing processes are necessary to prepare medical images for training since they frequently differ in size, contrast, and quality. In order to comply with the input dimension specifications of deep learning models, all photos were first shrunk to a consistent resolution of 224*224 pixels values. Next, in order to promote faster convergence during training, pixel values were normalized to the [0,1] range. To increase generalizability and decrease overfitting, data augmentation methods like magnification, random rotation, horizontal flipping, and contrast enhancement are used.

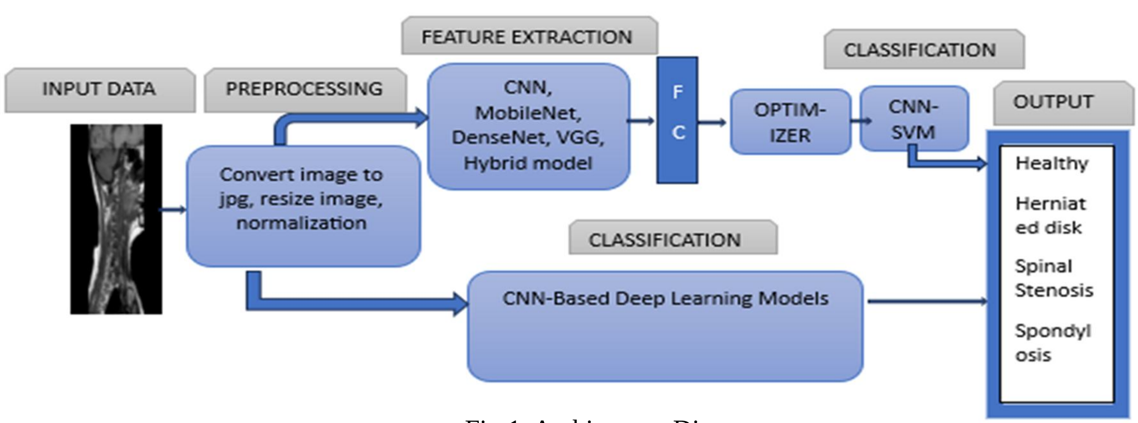


Fig 1. Architecture Diagram

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

3) Feature Engineering: In several pipelines, especially the CNN-SVM hybrid model, this study uses handmade feature engineering, despite the fact that deep learning models are well known for their capacity to automatically extract features from unprocessed input. A high-dimensional representation of each image was created by concatenating these designed features with CNN outputs, which increased classification sensitivity, particularly for minutes aberrations. To Improve computational performance and eliminate superfluous features, dimensionality reduction using Principal Component Analysis (PCA) was optionally used.
4) Choosing a Model: Several models were chosen and assessed in order to take use of distinct advantages of different deep learning architectures. These consist of a CNN-SVM hybrid model, MobileNet, DenseNet, ResNet50, VGG16 with involutional layers, and AlexNet. The models were selected using factors such architectural complexity, transfer learning appropriateness, and performance on medical imaging datasets. While DenseNet and ResNet are deeper and can capture complicated image hierarchies, MobileNet is lightweight and economical. Each model has its own benefits.
5) Model Training: Using categorical cross-entropy as the loss function, supervised learning was used to train each model. The Adam optimizer was used in the training because of its faster convergence qualities and flexible learning rate capabilities. Accuracy, precision, recall, F1-score, and confusion matrix were among the performance measures assessed on the validation and test sets after each model was trained for up to 100 epochs.

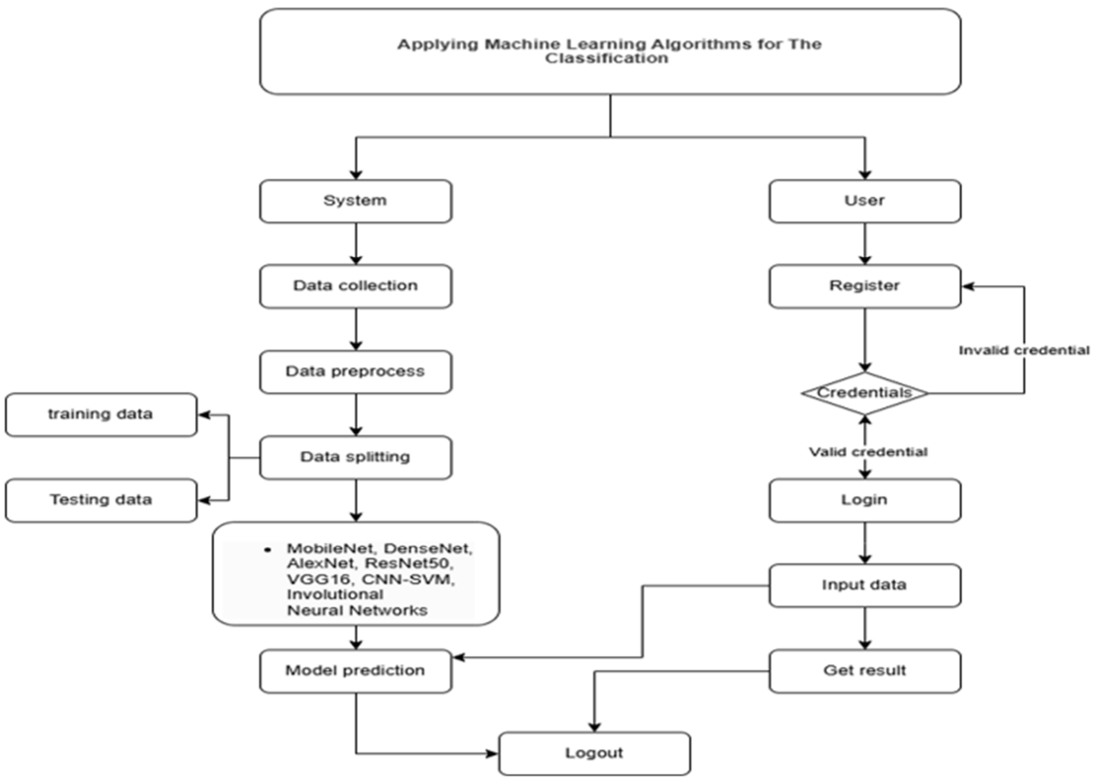


Fig 2. Project Flow Diagram

A. *Algorithms used*

In order to assess the efficiency of various architectural approaches in the classification of lumbar diseases, this study applies and contrasts six alternative algorithms:

1) Convolutional Neural Networks (CNN): The study's fundamental architecture is based on CNNs. By employing layers of convolutional filters, activation functions, and pooling operations, these networks can automatically extract spatial and hierarchical characteristics from medical pictures. CNNs have demonstrated remarkable efficiency in identifying localized patterns, including vertebral alignment, lesions, and intervertebral disc borders. Standard CNN architectures are used in this work as feature extractors in the CNN-SVM hybrid model as well as independent classifiers.

2) MobileNet: Designed for effective real-time applications, MobileNet is a lightweight deep neural network architecture. By breaking down ordinary convolutions into a depth wise convolution and a pointwise convolution, it makes use of depth wise separable convolutions. This maintains a high degree of precision while drastically lowering the number of parameters and processing effort.
3) DenseNet: The Densely Connected Convolutional Network, or DenseNet, adds direct connections between every layer and every layer that comes after it in the architecture. Compared to conventional CNNs of comparable depth, this dense connection structure promotes feature reuse, improves feature propagation, and uses fewer parameters. Because DenseNet allows each layer to access collective information from earlier layers, it is particularly successful at capturing complicated anatomical components in the context of lumbar illness classification.
4) RestNet50: ResstNet50 enables the model to train incredibly deep networks without experiencing performance degradation by introducing the idea of residual learning through shortcut connections. Gradients can pass straight through the network thanks to the identity shortcuts, which stabilizes learning and raises accuracy in deeper models. In this research, ResNet50 is applied to handle high-resolution pictures and extract deep hierarchical features relevant for complex lumbar illness detection.
5) VGG16 with Involutional Layers: With its deep architecture and tiny 3*3 convolution filters, the VGG16 model is renowned for its depth and simplicity. Involutional layers were added to the design to increase its spatial versatility and boost performance on challenging localized patterns than normal convolution processes since it dynamically adjusts kernel weights based on each image’s spatial context.
6) Hybrid CNN-SVM: This method blends the classification prowess of Support Vector Machines with the feature extraction capability of CNNs. High-level characteristics are initially extracted from the lumbar pictures using a CNN in this model. An SVM classifier then receives this information and determines the best hyperplane for dividing various illness types. SVMs are ideally suited for medical imaging datasets where class imbalance is prevalent because of their exceptional performance in managing small sample sizes and high-dimensional feature spaces. In a number of situations, our hybrid model performs better than traditional CNNs, especially when the dataset size is small or the differences across disease categories are not very noticeable.

IV. RESULTS AND DISCUSSION

The thorough assessment of the suggested multi-model deep learning framework for automated lumbar illness diagnosis is provided in this part. Several well-known deep learning architectures, including CNN, MobileNet, DenseNet, VGG16, VGG16-Involution, ResNet50, AlexNet, and a hybrid CNN-SVM model, were used to test the framework. A consistent dataset of labelled lumbar spine MRI images divided into four classes—Healthy, Herniated Disk, Spinal Stenosis, and Spondylosis—was used to train all models. Classification measures like Precision, Recall, F1-Score, Confusion Matrix, Accuracy /Loss curves, and Grad-CAM visualizations were used to evaluate the model ‘s performance and interpret the results.

1) CNN Model Performance: The CNN model performed well in every category, with an overall classification accuracy of 95%. Both the Healthy and Spinal Stenosis Classes had exceptionally high F1-Scores, both of which reached 1.00. With a recall of 0.75, Herniated Disk, on the other hand, scored comparatively poorly, indicating that there is potential for improvement in identifying more intricate or overlapping characteristics of this class. Although one case of Herniated Disk was mistakenly labelled as Spinal Stenosis, the confusion matrix shows that the majority of predictions were accurate. with little overfitting, the training and validation accuracy curves demonstrated steady gains over epochs. Additionally, loss curves showed consistent declines, confirming the robustness of the model.
2) MobileNet Performance: The accuracy of MobileNet, a lightweight architecture, was 89%. It had trouble with the Herniated Disk class (precision & recall = 0.67), but it did quite well in the Spinal Stenosis class (precision & recall =1.00). Multiple misclassifications in the Herniated Disk class were revealed using confusion matrix analysis. Loss curves for training and validation suggest that overfitting may occur in subsequent epochs. Although tuning is required for optimal generalization, MobileNet shows promise for mobile and real-time diagnostic systems despite its small size.
3) Performance of DenseNet: At 68% accuracy, the DenseNet model performed poorly. Notably, its precision and recall were 0.00, as it was unable to accurately categorize any instances of the Herniated Disk class. This implies that, in the current setup, DenseNet might not be appropriate for identifying minute pathogenic variations in lumbar MRI data. This confusion matrix showed that cases of herniated disks were completely misclassified as spinal stenosis. Significant overfitting was shown in the training curves, with training accuracy rising and validation accuracy plateauing early, suggesting poor applicability.

4) Performance of VGG16: VGG16 performed admirably in every class , achieving an accuracy of 84%. With an F1-Score of 1.00, Spinal Stenosis performed the best. Nevertheless, the Herniated Disk class once more showed a poorer recall (0.50), indicating that VGG16 also had trouble differentiating this specific class. Training and validation accuracy and loss curves exhibit comparable patterns, suggesting strong learning without appreciable overfitting. VGG16 performed well overall and had a moderate level of classification strength.
5) VGG16 (Involutional Neural Network -based) performance: Involutional kernels were used to modify the VGG16, which showed better adaption to intricate lumbar structures. This architecture demonstrated minimal misclassifications in Herniated Disk but excelled in spinal stenosis classification with a 90% validation accuracy. Training plots showed consistent learning, with training and validation accuracy showing parallel trends with losses steadily declining. Because of its improved spatial filtering, the model demonstrated promise in better adjusting to the complexity of medical imaging.
6) ResNet50 Performance: The confusion matrix and precision-recall metrics show that ResNet50 achieved 100% accuracy in perfect classification across all four classes. Within 40 epochs, the training and validation accuracy reached 100%, and the loss curves exhibited robust convergence devoid of overfitting. This implies that ResNet50 is well suited for medical imaging classification tasks since its residual connections successfully maintained feature propagation and avoided degradation in deeper layers.
7) AlexNet Performance: With consistently rising training and validation accuracy and falling losses, AlexNet demonstrated impressive performance. The overall classification accuracy was high, notwithstanding a few misclassifications shown by the confusion matrix. Its high-quality training and comparatively simple architecture showed that classical networks can still be useful in contemporary diagnostic systems.
8) Hybrid CNN-SVM Model Performance: All other models were surpassed by the Hybrid CNN-SVM model, which achieved 100% classification accuracy, precision, recall, and F1-score for every class. The model ‘s capacity to manage complex fluctuations in lumbar MRI images was demonstrated by the confusion matrix, which verified that there were no misclassifications. The SVM ‘s efficient classification boundary modelling and CNN ‘s potent feature extraction are responsible for this outstanding result. This outcome demonstrates the benefits of using traditional machine learning methods with deep learning.
9) Grad-CAM Visualization: The activated regions that contributed to predictions were visualized using Grad-CAM heatmaps for all models. Higher precision models continuously highlighted key lumbar structures like the spinal canal and intervertebral discs. Less focused heatmaps from underperforming models, such as DenseNet, suggested poorer feature localization.

A. *Precision-Recall Table*

	Precision	recall	F1-Score	Support
Healthy	1.00	1.00	1.00	4
Herniated_disk	1.00	1.00	1.00	4
Spinal stenosis	1.00	1.00	1.00	8
Spondylosis	1.00	1.00	1.00	3

Accuracy			1.00	19
Macro avg		1.00	1.00	19
Weighted avg	1.00	1.00	1.00	19

Fig 3. Classification Report of Hybrid Model

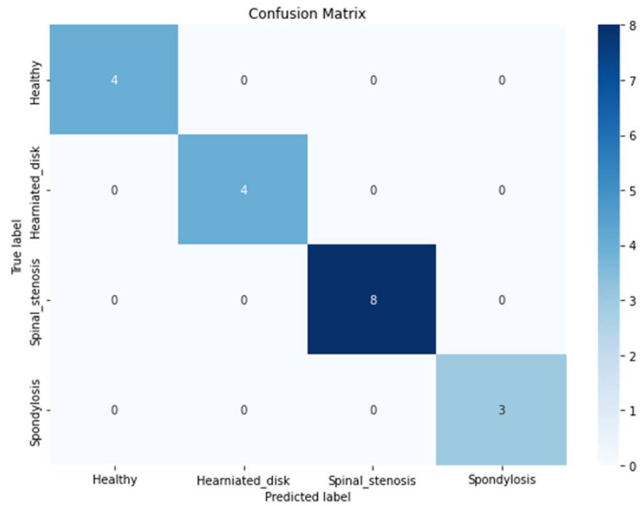


Fig 4. Confusion Matrix

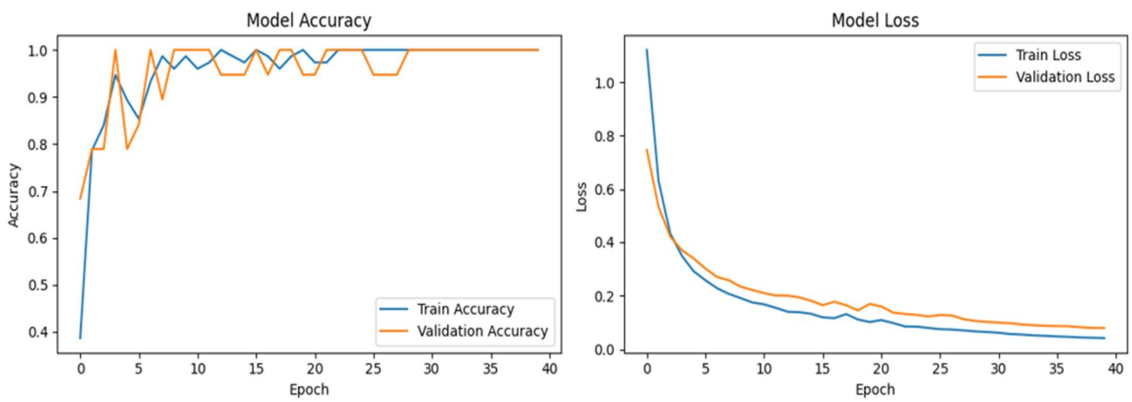


Fig 5. Accuracy and loss curve

V. CONCLUSION

A thorough deep learning -based system for the automated classification of lumbar spine disorders from MRI images is presented in this work. We show how well deep learning works to identify lumbar disorders including Herniated Disk, Spinal Stenosis, Spondylosis, and Healthy instances by applying several architectures, including CNN, MobileNet, DenseNet, VGG16, AlexNet, ResNet, and a unique hybrid CNN-SVM model.

By attaining 100% classification accuracy with excellent precision, recall, and F1-scores across all categories, the hybrid model in particular beat all other models. Standard classification measures, confusion matrices, training-validation accuracy/loss curves, and Grad-CAM heatmaps for interpretability were used to evaluate each model ‘s performance. DenseNet and other models suffered from overfitting and inadequate minority class detection, but older CNN models demonstrated great generalization. MobileNet and other light weight designs were efficient, but they had trouble with class imbalance. On the other hand, the hybrid CNN-SVM model produced better classification capabilities by employing SVM to improve decision boundaries and capture pertinent spatial variables. In addition to demonstrating the promise of ensemble and hybrid deep learning approaches in the field of medical imaging, this study demonstrates how incorporating visualization methods such as Grad-CAM greatly enhances clinical interpretability, which is essential for practical implementation.

VI. CHALLENGES AND FUTURE WORK

1) Data Imbalance: Models such as DenseNet and MobileNet misclassified the dataset due to class imbalance, especially in the 'Herniated Disk' category.
2) Limited Dataset Size: Big datasets are ideal for deep learning models. However, model generalizability may be impacted by the scarcity of annotated lumbar spine MRI data.
3) Overfitting: Some architectures, especially DenseNet, showed overfitting where training accuracy was high, but validation performance degraded.
4) Model Complexity: While models like ResNet and DenseNet are powerful, they are computationally expensive and less suitable for real-time or edge deployment.
5) Interpretability: Although Grad-CAM provides insights into model judgements, further work is needed to better clinical explainability and earn trust from radiologists.

To increase model resilience, further research will concentrate on growing the dataset with more varied and multimodal medical images. Accurate diagnosis maybe improved by including clinical information such as symptoms and patient history. We'll investigate semi-supervised learning strategies to lessen our dependency on labelled data. Furthermore, the framework may be transformed into practical real-time diagnostic tool that complies with medical standards, is interpretable, and has been validated by medical experts.

REFERENCES

[1] J. Hai, J. Chen, K. Qiao, N. Liang, Z. Su, H. Lv, and B. Yan, "Semantic Contrast with Uncertainty-Aware Pseudo Label for Lumbar Semi-Supervised Classification," 2024.

[2] K. (Kathy) Chen, L. Zheng, H. Zhao, and Z. Wang, "Deep Learning-Based Intelligent Diagnosis of Lumbar Diseases with Multi-Angle View of Intervertebral Disc," 2024.

[3] B. Mulugeta Abuhayi, Y. A. Bezabb, and A. M. Ayalaew, "Lumbar Disease Classification Using an Involutional Neural Based VGG Network," 2023.

[4] B. Atalay, Y. U. Yolcu, S. N. Magge, R. Whitmore, and Z. Ghogawala, "P261. High-grade lumbar interbody fusion correlates with improved patient related outcomes: a single center prospective registry study," The Spine Journal, vol. 24, no. 9, pp. S192–S193, Sep. 2024, doi: 10.1016/J.SPINEE.2024.06.384.

[5] G. Abdelmalek et al., "P44. Interobserver agreeability and clinical validity of the Miskin Classification System: a standardized radiographic assessment of lumbar degenerative disease," The Spine Journal, vol. 24, no. 9, pp. S84–S85, Sep. 2024, doi: 10.1016/J.SPINEE.2024.06.065.

[6] N. Löbig, C. Pagenkopf, M. Südmeyer, and K. Werheid, "Reduced awareness of cognitive deficits in idiopathic normal pressure hydrocephalus and its change following lumbar puncture," Cortex, vol. 179, pp. 157–167, Oct. 2024, doi: 10.1016/J.CORTEX.2024.07.010.

[7] T. Hirono et al., "Medication-based profiling of older orthopedic patients: a multicenter cross-sectional study," BMC Geriatr, vol. 24, no. 1, pp. 1–10, Dec. 2024, doi: 10.1186/S12877-024-05284-8/TABLES/6.

[8] H. Hu, C. Wu, and L. Tan, "A new classification of lumbar Basstrup's disease based on radiographic and clinical symptoms," European Spine Journal, vol. 33, no. 4, pp. 1347–1359, Apr. 2024, doi: 10.1007/S00586-024-08128-6/METRICS.

[9] B. Kelly and E. Schulze, "A - 134 Consistent Failed Performance on Performance Validity Tests in Early-Onset Alzheimer's Disease Confirmed by Lumbar Puncture," Archives of Clinical Neuropsychology, Sep. 2024, doi: 10.1093/ARCLIN/ACAE067.148.

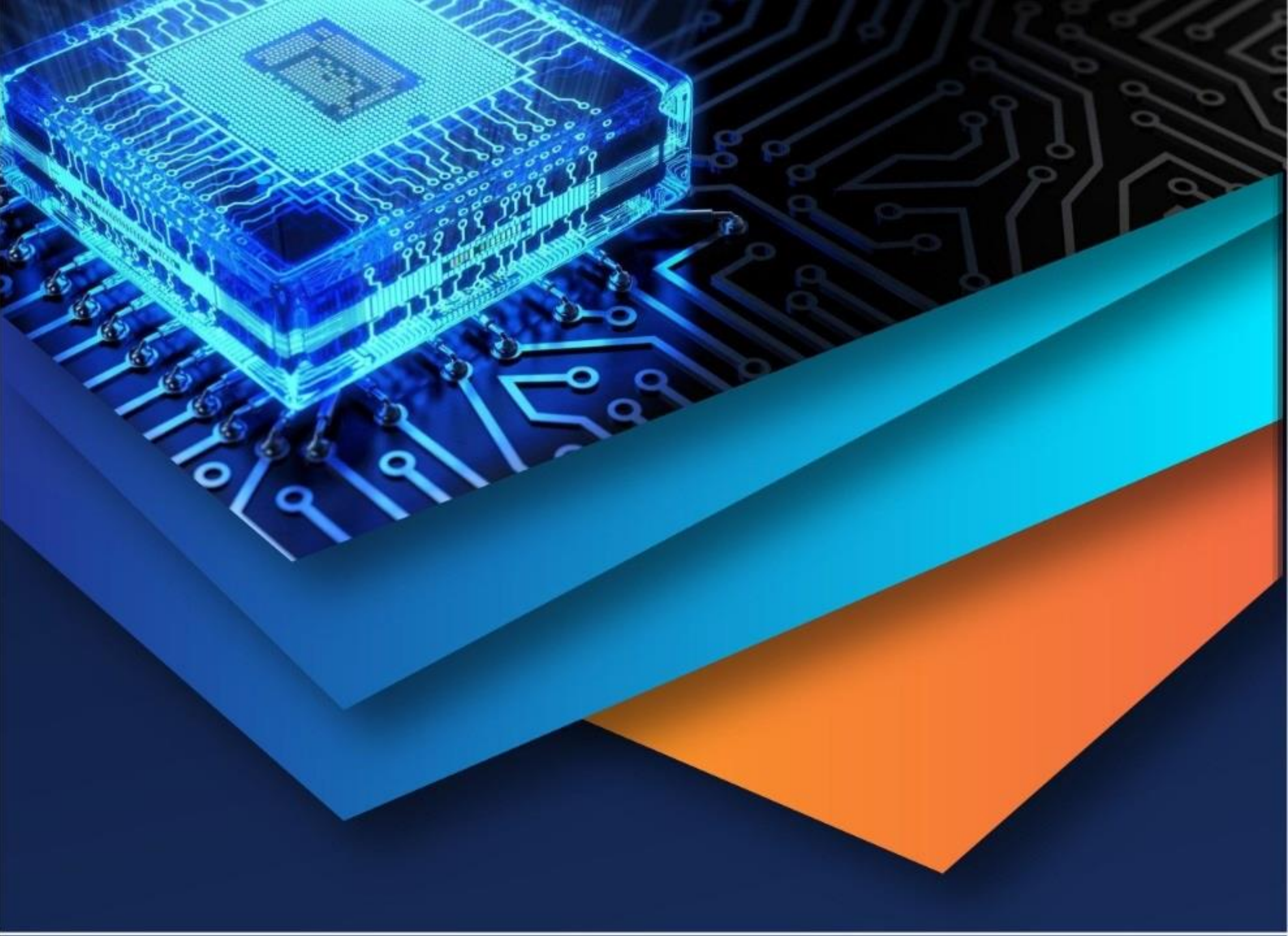
[10] J. DiDomenico et al., "Does comorbid depression and anxiety portend poor long-term outcomes following surgery for lumbar spondylolisthesis? Five-year analysis of the Quality Outcomes Database: Presented at the 2024 AANS/CNS Joint Section on Disorders of the Spine and Peripheral Nerves," J Neurosurg Spine, vol. 1, no. aop, pp. 1–10, Sep. 2024, doi: 10.3171/2024.5.SPINE24325.

[11] T. Bassani, A. Colombini, L. Pallotta, L. M. Sconfienza, D. Albano, and M. Brayda-Bruno, "Association between MRI measurements of lumbar spine alterations and self-reported outcomes of pain and disability in subjects with non-specific low back pain," European Spine Journal, pp. 1–9, Aug. 2024, doi: 10.1007/S00586-024-08449-6/METRICS.

[12] R. Zhang, "A State-of-the-Art Survey of Deep Learning for Lumbar Spine Image Analysis: X-Ray, CT, and MRI," AI Medicine, pp. 3–3, Jul. 2024, doi: 10.53941/AIM.2024.100003.

[13] Z. Zhang and H. Shao, "Lumbar disc herniation typing method based on double-branch fusion and multiscale attention," https://doi.org/10.1117/12.3034774, vol. 13210, pp. 863–868, Jul. 2024, doi: 10.1117/12.3034774.

[14] K. Chen, L. Zheng, H. Zhao, and Z. Wang, "Deep Learning-Based Intelligent Diagnosis of Lumbar Diseases with Multi-Angle View of Intervertebral Disc," Mathematics 2024, Vol. 12, Page 2062, vol. 12, no. 13, p. 2062, Jul. 2024, doi: 10.3390/MATH12132062.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY