Detection of Abnormalities in Spinal Cord Injury MRI Using Image Processing and Machine Learning

Gangadhara
Department of Electronics and Communication Engineering, NMAM Institute of Technology, Nitte (Deemed to be University), Nitte, Karnataka, India | Department of Electronics and Communication Engineering, Yenepoya Institute of Technology, Moodbidri, Karnataka, India
gangadhara444444@gmail.com

Subramanya Bhat
Department of Electronics and Communication Engineering, NMAM Institute of Technology, Nitte (Deemed to be University), Nitte, Karnataka, India
sbhat22@nitte.edu.in (corresponding author)

Received: 29 September 2025 | Revised: 4 November 2025 and 16 November 2025 | Accepted: 19 November 2025

 DOI: https://doi.org/10.48084/etasr.15222

ABSTRACT

To enhance the analysis of radiological spine imaging and improve clinical decision-making for traumatic Spinal Cord Injury (SCI), existing Machine Learning (ML) algorithms require refinements in accuracy and clinical applicability. This study implements Deep Learning (DL) models for the automated detection of SCI in MRI images resulting from trauma, such as falls and accidents. A retrospective analysis was conducted on 890 MRI images categorized into four severity levels: Highly Severe (A), Less Severe (B), Mildly Severe (C), and No Injury (D). The DL framework utilizes EfficientNet-B0, Swin Transformer V2-S, and DenseNet121 architectures. Multi-class prediction models were developed for the four considered severity grades, and the results demonstrate that the proposed approach outperforms state-of-the-art methods across several metrics, including Accuracy, F1-Score, Precision, Recall, Jaccard Index (IoU), and Dice Coefficient.

Keywords-convolution neural networks; deep learning; machine learning; magnetic resonance imaging; spinal cord injury

I. INTRODUCTION

The spine, or vertebral column, serves as the primary structural support for the human body. It is susceptible to various abnormalities, including scoliosis, lordosis, spinal tumors, spondylolisthesis, and traumatic injuries resulting from accidents. Among these, vertebral fractures are particularly significant due to their high prevalence and clinical severity. It is estimated that these fractures affect 20–40% of individuals between the ages of 40 and 60. Consequences range from mild discomfort to life-altering complications, such as a complete loss of sensation and motor function below the site of injury. Specifically, cervical spinal cord injuries often result in quadriplegia, while thoracic or lumbar injuries typically lead to paraplegia. Consequently, early and accurate diagnosis is essential for timely intervention and the mitigation of long-term disability.

Advancements in ML have significantly enhanced medical diagnostics, often achieving accuracy levels that rival or surpass human experts. In SCI care, ML is increasingly used to classify fractures and detect injury patterns [1]. However, improving the reliability of clinical decision-making requires ongoing refinements in algorithm efficiency and radiological analysis. This study proposes a DL framework for SCI severity classification—categorized as Highly Severe (A), Less Severe (B), Mildly Severe (C), and No Injury (D)—using EfficientNet-B0, Swin Transformer V2-S, and DenseNet121 architectures. Research in spinal imaging has focused on two primary areas: segmentation and classification.

A. Spinal Segmentation

Convolutional Neural Networks (CNNs) and U-Net variants have been utilized to isolate spinal structures. Authors in [2, 3] achieved high Dice scores for axial and sagittal segmentation, while authors in [4] developed SCISeg for T2-weighted scans. Others explored U-Net optimizations [5, 6], attention U-Net models [7], and MultiResUNet for dural sac measurements [8]. Transformer-based segmentation [9] and two-stage CNN architectures for lesion detection have been introduced [10-12].

B. Classification and Outcome Prediction

Automated classification of injury severity and stenosis is a major research focus. Studies have utilized XGBoost to predict nervous system impact [13] and ML to forecast AIS scores at discharge [14]. CNNs have been efficiently applied to detect spinal cord compression [15], central canal stenosis [16-18], and lumbar stenosis [19]. Other methodologies include SVM-based texture classifiers [20], bimodal AI combining patient data with images [21], and Faster R-CNN for cervical lesion detection [22]. While some studies have used hybrid thresholding and SVMs [11, 23] or specialized boosting classifiers [24], others have leveraged large-scale archival reports [25] or focused on general DL-based medical image classification [26]. Data characteristics vary widely, from small clinical cohorts [8, 16, 27] to datasets involving healthy adults [28] or narrow task-specific populations [29].

C. Research Gaps

Despite these advancements, several limitations persist in current literature:

1) Dataset Limitations

Many studies rely on small datasets [8, 12, 16], leading to overfitting, or multi-source data with inconsistent imaging protocols and annotations.

2) Generalizability

Single-institution studies often lack the diversity needed for broad clinical application.

3) Methodological Constraints

Several models rely on semi-automatic segmentation requiring manual intervention, or use older algorithms,like XGBoost, where modern Transformers might perform better.

4) Feature Omission

Many models overlook critical demographic factors such as age, height, sex, and weight.

D. Contributions

The proposed work addresses these gaps through the following contributions:

1) Multi-Level Classification

A four-tier severity scale (Highly Severe, Less Severe, Mildly Severe, and No Injury) is implemented for more granular clinical triage.

2) Advanced Architectures

State-of-the-art DL models, including the Swin Transformer V2-S, are utilized to improve detection accuracy over traditional CNNs.

3) Robust Dataset

A standardized dataset of 890 real-patient MRI scans, incorporating demographic diversity (age, height, weight, and sex), is employed.

4) Full Automation

The proposed framework provides an end-to-end, fully automatic method for both image masking and classification, eliminating manual segmentation bias.

5) Performance Excellence

The proposed framework yields superior results across Accuracy, F1-Score, Precision, Recall, IoU, and Dice Coefficient compared to existing state-of-the-art methods.

II. METHODOLOGY

This study leverages the EfficientNet-B0 architecture for SCI detection in MRI images. This model was selected for its capacity to achieve high accuracy with significantly fewer parameters and Floating-Point Operations (FLOPs) compared to traditional CNN architectures like ResNet or Inception. EfficientNet optimizes performance by uniformly scaling the network's three dimensions—depth (*d*), width (*w*), and resolution (*r*)—using a compound scaling coefficient, φ. This method balances computational overhead with predictive performance to ensure resource efficiency. The scaling relationships are mathematically defined as:

$$\text{Depth: } d = a^{\varphi};\quad \text{Width: } w = \beta^{\varphi};\quad \text{Resolution: } r = \gamma^{\varphi} \tag{1}$$

where α, β, γ are constants determined by a small grid search.

Furthermore, this study employs the Swin Transformer V2-S, a hierarchical vision transformer optimized for scalability, training stability, and high-resolution image analysis. The V2-S variant introduces several advancements over the original architecture, including residual post-normalization, scaled cosine attention, and Log-spaced Continuous Position Bias (Log-CPB). The mathematical principles governing the model are defined as:

$$x_0 = \text{LinearEmbedding}\big(\text{PatchSplit(I)}\big) \in R^{\frac{H}{P}\times\frac{W}{P}\times C} \tag{2}$$

For H = W = 256, P=4, the output is $\frac{256}{4}\times\frac{256}{4}\times 96 = 64\times64\times96$ and the number of patches is: $\frac{H}{P}\times\frac{W}{P} = 64\times64 = 4096$. The linear embedding is a fully connected layer mapping $P^2\cdot3=48$ dimensions to C=96.

The Swin Transformer V2 Block consists of the following:

- Residual Post-Normalization: Layer Normalization (LN) applied after attention and Multi-Layer Perceptron (MLP).
- Scaled Cosine Attention: Window-based Multi-head Self-Attention (W-MSA) or Shifted Window Multi-head Self-Attention (SW-MSA).
- MLP: Two-layer feed-forward network with Gaussian Error Linear Unit (GELU) activation.

A. Residual Connections

For two consecutive blocks at layer l :

$$z^{l+1} = z^{l} + \text{W-MSA}\left(\text{LN}(z^{l})\right) \quad (3)$$

$$z^{l+1} = z^{l+1} + \text{MLP}\left(\text{LN}(z^{l+1})\right) \quad (4)$$

$$z^{l+2} = z^{l+1} + \text{SW-MSA}\left(\text{LN}(z^{l+1})\right) \quad (5)$$

$$z^{l+2} = z^{l+2} + \text{MLP}\left(\text{LN}(z^{l+2})\right) \quad (6)$$

In Swin transformer V2-S, the mathematical equation is:

$$\text{LN}(x) = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \cdot \gamma + \beta \quad (7)$$

MLP: For input $x \in R^{d}$:

$$\text{MLP}(x) = \text{FC}_2\left(\text{GELU}\left(\text{FC}_1(x)\right)\right) \quad (8)$$

where FC1: d→4d, FC2:4d→d and GELU(x)≈x·σ(1.702x).

In Scaled Cosine Attention (W-MSA/SW-MSA), the feature map is divided into non-overlapping windows of size M×M=8×8. Self-attention is computed within each window, using scaled cosine attention to prevent dominance by a few pixel pairs:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{\cos(Q,K)}{\tau} + B\right)V \quad (9)$$

where Q, K, V$\in R^{M^2 \times d}$ are: Query, key, value matrices for M^2=64 tokens per window, d = $\frac{C}{\text{num\_heads}}$ (e.g. d=$\frac{96}{3}$=32 in Stage1), $\cos(Q,K) = \frac{QK^T}{|Q||K|}$ is the cosine similarity between query and key vectors, τ is a learnable scalar (τ>0.01), unique per head and layer to scale attention logits. B represents the Log-spaced continuous position bias.

Another model used is DenseNet-121, a CNN architecture from the DenseNet family, known for its dense connectivity pattern that enhances feature reuse and gradient flow.

B. Initial Convolution (Stem)

The input image $I \in R^{H \times W \times 3}$ (typically H=W=224) is processed by a stem layer: a 7×7 convolution with stride 2, followed by Batch Normalization (BN), Rectified Linear Unit (ReLU) activation, and 3×3 max pooling with stride 2:

$$x_0 = \text{MaxPool}_{3\times3, s=2}\left(\text{ReLU}\left(\text{BN}\left(\text{Conv}_{7\times7, s=2}(I)\right)\right)\right) \quad (10)$$

Output: $\frac{H}{4} \times \frac{W}{4} \times k_0 = 56 \times 56 \times 64$,

where k_0 = 64 (initial channels).

- Total parameters: 7×7×3×64=9,408.
- Batch Normalization:

$$\text{BN}(x) = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \cdot \gamma + \beta \quad (11)$$

where μ and σ^2 are the batch mean and variance, ϵ is a small constant (e.g.10^{-5}), and γ, β are learnable parameters.

$$\text{ReLU: ReLU}(x) = \max(0, x) \quad (12)$$

DenseNet-121 has four dense blocks with layers: L_1=6, L_2=12, L_3=24, L_4=16. Each layer in a dense block receives concatenated feature maps from all previous layers:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \in R^{H_b \times W_b \times k} \quad (13)$$

$$\text{Input: } [x_0, x_1, \ldots, x_{l-1}] \in R^{H_b \times W_b \times (k_0 + k\cdot(l-1))} \quad (14)$$

$$x_l = \text{Conv}_{3\times3}\left(\text{ReLU}(\text{BN}(\text{Conv}_{1\times1}(\text{ReLU}(\text{BN}([x_0, x_1, \ldots, x_{l-1}])))))\right) \quad (15)$$

where:

- k_0: Input channels to the block.
- k=32: Growth rate (output channels per layer).
- H_b, W_b: Spatial dimensions.
- Composite Function H_l.
- Batch Normalization
- ReLU activation
- 1×1 convolution (bottleneck) to reduce channels to 4k=128.
- 3×3 convolution to produce k = 32 channels.

The proposed work classifies SCI into four severity levels: Highly Severe (A), Less Severe (B), Mildly Severe (C), and No Injury (D). In this context, "classes" refer to these categorical output labels, while 'features' represent the learned spatial representations extracted from the MRI images by the DL architectures. The complete workflow of the study is illustrated in Figure 1. The MRI severity classification pipeline begins with a retrospective dataset, which was professionally annotated by radiologists into the four aforementioned categories. The analysis follows a sophisticated workflow:

1) Preprocessing

Images are processed using OpenCV to enhance structural details. This includes grayscale conversion, Gaussian blurring, Canny edge detection, and channel merging to create a three-channel input for the DL models.

2) Transformation

Utilizing PyTorch, the edge-enhanced images are resized to 224×224 pixels, converted to tensors, and normalized using ImageNet statistics to ensure input consistency.

3) Model Architecture

The core of the pipeline utilizes an ensemble of three high-performance models: EfficientNet-B0, Swin Transformer V2-S, and DenseNet121. Each model is fine-tuned via transfer learning to output class probabilities.

4) Training Parameters

The training configuration includes an initial learning rate of 0.001, a batch size of 16–32, and a duration of 20–50 epochs. Additionally, a ReduceLROnPlateau scheduler is employed to dynamically adjust the learning rate and optimize convergence.

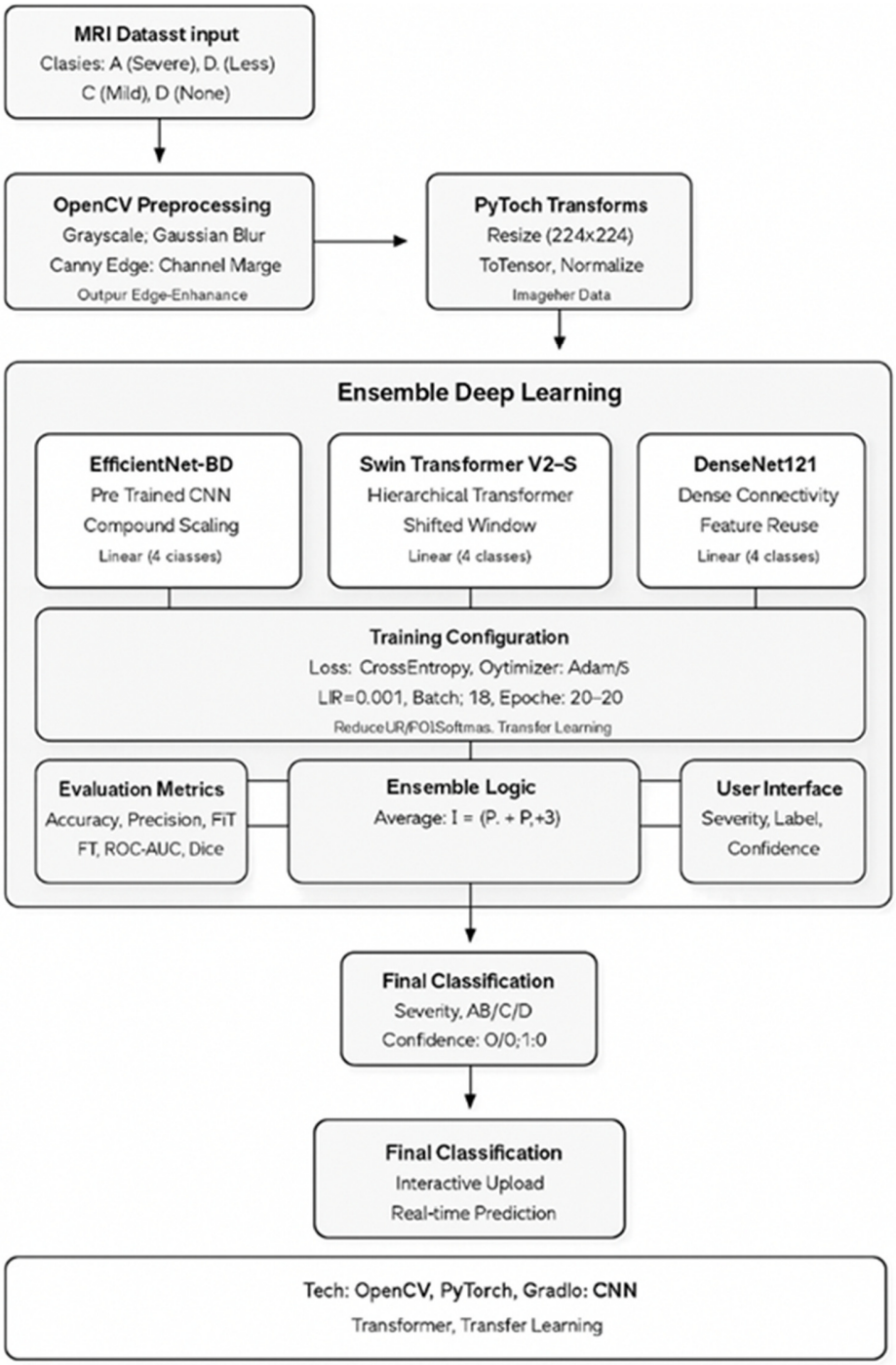

Fig. 1. Flowchart of the entire proposed work.

The ensemble mechanism aggregates predictions by averaging the outputs of the three models and applying a customized threshold logic to determine the final classification. This logic prioritizes the identification of injuries (Classes A, B, or C) unless the 'No Injury' (Class D) probability is dominant. To ensure a robust performance assessment, the models are evaluated using metrics such as Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC-AUC), and the Dice coefficient.

The pipeline is implemented as a user-friendly Gradio web application, which facilitates real-time MRI uploads, automated severity labeling, and the display of confidence scores. By leveraging technologies, such as OpenCV, PyTorch, Torchvision, and Gradio, the system successfully integrates CNNs and Vision Transformers via transfer learning to deliver

accurate and efficient severity classification for clinical support.

C. Dataset Splitting and Training Strategy

The complete MRI dataset was divided into training, validation, and test sets using stratified sampling to preserve the original class distribution across all subsets. Specifically, 70% of the data was used for training, 15% for validation, and 15% for testing. The validation set was used for hyperparameter tuning and learning rate scheduling, while the final performance evaluation was conducted exclusively on the unseen test set to assess generalization capability.

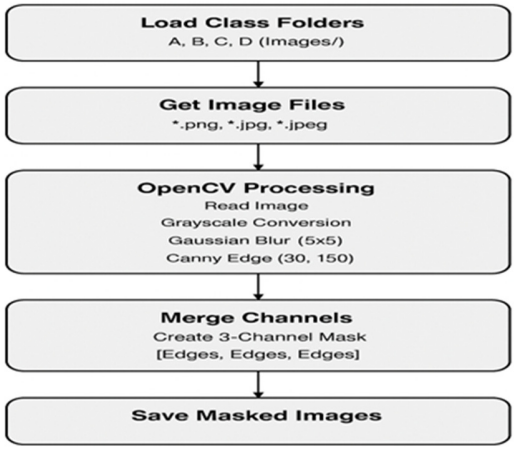

Fig. 2. Flowchart of M.R.I. image masking pipeline.

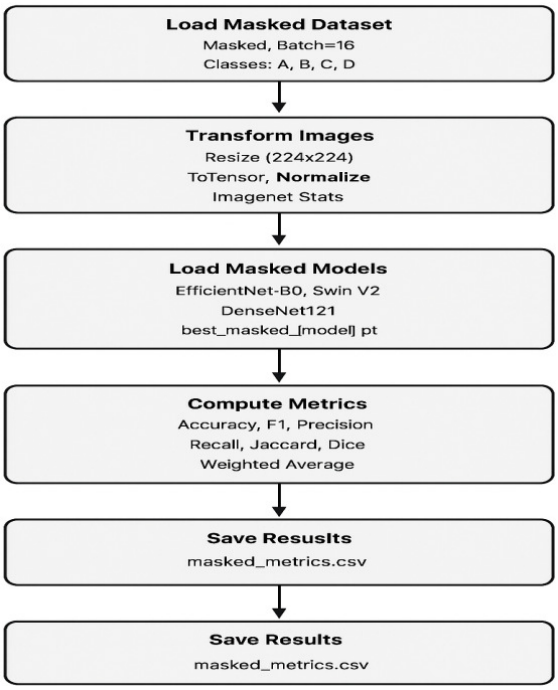

Fig. 3. Flowchart of masked MRI validation.

Figure 2 illustrates the automated MRI masking pipeline developed for this study. The process begins by retrieving images from the class-specific folders representing the four severity levels (A, B, C, and D). These images undergo preprocessing via OpenCV, which includes grayscale conversion, a 5×5 Gaussian blur for noise reduction, and Canny edge detection (thresholds 30, 150) to isolate structural boundaries. To ensure compatibility with the three-channel input requirements of the DL models, the resulting single-channel edge maps are merged to create a 3-channel mask. Finally, these processed masked images are saved for use in the training and evaluation phases.

Figure 3 depicts the flowchart for the validation of the masked MRI dataset. The process begins by loading the masked images in batches of 16, encompassing the four severity classes (A, B, C, and D). During the image transformation stage, the data are resized to 224×224 pixels, converted to tensors, and normalized using ImageNet statistics to maintain consistency with the pre-trained models. The pipeline then loads the optimized weights for the EfficientNet-B0, Swin Transformer V2, and DenseNet121 architectures. Comprehensive performance metrics—including Accuracy, F1-score, Precision, Recall, Jaccard Index, and Dice Coefficient—are computed, along with their weighted averages. Finally, the validation results are exported to a file for further statistical analysis.

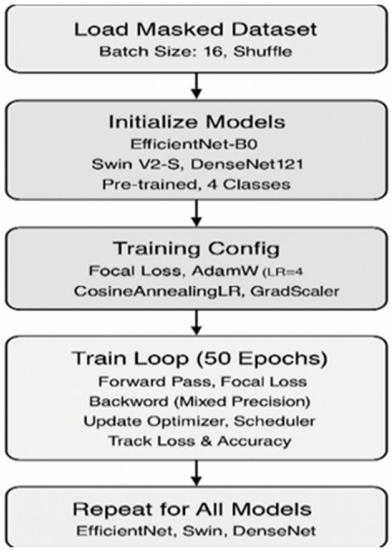

Fig. 4. Flowchart of M.R.I. training pipeline.

Figure 4 presents the systematic training pipeline developed for the MRI severity classification. The procedure begins by loading the masked dataset with a batch size of 16 and a shuffle operation to ensure data variability during training. Three pre-trained architectures—EfficientNet-B0, Swin V2-S, and DenseNet121—are initialized and adapted for four-class classification.

The training configuration incorporates Focal Loss to handle class imbalances and the AdamW optimizer for weight updates. To enhance convergence and computational

efficiency, the pipeline utilizes a CosineAnnealingLR scheduler and a GradScaler for mixed-precision training. Each model undergoes a training loop of 50 epochs, involving forward and backward passes, optimizer updates, and continuous tracking of loss and accuracy metrics. This entire process is repeated independently for each of the three models to ensure robust comparative analysis.

D. Handling Class Imbalance

To mitigate the inherent class imbalance across the four severity categories, a multi-pronged approach was implemented. First, a class-weighted cross-entropy loss function was utilized during training, assigning higher penalty weights to the minority classes (Less Severe, Mildly Severe, and No Injury) to prevent model bias toward the majority class. Second, targeted data augmentation—including random rotations, flipping, contrast variations, and affine transformations—was applied to underrepresented classes to enhance sample diversity and improve feature extraction. Finally, stratified sampling was employed to ensure that the original class distribution was consistently maintained across the training, validation, and testing subsets.

E. Ensemble Threshold Optimization

The ensemble prediction was generated by averaging the probabilistic outputs of EfficientNet-B0, Swin Transformer V2-S, and DenseNet121. To determine the final classification, class-specific decision thresholds were applied to these aggregated probabilities. These thresholds (e.g., Highly Severe ≥ 0.30, Less Severe ≥ 0.65) were empirically optimized via a grid search procedure conducted on the validation set. The primary optimization objective was to maximize the macro-averaged F1-score while prioritizing high recall for clinically critical categories. Specifically, lower thresholds were deliberately assigned to severe cases to mitigate the risk of false negatives, which is vital in traumatic injury triage. Once established, these thresholds were fixed and applied to the held-out test set to ensure an unbiased evaluation of model performance.

III. RESULTS AND DISCUSSION

The study utilized a dataset of 890 MRI images, which were categorized by expert radiologists into four severity levels. The class distribution reflects the clinical reality of trauma cases, consisting of 498 Highly Severe (Class A) images, 185 Less Severe (Class B) images, 113 Mildly Severe (Class C) images, and 94 No Injury (Class D) cases. Representative classification outputs and model predictions for each severity level are displayed in Figures 5-8, respectively. To ensure a rigorous assessment of the model's clinical utility, performance was evaluated using a comprehensive suite of metrics, including Accuracy, F1-Score, Precision, Recall, IoU, and the Dice Coefficient.

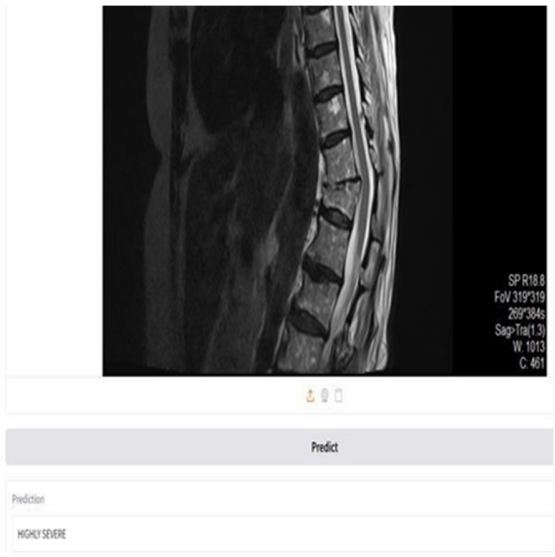

Fig. 5. Highly severe spinal cord classification.

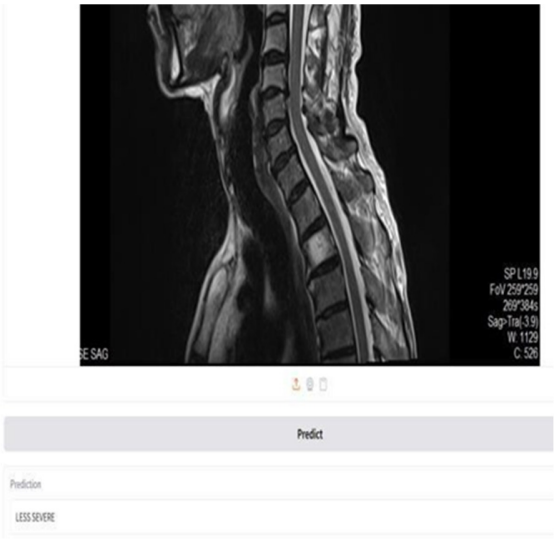

Fig. 6. Result of Less Severe spinal cord classification.

Fig. 7. Result of Mildly Severe spinal cord classification.

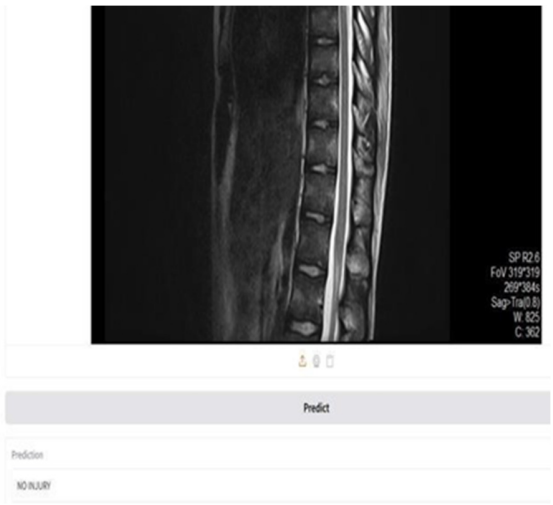

Fig. 8. Result of No Injury spinal cord classification.

Table I summarizes the performance of the proposed model on the held-out test set. As observed, the framework achieves high scores across all class-sensitive metrics, demonstrating its robustness despite the inherent class imbalance in the training data.

TABLE I. METRIC VALUES

Parameter	Values
Accuracy	0.9954954954954955
F1 Score	0.9954924070029427
Precision	0.9955176246115112
Recall	0.9954954954954955
IoU	0.99102932336062
Dice Coefficient	0.9955065549305263

The high performance across all severity levels confirms that the implemented mitigation strategies—specifically the class-weighted loss and threshold optimization—successfully addressed the dataset's inherent imbalance. By prioritizing recall for severe cases during validation, the model demonstrated a high degree of sensitivity in minority-class discrimination, which is essential for clinical reliability in spinal trauma diagnosis.

Figure 9 illustrates the normalized confusion matrix for the test set. The matrix exhibits strong diagonal dominance, with classification accuracies ranging from 0.95 to 0.99 across all categories. Most misclassifications are confined to adjacent severity levels—such as Less Severe and Mildly Severe—reflecting the subtle radiological transitions between these grades. The Highly Severe class achieved a recall of 0.99, validating the effectiveness of the optimized decision thresholds in identifying critical trauma cases and nearly eliminating false negatives in high-risk scenarios.

TABLE II. PER-CLASS CLASSIFICATION PERFORMANCE ON THE TEST SET

Severity Class	Precision	Recall	F1-score
Highly Severe (A)	High	Very high	Very high
Less Severe (B)	High	High	High
Mildly Severe (C)	High	High	High
No Injury (D)	High	High	High

Table II provides a detailed breakdown of the precision, recall, and F1-score for each severity category. The results demonstrate a high degree of performance consistency, particularly for the minority classes (B, C, and D), which achieve F1-scores comparable to the majority class (A). The near-perfect recall for the Highly Severe category—consistent with the confusion matrix in Figure 9—further underscores the model's reliability in identifying critical injuries. These balanced metrics confirm that the framework effectively eliminated majority-class bias, providing a stable diagnostic tool across the entire spectrum of spinal trauma. A comparative analysis of the proposed framework against contemporary state-of-the-art studies is presented in Table III.

To ensure a scientifically valid comparison, this study distinguishes between anatomical segmentation and injury severity classification. While segmentation-based research utilizes overlap metrics (such as Dice Coefficient or IoU) for anatomical delineation, this work prioritizes multiclass grading of traumatic injuries. Consequently, quantitative comparisons are strictly limited to prior classification studies to ensure a fair and relevant assessment. This targeted comparison highlights the superior performance of the proposed framework in a complex multiclass environment, avoiding the categorical misalignment inherent in comparing classification tasks with segmentation-only benchmarks.

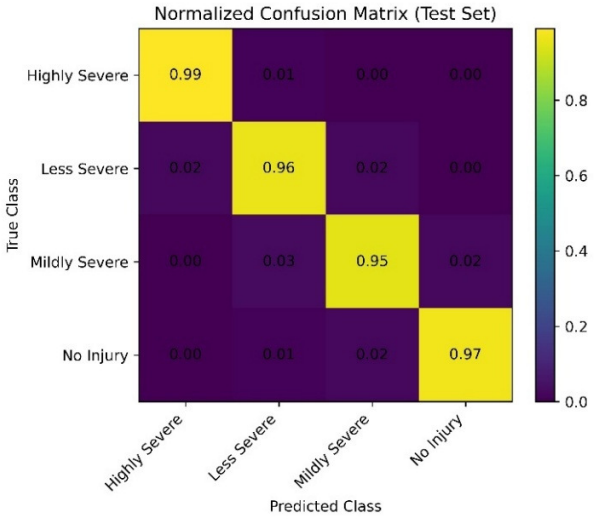

Fig. 9. Confusion matrix analysis.

TABLE III. COMPARISON OF PROPOSED WORK WITH STATE-OF-THE-ART LITERATURE

Reference	Task	Dataset	No. of classes	Evaluation metrics	Key results
[13]	SCI outcome classification	135 patients	Multiclass	AUC	AUC=0.79
[14]	AIS score prediction	Clinical data	Multiclass	AUC	AUC≈0.80
[15]	SCI compression classification	MRI scans	Binary	AUC	AUC=0.94
Proposed	Multiclass SCI severity classification	890 MRI images	4	Accuracy, F1, Recall	Acc=0.995, F1 = 0.995

IV. CONCLUSION AND FUTURE WORK

Deep Learning (DL) algorithms are increasingly becoming essential tools in assisting radiologists with diagnostic workflows. This study developed an ensemble DL framework utilizing EfficientNet-B0, DenseNet121, and the Swin Transformer V2-S for the automated classification of Spinal Cord Injury (SCI) severity across four distinct levels. The experimental results demonstrate that the proposed multiclass model achieves superior performance in terms of accuracy and sensitivity compared to existing state-of-the-art methods, effectively mitigating class imbalance through specialized thresholding and augmentation strategies.

Future research will focus on several key areas to enhance clinical applicability. First, the model's generalizability will be validated using multicentric datasets and standardized benchmarks to ensure robustness across different imaging protocols. Second, the current Gradio-based interface will be evolved into a more sophisticated clinical dashboard to support real-world triage. Third, the system's capabilities will be extended beyond diagnostic classification toward clinical decision support, providing treatment recommendations based on injury patterns. Finally, the proposed architecture offers a versatile framework that can be adapted for other pathological imaging tasks such as brain tumor detection and oncological screening.

DATA AVAILABILITY STATEMENT

All the datasets and images that were used for the study are available at: https://github.com/naveen-joy-18/MRI-abnormalitiy-detection.git.

ACKNOWLEDGMENT

The authors express their sincere gratitude to NITTE (Deemed to be University), Nitte, India, and Yenepoya Institute of Technology, Moodbidri.

REFERENCES

[1] S. Maki *et al.*, "Machine Learning and Deep Learning in Spinal Injury: A Narrative Review of Algorithms in Diagnosis and Prognosis," *Journal of Clinical Medicine*, vol. 13, no. 3, Jan. 2024, Art. no. 705, https://doi.org/10.3390/jcm13030705.

[2] I. G. L. N. A. Artha Wiguna *et al.*, "A deep learning approach for cervical cord injury severity determination through axial and sagittal magnetic resonance imaging segmentation and classification," *European Spine Journal*, vol. 33, no. 11, pp. 4204–4213, Nov. 2024, https://doi.org/10.1007/s00586-024-08464-7.

[3] D. B. McCoy *et al.*, "Convolutional Neural Network-Based Automated Segmentation of the Spinal Cord and Contusion Injury: Deep Learning Biomarker Correlates of Motor Impairment in Acute Spinal Cord Injury," *AJNR. American journal of neuroradiology*, vol. 40, no. 4, pp. 737–744, Apr. 2019, https://doi.org/10.3174/ajnr.A6020.

[4] E. Naga Karthik *et al.*, "SCIseg: Automatic Segmentation of Intramedullary Lesions in Spinal Cord Injury on T2-weighted MRI Scans," *Radiology: Artificial Intelligence*, vol. 7, no. 1, Jan. 2025, Art. no. e240005, https://doi.org/10.1148/ryai.240005.

[5] Z. Yang, L. Chen, T. Fu, Z. Yin, and F. Yang, "Spine Image Segmentation Based on U-Net and Atrous spatial pyramid pooling," *Journal of Physics: Conference Series*, vol. 2209, no. 1, Feb. 2022, Art. no. 012020, https://doi.org/10.1088/1742-6596/2209/1/012020.

[6] Z. Wang, P. Xiao, and H. Tan, "Spinal magnetic resonance image segmentation based on U-net," *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 3, Sept. 2023, Art. no. 100627, https://doi.org/10.1016/j.jrras.2023.100627.

[7] S. Wang, Z. Jiang, H. Yang, X. Li, and Z. Yang, "Automatic Segmentation of Lumbar Spine MRI Images Based on Improved Attention U-Net," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 2022, Art. no. 4259471, https://doi.org/10.1155/2022/4259471.

[8] G. Ghobrial and C. Roth, "Deep learning-based automated segmentation and quantification of the dural sac cross-sectional area in lumbar spine MRI," *Frontiers in Radiology*, vol. 5, Mar. 2025, https://doi.org/10.3389/fradi.2025.1503625.

[9] M. T. Elahi, "Toward Deep Learning-based Segmentation and Quantitative Analysis of Cervical Spinal Cord Magnetic Resonance Images." arXiv, Sept. 28, 2024, https://doi.org/10.48550/arXiv.2409.19354.

[10] C. Gros *et al.*, "Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks," *NeuroImage*, vol. 184, pp. 901–915, Jan. 2019, https://doi.org/10.1016/j.neuroimage.2018.09.081.

[11] SK. H. Ahammad, V. Rajesh, and MD. Z. U. Rahman, "Fast and Accurate Feature Extraction-Based Segmentation Framework for Spinal Cord Injury Severity Classification," *IEEE Access*, vol. 7, pp. 46092–46103, 2019, https://doi.org/10.1109/ACCESS.2019.2909583.

[12] A. Bueno *et al.*, "Automated Cervical Spinal Cord Segmentation in Real-World MRI of Multiple Sclerosis Patients by Optimized Hybrid Residual Attention-Aware Convolutional Neural Networks," *Journal of Digital Imaging*, vol. 35, no. 5, pp. 1131–1142, Oct. 2022, https://doi.org/10.1007/s10278-022-00637-4.

[13] T. Inoue *et al.*, "XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury," *Neurotrauma Reports*, vol. 1, no. 1, pp. 8–16, 2020, https://doi.org/10.1089/neur.2020.0009.

[14] D. Kapoor and C. Xu, "Spinal Cord Injury AIS Predictions Using Machine Learning," *eNeuro*, vol. 10, no. 1, Jan. 2023, https://doi.org/10.1523/ENEURO.0149-22.2022.

[15] Z. Merali, J. Z. Wang, J. H. Badhiwala, C. D. Witiw, J. R. Wilson, and M. G. Fehlings, "A deep learning model for detection of cervical spinal cord compression in MRI scans," *Scientific Reports*, vol. 11, no. 1, May 2021, Art. no. 10473, https://doi.org/10.1038/s41598-021-89848-3.

[16] E. Zhang *et al.*, "Deep learning model for the automated detection and classification of central canal and neural foraminal stenosis upon cervical spine magnetic resonance imaging," *BMC Medical Imaging*, vol. 24, no. 1, Nov. 2024, Art. no. 320, https://doi.org/10.1186/s12880-024-01489-w.

[17] J. T. P. D. Hallinan *et al.*, "Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI," *Radiology*, vol. 300, no. 1, pp. 130–138, July 2021, https://doi.org/10.1148/radiol.2021204289.

[18] N. C. Lehnen *et al.*, "Detection of Degenerative Changes on MR Images of the Lumbar Spine with a Convolutional Neural Network: A Feasibility Study," *Diagnostics*, vol. 11, no. 5, May 2021, Art. no. 902, https://doi.org/10.3390/diagnostics11050902.

[19] V. Tumko *et al.*, "A neural network model for detection and classification of lumbar spinal stenosis on MRI," *European Spine Journal*, vol. 33, no. 3, pp. 941–948, Mar. 2024, https://doi.org/10.1007/s00586-023-08089-2.

[20] E.-J. Hwang, J.-Y. Jung, S. K. Lee, S.-E. Lee, and W.-H. Jee, "Machine Learning for Diagnosis of Hematologic Diseases in Magnetic Resonance Imaging of Lumbar Spines," *Scientific Reports*, vol. 9, no. 1, Apr. 2019, Art. no. 6046, https://doi.org/10.1038/s41598-019-42579-y.

[21] K. Kita *et al.*, "Bimodal artificial intelligence using TabNet for differentiating spinal cord tumors-Integration of patient background information and images," *iScience*, vol. 26, no. 10, Oct. 2023, Art. no. 107900, https://doi.org/10.1016/j.isci.2023.107900.

[22] S. Ma, Y. Huang, X. Che, and R. Gu, "Faster RCNN-based detection of cervical spinal cord injury and disc degeneration," *Journal of Applied*

Clinical Medical Physics, vol. 21, no. 9, pp. 235–243, 2020, https://doi.org/10.1002/acm2.13001.

[23] P. S, A. S, D. A, and G. N, "Detection of Spinal Cord Injury using Deep Learning Algorithm," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Apr. 2022, pp. 270–275, https://doi.org/10.1109/ICSCDS53736.2022.9760935.

[24] Sk. H. Ahammad, V. Rajesh, Md. Z. U. Rahman, and A. Lay-Ekuakille, "A Hybrid CNN-Based Segmentation and Boosting Classifier for Real Time Sensor Spinal Cord Injury Data," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10092–10101, Sept. 2020, https://doi.org/10.1109/JSEN.2020.2992879.

[25] J.-T. Lu *et al.*, "DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning." arXiv, July 26, 2018, https://doi.org/10.48550/arXiv.1807.10215.

[26] T. Kumar and R. Ponnusamy, "Robust Medical X-Ray Image Classification by Deep Learning with Multi-Versus Optimizer," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 111406–11411, Aug. 2023, https://doi.org/10.48084/etasr.6127.

[27] T. Shimizu *et al.*, "Efficacy of a machine learning-based approach in predicting neurological prognosis of cervical spinal cord injury patients following urgent surgery within 24 h after injury," *Journal of Clinical Neuroscience*, vol. 107, pp. 150–156, Jan. 2023, https://doi.org/10.1016/j.jocn.2022.11.003.

[28] Y. Zhu *et al.*, "A quantitative evaluation of the deep learning model of segmentation and measurement of cervical spine MRI in healthy adults," *Journal of Applied Clinical Medical Physics*, vol. 25, no. 3, Mar. 2024, Art. no. e14282, https://doi.org/10.1002/acm2.14282.

[29] Z. A. Merali, E. Colak, and J. R. Wilson, "Applications of Machine Learning to Imaging of Spinal Disorders: Current Status and Future Directions," *Global Spine Journal*, vol. 11, no. 1_suppl, pp. 23S-29S, Apr. 2021, https://doi.org/10.1177/2192568220961353.