IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2024, accepted 21 July 2024, date of publication 23 July 2024, date of current version 1 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432803

RESEARCH ARTICLE

Inv-AlxVGGNets: Cervical Spine Disease Classification Using Concatenated Involutional Neural Networks With Residual Net

BINIYAM MULUGETA ABUHAYI, YOHANNES AGEGNEHU BEZABH, AND ALEKA MELESE AYALEW

Department of Information Technology, University of Gondar, Gondar 196, Ethiopia

Corresponding author: Biniyam Mulugeta Abuhayi (biniyam.mulugeta123@gmail.com)

ABSTRACT Cervical spine diseases, encompassing conditions like spondylolisthesis, disc degeneration, and cervical spinal stenosis, stand as significant contributors to global disability. Precise classification of these conditions is paramount for effective medical diagnosis. This paper introduces an innovative methodology aimed at addressing the limitations of traditional convolutional neural networks (CNNs) and pretrained models in this domain. We propose a novel approach dubbed Inv-AlxVGGNets, which leverages concatenated pretrained architectures AlexNet and VGG, augmented by involutional neural networks and residual layers. Unlike conventional CNNs that are location-specific and channel-agnostic, involutional neural networks offer enhanced Adaptability to diverse visual patterns in medical images. Focusing on a four-class cervical spine disease classification task utilizing MRI images, our study evaluates the performance of Inv-AlxVGGNets (AlexNet with INN and VGG with residual layer models) as well as machine learning algorithms. Our findings demonstrate superior performance in terms of accuracy, precision, recall, and AUC ROC values. Notably, Inv-AlxVGGNets achieves an impressive 98.73% accuracy on the testing set and 99.78% on the training set, underscoring its potential for precise cervical spinal disease classification. In a comparative analysis, we highlight that conventional CNNs entail over 133 million parameters, whereas Inv-AlxVGGNets require less than 8 million parameters, rendering them more efficient and resource-friendly. This reduced parameter count is particularly advantageous in resource-constrained scenarios, where computational resources and datasets may be limited. The promising results position Inv-AlxVGGNets as a valuable tool for precise cervical spine disease classification, offering implications for enhancing patient care in resource-constrained settings.

INDEX TERMS Cervical spine disease, involution neural network, convolutional neural network, deep learning, machine learning, classification, medical image processing.

I. INTRODUCTION

Spinal cord injuries (SCIs) have profound and enduring consequences, ranging from persistent disability to reduced life expectancy and compromised quality of life [1]. Not only do these injuries exact a toll on affected individuals, but they also impose significant financial burdens on healthcare systems [2], [3]. In a parallel domain, degenerative spinal conditions (DSCs) present a multifaceted spectrum of pathologies that profoundly impact health and well-being, affecting a substantial portion of the population at least once in their lifetime [4]. Within the realm of degenerative spinal conditions, cervical spine disease stands out for its complexity and prevalence. This category encompasses a variety of conditions, including osteoarthritis, degenerative disease, metastatic tumors, and osteomyelitis, each presenting unique

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

challenges in diagnosis and management [5]. Cervical discogenic disease, a form of osteoarthritis affecting the cervical spine, is particularly troublesome among older arthritic patients, often leading to symptomatic cervical radiculopathy or myelopathy [6]. Moreover, cervical degenerative disease, characterized by degenerative radiographic changes, can give rise to debilitating symptoms, further complicating clinical management [6]. Cervical spondylosis, another term for degenerative disease of the cervical spine, is prevalent and is primarily attributed to the natural aging process, manifesting as axial neck pain, radiculopathy, or myelopathy [7]. On the other end of the spectrum, osteomyelitis of the cervical spine, although rare, presents a serious infectious threat necessitating prompt diagnosis and aggressive surgical intervention for optimal outcomes [8]. Accurate classification of cervical spine disease is pivotal for determining appropriate treatment strategies and prognosticating outcomes. However, challenges abound, as exemplified by the complexities encountered in direct laryngoscopy in patients with cervical spine disease [7]. In the era of rapidly advancing technology, particularly in the domains of artificial intelligence (AI) and deep learning, there is burgeoning potential to revolutionize the detection and classification of spinal disease. While integrating such technologies into clinical practice poses challenges, ongoing research holds promise for enhancing patient outcomes [4]. Traditional diagnostic modalities, while effective, are not without limitations, often plagued by subjectivity and time constraints. The emergence of deep learning, particularly Involutional Neural Networks (INNs), presents an exciting opportunity to augment the precision and efficiency of cervical spine disease classification [9]. Involution layers within neural network architectures have demonstrated remarkable pattern-recognition capabilities in medical imaging datasets(lumbar bm), laying the foundation for a more nuanced approach to classification. In this study, we propose an innovative model, Inv-AlxVGGNet, which integrates elements from AlexNet and modified VGG architectures with residual layer based on involution neural networks, aiming to overcome the limitations inherent in traditional convolution networks. This approach holds the potential to enhance the accuracy and time complexity of learned features, thereby facilitating more accurate classification of cervical spine diseases. Contributions of this work include

- The study introduces a novel neural network architecture that combines the strengths of modified VGG with residual layer and AlexNet models with involutional neural networks.
- a novel approach to cervical spine disease classification using involutional neural networks.
- The introduction of the Inv-AlxVGGNet, and the demonstration of Inv-AlxVGGNet' ability to achieve accurate classification with fewer parameters than traditional CNNs.
- The paper also evaluates the performance of InvNets on a four-class cervical disease classification problem and underscores their potential for medical image analysis tasks, particularly in resource-constrained scenarios.

The subsequent sections of the paper are organized as follows: Section II presents literature review, section III details the materials and methods employed, including data collection and analysis procedures. Section IV presents the study findings supported by tables and figures, followed by Section V, which interprets the results, discusses their implications, and addresses study limitations. Section VI summarizes the main findings, reiterates their significance, highlights study limitations, and suggests avenues for future research.

II. LITERATURE REVIEW

The utilization of deep learning techniques in medical image classification has seen significant advancements in recent years. In this section, we review several studies that have explored various approaches for diagnosing and classifying cervical spine diseases, providing valuable insights into the potential of deep learning methodologies in this domain. One notable study by [8] introduced a computer-aided diagnosis system based on deep learning to classify cervical spine injuries as fractures or dislocations with remarkable accuracy. The model, leveraging deep learning architectures such as AlexNet and GoogleNet, achieved an impressive accuracy score of 99.56%, demonstrating its efficacy in clinical settings. Furthermore, the incorporation of Saliency maps enhanced the spatial understanding of specific classes, contributing to improved diagnostic precision. Similarly, [9] focused on developing a fully automated artificial intelligence-aided method for cervical vertebral maturation (CVM) classification using convolutional neural networks (CNNs). Four CNN models, including VGG16, GoogLeNet, DenseNet161, and ResNet152, were evaluated, with ResNet152 emerging as the most effective model for CVM classification. This study underscores the potential of deep learning approaches in automating complex clinical tasks such as CVM assessment. Addressing a specific pathological condition related to the cervical spine, [10] proposed a novel model for diagnosing foraminal stenosis using only X-ray images. By integrating data preprocessing techniques and transfer learning with a pre-trained ResNet50 model, the proposed approach demonstrated significant improvements in diagnostic accuracy, offering a cost-effective alternative to MRI examinations. Moreover, [4] introduced a deep learning model aimed at enhancing the early detection and screening of degenerative spinal conditions (DSCs). Trained on a dataset of spinal X-ray images, the model achieved an overall accuracy of 89% in classifying various DSCs, showcasing its potential for aiding clinicians in timely diagnosis and treatment planning. Furthermore, [11] evaluated the effectiveness of vision transformers (ViT) for detecting cervical spine fractures, achieving a notable accuracy of 98%. This study highlights the utility of novel machine learning architectures in streamlining diagnostic processes while ensuring interpretability and ease of training. Additionally, [12] proposed a deep learning model for detecting lumbar degenerative disease and assessed its generalization ability for detecting cervical degenerative disease using transfer learning. The model exhibited robust performance in both internal and external validation datasets, underscoring its potential for clinical implementation. In the realm of medical image classification beyond cervical spine diseases, various studies have demonstrated the efficacy of deep learning models such as AlexNet and modified versions thereof [13], [14] [15]. These studies showcase the versatility of deep learning approaches in tackling diverse medical imaging challenges, ranging from pneumonia detection to cervical cancer classification. Moreover, concatenated models demonstrated promising results in improving classification accuracy compared to traditional CNN models. This highlights the importance of exploring innovative architectures and techniques to enhance the diagnostic capabilities of deep learning models. While previous research has made significant strides in cervical spine disease detection and classification, our work contributes to this body of literature in several key aspects. Firstly, we propose a novel approach by integrating involutional neural networks with the concatenated modified AlexNet and VGG architecture, offering a comprehensive solution to address the limitations of traditional convolutional networks. Secondly, our study focuses specifically on multiclass classification of cervical spine diseases, providing a detailed analysis of our model's performance in a clinically relevant context. Furthermore, meticulous attention to dataset selection and preprocessing techniques ensures the robustness and generalization capability of our model. The related work underscores the potential of deep learning methodologies in advancing cervical spine disease diagnosis and classification. Our proposed Inv-AlxVGGNet model represents a significant advancement in this domain, offering superior performance, reduced parameter count, and enhanced adaptability to spatial variations. These findings hold promise for improved diagnostic accuracy and patient care in the field of cervical spine pathology.

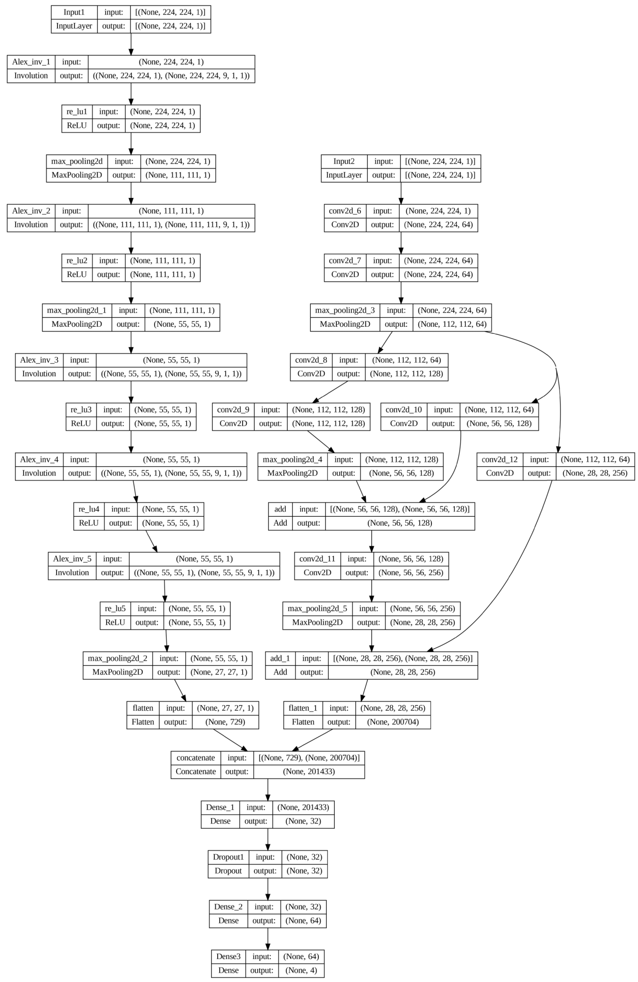


FIGURE 1. Proposed system model.

TABLE 1. Dataset before and after data augmentation.

Class	before data augmentation	after data augmentation
Healthy Cervical Spine	216	865
Cervical Spondylosis	184	735
Cervical Herniated disc	190	760
Cervical Spine Stenosis	202	810

III. MATERIALS AND METHODS

A. DATASET AND IMAGE PREPROCESSING

The dataset utilized in this study Collected from the University of Gondar's specialized medical College and comprises a collection of cervical spine MRI images.this dataset contains an anonymized clinical MRI study or set of scans of 423 patients with symptomatic back and neck pain. Each patient data can have one or more MRI studies associated with it. Each study contained slices, individual images taken from sagittal or axial views of the upper vertebrae of spine. These images are categorized into four groups: those depicting Healthy cervical spine,Cervical spondylosis, Cervical herniated disc, and Cervical spinal stenosis, as depicted in Figure 3 and Table 1. We affirm that all procedures performed in this study were in accordance with ethical standards and comply with the 1964 Helsinki Declaration. Patient identification data were not collected. **Data Splitting:** The dataset was split into training, validation, and testing subsets utilizing the 'train_test_split' function from ScikitLearn. This partitioning enabled the evaluation of the model's performance across varied data subsets, ensuring its robustness. **Image Resizing:** In order to standardize input dimensions, all images were resized to a consistent size of 224 × 224 pixels using the 'target_size' parameter within the Keras ImageDataGenerator. **Label Encoding:** Categorical labels were encoded into numerical format using one-hot encoding, facilitating the utilization of categorical

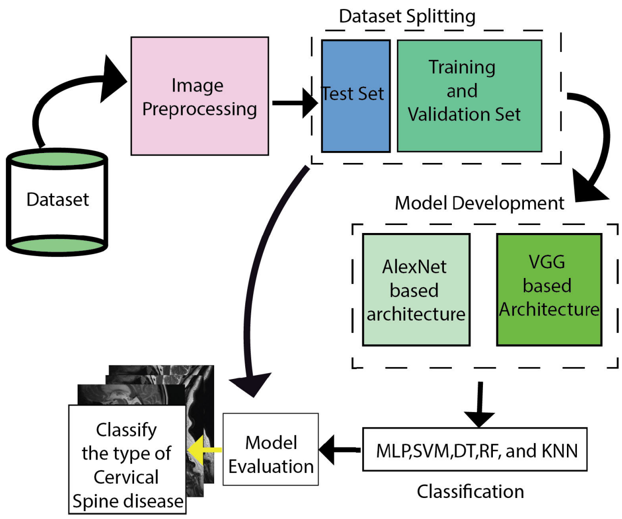


FIGURE 2. Proposed model architecture.

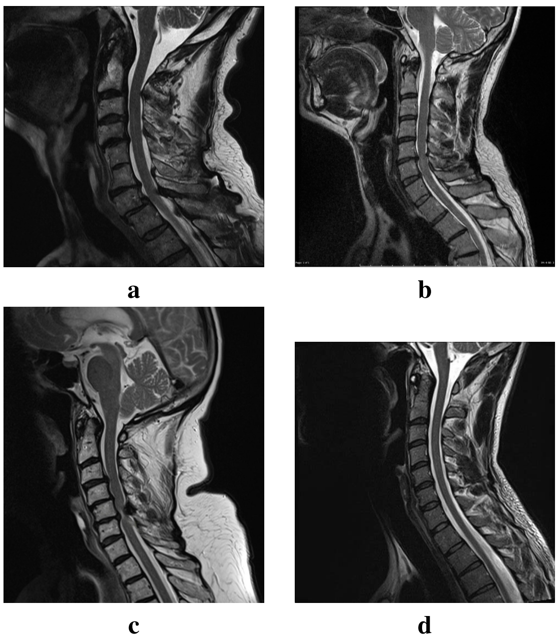


FIGURE 3. MRI images of cervical spine diseases and a healthy cervical spine: (a) cervical herniated, (b) cervical spinal stenosis, (c) cervical spondylosis, and (d) healthy cervical spine.

cross-entropy loss during model training. **Data Augmentation:** To enhance the model's generalization capabilities and improve its robustness, various data augmentation techniques were employed during training. These techniques were seamlessly integrated into the data pipeline using the Keras ImageDataGenerator and included random rotations, flips, and zoom. By artificially increasing the diversity of the training data, data augmentation aided the model in learning more effectively.

B. THE ARCHITECTURE OF INVOLUTIONAL NEURAL NETWORK

Involutional neural networks (INV-Nets) represent a paradigm shift from traditional convolutional neural networks (CNNs) by introducing a novel approach to feature extraction. Unlike conventional CNNs, which rely on spatial-agnostic kernels, INV-Nets utilize location-specific and channel-agnostic involution kernels. These involution kernels dynamically adapt to specific spatial positions within the input tensor, enabling more effective feature extraction. By generating each kernel based on the input tensor, INV-Nets can capture intricate details and subtle patterns, essential for tasks such as image classification [16]. This adaptability makes INV-Nets particularly well-suited for medical image analysis, where precise feature extraction is crucial for accurate diagnosis. Moreover, the reduced parameter count of INV-Nets contributes to enhanced computational efficiency, making them suitable for resource-constrained environments. INV-Nets offer a promising alternative to traditional CNNs, with the potential to significantly improve diagnostic accuracy and efficiency in various applications, including medical imaging tasks.

C. THE ARCHITECTURE OF CONVOLUTIONAL NETWORK

Convolutional neural networks (CNNs) are deep learning architectures commonly used for image processing and computer vision tasks. They consist of several layers, including the convolution layer, pooling layer, fully connected layer, and non-linear layer [17]. The convolutional layer applies kernel filters to extract fundamental features from input images [18]. The pooling layer combines successive convolutional layers and downsamples the feature maps [19]. The fully connected layer is responsible for generating the final output of the CNN. CNNs use activation functions such as Sigmoid, Tanh, ReLU, Leaky ReLU, Noisy ReLU, and Parametric Linear Units to define the output of the neural network [20]. Popular CNN architectures include LeNet, AlexNet, and VGGNet. Efficient hardware and algorithmic optimizations are necessary for real-time CNN inference, as CNNs require large computational and energy resources.

D. PROPOSED MODEL ARCHITECTURE

Our methodology propose a novel methodology for cervical spine disease classification by leveraging the output features from both the AlexNet-inspired branch utilizing Involution instead of traditional CNN, and the VGG-inspired branch employing residual layers. These features are then concatenated to serve as input to fully-connected layers for classification. Our aim is to harness the combined strengths of both architectures to achieve superior performance in cervical spine disease classification tasks (see Figure 1 and Figure 2). The integration of modified VGG with an AlexNet-based involution neural network offers several compelling benefits: **Feature Enrichment:** By combining the features learned by both models, our approach enriches the representation of the input data. While the VGG model captures certain aspects of the input, the involution operations in AlexNet provide adaptive processing of spatial information, resulting in a more comprehensive feature set [21]. **Improved Performance:** The concatenation of the output features from both branches

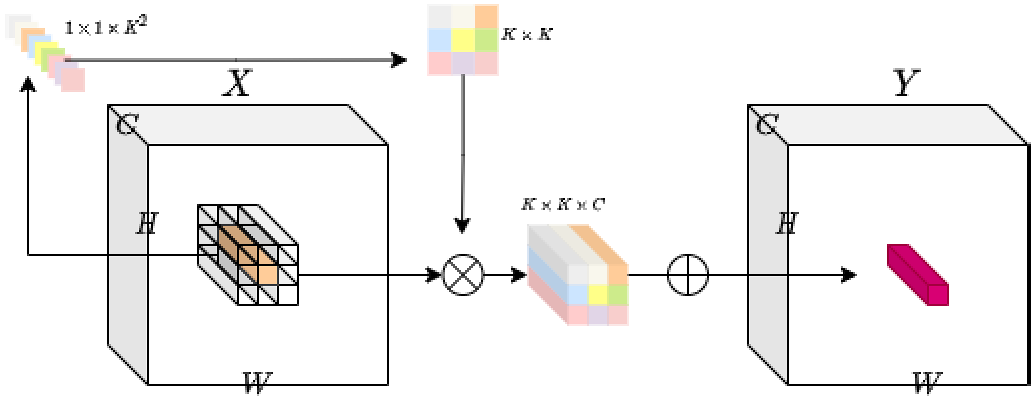


FIGURE 4. Involution Neural Network architecture.

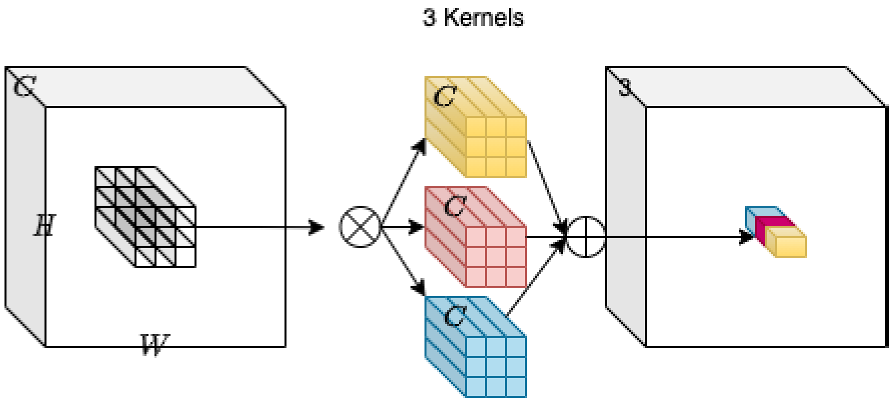


FIGURE 5. Convolution Neural Network architecture.

has the potential to significantly enhance performance in tasks such as image classification. This integration allows the model to exploit the unique strengths of each architecture, leading to improved accuracy and discriminative power. **Increased Model Capacity:** Concatenating the output features increases the model's capacity to learn more complex features, which is particularly advantageous for large and diverse datasets. This expanded capacity enables the model to capture subtle nuances in the input data, improving its ability to generalize effectively. Robustness to Variations: The involution operations employed in the AlexNet-based branch enhance the network's robustness to variations in the input data, such as differences in scale or rotation. This resilience is crucial for tackling the inherent challenges in image recognition tasks, ensuring reliable performance across diverse real-world scenarios. This section provides a detailed overview of the architectural components and their integration to form a robust framework for feature extraction and classification.

1) VGG-INSPIRED LAYERS

VGG, a widely recognized convolutional neural network (CNN) architecture, has showcased exceptional performance across a spectrum of applications, spanning from the detection of plant diseases in agriculture [22], [23] to the identification of lung and breast cancer severities [24], [25]. Among its variants, VGG-16 stands out for its effectiveness in image classification tasks, owing to its utilization of Rectified Linear Units (ReLU) activation function and a stack of convolutional layers augmented with flatten, normalization, dense, and drop-out layers. Often complemented by transfer learning and deep learning techniques, VGG-16 has been instrumental in achieving high accuracy rates in various domains. In our study, we propose an enhanced VGG model that integrates residual layers, marking a significant evolution in neural network design. This novel architecture combines the depth and comprehensive feature extraction prowess of the VGG model with the gradient-preserving benefits of residual connections. These skip connections facilitate the training of deeper networks by enabling gradients to flow more freely, thereby addressing the vanishing gradient problem. Notably, our proposed architecture not only accelerates convergence and enhances accuracy during training but also exhibits greater resilience against overfitting, making it particularly suitable for intricate image processing tasks such as medical image analysis. Implemented using modern deep learning frameworks, this modified VGG model represents a notable advancement in the field, offering a potent and efficient solution for diverse image-related applications. The VGG-inspired model architecture begins with an input layer defining the shape of input images as (224, 224, 1). It follows the characteristic VGG pattern, starting with Block 1 consisting of two convolutional layers with 64 filters each, followed by max-pooling to downsample the feature maps. Block 2 builds upon this, increasing the number of filters to 128 while maintaining the convolutional and max-pooling structure. Notably, there's a residual connection from Block 1 to enhance gradient flow. Block 3 introduces a convolutional layer with 256 filters followed by max-pooling, excluding the last three convolutional layers of the original VGG. Another residual connection from Block 1 is added to Block 3. The architecture concludes with two fully connected layers with 512 and 1024 neurons, respectively, and an output layer with 4 neurons for classification tasks, employing softmax activation. This model combines the effectiveness of VGG's stacked convolutional layers with the integration of residual connections, offering a robust framework for image classification tasks.

2) ALEXNET-INSPIRED LAYERS

The provided model architecture draws inspiration from AlexNet, a groundbreaking convolutional neural network (CNN) renowned for its pivotal role in advancing deep learning for image classification tasks. AlexNet gained prominence by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. In this architecture, the input layer is configured to accept images with dimensions of 224 × 224 pixels. Departing from traditional convolutional layers, the model incorporates Involutional Layers, a novel self-attention mechanism designed to capture long-range dependencies within feature maps, followed by Rectified Linear Unit (ReLU) activation functions to introduce non-linearity. Max pooling layers with a (3, 3) pool size and (2, 2) strides are interspersed after every two Involution layers to downsample feature maps effectively. The architecture also includes fully connected layers with 4096 neurons each, activated by ReLU functions, followed by dropout layers with a dropout rate of 0.5 to mitigate overfitting. Finally, the model concludes with a dense output layer comprising four neurons, representing the classification classes, and utilizes the softmax activation function to output class probabilities. By integrating Involution layers and retaining the core principles of AlexNet, this architecture offers a promising approach for image classification tasks, showcasing advancements in deep learning methodologies.

3) INVOLUTION LAYERS

Innovatively, instead of traditional CNN convolutions, involution layers are proposed. These layers dynamically learn convolutional patterns within local receptive fields, effectively reducing the number of parameters compared to standard convolutions while maintaining efficiency. Key aspects of involution layers include: Dynamic Convolutional Patterns: Involution layers adaptively learn convolutional weights within small kernel sizes. Parameter Efficiency: The use of involution contributes to model efficiency by reducing parameter count while preserving performance.

TABLE 2. Parameter configuration of proposed model.

Layer (type)	Output Shape	Param
Input1 (InputLayer)	(None, 224, 224, 1)	0
Alex_inv_1 (Involution)	(None, 224, 224, 1)	24
max_pooling2d (MaxPooling2D)	(None, 111, 111, 1)	0
Input2 (InputLayer)	(None, 224, 224, 1)	0
Alex_inv_2 (Involution)	(None, 111, 111, 1)	24
conv2d_6 (Conv2D)	(None, 224, 224, 64)	640
conv2d_7 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d_1 (MaxPoolinng2D)	(None, 55, 55, 1)	0
max_pooling2d_3 (MaxPoolinng2D)	(None, 112, 112, 64)	0
Alex_inv_3 (Involution)	(None, 55, 55, 1)	24
conv2d_8 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_9 (Conv2D)	(None, 112, 112, 128)	147584
Alex_inv_4 (Involution)	(None, 55, 55, 1)	24
max_pooling2d_4 (MaxPooling2D)	(None, 56, 56, 128)	0
conv2d_10 (Conv2D)	(None, 56, 56, 128)	8320
add (Add)	(None, 56, 56, 128)	0
Alex_inv_5 (Involution)	(None, 55, 55, 1)	24
conv2d_11 (Conv2D)	(None, 56, 56, 256)	295168
max_pooling2d_5 (MaxPooling2D)	(None, 28, 28, 256)	0
conv2d_12 (Conv2D)	(None, 28, 28, 256)	16640
max_pooling2d_2 (MaxPooling2D)	(None, 27, 27, 1)	0
add_1 (Add)	(None, 28, 28, 256)	0
flatten (Flatten)	(None, 729)	0
flatten_1 (Flatten)	(None, 200704)	0
concatenate (Concatenate)	(None, 201433)	0
dense_1 (Dense)	(None, 32)	6445888
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 64)	2112
dense_4 (Dense)	(None, 4)	260
Total params:	—	7027516
Trainable params:	—	7027506
Non-trainable params:	—	10

4) MODEL CONCATENATION

The output features from both the AlexNet-inspired and VGG-inspired branches are concatenated, serving as input to fully-connected layers for classification. By concatenating the strengths of both architectures, our approach aims to achieve superior performance in cervical spine disease classification tasks.

E. MODEL TRAINING

Dataset Splitting: The dataset was divided into training, validation, and testing sets in a 60:20:20 ratio, respectively, to train, validate, and test the deep learning model. Stratified sampling was employed to ensure balanced class distributions across the sets. **Training Procedure:** The model was trained using the extracted features of the training set with the Adam optimizer and categorical cross-entropy loss function. Hyperparameters such as learning rate, batch size, and number of epochs were fine-tuned through empirical experimentation to optimize model performance. During training, the model was trained using the extracted features of the training set. The validation set was used to evaluate the model's performance and prevent overfitting by monitoring its performance. After training, the model was tested on the testing set to provide an unbiased estimate of its performance on unseen data. To achieve this splitting, the dataset was initially divided into a training set (60%) and a temporary set containing the remaining data (40%). Stratified sampling was applied to ensure that the class distribution was preserved in both sets. Subsequently, the temporary set was further divided into a validation set (20% of the original dataset) and a test set (20% of the original dataset) using a 50:50 split. This stratified approach was maintained to ensure that each class was proportionally represented in the training, validation, and test sets, preserving the integrity of the original dataset's class distribution. **Early Stopping:** To prevent overfitting, early stopping was employed based on the validation loss. Training was halted when the validation loss failed to decrease over a predefined number of epochs.

F. EVALUATION TECHNIQUES

Accurate assessment methodologies are crucial for validating the classification efficacy of a system, especially in image classification tasks where subtle distinctions are crucial. In this investigation, various evaluation metrics were carefully selected to gauge the precision and efficiency of our proposed model for cervical spine disease classification. The initial metric employed was accuracy, a fundamental measure that quantifies the proportion of accurately classified images within the test set. Accuracy serves as a reliable indicator of overall classification performance, offering a quick and intuitive assessment of the model's effectiveness. Additionally, precision, recall, and the F1 score were thoroughly evaluated, using the confusion matrix to provide deeper insights into classification performance. Precision measures the proportion of correctly classified positive instances out of all instances classified as positive, offering insights into the model's ability to minimize false positives. Recall, or sensitivity, quantifies the proportion of correctly classified positive instances out of all actual positive instances, providing valuable information on the model's ability to detect true positives. The F1 score, which harmonizes precision and recall, offers a balanced measure of the model's accuracy, particularly valuable in scenarios with imbalanced class distributions. Collectively, these evaluation metrics offer a detailed and comprehensive assessment of the proposed cervical spine disease classification model, ensuring its suitability for image classification tasks where accuracy and reliability are crucial.

Accuracy: Accuracy is the proportion of correct predictions made by the model out of all predictions made. It measures how well the model classifies the samples. The formula for accuracy is:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

Recall (Sensitivity): Recall is the fraction of positive instances that are correctly identified. It can be represented as:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Precision:

Precision is the fraction of positive predictions that are actually correct. It can be represented as:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

F1 Score: F1 Score is the harmonic mean of precision and recall. It is a single number that balances both the precision and recall. It can be represented as:

$$F_1 Score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

where True Positives (TP): The number of instances that were correctly classified as positive by the model. True Negatives (TN): The number of instances that were correctly classified as negative by the model. False Positives (FP): The number of instances that were incorrectly classified as positive by the model. False Negatives (FN): The number of instances that were incorrectly classified as negative by the model. Total Samples: The total number of instances in the dataset.

AUC:

Area Under the Curve (AUC) is a commonly used performance metric machine learning classification problems, which evaluates the overall performance of a classifier. In binary classification, a classifier outputs a predicted probability for each sample to belong to one of two classes, positive or negative. The AUC is the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) for different classification thresholds.

The equation for the ROC curve is:

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$

where TP, TN, FP, and FN are the number of True Positives, True Negatives, False Positives, and False Negatives, respectively.

The equation for the AUC can be written as:

$$AUC = \int_{-\infty}^{\infty} [\frac{TP}{TP + FN} - \frac{FP}{FP + TN}] ds$$

G. CLASSIFICATION

To classify cervical spine disease, we employed a combination of Multi Layer Perceptron (MLP), and traditional machine learning classifiers such as SVM, KNN, and RF techniques on top of Inv-AlxVGGNets. **MLP:** MLP-like architectures have been used in image classification tasks, particularly in medical image classification and hyperspectral image classification. These MLP-like models are simple, computationally efficient, and have strong generalization capabilities. They require better input features from the image, which can be achieved by using convolutional layers and Involutional layers and complex transformations [26].

The integration of machine learning, including MLP, in medical image analysis can enhance diagnostic accuracy and improve healthcare [27]. These studies highlight the potential applications and techniques of MLP in medical imaging, demonstrating its effectiveness in disease diagnosis, surgical planning, and prognosis assessment. **SVM:** Support Vector Machine (SVM) is a popular technique used in image classification. SVM is a mathematical model for classification and regression that has been widely studied and improved over the years. It has been applied in various image processing approaches and algorithms, including vehicle classification and medical image classification [28], [29]. SVM has also been used in combination with deep neural networks for image classification tasks, where the features learned by the neural network are transferred to the SVM for prediction. SVM has shown promising results in terms of accuracy for image classification, with studies reporting accuracies of greater than 90% [30]. The selection of an appropriate approach and the quality of the labels significantly impact the robustness of the SVM model. Overall, SVM is a powerful tool for image classification and continues to be an active area of research. **KNN:** NN algorithm has been widely used in image classification. It is one of the oldest, simplest, and accurate algorithms for pattern classification and regression models. Researchers have also improved the KNN algorithm to enhance its accuracy. One study optimized the selection strategy of K value using genetic algorithm, resulting in a nearly 10% improvement in classification accuracy [31]. Another study proposed a CNN-KNN architecture for brain tumor detection and classification. The CNN-KNN method effectively detects and classifies various forms of tumors with a promising accuracy of 95.7% [32]. Among the two types of classifiers used in the CNN-KNN architecture, KNN performed the best accuracy. **RF:** Random forest models have been used in image classification in various studies. For example, Winterauer et al. developed a random decision forest model for fast identification of microplastics in environmental samples using Fourier-transform infrared spectra [33]. Other study proposed a breast cancer detection model based on ResNet and random forest models to improve the efficiency and accuracy of image processing and classification [34]. Reference [34] presented a Self-Attention Random Forest (SARF) model for the classification of breast cancer histopathological images, achieving high accuracy and outperforming other methods. Additionally, Gupta and Kumar used a random forest classifier to classify emotions in emojis and small texts, achieving a high classification accuracy [18], [35]. These studies demonstrate the effectiveness of random forest models in image classification tasks.

H. HARDWARE AND SOFTWARE SPECIFICATIONS

In this study, our hardware setup included an Intel(R) Core(TM) i5-10210U Multi core CPU with Maximum Turbo Frequency of 4.2GHz. The software stack comprised Python version 3.9.13 as the primary programming language, along with key libraries and frameworks. Specifically, we utilized Keras version 2.11.0 for deep learning model development and training, scikit-learn version 1.0.2 for various machine learning tasks, and OpenCV for image processing and computer vision applications. These software versions were chosen for their compatibility and functionality, ensuring robustness and efficiency throughout the experimentation process.

TABLE 3. Results of individual models with convolutional neural network.

classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
AlexNet	82.73	83.08	81.64	82.35
VGG	84.62	83.13	82.78	82.95

IV. RESULTS

In the results section, we present an in-depth analysis of our proposed approach's performance. We evaluated key performance metrics, including training and validation accuracy, as well as loss. The confusion matrix was employed to examine the behavior of distinct classes. Additionally, we generated ROC curve plots to visualize the model's performance at various threshold levels. To provide context for our findings, we compared our results with other state-of-the-art approaches, highlighting the strengths and advantages of our proposed method.

A. RESULTS OF INDIVIDUAL ARCHITECTURES

We conducted an analysis of both AlexNet-inspired and VGG-inspired models based on Involutional Neural Networks (INNs) and Convolutional Neural Networks (CNNs) to evaluate their impact on the overall model performance. For each feature extractor, we assessed its effectiveness using a fully connected (FC) classifier. Initially, we utilized the original architectures of AlexNet and VGG16. The outcomes are presented in Table 3 and Table 4. The VGG model demonstrated better performance when utilizing a fully connected layer combined with a SoftMax classifier, achieving a testing accuracy of 84.62%. Subsequently, we tested individual models of our proposed architecture, which include an involution-based AlexNet and a VGG-inspired model with residual layers. The results showed that the AlexNet model performed slightly better than the VGG-inspired model, achieving an accuracy score of 88.51%. This improvement over the original AlexNet and VGG models without any alterations suggests the effectiveness of incorporating involutional neural networks and residual layers in our architecture.

B. RESULTS BEFORE AND AFTER DATA AUGMENTATION

We conducted experiments to evaluate the impact of data augmentation techniques on the performance of our proposed model. Initially, we assessed the performance of Convolutional Neural Networks with Residual layers and Involutional Neural Networks without applying any data augmentation. The results, as depicted in the Figure 6 and summarized in

TABLE 4. Results of individual models with involutional neural network and residual layers.

classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
AlexNet with INN	88.51	88.08	87.64	87.85
VGG with Residual layers	87.21	86.11	86.88	86.49

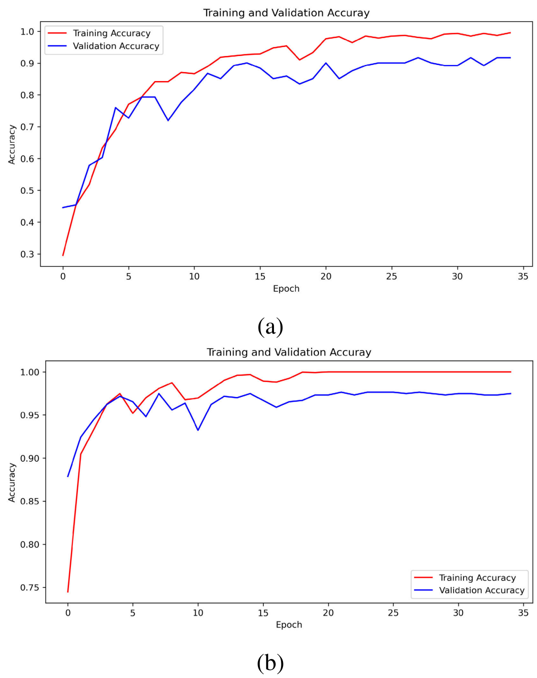


FIGURE 6. Learning curve of proposed model before data augmentation (a) and after data augmentation (b).

Table 5, revealed comparable accuracies between the models with and without data augmentation. However, a noticeable gap in the learning curves between training and validation results indicated the presence of overfitting. Overfitting occurs when a model performs well on the training data but poorly on the validation or testing dataset, reflecting its limited generalization ability. To mitigate this issue, we applied data augmentation techniques such as flipping, rotation, and zooming to the dataset. The subsequent results demonstrated improvements in the model, as evidenced by enhanced validation accuracy and reduced disparity between training and validation accuracies. This enhancement suggests an improvement in the models' generalization ability, addressing the overfitting problem effectively and increase the performance of the proposed model.

C. RESULTS AFTER IDENTIFYING OPTIMAL HYPER PARAMETERS

After we have experimented with different hyperparameters such as learning rate, loss function and activation functions as shown in Table 6 and Figure 5 the proposed model achieved

TABLE 5. Results of proposed study before and after data augmentation.

classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Proposed model without Augmentation	87.29	86.01	87.10	86.55
Proposed model with Augmentation	92.62	91.75	90.89	91.32

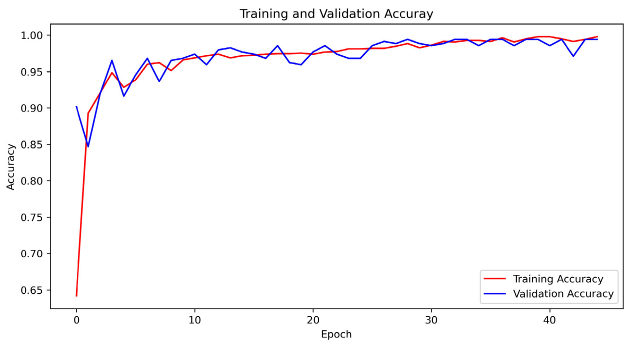
TABLE 6. Optimal hyperparameters for the proposed model.

list of hyperparameters	optimal values
Learning Rate	0.01
optimizer	adam
loss function	categorical cross entropy
activation function	ReLu and Softmax
batch size	32
EarlyStopping	monitor=loss, patience=2

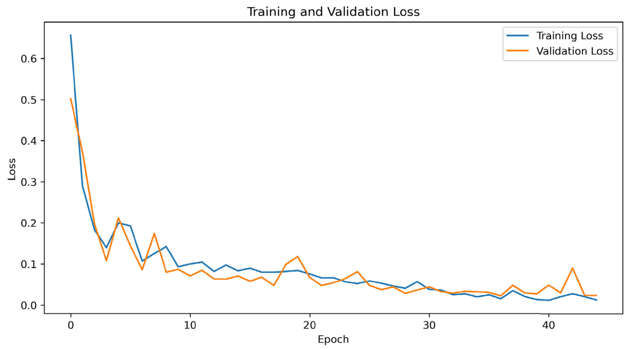
99.78% training accuracy and a 98.73% testing accuracy after trained for 40 epochs. As we can see from Figure 7 (a) the model shows no overfitting problem. And Figure 7 (b) shows the loss for training and validation which indicates training loss focuses on how well the model fits the training data, while validation loss evaluates its performance on unseen data. Monitoring both these metrics helps us understand the model's behavior during training and aids in preventing overfitting. and the proposed model was able to achieve much less than 1%. A shown in Figure 7(c) the results have further been explained through the use of confusion matrix which summarizes the performance of a the model on a set of test data. It provides a clear breakdown of accurate and inaccurate predictions based on the model's output. And in our model as we can see from the figure out of 173 images of healthy cervical spine sets 173 were correctly classified as healthy cervical spine and out of 147 images of cervical spondylosis only one of them were misclassified as cervical spine stenosis. and from 152 images of cervical herniated disc one image were misclassified as cervical herniated disc and finally from 162 images of cervical spinal stenosis 6 images were classified as cervical spondylosis. This meticulous analysis not only validates the high accuracy metrics reported but also offers valuable insights into the model's strengths and areas for potential improvement, ultimately enhancing our understanding of its performance characteristics and informing future optimization efforts.

D. RESULTS OF MACHINE LEARNING CLASSIFIERS

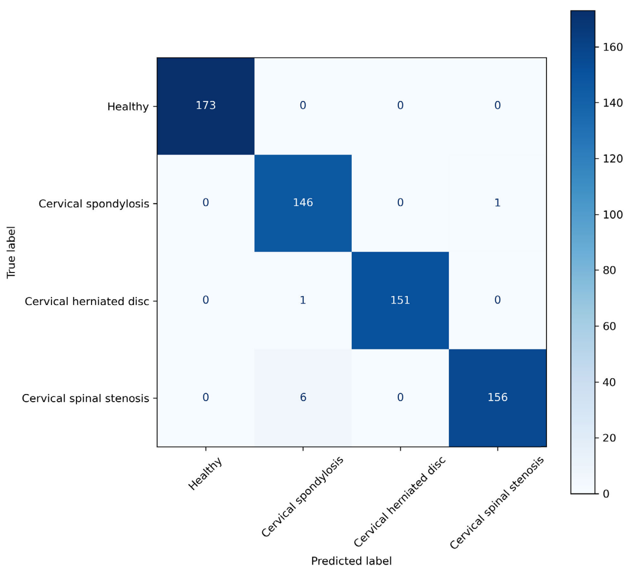
In this section, we embarked on a comprehensive exploration of classical machine learning classifiers, including Support Vector Machine (SVM), Random Forest, Decision Tree, and K-Nearest Neighbors (KNN). Our evaluation methodology leveraged Receiver Operating Characteristics (ROC) Curves and the Area Under the Curve (AUC), as illustrated in the



(a)



(b)



(c)

FIGURE 7. **Learning curve of proposed model (a) training and validation accuracy, (b) training and validation loss, and (c) confusion matrix.**

accompanying Figure 8 and Table 8, to provide nuanced insights into classifier performance. As shown in Table 7 by employing Grid Search, we meticulously fine-tuned various hyperparameters for each classifier to optimize predictive accuracy. Notably, for SVM, we identified optimal parameters, including a regularization parameter (C) of 44216.28, a gamma value of 1000, and an RBF kernel. The Decision Tree classifier was trained with a maximum leaf node of

TABLE 7. **Hyperparameters of machine learning classifiers.**

Classifier	Range for GreadSearchCv	optimal Hyper-parameters
SVM	C: logspace of (-2, 10, 10), gamma: logspace of (-9, 3,10)	C=44216.28, gamma=1000, RBF kernel
Random Forest	Max depth: [4,5,6,7,8], Max Features: ['auto', 'sqrt', 'log2'], Estimators: [10,30,40,50,100,200,500], Criterion: ['gini', 'entropy']	Criterion: entropy, max depth: 8, max features: log2, estimators: 42
Decision Tree	Max leaf nodes: [2, 100], Min sample split: [2,4,6,8]	Max leaf nodes: 95, min sample split: 6
K-Nearest Neighbors	K: [1, 20]	K=7

TABLE 8. **Results of machine learning classifiers.**

classifier	Accuracy (%)	AUC (%)
SVM with AlexNet	92.37	93
SVM with VGG	90.44	91
SVM with Proposed	95.23	96
RF with AlexNet	89.30	90
RF with VGG	83.23	86
RF with Proposed	94.66	95
KNN with AlexNet	86.98	88
KNN with VGG	83.58	85
KNN with Proposed	92.12	93
DT with AlexNet	72.01	75
DT with VGG	71.98	73
DT with Proposed	85.22	86

95 and a minimum sample split of 6, while the Random Forest classifier utilized an entropy criterion, a maximum depth of 8, log2 maximum features, and 42 estimators. Similarly, KNN was trained with a k value set to 7. Initially, our experimentation focused on evaluating classifier perforance of each individual models AlexNet and VGG with Involutional Nets and residual layers, yielding AUC scores of 0.90, 0.88, 0.75, and 0.93 for Random Forest, KNN, Decision Tree, and SVM for AlexNet with INNs, respectively. And for VGG with residual layers an AUC scores of 0.86, 0.85, 0.73, and 0.91 for Random Forest, KNN, Decision Tree, and SVM respectively was achieved. Subsequently, upon concatenating both architectures and incorporating featured data with ML classifiers, we observed notable improvements, with AUC scores of 0.96, 0.86, 0.93, and 0.95, for SVM, DT, KNN, and RF classifiers respectively. This underscored the efficacy of concatenating different architectures in enhancing classifier performance. These results underscored the pivotal role of involutional neural networks, residual layers and integration of different modes in enhancing classifier robustness and overall predictive accuracy. By systematically evaluating the impact of various methodologies, our study provides valuable insights into effective strategies for optimizing classifier performance in complex data domains, ultimately advancing the state-of-the-art in machine learning applications.

E. COMPARATIVE ANALYSIS OF TIME COMPLEXITY

In this section, we investigated the impact of the number of parameters on the time complexity of Convolutional Neural Networks (CNNs) and involutional neural networks in image

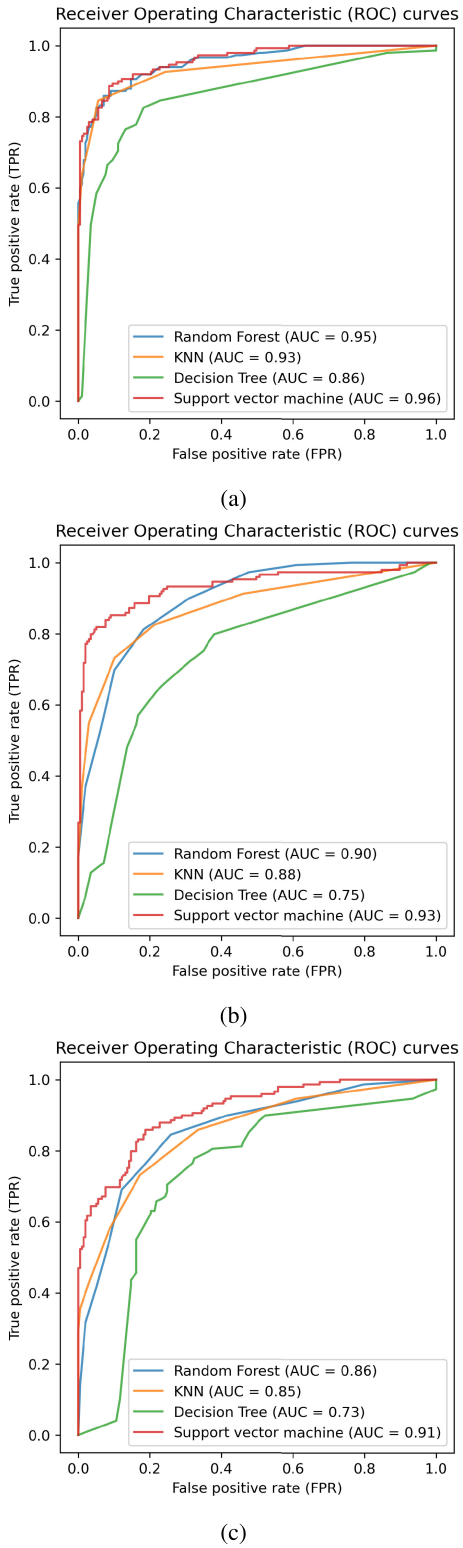


FIGURE 8. ROC curves for four different classification algorithms SVM, KNN, RF, and DT with AlexNet-inspired model(b), VGG-inspired model (c) and Proposed model (a).

classification tasks. We started by training a CNN model on a dataset of 224 by 224 MRI images using individual architectures of AlexNet and VGG models that is based on Covolutional Neural Networks and record the training time. we then asses the results of concatenated AlexNet and VGG Models that is based on Convolutional Network. As shown in Table 9 our results showed that the time it took to train concatenated model was higher compared to individual models and Alex model took less time to train compared to individual and concatenated models, it took 79 sec/epoch to train the model with 46763396 total parameter count. The increase in training time can be attributed to the increase in the number of computations required to process the larger number of parameters. Therefore, it is crucial to carefully consider the trade-off between training time and model accuracy when deciding on the appropriate CNN model. Next, we investigated the impact of increasing the number of parameters on the time complexity of our proposed concatenated Involutional Neural Network model with residual layers for cervical spine disease classification. We compared three different versions of the model, each with different number of layers and parameters. The first model (AlexNet with INNs) had 25759878 parameters. The second model (VGG with residual layers) had 103870660 parameters. The third model (concatinated AlexNet and VGG model) had 7027516 parameters, which was the proposed model. As expected, we observed an increase in the time required to train the models as the number of parameter increased. The first model (AlexNet based INN) took 57 sec/epoch to train, while the second (VGG with Residual layers) and third model (concatenated) took 181 sec/epoch and 21 sec/epoch, respectively. This increase in training time can be attributed to the increase in the number of computations required to process the larger number of parameters. However, it is important to note that while increasing the number of layers and parameters may improve the accuracy of the model, it may not necessarily result in a proportional increase in performance.

TABLE 9. Impact of model parameters on CNN and INN training time and accuracy.

Model	Accuracy (%)	Parameter	Time (sec/epoch) CPU
AlexNet	90.11	46763396	79
VGG	89.09	138357544	201
Concatenated (AlexNet and VGG)	91.75	133130212	190
AlexNet based on Involution	94.91	25759878	57
VGG with Residual layers	92.36	103870660	181
Proposed	98.73	7027516	21

V. DISCUSSION

This paper introduces an innovative approach to cervical spine disease classification through the integration of Involutional Neural Networks (InvNets) with the concatenated AlexNet and VGG architectures with residual layers.

Our research addresses a four-class cervical spine disease classification problem, with the objective of distinguishing between various cervical spine disease types based on MRI data. A comparative analysis of InvNets and traditional Machine Learning classifiers is presented, employing diverse evaluation parameters. Our results indicate that the proposed model (Inv-AlxVGGNets) surpasses conventional machine learning classifiers in terms of accuracy, precision, recall, and other values. This enhanced performance is attributed to the distinctive features of InvNets. In contrast to conventional CNNs that utilize spatial-agnostic and channel-specific convolution kernels, InvNets employ location-specific and channel-agnostic involution kernels. This design enables the network to adapt to varied visual patterns across spatial locations, augmenting its capacity to capture intricate features in medical images. The evaluation outcomes of Inv-AlxVGGNets demonstrate an impressive accuracy rate of 98.73%. This noteworthy accuracy, coupled with a significantly reduced parameter count, underscores the efficacy of InvNets for tasks in medical image analysis. Particularly in resource-constrained environments, InvNets emerge as a promising solution for accurate cervical spine disease classification. Comparative analyses with alternative machine learning methods underscore the superiority of our proposed model. Decision Trees, KNN, and SVM methods exhibit lower accuracy rates and potential overfitting concerns. In contrast, our hierarchical approach, amalgamating concatenated AlexNet and VGG architecture with InvNets, achieves a balanced trade-off between accuracy and computational efficiency.

Our research findings hold significant practical implications, particularly for regions with fewer doctors and specialists. By leveraging advanced deep learning techniques, our methodology for cervical spine disease classification offers several potential benefits for improving healthcare access and delivery in underserved areas. Firstly, our methodology can serve as a valuable tool for expanding access to specialized healthcare services in regions where there is a scarcity of doctors and specialists. By automating the classification of cervical spine diseases, healthcare facilities in underserved areas can provide timely diagnoses without relying solely on the availability of specialized professionals. This can help bridge the gap in healthcare access between urban and rural areas, ensuring that patients in remote regions receive the care they need. In addition to expanding access to specialized care, our methodology can also optimize workflow efficiency and empower primary care providers in underserved areas. With limited availability of specialists, primary care providers often bear the responsibility of diagnosing and managing a wide range of health conditions, including cervical spine diseases. Our methodology can provide them with a reliable tool for cervical spine disease classification, enhancing their diagnostic capabilities and confidence in managing patients with spinal conditions. Moreover, our methodology supports telemedicine initiatives by enabling remote consultations and diagnosis. Integrated into telemedicine platforms, our model allows primary care providers in remote regions to securely transmit MRI images for automated classification by our model. This facilitates timely diagnosis and treatment recommendations without the need for patients to travel long distances to see specialists in person. Furthermore, our methodology can contribute to capacity building and training efforts in underserved areas by exposing local healthcare providers to advanced deep learning techniques. By implementing our model in healthcare facilities, local providers can expand their skill sets in medical image analysis, ultimately strengthening the healthcare workforce and improving patient care over the long term.

However, it is essential to acknowledge several limitations of our study. Firstly, the dataset used was sourced from a single institution, potentially limiting the generalizability of our findings. Future research should aim to validate our approach on larger, more diverse datasets encompassing multiple demographics and clinical settings. Secondly, our study focused exclusively on MRI images for cervical spine disease classification. Expanding our investigation to include other imaging modalities, such as CT scans or X-rays, could provide a more comprehensive understanding of our model's performance across different clinical scenarios. Furthermore, while our model showcased superior performance compared to traditional machine learning classifiers, it is essential to consider the computational resources required for training and inference. Future research should explore optimization techniques to enhance the efficiency and scalability of our model, making it more feasible for deployment in real-world clinical settings.

Despite these considerations, the proposed model's robust performance and reduced computational requirements position it as a promising candidate for practical applications in the medical domain. Future research directions include developing methods for model interpretability, integrating multiple imaging modalities, exploring transfer learning techniques, and conducting prospective studies to evaluate the clinical impact of our model in real-world healthcare settings. Collaborating with healthcare practitioners to assess the model's effectiveness in clinical decision-making and patient outcomes is crucial for validating its utility and facilitating its integration into routine clinical workflows.

VI. CONCLUSION

In this study, we have presented an innovative methodology, Inv-AlxVGGNets, for precise classification of cervical spine diseases using MRI images. By leveraging concatenated pretrained architectures AlexNet and VGG, augmented by involutional neural networks and residual layers, we have addressed the limitations of traditional convolutional neural networks (CNNs) and pretrained models in this domain. Our study demonstrates superior performance of Inv-AlxVGGNets in terms of accuracy, precision, recall, and AUC ROC values, achieving an impressive 98.73% accuracy on the testing set and 99.78% on the training set. Importantly, Inv-AlxVGGNets requires significantly fewer parameters compared to conventional CNNs, making it more efficient and resource-friendly, especially in resource-constrained settings. These promising results highlight the potential of Inv-AlxVGGNets as a valuable tool for precise cervical spine disease classification, with implications for enhancing patient care in various healthcare settings.

DECLARATIONS

- Conflict of interest
  On behalf of all authors, the corresponding author states that there is no conflict of interest.
- Consent for publication
  The authors declare that they are in agreement with this submission and for the paper to be published if accepted.
- Funding
  The authors have no funding to report
- Data Availability
  data will be available upon reasonable request

REFERENCES

[1] M. Qasheesh, M. A. Shaphe, A. Iqbal, and A. H. Alghadir, "Association of psychological variants with functional outcomes among people with spinal cord injury," *Sci. Rep.*, vol. 11, no. 1, p. 20325, Oct. 2021.

[2] R. Thietje, M. H. Pouw, A. P. Schulz, B. Kienast, and S. Hirschfeld, "Mortality in patients with traumatic spinal cord injury: Descriptive analysis of 62 deceased subjects," *J. Spinal Cord Med.*, vol. 34, no. 5, pp. 482–487, Sep. 2011.

[3] A. Ackery, C. Tator, and A. Krassioukov, "A global perspective on spinal cord injury epidemiology," *J. Neurotrauma*, vol. 21, no. 10, pp. 1355–1370, Oct. 2004.

[4] M. S. Ghauri, A. J. Reddy, N. Tak, E. A. Tabaie, A. Ramnot, P. R. Esfahani, N. Nawathey, and J. Siddiqi, "Utilizing deep learning for X-ray imaging: Detecting and classifying degenerative spinal conditions," *Cureus*, vol. 15, no. 7, p. 41582, 2023.

[5] I. Calder, "Cervical spine disease and anaesthesia," *Anaesthesia Intensive Care Med.*, vol. 9, no. 6, pp. 227–230, Jun. 2008.

[6] M. W. Kim, C.-N. Kang, and S. H. Choi, "Update of the natural history, pathophysiology, and treatment strategies of degenerative cervical myelopathy: A narrative review," *Asian Spine J.*, vol. 17, no. 1, pp. 213–221, Feb. 2023.

[7] I. Calder, J. Calder, and H. A. Crockard, "Difficult direct laryngoscopy in patients with cervical spine disease," *Anaesthesia*, vol. 50, no. 9, pp. 756–763, Sep. 1995.

[8] S. M. Naguib, H. M. Hamza, K. M. Hosny, M. K. Saleh, and M. A. Kassem, "Classification of cervical spine fracture and dislocation using refined pre-trained deep model and saliency map," *Diagnostics*, vol. 13, no. 7, p. 1273, Mar. 2023.

[9] H. Li, Y. Chen, Q. Wang, X. Gong, Y. Lei, J. Tian, and X. Gao, "Convolutional neural network-based automatic cervical vertebral maturation classification method," *Dentomaxillofacial Radiol.*, vol. 51, no. 6, Sep. 2022, Art. no. 20220070.

[10] J. Park, J. Yang, S. Park, and J. Kim, "Deep learning-based approaches for classifying foraminal stenosis using cervical spine radiographs," *Electronics*, vol. 12, no. 1, p. 195, Dec. 2022.

[11] P. Chłąd and M. R. Ogiela, "Deep learning and cloud-based computation for cervical spine fracture detection system," *Electronics*, vol. 12, no. 9, p. 2056, Apr. 2023.

[12] W. Yi, J. Zhao, W. Tang, H. Yin, L. Yu, Y. Wang, and W. Tian, "Deep learning-based high-accuracy detection for lumbar and cervical degenerative disease on T2-weighted MR images," *Eur. Spine J.*, vol. 32, no. 11, pp. 3807–3814, Nov. 2023.

[13] R. Vinoth, S. Subalakshmi, and S. Thamaraichandra, "Pneumonia detection from chest X-ray using AlexNet image classification technique," in *Proc. 7th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2023, pp. 1307–1312.

[14] M. Zhao, L. Zhang, J. Wang, and C. Xia, "Automated classification of cervical image based on deep neural network," in *Proc. IEEE 11th Data Driven Control Learn. Syst. Conf. (DDCLS)*, Aug. 2022, pp. 1–6.

[15] S. Kollem, K. R. Reddy, C. R. Prasad, A. Chakraborty, J. Ajayan, S. Sreejith, S. Bhattacharya, L. M. I. L. Joseph, and R. Janapati, "AlexNet-NDTL: Classification of MRI brain tumor images using modified AlexNet with deep transfer learning and Lipschitz-based data augmentation," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 4, pp. 1306–1322, Jul. 2023.

[16] D. Li, J. Hu, C. Wang, X. Li, Q. She, L. Zhu, T. Zhang, and Q. Chen, "Involution: Inverting the inherence of convolution for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12316–12325.

[17] C. John, J. Sahoo, M. Madhavan, and O. K. Mathew, "Convolutional neural networks: A promising deep learning architecture for biological sequence analysis," *Current Bioinf.*, vol. 18, no. 7, pp. 537–558, Aug. 2023.

[18] V. Mishra and L. Kane, "Self-build deep convolutional neural network architecture using evolutionary algorithms," in *Proc. 4th Int. Conf. Comput. Commun. Technol. (IC3T)*. Cham, Switzerland: Springer, 2023, pp. 463–471.

[19] P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. ul Haq, "Understanding of convolutional neural network (cnn): A review," *Int. J. Robot. Control Syst.*, vol. 2, no. 4, pp. 739–748, 2022.

[20] A. Ther, B. K. Pandit, and A. Banerjee, "VLSI architecture of generalized pooling for hardware acceleration of convolutional neural networks," in *Proc. 11th Int. Symp. Electron. Syst. Devices Comput. (ESDC)*, vol. 1, May 2023, pp. 1–5.

[21] K.-H. Chan, S.-K. Im, and W. Ke, "Multiple classifier for concatenate-designed neural network," *Neural Comput. Appl.*, vol. 34, no. 2, pp. 1359–1372, Jan. 2022.

[22] S. Balaji, B. N. Shivacharan, A. Nissar, and S. V. Bhaskar, "Plant infirmity detection using Vgg-16 convolutional neural network," in *Proc. Int. Conf. Comput. Intell. Sustain. Eng. Solutions (CISES)*, Apr. 2023, pp. 485–491.

[23] C. S. Sannasi and H. Rajaguru, "Deep learning-based breast cancer detection using VGG-NiN architecture," *AIP Conf. Proc.*, vol. 2725, Apr. 2023, Art. no. 020001.

[24] P. H. S. Prasad, N. M. V. S. Daswanth, C. V. S. P. Kumar, N. Yeeramally, V. M. Mohan, and T. Satish, "Detection of lung cancer using VGG-16," in *Proc. 7th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Feb. 2023, pp. 860–865.

[25] N. H. Hider, A. S. B. M. Khairuddin, and E. B. Hanafi, "VGG classification model for lung cancer diagnosis," in *Proc. AIJR*, 2022, pp. 57–62.

[26] M. Zhang, G. Wen, J. Zhong, D. Chen, C. Wang, X. Huang, and S. Zhang, "MLP-like model with convolution complex transformation for auxiliary diagnosis through medical images," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 9, pp. 4385–4396, Sep. 2023.

[27] T. T. Jidney, A. Biswas, N. M. A. Al, I. Hossain, M. J. Alam, S. Talukder, M. Hossain, and M. A. Ullah, "Automl systems for medical imaging," in *Data Driven Approaches on Medical Imaging*. Cham, Switzerland: Springer, 2023, pp. 91–106.

[28] H. Oki, K. Yoshida, and S. Okura, "Image classification with SVM for CMOS sensor generating a vector image," *J. Signal Process.*, vol. 27, no. 4, pp. 97–101, 2023.

[29] I. O. Joshua, M. O. Arowolo, M. O. Adebiyi, O. R. Oluwaseun, and K. A. Gbolagade, "Development of an image processing techniques for vehicle classification using OCR and SVM," in *Proc. Int. Conf. Sci., Eng. Bus. Sustain. Develop. Goals (SEB-SDG)*, vol. 1, Apr. 2023, pp. 1–9.

[30] H. Zhang, "Image classification method based on neural network feature extraction," in *Proc. 6th Int. Conf. Electron. Inf. Technol. Comput. Eng.*, Oct. 2022, pp. 1696–1699.

[31] Y. Zhu, J. Wang, and X. Li, "Research on ga-knn image classification algorithm," in *Proc. 4th Int. Conf. Artif. Intell. Adv. Manuf. (AIAM)*, Aug. 2022, pp. 278–282.

[32] S. Shanjida, Md. S. Islam, and M. Mohiuddin, "MRI-image based brain tumor detection and classification using CNN-KNN," in *Proc. IEEE IAS Global Conf. Emerg. Technol. (GlobConET)*, May 2022, pp. 900–905.

[33] J. Valls-Conesa, D. J. Winterauer, N. Kröger-Lui, S. Roth, F. Liu, S. Lüttjohann, R. Harig, and J. Vollertsen, "Random forest microplastic classification using spectral subsamples of FT-IR hyperspectral images," *Anal. Methods*, vol. 15, no. 18, pp. 2226–2233, 2023.

[34] T. Rui, W. Tianyi, X. Yifan, S. Hongji, and T. T. Toe, "Breast image classification based on ResNet and random forest multilayer classifier model," in *Proc. 5th Int. Conf. Bio-Eng. Smart Technol. (BioSMART)*, Jun. 2023, pp. 1–6.
[35] J. Li, J. Shi, J. Chen, Z. Du, and L. Huang, "Self-attention random forest for breast cancer image classification," *Frontiers Oncol.*, vol. 13, Feb. 2023, Art. no. 1043463.

BINIYAM MULUGETA ABUHAYI received the B.S. degree in electrical and computer engineering and the M.Sc. degree in information technology from the University of Gondar, Gondar, Ethiopia, in 2016 and 2022, respectively. He is currently working with the University of Gondar. He is the author of two books, and many articles. His research has been published in a number of reputable international conferences, and major international journals. His research interests include machine learning, natural language processing, image processing, and computer vision. He serves as a reviewer for numerous reputable international journals.

YOHANNES AGEGNEHU BEZABH received the B.S. degree in information technology from Jijiga University and the M.Sc. degree in information technology from the University of Gondar, Gondar, Ethiopia. He is currently working with the University of Gondar. He is the author of multiple articles. His research interests include research computer vision, image processing, machine learning, deep learning, and data science.

ALEKA MELESE AYALEW received the B.Sc. degree in information technology from Adigrat University, Ethiopia, in 2017, and the M.Sc. degree in information technology from the University of Gondar, Ethiopia, in 2021. Since 2018, he has been sharing his knowledge and expertise as a Lecturer with the Department of Information Technology, University of Gondar. With a strong academic background, he has contributed to the field by publishing research in international reputable journals and presenting at conferences. In addition, he actively participates as a reviewer for several international journals. His research interests include wide range of topics including computer vision, image processing and analysis, the IoT, artificial intelligence, machine learning, deep learning, and data science.