JOURNAL OF
EMERGING INVESTIGATORS

Article

The effects of image manipulation on classification of cervical spondylosis X-ray images using deep learning

Mahika Reddy[1], Nelson Nishio[2], Alyssa Chou[3], Jason Nishio[2], Bashar Kabbarah[4], Avni Saxena[5], Larry N. McMahan[6]

[1] Dougherty Valley High School, San Ramon, California
[2] BASIS Independent Silicon Valley Upper School, San José, California
[3] Walton High School, Marietta, Georgia
[4] Monte Vista High School, Danville, California
[5] Irvington High School, Fremont, California
[6] Department of Computer Science and Engineering, Aspiring Students Directed Research Program, Fremont, California

SUMMARY

Cervical spondylosis is a prevalent degenerative disorder of the cervical spine and a leading cause of chronic neck pain and neurological impairment worldwide. Diagnostic imaging, particularly X-rays, remains the first-line tool for detection; however, reliance on manual interpretation contributes to delays, variability, and potential diagnostic error. This study addresses the challenge of accurately diagnosing cervical spondylosis from X-rays by assessing the effects of various preprocessing procedures on the efficacy of classical convolutional neural networks (CNNs) for image classification. To determine whether preprocessing improves diagnostic outcomes, we trained CNNs with transfer learning on the Cervical Spine X-ray Atlas (CXSA) dataset, applied techniques such as wide-area cropping and color enhancement, and measured classification accuracy, precision, and F1-scores. We hypothesized that image processing via the wide cropping of images would significantly increase the performance of our model when tasked with detecting cervical spondylosis in comparison with other typical preprocessing techniques. Our most effective approaches, wide-area cropping and color enhancement, achieved a maximum accuracy of 95.73%, with wide-area cropping rooting out the most false negatives and positives. These findings demonstrate that image preprocessing improves diagnostic accuracy and efficiency, offering potential for clinical translation. More broadly, the methods developed here could be extended to other musculoskeletal conditions, supporting more reliable, individualized treatment plans and advancing the role of deep learning in medical diagnostics.

INTRODUCTION

Medical imaging, especially X-ray, is crucial in diagnosing various spinal conditions, particularly those occurring in the cervical spine, which consists of the C1 through C7 vertebrae (1). Accurate interpretation of X-ray scans is essential for effective treatment planning. However, many spinal abnormalities, including congenital deformities, degenerative conditions like cervical spondylosis, and traumatic injuries, like fractures or dislocations, manifest as progressive changes in bone structure that are difficult to discern visually during initial onset (2). Cervical spondylosis is a degenerative disease that occurs when the cartilage and bones in the neck wear down over time, and it often results in chronic neck pain (2). This degeneration is especially an issue for elderly people, with around half of adults over the age of 40 and 85 % of adults over the age of 60 affected by cervical spondylosis (2). Therefore, there is a great need for studies which focus on detecting this condition. Furthermore, conditions such as herniated discs, osteoarthritis, and spinal stenosis may cause mild misalignments or changes in bone density that are easily overlooked in conventional imaging (3). The cervical spine's complex anatomy, including its role in supporting the skull and enabling a wide range of motion, complicates the interpretation of imaging further. Subtle pathologies in this region can lead to serious outcomes, including chronic pain, neurological deficits, and impaired mobility, if not identified and treated promptly (4). While X-rays are widely used due to their accessibility and cost-effectiveness, their limitations in contrast resolution and inability to capture soft tissue abnormalities highlight the need for advanced diagnostic methods.

The purpose of this study is to utilize classical deep learning frameworks, namely convolutional neural networks (CNNs), in order to streamline the process of identifying cervical spine conditions from X-rays. Deep learning has been used in efforts to automate and improve the accuracy of tasks done by humans; thus, we chose to apply this concept to the biomedical field. This study focuses on image classification and a popular high-performing deep learning framework, CNN. A standard CNN is made up of three primary elements: convolutional layers, pooling layers, and fully connected layers (5). Convolutional layers use learnable filters to scan input images, capturing features, such as edges, textures, and increasingly intricate patterns, as the network progresses (5). Pooling layers, such as those performing max pooling, shrink the spatial size of the data, preserving essential information while reducing the computational load and mitigating the risk of overfitting (5). Finally, fully connected layers integrate the extracted features from earlier stages to generate final predictions (5). Additionally, CNNs can automate feature engineering, eliminating the need to manually extract features to put together a pattern.

https://doi.org/10.59720/25-135

This capability not only optimizes the classification process but streamlines it, allowing for a greater volume of cases to be processed with higher efficiency (6). These components allow CNNs to identify patterns, often performing complex tasks with a higher accuracy than the average human. With this prior knowledge in hand, the purpose of this study was to evaluate the effects of color and image manipulation on the classification of cervical spondylosis X-rays through CNNs.

The CNN model used in this study is one at the forefront of deep learning used in medical image classification, the ResNet50. ResNet50, a model given by Microsoft, consists of 50 layers, with convolutional layers, activation layers, and residual blocks (15). ResNet50 incorporates residual connections that help prevent the vanishing gradient problem, which is a phenomenon where gradients shrink significantly as they backpropagate through the network (7). With this feature, the model can be effectively trained even at considerable depth (7). Its balanced size, computational efficiency, and high compatibility with data augmentation allows the model to generalize well and maintain success in binary classification tasks, like it was tasked with in our study. CNNs have been successfully applied to a wide range of spinal imaging tasks, consistently demonstrating strong diagnostic performance across modalities (8–13). For instance, studies using MRI and CT scans have achieved high accuracy in detecting conditions such as herniated discs, cervical cord lesions, and spinal fractures, often exceeding 90% with models such as ResNet50, VGG19, and MobileNetV2 (8–13). These findings highlight the utility of deep learning in spinal diagnostics but also underscore a key limitation: most prior work has focused on MRI or CT data, which are not always available in resource-limited clinical settings. By contrast, our study advances this field by applying CNNs to cervical spine X-rays, which are more accessible and cost-effective, thereby extending the potential impact of AI-driven diagnostic tools. However, many of these studies focus on different spinal conditions or imaging types, and often rely on standardized toy datasets, reducing their applicability to real-world situations where image quality and conditions can vary. Research specifically using CNNs to analyze cervical spine X-rays—a more accessible and affordable imaging method—is still quite rare.

Our study fills this gap by investigating whether CNNs, specifically the ResNet50 model, can effectively classify cervical spinal conditions using X-ray images. We used a newly released public dataset published in 2024 that has not been studied before, and applied intricate preprocessing techniques to account for the challenges of imperfect real-world images.

We hypothesized that image processing via the wide cropping of images would significantly increase the performance of our model when tasked with detecting cervical spondylosis in comparison with other typical preprocessing techniques. To test this, we trained ResNet50 and VGG19 on both original and modified datasets, applied a variety of preprocessing techniques, and compared performance across accuracy, precision, and F1-scores.

Our findings demonstrate that both wide-area cropping and color enhancement independently improved classification accuracy to 95.73%, with wide-area cropping providing the greatest precision and clinical utility. These results suggest that preprocessing strategies tailored to medical imaging can significantly reduce diagnostic errors such as false positives and false negatives. In turn, this highlights the potential of carefully optimized CNN pipelines to enhance the reliability of automated cervical spondylosis diagnosis, while future work should focus on expanding dataset diversity and exploring hybrid or ensemble modeling strategies to further advance clinical applicability.

RESULTS

We sought to understand the role of image processing techniques and their combinations on improving cervical spondylosis detection. Thus, deep learning models were trained on versions of the dataset after applying different image processing techniques and combinations to experimentally improve performance (**Figure 1**). The dataset consisted of 181 healthy spinal x-rays and 4782 diseased x-rays and was divided into 75% for training purposes and 25% for testing purposes using a fixed random seed for consistent comparison across all processed versions of the dataset (**Figure 2**).

To evaluate the performance before and after applying

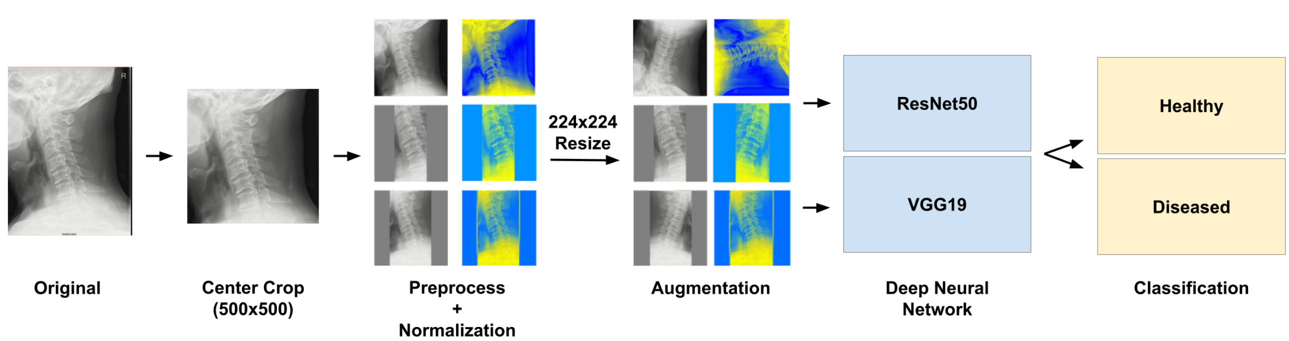

Figure 1: Workflow for cervical spinal X-ray image classification using deep neural networks. The process began with the original X-ray images, which underwent a 500x500 center crop to standardize dimensions. Next, the cropped images were subjected to preprocessing and normalization, using various configurations of cropping, color mapping, and intensity standardization, depending on the specific process. These processed images were then resized to 224x224 pixels to match the input size requirements of the deep neural networks. Augmentation techniques, such as rotation, flipping, and scaling, were applied to enhance the dataset's diversity and improve model robustness. Two pre-trained deep neural networks, ResNet50 and VGG19, were employed for feature extraction and classification, and categorized the images into healthy or diseased classes.

https://doi.org/10.59720/25-135

Classification	Healthy	Diseased
Count	181	4782
Image		

Figure 2: Distribution of dataset images across classification categories. The dataset contains 181 images classified as healthy and 4782 images classified as diseased. Example X-ray images from each category are shown.

image processing, we evaluated the trained models' performances using performance metrics like accuracy, precision, recall, and F1-score (**Table 1**). First, two deep learning CNNs, ResNet50 and VGG19, were trained over 30 epochs on the original dataset with no image processing as a control group (**Figure 3**). Since the ResNet50 model's accuracy was superior to the VGG19 model by approximately 1.5%, ResNet50 was used on the rest of the deep learning model tests. The accuracy of the model without prior image modification was 94.19%, which increased to 95.73% during both instances in which wide cropping and color modification were separately applied (**Table 1**). Color enhancement amplifies contrast and sharpens subtle variations in grayscale radiographs, making pathological features such as abnormal curvatures or density shifts more distinguishable to the network (**Figure 4**). Wide-area cropping, on the other hand, reduces irrelevant background information while ensuring that the cervical region remains in focus, thereby limiting noise and improving the consistency of inputs (**Figure 4**). By applying these preprocessing steps, we aimed to isolate diagnostically relevant features and minimize distractions, enabling the models to learn more robust and clinically meaningful patterns. Utilizing two proportion *z* tests, we compared the accuracy of our model when run on the original dataset in comparison to five, distinctly preprocessed, versions of the dataset: cropped, wide cropped, color enhanced, color enhanced and cropped, and color enhanced and wide cropped. When our data was preprocessed through color enhancement and regular cropping, the model produced an accuracy of 94.76% (p = 0.269); when the data was preprocessed by color enhancement and wide cropping, the model produced an accuracy of 94.92% (p = 0.213). When our dataset went through only regular cropping, the model produced an accuracy of 95.48% (p = 0.0732). Although this was a great improvement from the previous two methods, it was still not statistically significant. When wide cropping and color enhancement were tested individually, their shared accuracy was 95.73%, (p = 0.0406). The improvement in accuracy by the use of these preprocessing methods is statistically significant. Although this showcases how important these two methods are in the study of machine learning, this study particularly engages with the intersection of machine learning and the medical domain; thus, it was important to take into account more than just accuracy. In terms of precision, wide cropping shows better results with 93.09% precision versus 92.55% precision from the color modification, and the same occurs when using F1-score as a metric with scores of 94.24% and 94.11% respectively (**Table 1**).

DISCUSSION

To determine the most effective preprocessing techniques for improving convolutional neural network performance in cervical spondylosis classification, we trained and evaluated two candidate models, ResNet50 and VGG19. For each model, we measured accuracy, precision, and F1-scores across multiple image manipulation strategies. We first conducted trials using the original dataset and modified duplicates, when necessary, with our candidate models, ResNet50 and VGG19, at least one of each. The model with the highest base accuracy is the ResNet50 model at 94.19%, which we optimized through image color manipulation and cropping to a maximum accuracy of 95.73%. In addition to the higher parameter efficiency and deep structure of ResNet50, we attribute the high base accuracy to crucial consistencies in the dataset, such as the same imaging angles and regions. ResNet50 outperforms VGG19 for this task, which has a base accuracy of 92.74%, using the most optimal hyperparameters. Our most effective image preprocessing procedure combinations were color manipulation and cropping with a manually widened area—performed independently on separate, modified dataset copies. Both color manipulation and widened area cropping increase the consistency of data and remove unnecessary visual details that may pose a distraction, allowing the model to extract and evaluate features with higher accuracy. They both achieved accuracies of 95.73%. Cropping with a manually widened area, however, achieved higher precision and F1-scores, so this processing procedure offers the best performance and practicality for clinical use in this study. High accuracy and precision scores indicate that the model minimizes both false positives and false negatives, which is critical in medical diagnostics. Reducing false negatives lowers the risk of missed diagnoses, while reducing false positives prevents unnecessary treatments and patient anxiety. However, when both are applied

https://doi.org/10.59720/25-135

Image Enhancement	Model	Accuracy (%)	Precision (%)	F1-score (%)	P-value
Original	ResNet50	94.19	93.08	93.62	n/a
	VGG19	92.74	92.89	92.81	n/a
Cropped	ResNet50	95.48	92.54	93.99	0.0732
Wide Cropped	ResNet50	95.73	93.09	94.24	0.0406
Colored	ResNet50	95.73	92.55	94.11	0.0406
Colored, Cropped	ResNet50	94.76	92.77	93.73	0.269
Colored, Wide Cropped	ResNet50	94.92	92.52	93.70	0.213

Table 1: Comparison of model performance using different image processing techniques. Statistics of the deep CNNs, ResNet50 and VGG19, when using different image processing techniques and their combinations, including accuracy, precision, recall, and F1-score. We computed the *p*-value using a paired *t*-test, which measures the probability of obtaining the observed difference under the null hypothesis (no image enhancement). The test statistic was derived from the mean difference in accuracy divided by the standard error of the differences across folds.

together, accuracy decreases to 94.92%, which we theorize is because of excessive loss of necessary information due to over-processing, exacerbated by an increase in data ambiguity after data augmentation. This finding supported our hypothesis that image processing is key for cervical spondylosis classification tasks—specifically, wide image cropping is a superior form of image manipulation. This further supports the idea that image cropping not only produces a higher accuracy but also reduces the risk of false positives and negatives, which is particularly important as missing a cervical spondylosis diagnosis could delay treatment and increase the risk of neurological complications. Preprocessing with color manipulation and cropping, individually, as proven with aforementioned results, can be used to increase the accuracy of cervical spondylosis diagnosis and add to the researched advantages of incorporating deep learning into medical procedures.

The Cervical Spine X-ray Atlas (CSXA) dataset used in this study contained a severe imbalance between abnormal and normal spinal X-rays, with 4782 symptomatic X-rays versus only 181 asymptomatic X-rays (1). In order to combat this issue, disparity methods, such as image augmentation and oversampling, were used. However, these techniques were only able to bridge the gap to a certain extent. Augmentation generates altered versions of existing images, and oversampling reuses the same limited samples, neither of which introduces new biological variability. As a result, the model remains vulnerable to overfitting and struggles to generalize to unseen asymptomatic cases. As the majority of research done in the field of spinal injuries utilizes privately sourced data, a provision not readily available to us, this uneven dataset limited the possible accuracy of our model. In the future, to increase said accuracy, it would be optimal to find a larger and more diverse dataset, which would allow our model to more accurately identify asymptomatic spinal X-rays.

This study utilized only two classical deep learning models. As a result, when tasked with identifying a spinal condition from X-rays, the product included a lengthy runtime and moderate accuracy. While this study aims to achieve accurate results regarding the classification of spines, it also aims to streamline the process, allowing for quicker results that could be applied as medical diagnosis in real-time. Using a single classical model limits our ability to achieve this goal, as all data is processed through this model bit by bit. Working towards this goal in the future could include utilizing a larger variety of CNN architectures, such as EfficientNet, which is designed for optimal performance with fewer parameters and computations, making it both accurate and efficient for training and inference. Combined with VGG's straightforward architecture, which we utilized in this study, we could capture detailed hierarchical features, which can enhance accuracy in specific scenarios. By combining these models, either through ensemble methods or hybrid approaches, we could achieve quicker and more accurate results. Another way results could be optimized in the future is through the use of a quantum machine learning model. As stated earlier, the classical model utilized in this study processed all data bit by bit. Quantum machine learning models are capable of parallel processing through the use of qubits, which can result in multiple states simultaneously, rather than the classical "bit by bit" processing that only allows bits to be one state at a time. This capability can lead to faster computation on complex tasks such as the one in this study. Thus, an adaptation of this study in the future could include the testing of a quantum machine learning model tasked with the same duty of identifying cervical spinal abnormalities. From what we know about quantum versus classical machine learning, this new

https://doi.org/10.59720/25-135

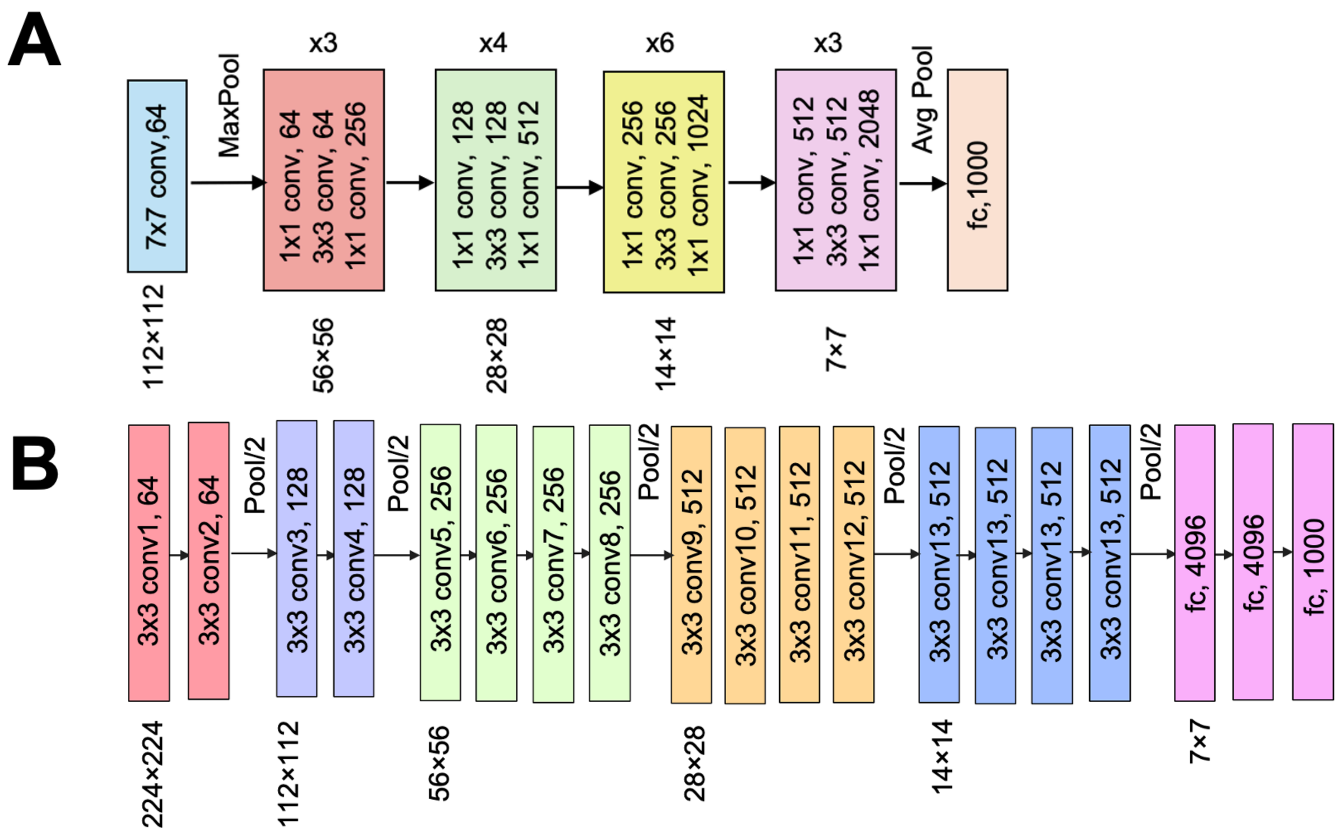

Figure 3: Overview of deep learning architectures used in this study. (A) ResNet50: A residual network featuring convolutional layers with skip connections to mitigate vanishing gradients, structured into blocks of increasing depth and complexity. (B) VGG19: A deep convolutional neural network characterized by sequential 3×3 convolutional layers and fully connected layers, designed for high-resolution feature extraction and classification. Both architectures are evaluated for their performance in cervical spinal X-ray classification tasks.

quantum model would most likely serve to improve not only accuracy but also streamline the identification process as a whole.

In summary, this study demonstrates that targeted preprocessing strategies, particularly wide-area cropping, can significantly enhance the diagnostic accuracy and clinical applicability of CNN-based models for cervical spondylosis classification. These results align with and extend current knowledge on the role of preprocessing in optimizing deep learning pipelines, highlighting that even modest manipulations of medical imaging data can yield substantial improvements in model performance. By reducing false positives and negatives, wide-area cropping contributes not only to improved accuracy but also to greater clinical reliability, an essential factor in preventing diagnostic delays and mitigating risks of neurological complications. Although limitations such as dataset imbalance and reliance on classical CNN architectures constrain the present study, our findings underscore the importance of methodological refinement in advancing the field of computer-aided diagnosis. This work supports the integration of optimized CNN models into medical diagnostics as a means to enhance precision, efficiency, and patient outcomes, contributing to the broader objective of leveraging artificial intelligence to improve human health.

MATERIALS AND METHODS

Data Preparation

The dataset used in this study is the CSXA, made publicly available by the Dongzhimen Hospital of Beijing University of Chinese Medicine (1). The dataset comprises 4,963 spinal X-ray images in PNG format, featuring 4,782 scans from patients showing symptoms of cervical pain or cervical spondylosis, and 181 scans from individuals without symptoms. Before using deep learning CNNs to detect cervical spondylosis, the Cervical Spine X-ray Atlas dataset used was cleaned up to ensure smooth processing. The dataset ensures that key demographic information—identity, gender, and age—is embedded directly into the filename, simplifying data management and analysis (1). The dataset comes with a Microsoft Excel Spreadsheet that includes 5000 rows with distinct sequence numbers, despite only having 4963 patients. After analysis, we excluded the 37 datapoints lacking images. We also found a number of datapoints containing mismatched age labels. Since the focus of this study does not concern patient age, the age label on the spreadsheet was corrected to match the image file name, regardless of its accuracy. The cervical spine dataset contains images of both healthy spines and those with cervical spondylosis. However, it predominantly consists of symptomatic cases, so data augmentation was used not only to balance the dataset but also to expand its size. This approach introduces real-world variability, helping to reduce overfitting and improve model generalization. For this study, the data were divided

https://doi.org/10.59720/25-135

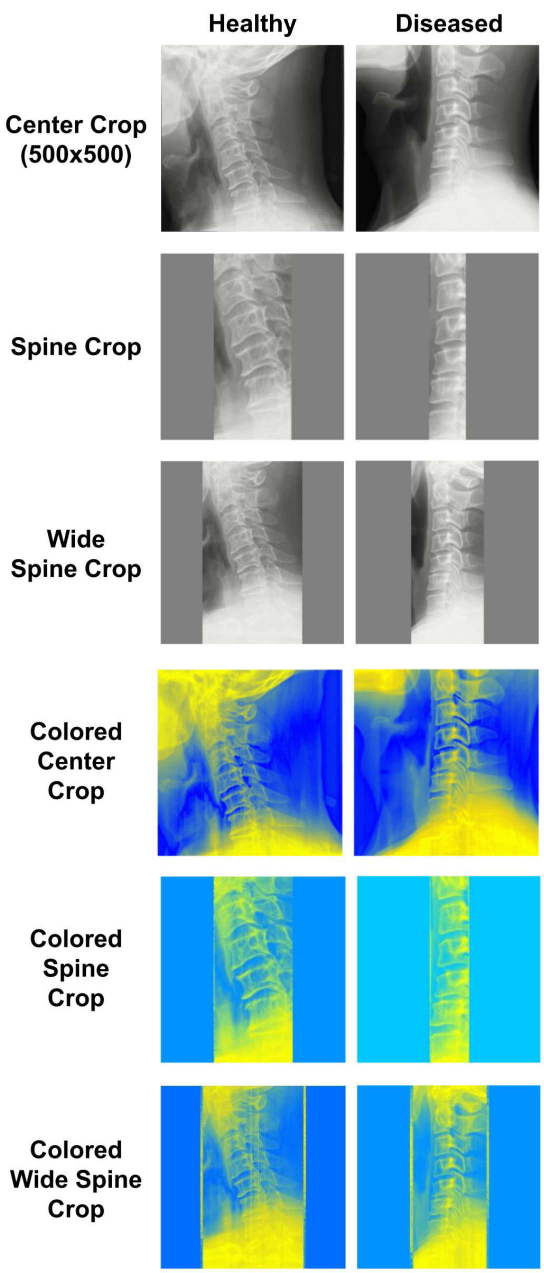

Figure 4: Overview of preprocessing methods applied to cervical spinal X-ray images for Convolutional Neural Network (CNN) input. The dataset underwent various image processing techniques, including cropping, segmentation, and color mapping, to enhance model accuracy. Preprocessing focused on standardizing anatomical regions and balancing pixel intensity variations between images. These techniques generated six dataset versions, incorporating combinations of grayscale manipulation, segmentation cropping, and augmented color mapping for evaluation.

into training and testing sets, with 75% (3,722 images) used for training and the remaining 25% (1,241 images) reserved for testing.

Preprocessing

The Cervical Spine X-ray Atlas (CSXA) dataset underwent basic modification and augmentation prior to training to ensure that files were processable by the deep learning models and to minimize class imbalance **(Figure 2)**. For individual images, inconsistent image dimensions, footnotes, labels, as well as presumably unintentional borders from the source X-rays were resolved by first cropping to 500 × 500 pixels, while maintaining the center of the original file. The images were then resized to 224 × 224 pixels, the optimal input size for most CNN architectures. Additionally, the alpha channel was removed because of its redundancy, given all pixels are at full opacity, reducing pixels to three channels. Another issue within the dataset was an insufficiency of asymptomatic patient data, originally comprising only 3.60% of the training set and 3.79% of the testing set. Using synthetic minority oversampling technique (SMOTE), randomly selected image combinations were generated from the smaller class, resulting in 3454 more images to append to the aforementioned class within the training set and match the class size of symptomatic images. Prior to oversampling, we also cleaned the data by removing images that did not have a corresponding label and vice versa.

To test our hypothesis that the image processing techniques—especially the area cropping—are a key factor for model performance, combinations of image enhancements were subsequently applied and tested for effectiveness through model accuracy. The first procedure incorporated the preprocessing procedure by another study, which targets instances in the dataset where the image colors are inverted, presenting a white background and dark foreground, instead of the originally more frequent dark background and white foreground (14). For each image, an inverted and equalized version was created by manipulating the grayscale pixel values before stacking the images in different orders to create two three-channel new image files. In both versions, the original grayscale values were in the first channel, while the equalized and inverted values were swapped between the second and third channels. This procedure further balances classes by reducing the large value differences between white-background and dark-background images by producing those of intermediate values. The second procedure uses available JavaScript Object Notation (JSON) data from the source dataset, including pixel coordinates of each spinal segment, to crop out unnecessary objects in images, such as parts of the skull. The maximum and minimum x and y coordinates from each image determine a rectangular cropping area, which has been enlarged up to 200 pixels vertically and 100 pixels horizontally to contain the entire spine. Two dataset duplicates, cropped accordingly, were produced, one in which the cropping area was enlarged. In sum, a total of five versions of the dataset were used to assess the effectiveness of the proposed image processing techniques **(Figure 3)**.

To enhance the model's reliability and adaptability, we employed data augmentation techniques during training, including random horizontal flips, vertical flips, and rotations. The images are first given random horizontal and vertical flips, with a probability of 50% each. For the random rotations, we applied two sequential random rotations to the input images. The first rotation was randomly selected from a range of -20 to +20 degrees, while the second rotation was randomly chosen from a range of -60 to +60 degrees. Using two sequential rotations rather than one allows for angles closer to zero degrees to have higher probability, rather than all angles from -80 to +80 degrees having uniform probability distributions. This is especially important because we expect that most spinal x-ray will be oriented closer to a vertically straight angle, and using this combination of rotations allows us to handle outlier images with the spine oriented in a wider angle. Using augmentation ultimately introduced a wider range of orientations for the images, potentially reducing overfitting and improving the model's ability to handle real-world variations in the orientation of the cervical spine in unseen data.

Finally, to standardize image data for model training, the mean and standard deviation of the dataset were calculated using an online approach. The batches of images were iteratively processed and the sum of pixel values and squared pixel values were calculated. Using these accumulated values, the mean and standard deviation for each color channel were then computed. This process allowed for efficient calculation without requiring the entire dataset to be loaded into memory. The images from each version of the dataset were then normalized to these values to ensure consistent input scaling, improving convergence during training and optimizing the model's learning process by reducing skewness.

Convolutional Neural Networks

The CNN models are designed to learn hierarchical features from the spinal X-ray images through a series of convolutional and pooling layers. The VGG19 and ResNet50 architectures were tested, leveraging those convolutional layers to extract spatial patterns from the input images (**Figure 4**).

Transfer learning is a method that leverages a model trained on one task as a starting point for learning a different, but related, task. In this study, transfer learning is used on both the ResNet50 and VGG19 model, which are pre-trained on ImageNet, a dataset consisting of over 14 million labeled images across thousands of categories (15).

Using transfer learning, the pre-trained ResNet50 model trained on a large dataset of cervical spine X-ray images, which included images with and without spinal conditions. To enhance the reliability of the analysis, the dataset was preprocessed to isolate the spine region in each X-ray; image augmentation techniques were applied to balance the severe difference in representation of images with and without spinal abnormalities.

The images are processed through the model's layers, with training conducted over 30 epochs. The initial learning rate is set to 0.0003 and is reduced by half every 5 epochs to optimize performance. The CrossEntropyLoss function is used as the criterion, and the models are optimized using the Adam optimizer.

Metrics

Model performance was evaluated using accuracy, recall, precision, and F1 scores. Accuracy refers to how often the model was correct across all diagnoses, while precision focuses directly on how many of the model's positive predictions were correct. Recall refers to how many of the positive cases the model was able to correctly detect, which functions to call out missing positives. A lot of value was placed into the F1 score which serves to balance precision and recall, taking into account the model's ability to make accurate positive predictions but simultaneously not miss too many real ones.

To determine whether improvements in model accuracy were statistically significant, we used a two-proportion z-test. This test compares the proportions of correct classifications between two models, typically the baseline model using the original dataset and a modified model using specific preprocessing techniques. By treating the number of correct predictions as "successes" out of the total predictions, we applied the z-test to evaluate whether the observed differences in accuracy could have occurred by chance. The resulting p-values reflect the probability that the accuracy improvements were due to random variation rather than the preprocessing methods themselves.

ACKNOWLEDGMENTS

We would like to thank the Aspiring Scholars Directed Research Program for the opportunity to conduct this research.

Received: October 31, 2024
Accepted: June 7, 2025
Published: January 8, 2026

REFERENCES

1. Ran, Yu, *et al.* "A High-Quality Dataset Featuring Classified and Annotated Cervical Spine X-ray Atlas." *Scientific Data*, vol. 11, no. 1, 13 June 2024. https://doi.org/10.1038/s41597-024-03383-0.
2. Margetis, Konstantinos and Prasanna Tadi. "Cervical Spondylosis." *National Library of Medicine*, 2 Aug. 2025. https://www.ncbi.nlm.nih.gov/books/NBK551557/. Accessed 17 Dec. 2024.
3. "Understanding the Benefits and Limitations of a Spine X-ray." *Atlas Brain & Spine*. https://atlasbrainspine.com/spine-xray-in-upland-ca/can-a-x-ray-see-a-herniated-disc-understanding-the-benefits-and-limitations-of-a-spine-xray/. Accessed 25 Sep 2024.
4. Swartz, Erik E. *et al.* "Cervical spine functional anatomy and the biomechanics of injury due to compressive loading." *Journal of athletic training*, vol. 40, no.3, 10 Jan. 2025.
5. Alzubaidi, Laith, *et al.* "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." *Journal of Big Data*, vol. 8, no. 1, 31 Mar. 2021. https://doi.org/10.1186/s40537-021-00444-8.
6. Shaheen, Fatma, *et al.* "Impact of Automatic Feature Extraction in Deep Learning Architecture." *IEEE Xplore*, 1 Nov. 2016, pp. 1-8. https://doi.org/10.1109/DICTA.2016.7797053.
7. Wilkerson, Matthew, *et al.* "Benchmarking ResNet50 for Image Classification on Diverse Hardware Platforms." *The ITEA Journal of Test and Evaluation*, vol. 45, no. 3, June 2024. https://doi.org/10.61278/itea.45.3.1008.
8. Mohanty, Ricky, *et al.* "NAMSTCD: A Novel Augmented Model for Spinal Cord Segmentation and Tumor Classification Using Deep Nets." *Diagnostics*, vol. 13, no. 8, Apr. 2023. https://doi.org/10.3390/diagnostics13081417.
9. Ma, Shaolong, *et al.* "Faster RCNN-Based Detection of Cervical Spinal Cord Injury and Disc Degeneration." *Journal of Applied Clinical Medical Physics*, vol. 21, no. 9, 1 Sept. 2020, pp. 235–243. https://doi.org/10.1002/acm2.13001.

https://doi.org/10.59720/25-135

10. Merali, Zain, *et al*. "A Deep Learning Model for Detection of Cervical Spinal Cord Compression in MRI Scans." *Scientific Reports*, vol. 11, no. 10473, 18 May 2021, https://doi.org/10.1038/s41598-021-89848-3.
11. Bezabh, Yohannes Agegnehu , *et al*. "Classification of Cervical Spine Disease Using Convolutional Neural Network." *Multimedia Tools and Applications*, vol. 83, 27 Mar. 2024, pp. 88963–88979. https://doi.org/10.1007/s11042-024-18970-x.
12. Lee, Gun Woo, Shin, Hyunkwang, and Chang, Min Cheol "Deep Learning Algorithm to Evaluate Cervical Spondylotic Myelopathy Using Lateral Cervical Spine Radiograph." *BMC Neurology*, vol. 22, 20 Apr. 2022. https://doi.org/10.1186/s12883-022-02670-w.
13. Paul, Showmick Guha, *et al*. "A Real-Time Deep Learning Approach for Classifying Cervical Spine Fractures." *Healthcare Analytics*, vol. 4, 2023, no. 100265, 2023. https://doi.org/10.1016/j.health.2023.100265.
14. Ruchi, *et al*. "Lumbar Spine Disease Detection: Enhanced CNN Model with Improved Classification Accuracy." *IEEE Access*, vol. 11, 2023, pp. 141889–141901. https://doi.org/10.1109/access.2023.3342064.
15. Liu, Shanshan, *et al*. "A Machine Learning Based Quantification System for Automated Diagnosis of Lumbar Spondylolisthesis on Spinal X-rays." *Heliyon*, vol. 10, no. 17, 2024. https://doi.org/10.1016/j.heliyon.2024.e37418.
16. Fard, Ali Pourramezan, *et al*. "Sagittal Cervical Spine Landmark Point Detection in X-ray Using Deep Convolutional Neural Networks." *IEEE Access*, vol. 10, 2022, pp. 59413–59427. https://doi.org/10.1109/access.2022.3180028.
17. He, Kaiming, *et al*. "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Appendix

All preprocessing and deep learning code for this project can be found in github.com/nknishio/Cervical-Spondylosis.