ABSTRACT
Deep learning algorithms have shown significant potential for early disease identification and prevention, becoming increasingly popular in recent years. Cervical spine injuries require immediate diagnosis to ensure adequate treatment. Numerous methods have been projected; but, they repeatedly lack accurateness in detecting minor fractures and may surge the false positive rate due to the limitations of single network-based classifiers. Moreover, the shortage of publicly obtainable spine data makes automated cervical spine fracture detection ominously more challenging to attain. To address these issues, we suggest a new and vigorous method called the Spine-Ensemble Deep Learning Network (Spine-EDLNet). Ensemble learning subsidises a vital role in extracting powerful features from image data. Our model harnesses the strengths of three pretrained deep learning networks: EfficientNetV2, InceptionNetV3, and VGG16, united with a majority voting mechanism. Tailored layers are combined into each model to improve fracture classification. Besides, comprehensive data augmentation and preprocessing practices are applied to the dataset before training, successfully overcoming the dataset obtainability challenge. Investigational results validate that Spine-EDLNet outperforms previous models, attaining a maximum accuracy of 99.6%. This methodology purposes to optimize diagnostic accuracy, robustness, and generalization.

Keywords: Ensemble Method, Cervical Spine Fractures, Deep Learning, Transfer Learning, Binary Classification

1. INTRODUCTION
Globally, Eight million spinal fractures occur annually, with a notable prevalence in the cervical spine, making it one of the most frequently impacted regions [1]. In North America, injuries to the cervical spine cause more than 1 million admissions to hospitals for emergency treatment annually [2]. Seven stacked vertebrae, or stacks of bones, make up the cervical portion of the spine and are designated C1, C2, C3, C4, C5, C6, and C7. Injuries to the Cervical spine affect the neck, discs located amongst vertebrae, the joint muscles, and the ligaments [3]. Fractures or dislocations of the cervical spine can happen as a result of intense strain from car crashes, trips and falls, and sports-related injuries [4].

The elderly generation has a much higher rate of spinal fractures caused by osteoporosis and deteriorating health conditions [5]. Cervical spine fractures can have serious neurological effects. These injuries have the potential to cause instability and compression of the spinal cord underneath,
thereby worsening the potential complications [6-8]. A significant rate of illness and mortality has been associated to cervical spine injuries [9]. As a result, primary detection and treatment of cervical spine fractures (CSFx) is essential to prevent additional impairments [10, 11].

Deep learning techniques have become essential in healthcare for locating fractured and distressed regions in any part of the body. The efficacy of deep learning algorithms is evident in image classification, object detection, pattern recognition, and medical image analysis. Diverse methodologies have been suggested to identify cervical spinal fractures. Certain algorithms make use of machine learning approaches, which conventionally depend on time-intensive manual feature extraction. However, a variety of deep learning techniques have emerged with the advances in convolutional neural networks. These CNN-based techniques greatly increase the accuracy and effectiveness of fracture identification as compared to conventional methods through automatically extracting information from medical image data. Cervical spine fracture classification in images is an intricate process that usually requires specialized medical expertise. Moreover, the unavailability of diverse public datasets for training deep learning models presents a significant challenge.

Despite extensive research on detecting cervical spine fractures, very little attention has been given to ensemble approaches. To the best of our knowledge our proposed work is novel, which performs data augmentation techniques along with ensemble method for the effective detection. The main offerings of this research is as follows:

**Our Proposed Model’s contribution**

- The paper introduced an improved deep learning model called Spine-Ensemble Deep Learning Network (Spine-EDLNet), which utilizes transfer learning and ensemble learning approaches to identify cervical spine fractures from CT scans.
- This work implemented a cutting-edge cervical fracture CT scans dataset sourced from the RSNA Cervical Spine Fracture Detection Challenge 2022.
- Our Model employed ensemble learning by combining three TL-based deep learning models: VGG16, EfficientNetV2, and InceptionNetV3, along with our customized layers. This approach integrated the strengths of multiple models, resulting in improved performance in detecting cervical spine fracture due to enriched feature extraction.
- Ensemble learning helped overcome the deficiencies of individual models, leading to more reliable results. By aggregating the predictions of multiple models through majority voting scheme, our proposed model achieved enhanced accuracy and robustness in classifying cervical spinal fractures.
- Our suggested approach has been validated by comprehensive experiments that proved the superiority of Spine-EDLNet over existing methods.

The remaining paper is organized as: section 2 describes related studies, section 3 demonstrates methodology, section 4 explains the results, and section 5 describes the conclusion and future work.

**2. LITERATURE REVIEW**

Artificial intelligence (AI) has been widely applied in healthcare in recent years. It has frolicked a vigorous role in saving uncountable lives by, for example, diagnosing cancer in its initial stages or distinguishing benign and malignant tumors [12]. The consequences of AI are sometimes more precise compared to those of trained professionals, henceforth assuring an immediate investigation. AI offers quick detection, and in certain circumstances, its deductions are far more exact compared to the findings of professionals [12].

Numerous inquiries have been done to decide how well AI performs in identifying fractures [13]. It has been used to find fractures in many anatomical regions such as the humerus, hip, distal radius,
hand, wrist, and ankle using radiographs [14-21]. Moreover, it has also been applied in diagnosing thoracic and lumbar spine injuries with dual-energy x-ray absorptiometry (DXA) [22]. Also, AI has also been successful in sensing calcaneal fractures and fractures in the vertebral bodies of the thoracic and lumbar regions through computed tomography scans [23-26].

In emergency conditions, plain cervical spine capturing is a crucial and often-used method for diagnosing cervical spinal fractures. Anteroposterior and lateral cervical spine radiographs are frequently collected in severely injured patients, facilitating quicker diagnosis and management of numerous serious ailments [27, 28]. However, a complementary CT scan is recommended since plain scans are unable to accurately observe the cervical spine following significant damage [29, 30]. Computed tomography scan is one of the most advantageous and commonly used imaging techniques for detecting fractures [31].

Computed tomography rather than traditional radiography is constantly employed for imaging diagnosis of adult spine fractures [32] nowadays. Nevertheless despite improvements in artificial machine learning (ML), very few techniques are available for identifying fractures in CT images of the cervical spine [31]. In [2], authors employed a three-dimensional ResNet-101 Deep CNN [33], trained on 222 fractured and 990 healthy instances. This method performed well in terms of AURPC and AUROC metrics on the validation dataset which contained 37 fractured and 98 healthy instances. The model attained an AUPRC of 0.52 and AUROC of 0.87 in terms of image-level evaluation. The AUPRC and AUROC were both 0.82 when assessed at the case level. In [31] authors proposed a ResNet-50[33] combined with a Bidirectional LSTM [34] Machine learning model, encompassing the properties of a Deep neural network, for automated categorization of cervical spine fractures. The network is trained and validated employing a labelled dataset of 3,666 computed tomography (CT) images containing 729 positive and 2,937 negative samples. The model showed a categorization accuracy of 70.92% on the balanced (104= Positive and 104=Negative) and 79.18% on the imbalanced (Positive =104 and Negative = 419) test dataset, respectively.

The suggested model integrated both spatial and temporal features to categorize fractures of the cervical spine. In [35] authors employed a two-stage pipeline for fracture detection. In the first phase, segmentation was carried out using UNet-EfficientNet, and in the second stage, detection was performed using CrackNet-LSTM. The overall accuracy of this approach reached 94.9%. To lessen serious outcomes, [36] emphasized on the urgent early diagnosis of upper cervical spine fractures, especially in segment C1. This study involved Efficient Net DNN models from B0 to B7 and a dataset containing over 350 image slices, where EfficientNet B6 demonstrated the highest accuracy. It attained a training accuracy of 98.25%, a testing accuracy of 99.25%, and a validation accuracy of 99.4%.

3. PROPOSED METHODOLOGY
This section describes all stages of the proposed method, namely Spine-EDLNet, from acquiring data to assessing the resultant outcome. The suggested methodology includes several key phases, such as data preprocessing, which involves image cropping, noise elimination, data augmentation, and data rescaling. To classify cervical spine fractures into normal and fractured classes, preprocessed data is fed into our chosen pre-trained models, EfficientNetV2, InceptionV3, and VGG16, combined in an ensemble fashion. Additionally, for improved classification performance, customized layers such as Max Pooling, Global Average Pooling, Batch Normalization, Dropout, Dense, and finally Sigmoid layers are added to each pretrained model separately. Furthermore, a majority voting block is added to combine the outcomes of the three customized pretrained models, enhancing the model's accuracy and generalizability. Figure 1 shows a general flow of our proposed framework.
The key motivation for selecting these deep learning models lies in their extensive features extraction, employing architecture adaptability, transfer learning advantages, design flexibility, and strong support from the DL communities [37]. These techniques excel at extracting and understanding complex characteristics and patterns from scans, which is essential for accurately identifying cervical spine fractures. Additionally, we evaluated pre-trained models depending on validation accuracy, validation loss, and overall performance. The detection performance of these pretrained deep learning models was extensively evaluated both individually and in ensemble configuration. The flow diagram of the proposed model is shown in Figure 2.

3.1. Data preprocessing
Data Prepossessing improves model performance and reduces computational time. We have employed several preprocessing techniques to prepare our dataset [38]. we have resized all input images to a standardized dimension of (224x224). The process of rescaling images aids in reducing training time and computing overhead for models. It also ensures uniformity in input image sizes for model, reducing complexity and accelerating the training process. Moreover, data standardization has been instrumental to ensure that our dataset is used for optimal learning. We have converted CT images to grayscale, applied histogram equalization to enhance contrast, normalized pixel values using mean and standard deviation, and rescaled pixel values to the range [0-255]. These normalization steps have standardized our input data, making it suitable for training our model. The pre-processed data is displayed in Figure 3
3.2 Data Augmentation
Data augmentation enhances dataset diversity and eventually model efficacy. In our study, we employed several data augmentation techniques. These techniques include zooming (0.2), rotation (15 degree), horizontal (0.1) and vertical (0.1) shifting, shear transformations (0.2), horizontal flipping, and fill mode adjustment. Figure 4 displays a selection of augmented images.

3.3 Pretrained- DL models
Transfer learning, a highly popular approach, significantly improves model performance. This method involves employing knowledge gained from completing one task to enhance performance on another associated task. The following pre-trained base DL networks were chosen for the work: InceptionNetV3, VGG16, and EfficientNetV2. Each of these models will be discussed in detail below.

3.3.1 VGG16
Our proposed framework involves utilizing a pretrained VGG16 [39] model followed by a series of customized layers. VGG16 was released in 2014 at the University of Oxford by the Visual Geometry Group. With its 16 layers, VGG16 follows an established architecture. Following pre-processing, the collected features are sent into a stacked Convolution layer with a constant stride of 1 and 3x3 receptive-field filters. Next, 5 max-pooling layers, each having 2x2 filter and a stride of 2, are used to apply spatial pooling. The architecture is completed with a SoftMax layer for the result and two fully connected layers after the final convolutional layer.
3.3.2 InceptionNetV3
InceptionNetV3[40] is a CNN considered for classification tasks. It is pretrained using the ImageNet dataset and contains 48 deep layers. It is an improved version of inception architecture and employs label smoothing. It was designed to provide efficient network without requiring an unnecessary number of training parameters. The InceptionV3 model uses batch normalization and label propagation through factorized 7 × 7 convolutions with an auxiliary classifier. In addition to its present design, numerous other modified layers have been incorporated. The InceptionV3 network has 525,313 trainable parameters and a total of 22,328,609 parameters.

3.3.3. EfficientNetV2
EfficientNetV2 [41] is an erudite convolutional neural network architecture famous for its high performance and computational efficacy. It stimulates advanced techniques such as progressive learning, which vigorously regulates image sizes through training, and an improved compound scaling method to balance model resolution, depth, and width professionally. This design surpasses in both precision and speed, applying fewer parameters and smaller convolutions while keeping outstanding performance. Its elastic design makes it well-suited for transfer learning, garnering noteworthy support from the deep learning communal.
3.3.4. Our customized Layers
We involved pre-trained deep learning models, authorising their weights remain frozen to hold their learned features. This method halts re-training of the base models, allowing them to practise on their current knowledge while focusing on learning novel features from employed data. Complementary, tailored layers are added to enlarge the models' classification performance. These additional layers are strategically chosen such as: MaxPooling2D was utilized to excerpt essential features from input images expertly, minimalizing computational complexity. Gaussian blur (1.5) and Gaussian noise (0.25) were applied for regularization, efficiently dropping image noise and enhancing data clarity. Also, GlobalAveragePooling2D aggregated features largely across the spatial dimensions of the input sample. Dense layers equipped with ReLU activation function allowed comprehensive feature extraction, familiarizing non-linearity. Batch normalization was utilized to soothe learning, while additional noise regularization layers additionally refined the noise reduction process. Dropout layers were employed to alleviate overfitting, promoting generalizability in classification tasks. Lastly, a sigmoid layer was added to guarantee exact classification of fractures grounded on the extracted features. Henceforth, these customized layers were applied independently to each model, enhancing their classification performance.

3.4. Ensemble Deep learning (EDL)
Due to its non-linear nature, deep Learning (DL) networks provide a high degree of flexibility when working with small training datasets. However, due to fine-tuning, which uses random methods and changes in weight settings during training, they might demonstrate some variance [41]. The model therefore generates inconsistent predictions as a result of this variance. EL method, which train several Deep Learning models rather than just one, have become popular as a solution to this problem [42]. The final output is obtained by summing up the predictions made by each of these individual models. Through the usage of several model’s predictions, EL captures more detailed characteristics from images and yields reliable classification results. EL improves the overall model’s output through leveraging the combined knowledge of numerous models. Existing literature primarily employs a single CNN for spinal fracture detection. However, there has been minimal attention given to the use of EL approaches for cervical spinal fracture detection and classification.

In this research, we investigate the use of ensemble Deep Learning approach. We proposed an ensemble method that combines three pre-trained deep learning models to mine comprehensive features while enhancing overall working, as shown in Figure 8.
Figure 8: The Process of Ensemble learning and Majority Voting Scheme

3.5. Voting Scheme
Following the acquisition of predictions from three pre-trained deep learning models, we implemented a majority voting scheme. This technique is commonly employed in classification tasks based on ensemble learning [43]. The three pre-trained models: EfficientNetV2, VGG16, and InceptionNetV3 categorized cervical spine data as either normal or fracture based on their respective features. The class with the maximum number of votes is then selected as the output class.

4. Experimental Result
4.1. Experimental Protocols
The research was performed utilizing the Python, and the Google Collaboratory code editor. Experiments were performed on a machine equipped with an Intel(R) Core (TM) i7-4700HQ processor and 8 GB of RAM.

4.2. Dataset
This research utilized the 'spine fracture detection from CT scans obtained from Kaggle'. The dataset is separated into 'Train' and 'Validation' sets, each containing both normal and fractured cervical spine images. It comprises in total 4200 images, evenly distributed with 1900 fractured images and 1900 normal images within the train folder, as shown in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Normal</th>
<th>Fracture</th>
<th>Total Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1900</td>
<td>1900</td>
<td>3800</td>
</tr>
<tr>
<td>Testing</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
</tbody>
</table>

Simultaneously, the test dataset consists of 400 photos, with 200 belonging to each class, namely fractured and normal. Moreover, 10 percent of the training images are randomly selected for validation process.

4.3 Evaluation Metrics
A number of metrics, comprising F1-score, Precision, Recall, and Accuracy, are utilized to evaluate the usefulness of our suggested model. The equations of all metrics are given below.

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}, \quad (1) \\
\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}}, \quad (2) \\
\text{Recall} = \frac{\text{TP}}{\text{TP + FN}}, \quad (3) \\
\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (4)
\]
Where TP denotes the count of correctly classified fracture cases, FP signifies the number of incorrectly identified fracture cases, TN indicates the accurately categorized non-fracture cases, and FN represents the erroneously classified non-fracture cases.

4.4 Results
In this section, we explain the outcomes by our proposed model. We trained and tested our model Spine-EDLNet by first training individual networks such as VGG16, EfficientNetV2, and InceptionNetV3. We merged the predictions of these individual models to form our ensemble approach. The outcomes showed significant improvements over the individual models. Results are shown in Table 2. Among all individual models, our proposed Spine-EDLNet performed the best in spinal fracture identification, achieving the highest precision (99.8%), accuracy (99.5%), recall (99.9%), and F1-score (99.7%), as shown in Figure 9. This highlights how ensemble learning effectively utilizes the various characteristics of individual models to improve accuracy and reliability in detection tasks.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG16</td>
<td>96.9%</td>
<td>97.5%</td>
<td>98.9%</td>
<td>97.3%</td>
</tr>
<tr>
<td>EfficientNetV2</td>
<td>97.6%</td>
<td>99.6%</td>
<td>99.6%</td>
<td>99.6%</td>
</tr>
<tr>
<td>InceptionNetV3</td>
<td>98.6%</td>
<td>98.3%</td>
<td>98.7%</td>
<td>98.6%</td>
</tr>
<tr>
<td>Proposed Spine-EDLNet</td>
<td>99.5%</td>
<td>99.8%</td>
<td>99.9%</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

**Table 2: Performance Evaluation of our Model**

Figure 9. Performance Analysis of our Models

4.5 Comparison with existing techniques
In this section, a comparative study is performed for cervical spine fracture detection with our proposed model Spine-EDLNet. The statistics are reported in Table 3. Researchers for the spine fracture detection have proposed several approaches attaining considerable outcomes. For example, RESNet50-LSTM [31] achieved an accuracy of 79.18% on the 3666CT dataset, indicating its capability but having room for improvement. Another model CrackNet-LSTM [36] confirmed significant performance, achieving an accuracy of 94.9% on the 2019 CT dataset. Then, DeIT-T16 [44] showcased significant performance, providing a robust accuracy of 98% on the 2019 CT dataset, suggesting its suitability for fracture detection tasks.
Table 3: Performance Evaluation Against Existing Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESNet50-LSTM[31]</td>
<td>79.18%</td>
<td>3666CT</td>
</tr>
<tr>
<td>CrackNet-LSTM[36]</td>
<td>94.9%</td>
<td>2019 CT</td>
</tr>
<tr>
<td>DeIT-T16[44]</td>
<td>98%</td>
<td>2019 CT</td>
</tr>
<tr>
<td>EfficientNet B6 [45]</td>
<td>99.2%</td>
<td>2019 CT</td>
</tr>
<tr>
<td>GoogleNet [46]</td>
<td>99.56%</td>
<td>2009 x-ray</td>
</tr>
<tr>
<td>Proposed Spine-EDLNet</td>
<td>99.6%</td>
<td>4200 CT</td>
</tr>
</tbody>
</table>

However, our proposed Spine-EDLNet model outperforms all existing models, achieving a considerable accuracy of 99.6% on an extensive 4200 CT dataset. This remarkable performance confirms the efficacy and reliability of our model in accurately identifying cervical spine fractures, indicating its potential for advancement in medical imaging diagnostics. Figure 7. depicts the performance comparison with existing deep learning techniques. It is evident that our model surpassed all detectors in terms of accuracy.

Figure 10. Comparative Analysis among Existing Techniques

5. Conclusion
Cervical spine injuries necessitate immediate diagnosis to ensure adequate treatment. Therefore, in this work, we propose a novel and robust method called the Spine-Ensemble Deep Learning Network (Spine-EDLNet). The proposed model contributes an essential role in extracting powerful features from image data due to ensemble learning approach. More specifically, model harnesses the strengths of three pre-trained deep learning networks: EfficientNetV2, InceptionNetV3, and VGG16, combined with a majority voting mechanism. Moreover, customized layers are attached into each model to enhance fracture classification. Furthermore, comprehensive data augmentation and pre-processing techniques are employed to the dataset before training, effectively overcoming the dataset availability challenge. Experimental results demonstrate that Spine-EDLNet outperforms previous models, achieving a maximum accuracy of 99.6%. This methodology aims to optimize diagnostic accuracy, robustness, and generalization. However, we noticed that when we fed blurry scans, the performance of the proposed model was degraded. Therefore, in future we aim to improve the performance on unseen samples by leveraging new datasets and fine-tuning the model.

References