

Deep Learning Algorithm Trained on Cervical Magnetic Resonance Imaging to Predict Outcomes of Transforaminal Epidural Steroid Injection for Radicular Pain from Cervical Foraminal Stenosis

Ming Xing Wang ^{1,*}, Jeoung Kun Kim ^{1,*}, Min Cheol Chang ²

¹Department of Business Administration, School of Business, Yeungnam University, Gyeongsan-si, Republic of Korea; ²Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University, Daegu, Republic of Korea

*These authors contributed equally to this work

Correspondence: Min Cheol Chang, Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Daegu, 705-717, Republic of Korea, Tel +82-53-620-4682, Email wheel633@ynu.ac.kr

Purpose: A convolutional neural network (CNN) is one of the representative deep learning (DL) model that is especially useful for image recognition and classification. In the current study, using cervical axial magnetic resonance imaging (MRI) data obtained prior to transforaminal epidural steroid injection (TFESI), we developed a CNN model to predict the therapeutic outcome of cervical TFESI in patients with cervical foraminal stenosis.

Patients and Methods: We retrospectively recruited 288 patients with cervical foraminal stenosis who received cervical TFESI due to cervical radicular pain. We collected single T2-axial spine MR image obtained from each patient. The image showing narrowest width of the neural foramen in the level at which TFESI was performed was used for input data. A “favor outcome” was defined as a $\geq 50\%$ reduction in the NRS score at 2 months post-TFESI vs the pretreatment NRS score. A “poor outcome” was defined as a $< 50\%$ reduction in the NRS score at 2 months post-TFESI vs the pretreatment score.

Results: The area under the curve of our developed model for predicting therapeutic outcome of cervical TFESI in patients with cervical spinal stenosis was 0.801.

Conclusion: We showed that a CNN model trained using cervical axial MRI could be helpful for predicting therapeutic outcome after cervical TFESI in patients with cervical foraminal stenosis.

Keywords: convolutional neural network, artificial intelligence, deep learning, spinal stenosis, cervical spine, transforaminal epidural steroid injection

Introduction

Cervical foraminal stenosis is a disorder marked by neural foraminal narrowing and a common cause of radicular pain in the upper extremity.¹ Facet and ligament hypertrophy, degenerative bony spurs, and laterally herniated disc are the main factors causing narrowing of the neural foramen, which result in mechanical compression of the cervical nerve root.^{2,3} Also, compression of the cervical nerve root induces an inflammatory response, in which various inflammatory mediated cells and proinflammatory cytokines are associated and cause cervical radicular pain in patients with cervical foraminal stenosis.⁴ For controlling cervical radicular pain from cervical foraminal stenosis, various conservative treatments, such as oral medication, physical therapy, and injection procedure, are being applied.⁴⁻⁷ Transforaminal epidural steroid injection (TFESI) is one of the most effective therapeutic methods for alleviating radicular pain induced by cervical foraminal stenosis.⁴ Corticosteroids inhibit the synthesis of various proinflammatory mediators. Positive therapeutic effect of TFESI in patients with cervical foraminal stenosis has been demonstrated in several previous clinical trials.^{4,8,9}

The ability to predict therapeutic outcomes is important because it allows the treatment plan for cervical radicular pain. Previously, some studies investigated the therapeutic outcome of TFESI based on clinical information and imaging findings in patients with cervical and lumbar radicular pain.^{4,10,11} However, little is known about factors affecting therapeutic outcome of TFESI in patients with cervical foraminal stenosis. In 2017, Kim et al evaluated the treatment outcome of TFESI according to the severity of cervical foraminal stenosis seen in cervical axial magnetic resonance imaging (MRI), but they could not find any significant difference in therapeutic outcome following severity of stenosis.⁷

Machine learning (ML) is a computer algorithm that can automatically learn from data without the need for explicit programming.^{12,13} ML is known for its ability to overcome the limitations of existing image analysis techniques and enable breakthroughs in the field of image analysis.^{12,13} Deep learning (DL) is an advanced ML approach that involves the use of a large number of hidden layers to build artificial neural networks with structures and functions similar to those of the human brain.¹⁴ Traditional ML techniques are outperformed by the DL technique, which learns from unstructured and perceptual image data. A convolutional neural network (CNN) is a typical DL model that is extremely useful for image recognition and classification.¹⁵

In the current study, we used cervical axial MRI as input data to train a CNN model to predict therapeutic outcomes after cervical TFESI in patients with chronic cervical radicular pain induced by cervical foraminal stenosis.

Materials and Methods

Subjects

A total of 288 consecutive patients who visited the spine center of a university hospital for cervical radicular pain due to cervical foraminal stenosis and underwent cervical TFESI between January 2013 and December 2021 (mean age = 53.7 ± 11.2, M:W = 166:122, injection levels C5:C6:C7:C8 = 7:135:136:10, right: left = 159:129) were retrospectively recruited for this study. The inclusion criteria for this study were as follows: (1) age 20–79 years; (2) received single-level cervical TFESI for segmental pain that radiated to the upper extremity due to cervical foraminal stenosis; (3) grade 1 or 2 cervical foraminal stenosis based on the classification method of Kim et al¹⁶ (grade 1 refers to moderate cervical foraminal stenosis – the narrowest width of the neural foramen is 51–100% of the width of the extraforaminal nerve root at the level of the anterior margin of the superior articular process; grade 2 refers to severe cervical foraminal stenosis – when the width of the neural foramen is ≤50% of the width of the extraforaminal nerve roots); (4) having a ≥ 3-month history of a symptomatic cervical radicular pain score of >3 on a numerical rating scale (NRS-11; 0 = no pain; 10 = the worst pain) prior to TFESI; (5) ≥50% temporary pain relief following a diagnostic nerve block with 1 mL of 2% lidocaine; and (6) MRI findings corresponding to the clinical presentations. We excluded patients having peripheral neuropathy, cervical myelopathy or infection on the spine. The study protocol was approved by the institutional review board of Yeungnam university hospital. Patient consent to review their medical records was not required by the Institutional Review Board of Yeungnam university hospital due to the retrospective nature of this study. To ensure patient data confidentiality, all personal information obtained from the medical records was anonymized and strictly protected in accordance with the regulations and guidelines offered by the Institutional Review Board of Yeungnam university hospital. This study complied with the ethical standards of the Declaration of Helsinki.

Transforaminal Epidural Steroid Injection

An aseptic technique was adopted for the cervical TFESI procedure. Patients were placed in a supine position under C-arm fluoroscopy (Siemens, Erlangen, Germany). To focus the target, the C-arm was rotated toward the region and the craniocaudal angle was controlled for focusing the intervertebral foramen. A 26-gauge 90-mm spinal needle with a bend at the tip was inserted into the skin and advanced to the anterior half of the superior articular process at the cervical spine. Next, the depth of the needle tip was checked using the anterior-posterior view and lateral view of the C-arm. A test dose of contrast medium (0.2–0.3 mL) was injected to determine whether the needle tip was placed at the proper location. Then, further injection of contrast medium was performed under real-time fluoroscopic monitoring. Subsequently, 20 mg of dexamethasone mixed with 1.5 mL of normal saline was injected. Cervical TFESI was conducted once for each patient.

Images Used for the Deep Learning Algorithm (Input Data)

Single T2-axial spine MR image obtained from each patient was used in our study. The image showing narrowest width of the neural foramen in the level at which TFESI was performed was used for input data. MR images obtained prior to cervical TFESI were used to develop the DL algorithm. We cut the MR image into left and right halves, and used the image on the side where TFESI was conducted to develop the DL algorithm (Figure 1). In case of Lt. half MR image, we inverted Lt. and Rt. (ie, Lt. half MR image was set to Rt. half MR image). Then, we cut the image with a rectangle shape containing cervical foramen and disc, superior articular process, and facet around cervical foramen (Figure 1).

Measurement of Therapeutic Outcome (Output Data)

Pain severity was assessed at pretreatment and 2-month follow-up after cervical TFESI. It was assessed by the numeric rating scale (NRS) (0 = no pain; 10 = worst pain). The NRS data were collected via chart review. A “favor outcome” was defined as a $\geq 50\%$ reduction in the NRS score at 2 months post-TFESI vs the pretreatment NRS score. A “poor outcome” was defined as a $< 50\%$ reduction in the NRS score at 2 months post-TFESI vs the pretreatment score. To validate the change in pain reduction, NRS scores were evaluated by assessing the difference between the pretreatment NRS scores and the 2-month post-TFESI scores (change in NRS [%] = [pretreatment NRS score – 2 months post-TFESI NRS score]/pretreatment NRS score \times 100).

Deep Learning Algorithms

Python 3.8.10, SciKit-Learn 0.24.2, and TensorFlow 2.10.1 with Keras were used to develop the CNN model for predicting the cervical TFESI outcomes. We trained pre-trained CNN models separately using three state-of-the-art CNN models, including EfficientNetV2B0, B1, B2 and compared their performances. The EfficientNetV2B0 CNN model was selected for the development of the model for predicting therapeutic outcome after cervical TFESI in patients with cervical foraminal stenosis. Table 1 shows the details of proposed model and Figure 1 summarizes the model training process.

Statistical Analysis

The statistical analyses were performed using Python 3.8.10 and Scikit-Learn version 0.24.2. Receiver operating characteristic curve analysis was performed, and the area under the curve (AUC) was calculated. The 95% confidence interval (CI) for the AUC was calculated as described by DeLong et al.¹⁷ Scikit-Learn was used to calculate the receiver operating characteristic curve and AUC.

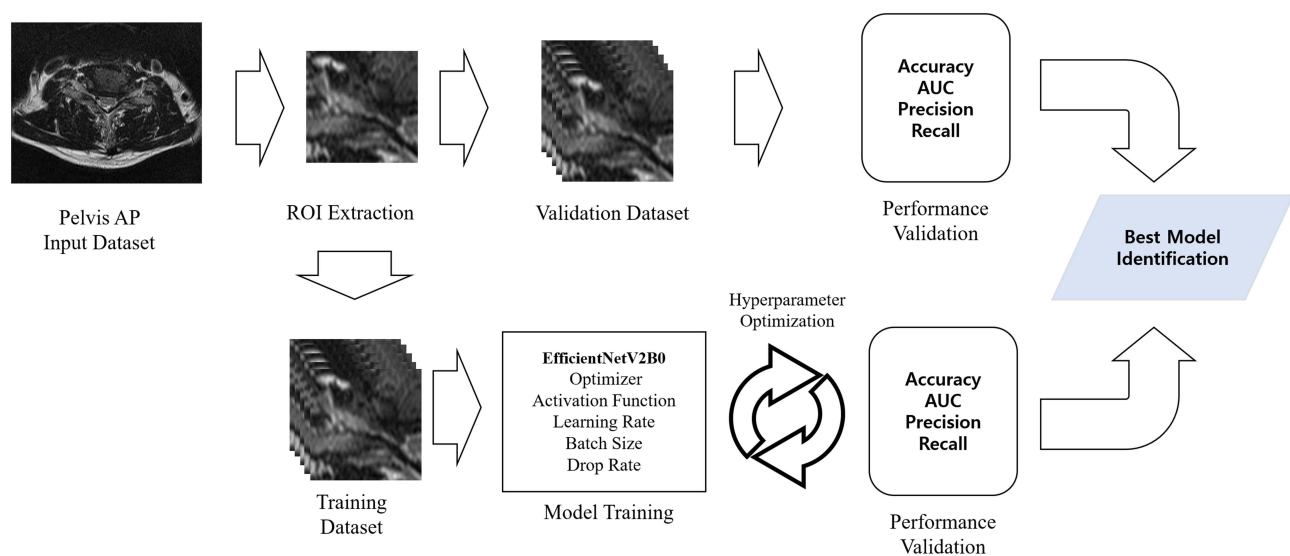


Figure 1 Diagram of the process for the development of the deep learning model for predicting the therapeutic outcome after transforaminal epidural steroid injection in patients with cervical foraminal stenosis.

Table 1 Layer Types and Parameters in Our Developed Model

Layer(Type)	Output Shape	Parameters
EfficientNetV2B0	4x4x1280	245,760
GlobalAveragePooling2D	1280	0
Dense	1024	1,311,744
Dropout	1024	0
BatchNormalization	1024	4096
Dense	1	1025
Total params: 7,236,177		
Trainable params: 7,173,521		
Non-trainable params: 62,656		

Results

The performance evaluation of our deep learning model yielded promising results. The validation accuracy was 79.3%, indicating high accuracy in predicting outcomes. Furthermore, the AUC reached 0.802 (95% CI, 0.682–0.923), demonstrating the capability of the model to distinguish between positive and negative cases. During the training phase, the model achieved an impressive 89.6% accuracy, along with a remarkable AUC of 0.981 (95% CI, 0.969–0.993) (Figure 2). Comprehensive details of the architecture of the model and its performance metrics are listed in Table 2. In Appendix 1, we provide a concise overview of the performance metrics for the deep learning models utilized in this study, serving as a point of reference for evaluating the performance of the proposed model.

Figure 3 illustrates the confusion matrix, which presents the correct classifications and misclassifications made using our deep learning model. The intensity of each cell's color corresponds to the frequency of cases within it, with darker

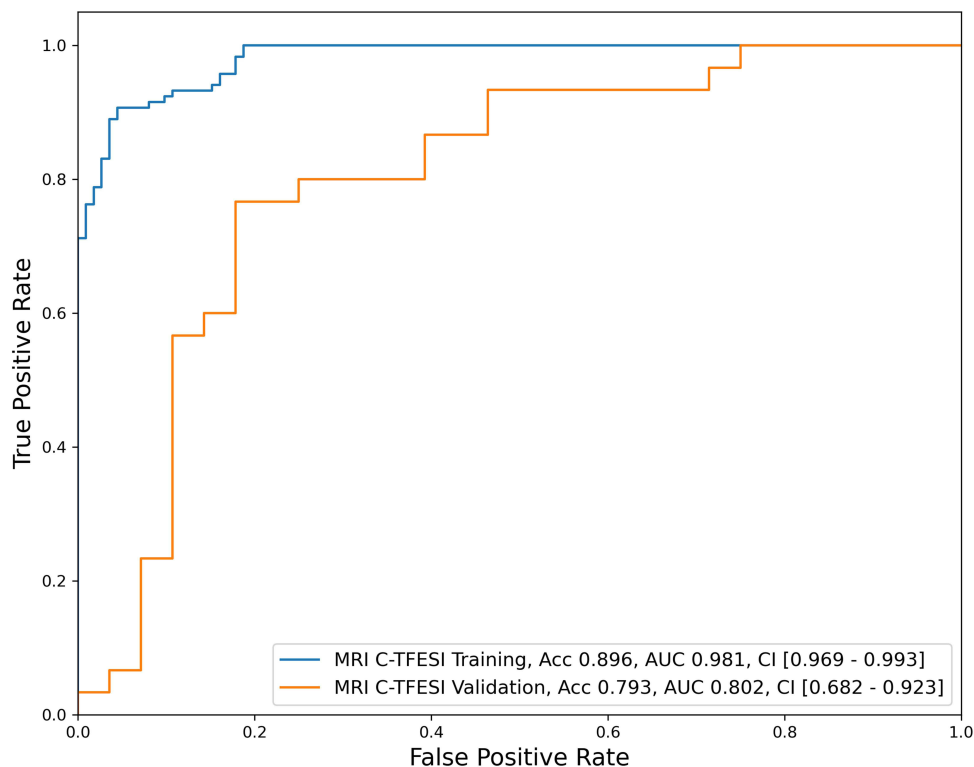


Figure 2 Receiver operating characteristic curve for the validation and test dataset of our developed model.

Table 2 Details of Performed of Our Developed Deep Learning Model

Sample size (patients) Sample ratio(patients)	230, 79.9% for training, 58, 20.1% for validation, total 288 Favor: 148, 51.4%; poor:140, 48.6% Favor: 118, 51.3%; poor:112, 48.7% for training Favor: 30, 51.7%; poor: 28, 48.3% for validation				
Model details	<ul style="list-style-type: none"> • EfficientNetV2B0 CNN model with transfer learning (freeze all layers without top) • RMSProp optimizer, ReLU activation • Learning rate 1e-05, batch size 8, training epoch 26 • Batch Normalization and dropout for regularization • ROI image resized to (100 × 100) • Training accuracy: 89.6%, AUC 0.981 with 95% CI [0.969–0.993] • Validation accuracy: 79.3%, AUC 0.802 with 95% CI [0.682–0.923] 				
Model performance (validation data)	Class	Precision	Recall	F1-score	Support
	Poor(0)	0.767	0.821	0.793	28
	Favor(1)	0.821	0.767	0.793	30
	Macro average	0.794	0.794	0.793	58

Abbreviations: CNN, convolutional neural network; RMSProp, root mean squared propagation; ReLU, rectified linear units; ROI, region of interest; AUC, area under the curve; CI, confidence interval.

shades indicating a higher count. Among the 58 validation data points, the model accurately predicted the outcomes of 46 cases and misclassified 12 cases. The misclassifications consisted of five false positives and seven false negatives, indicating areas where the model could be further optimized.

Ablation Study

An ablation study was conducted to gain insight into the individual contributions of the components in the prediction model for TFESI outcomes. Three models were examined: the original model, which incorporated regularization through a pretrained EfficientNetV2B0 CNN model with Dropout and BatchNormalization layers; the pretrained model without regularization, in which the Dropout and BatchNormalization layers were excluded; and the fully trained model without regularization. The models were evaluated using accuracy and AUC metrics and validation datasets. Table 3 presents detailed findings of the ablation study.

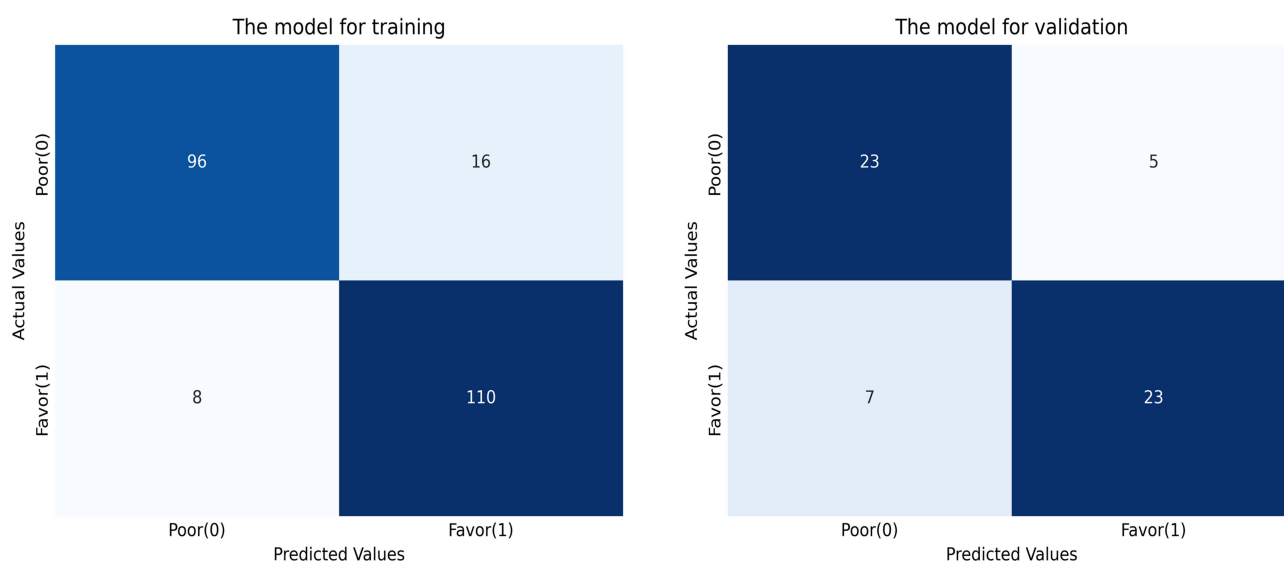


Figure 3 The correct classification and misclassification cases of our deep learning model. The intensity of each cell's color corresponds to the frequency of cases within it, with darker shades indicating a higher count.

Table 3 Ablation Study of the Prediction Model

Model	Model Details and Performance
Original model (pretrained model with regularization)	<ul style="list-style-type: none"> • Pretrained EfficientNetV2B0 CNN model using ImageNet data • Dropout and BatchNormalization layers adopted for regularization • Accuracy: 89.6% on training data, 79.3% on validation data • AUC: 0.981 on training data, 0.802 on validation data
Pretrained model without regularization	<ul style="list-style-type: none"> • Pretrained EfficientNetV2B0 CNN model using ImageNet data • Dropout and BatchNormalization layers removed • Accuracy: 93.9% on training data, 72.4% on validation data • AUC: 0.979 on training data, 0.692 on validation data
Fully trained model without regularization	<ul style="list-style-type: none"> • Fully trained EfficientNetV2B0 CNN model • Dropout and BatchNormalization layers removed • Accuracy: 77.4% on training data, 63.8% on validation data • AUC: 0.837 on training data, 0.679 on validation data

The ablation study revealed distinct performance variations among the three models. The original pretrained model, which included Dropout and BatchNormalization layers, achieved a 79.3% accuracy and an AUC of 0.802 on the validation dataset. In contrast, the pretrained model without regularization showed decreased accuracy to 72.4% on the validation data, with an AUC of 0.692. Remarkably, the fully trained model without regularization and pretrained parameters yielded a 63.8% accuracy and an AUC of 0.679 on the validation data.

The ablation study results highlighted the impact of pretrained parameters and regularization techniques on the performance of the prediction model outcomes. The original pretrained model, which employed Dropout and BatchNormalization layers, exhibited a reasonably high accuracy and AUC for the validation data. However, when these regularization layers were removed from the pretrained model, there was a noticeable decrease in performance, indicating the importance of Dropout and BatchNormalization in reducing overfitting and enhancing generalization. Furthermore, the fully trained model without regularization and pretrained parameters exhibited the lowest accuracy and AUC, emphasizing the significance of incorporating pretrained parameters and appropriate regularization techniques for optimal model performance.

Discussion

In this study, we developed the algorithm that can predict therapeutic outcome after cervical TFESI in patients with cervical foraminal stenosis using a cervical axial MRI as input data. The AUC of our model, evaluated with the validation dataset, was 0.801 with regarding predicting the therapeutic outcome at 2 months after cervical TFESI in patients with cervical foraminal stenosis.¹⁸ Considering the AUC ranging from 0.8 to 0.9 are generally considered to be excellent, our DL model trained using axial cervical MRI as input data would be helpful for clinicians in predicting therapeutic outcome of cervical TFESI for alleviating cervical radicular pain due to cervical foraminal stenosis.

In clinical practice, the severity of cervical foraminal stenosis is primarily determined using T2-axial cervical spine MR images. In 2015, Kim et al developed an MRI grading system to classify the severity of cervical foraminal stenosis based on the narrowest width of each neural foramen in a T2-axial cervical spine MRI.¹⁶ The developed system showed sufficient interobserver and intraobserver agreement. Additionally, in 2018, Kim et al assessed the effectiveness of TFESI according to the severity of cervical foraminal stenosis.⁴ They utilized a single T2-axial spine MR showing the narrowest width of the neural foramen at the level at which TFESI was performed for each patient. Considering the clinical situation, methodologies, and results of previous studies, we used only one axial MR image showing the narrowest width of the neural foramen at the TFESI level as the input data.

A deep neural network (DNN), also known as deep learning, contains a feed-forward neural network with multiple hidden layers and uses backpropagation, which provides greater capability than a traditional shallow neural network.^{14,15,19} A CNN is one of several DNN models.^{15,19,20} It takes multiple channels of two-dimensional data as input and transforms them repeatedly using convolution and pooling operations.^{15,20,21} These processes enable the

extraction of valuable features from the input image data. CNN model is being widely used in processing image data and recognizing pattern of image data. We think that our CNN model could recognize the degree of narrowing of the neural foramen and the degree of degeneration in disc and facet joint that may affect therapeutic outcome after cervical TFESI. However, due to the nature of DNN, we are not able to know which factors in cervical MRI were considered or weighted important in determining the therapeutic outcome after cervical TFESI.

So far, only 1 study attempted to find a prognostic factor that determines therapeutic outcome after TFESI in patients with cervical spinal stenosis.⁴ In 2017, Kim et al retrospectively recruited 53 patients with cervical radicular pain due to cervical foraminal stenosis.⁴ Kim et al divided the recruited patients into 2 groups (22 patients with non-severe foraminal stenosis and 31 patients with severe foraminal stenosis) following the severity of foraminal stenosis in cervical axial MRI. The patients in both groups showed a significant pain reduction at 2 weeks and at 1, 2, 3 months after TFESI. However, the degree of pain reduction following TFESI was not significantly different between patients with non-severe foraminal stenosis and those with severe foraminal stenosis. The reason why Kim et al could not find a factor influencing prognosis after TFESI would be that the binary classification they used was not sensitive and detailed enough to reflect the difference in the degree of narrowing of neural foramen between each patient. On the contrary, we think that our DL model recognized the difference in spinal structure shown on cervical axial MRI of each patient.

In conclusion, we showed that a CNN model trained using cervical axial MRI could be helpful for predicting therapeutic outcome after cervical TFESI in patients with cervical foraminal stenosis. Although the AUC of our developed model was 0.801 and interpreted as excellent performance, its predicting capacity is not accurate enough for clinical use. Accordingly, it can still be used as a supplementary tool for clinicians to predict therapeutic outcome of cervical TFESI in patients with cervical foraminal stenosis. We think that, for increasing predicting capacity, the combined use of spine MRI and clinical data as input data would be helpful. Also, we used a small number of cervical MR images to train the DL algorithm. If a larger number of input image data are used, the accuracy of DL model can be increased. In addition, we used cervical MR image obtained from a single center. For increasing the generalizability of DL model, MRI images obtained from other external hospitals for clinics are required.

Another limitation of this study is the lack of cross-validation using alternative approaches other than deep learning-based methods. To improve the generalizability of our findings, future research should incorporate a comparative analysis using the handcrafted features approach in conjunction with ML classifiers. This approach has received considerable attention in recent literature and may provide valuable insights for further validation.

Acknowledgments

Ming Xing Wang and Jeoung Kun Kim contributed equally to this work as co-first authors.

Funding

This study was supported by a National Research Foundation of Korea grant funded by the Korean government (grant no. NRF-2022R1F1A1072553).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Ko S, Choi W, Lee J. The prevalence of cervical foraminal stenosis on computed tomography of a selected community-based Korean population. *Clin Orthop Surg*. 2018;10:433–438.
2. Abbed KM, Coumans JV. Cervical radiculopathy: pathophysiology, presentation, and clinical evaluation. *Neurosurgery*. 2007;60:S28–34.
3. Yousem DM, Atlas SW, Goldberg HI, Grossman RI. Degenerative narrowing of the cervical spine neural foramina: evaluation with high-resolution 3DFT gradient-echo MR imaging. *AJNR Am J Neuroradiol*. 1991;12:229–236.
4. Kim MS, Lee DG, Chang MC. Outcome of transforaminal epidural steroid injection according to severity of cervical foraminal stenosis. *World Neurosurg*. 2018;110:e398–e403.
5. Chang MC. Conservative treatments frequently used for chronic pain patients in clinical practice: a literature review. *Cureus*. 2020;12:e9934.
6. Lee SH, Choi HH, Roh EY, Chang MC. Effectiveness of ultrasound-guided pulsed radiofrequency treatment in patients with refractory chronic cervical radicular pain. *Pain Physician*. 2020;23:E265–E272.

7. Patel EA, Perloff MD. Radicular pain syndromes: cervical, lumbar, and spinal stenosis. *Semin Neurol.* 2018;38:634–639.
8. Binler D, House LM, Mattie R, et al. The reliability of a grading system for digital subtraction imaging quality during cervical transforaminal epidural steroid injection. *Pain Med.* 2020;21:3126–3132.
9. Lee DG, Ahn SH, Lee J. Comparative effectiveness of pulsed radiofrequency and transforaminal steroid injection for radicular pain due to disc herniation: a prospective randomized trial. *J Korean Med Sci.* 2016;31:1324–1330.
10. Chang MC, Lee DG. Outcome of transforaminal epidural steroid injection according to the severity of lumbar foraminal spinal stenosis. *Pain Physician.* 2018;21:67–72.
11. Shrestha P, Subba L, Agrawal P, Lohani S. Outcome of transforaminal epidural steroid injection for lumbar radiculopathy: initial three-year experience at Upendra Devkota Memorial-National Institute of Neurological and Allied Sciences, Nepal. *Chin Neurosurg J.* 2020;6:6.
12. Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–1930.
13. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2:230–243.
14. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data.* 2021;8:53.
15. Kim JK, Choo YJ, Shin H, Choi GS, Chang MC. Prediction of ambulatory outcome in patients with Corona radiata infarction using deep learning. *Sci Rep.* 2021;11:7989.
16. Kim S, Lee JW, Chai JW, et al. A new MRI grading system for cervical foraminal stenosis based on axial T2-weighted images. *Korean J Radiol.* 2015;16:1294–1302.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–845.
18. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5:1315–1316.
19. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci.* 2021;2:420.
20. Kim JK, Wang MX, Chang MC. Deep learning algorithm trained on lumbar magnetic resonance imaging to predict outcomes of transforaminal epidural steroid injection for chronic lumbosacral radicular pain. *Pain Physician.* 2022;25:587–592.
21. Shin H, Kim JK, Choo YJ, Choi GS, Chang MC. Prediction of Motor Outcome of Stroke Patients Using a Deep Learning Algorithm with Brain MRI as Input Data. *Eur Neurol.* 2022;85:460–466.

Journal of Pain Research

Dovepress

Publish your work in this journal

The Journal of Pain Research is an international, peer reviewed, open access, online journal that welcomes laboratory and clinical findings in the fields of pain research and the prevention and management of pain. Original research, reviews, symposium reports, hypothesis formation and commentaries are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-pain-research-journal>