SERIOUS DISC BULGE DETECTION IN AXIAL MR IMAGE OF LUMBAR SPINE USING CNN AND IMAGE EXPLAINER

Paul Chukwurah [1], Daniel Chukwurah [2], Uthman Oyebanji [3], Ala Al-Kafri [4], Mohammad Alkasasbeh [5]

[1] The Owl Therapy Centre, Cardiff, United Kingdom
[2] Afrinvest West Africa, Lagos, Nigeria
[3] SSE Renewables, Glasgow, United Kingdom
[4] Teesside University, Middlesbrough, United Kingdom
[5] East Lancashire Hospitals NHS Trust, Burnley, United Kingdom

ABSTRACT

Disc bulge happens when the nucleus pulposus pushes outward through the annulus fibrosusand progresses over time, which can resultin disc degeneration problems such as spinal stenosis and sciatica. Serious bulges on the disc can put pressure on the surrounding nerve roots, sending pain down the spine and into other body regions. In this paper, a convolutional neural network (CNN) model wasdeveloped to diagnose composite axial MRI scans. The dataset used comprises 515 patients who reported lower back pain. It includes the last 3 lumbar spine discs, D3(L3-L4), D4(L4-L5), and D5(L5-S1)for each of the patients. The model achieved remarkable accuracy, recall, precision and F1scoreof89%. Local Interpretable Model-Agnostic Explanations (LIME) was alsoimplemented to explain the model's decision, henceremoving the black box problem generally associated with AImodels.This ensures the model provides interpretable insights, making it accurateand reliable.

KEYWORDS

Artificial Intelligence, Convolutional Neural Networks, Disc Bulge, Interpretable Diagnosis.

1. INTRODUCTION

Magnetic resonance imaging (MRI) serves as a cornerstone in diagnosing lumbar disorders, offering high-quality images without ionising radiation. Among the common lumbar intervertebral disc (IVD) injuries detected through MRI is disc bulge, which frequently manifests as low back pain and leg numbness. Lower Back Pain (LBP) stands as a primary cause of global disability, imposing significant socioeconomic burdens. Its diagnosis and treatment require a multidisciplinary, customised approach involving numerous outcome indicators, imaging data, and developing technologies. The increasing data generated throughout this process has accelerated the development of artificial intelligence (AI) methods to augment clinical decision-making and improve patient care [1].

Disc bulge occurs when the inner component of the intervertebral disc protrudes from its outer wall. This condition usually develops over time and can cause other disc degeneration conditions,

David C. Wyld et al. (Eds): CSIT, AMLA, IPDCA, NLPA, AIS, IPPR, SPTM – 2025
pp. 73-84, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151406

such as spinal stenosis [2]. Serious bulges can put pressure on the surrounding nerve roots, causing pain to travel down the back and other parts of the body depending on their position in the spinal column. This paper focuses on detecting *serious* disc bulges from MRI scans, so a CNN model is built to detect serious disc bulges on axial MR images. CNNs can recognise complex patterns and characteristics inside images using big datasets and deep learning approaches, allowing them to generate accurate predictions and assist physicians with diagnosis. By automating the diagnosis of serious disc bulges, AI models can help radiologists and doctors uncover anomalies that could otherwise be overlooked or misconstrued. This, in turn, can result in early detection of more precise diagnoses and better patient outcomes.

In addition to building the AI model, this paper underlines the significance of interpretability in AI-driven medical diagnostics. Due to the black-box nature of deep learning algorithms, clinicians frequently struggle to interpret AI-generated data. As a result, this study used an image explainer to provide clear insights into the AI's decision-making process in a bid to ensure the model is not only accurate but reliable. Interpretable diagnosis increases clinicians' trust and knowledge by explaining how AI-generated predictions are made, hence facilitating AI inclusion into clinical practice.

2. Related Works

2.1. Diagnosis of Lumbar

Lehnen et al. (2021) evaluated a convolutional neural network (CNN) trained on several MR imaging characteristics of the lumbar spine for detecting degenerative alterations. They examined 146 patients' lumbar spine MRIs with CNN to identify vertebrae, discs, and diseases such as disc herniation, bulging, spinal stenosis, nerve root compression, and spondylolisthesis. The CNN obtained complete accuracy in disc recognition and labelling, as well as moderate to high accuracy in disc herniations, extrusions, bulging, stenoses, nerve compressions, and spondylolisthesis. The work demonstrates that employing a single comprehensive CNN to automatically diagnose numerous lumbar spine degenerative alterations is feasible, with good diagnostic accuracy for clinically important findings [3]. *Al-Kafri et al. (2019)* offered a method for doctors to detect lumbar spinal stenosis by semantic segmentation and delineation of lumbar spine MRI scans using deep learning. Their dataset consists of 515 MRI examinations of patients with symptomatic back pain that were annotated by professional radiologists. They created a ground truth dataset to train and test segmentation methods, as well as innovative measures to evaluate dataset quality. The authors tested *SegNet* for semantic segmentation and evaluated the outcomes using contour and region-based metrics. They discovered that their methodology produced extremely good results, comparable to manual labelling, and had a great interrater agreement. Representative delineation results demonstrated accuracy appropriate for computer-aided diagnosis [4].

Ma (2015) proposed a novel pathological classification of lumbar disc protrusion, dividing it into four kinds based on intraoperative findings. The "damage-herniation" kind, which is most likely caused by injury, has soft herniation and easily detachable disc tissue, indicating that minimally invasive endoscopic surgery is likely to be successful. The "degeneration-protrusion" kind, which is distinguished by hard protrusions and degenerative alterations, may necessitate nerve decompression and posterior wall removal, whereas disc excision may not be required[5]. *Pan et al. (2021)* suggested an automated approach for identifying lumbar disc bulge and herniation using deep convolutional neural networks (CNNs) on magnetic resonance (MR) images. The study's goal was to simplify the interpretation of MR images, lowering radiologists' burden. The system locates vertebral bodies and discs with 100% accuracy in cross-validation. It diagnoses axial lumbar disc MR images as normal, bulging, or herniated, with accuracies ranging from 84.2% to 92.7% for various intervertebral disc levels. The created system improves diagnostic efficiency, standardises reports, and has potential uses beyond disc problems, such as recognising lumbar anomalies and cervical spondylosis [6]. *Kim et al. (2022)*developed CNN models for diagnosing severe central lumbar spinal stenosis (LSS) using radiography, and they evaluated radiological diagnostic features using gradient-weighted class activation mapping. Based on formal MRI reports, participants were divided into two groups: severe central LSS and healthy control, and radiographs were taken for both. A CNN-based transfer learning system was used to determine if radiographic findings were LSS or normal. The VGG19 model was the most accurate (82.8%), with an area under the receiver operating characteristic curve (AUROC) of 90.0%. Grad-CAM detected characteristics such as reduced disc height, limited foramina, short pedicle, and hyperdense facet joint[7].

Lin et al., (2024) used five machine learning models to identify important variables leading to disc herniation and bulge [8]. The best model achieved an f1-score of 78% for detecting disc bulge which gives room for improvement. *Al Masarwehet al.*, (2024) used AI to detect disc herniation in the lumbar spine, and the model achieved a maximum accuracy of 70% using CNN [9].

2.2. Interpretable Diagnosis

Hui et al. (2022) conducted a systematic review to investigate the potential use of explainable artificial intelligence (XAI) in healthcare. They found 99 high-quality articles on various XAI approaches, including SHAP, LIME, GradCAM, and rule-based systems. The review underlined the need for more attention to detecting irregularities in 1D biosignals and finding critical clinical text. By resolving these shortcomings, the XAI research community may boost trust in AI models and encourage their incorporation into healthcare systems [10]. *Jeong-Woon et al. (2023)* created a deep learning model to estimate bone mineral density (BMD) from CT scans, which had a correlation coefficient of .90 with DXA-measured BMD. With a maximum F1 score of .875 in abnormal/normal categorisation, the model shows potential as an auxiliary tool in clinical practice. Explainable AI approaches indicated that the network concentrated on tissues surrounding the vertebral foramen [11]. *Otaki et al. (2022*) developed a deep learning model named CAD-DL, which surpassed traditional methods in detecting obstructive coronary artery disease (CAD) using SPECT myocardial perfusion imaging (MPI). When CAD-DL is integrated with clinical software, doctors can receive quick results and explanations of predictions. When tested, CAD-DL outperformed quantitative analysis and visual reading in terms of diagnostic accuracy (AUC: 0.83). This work demonstrates how explainable AI may be used in clinical settings to improve diagnostic confidence in the diagnosis of CAD after MPI [12].

3. METHODOLOGY

In this section we discussed the steps employed in developing the convolutional neural network (CNN) for serious disc bulge detection in lumbar spine MRIs and the use of image explanation techniques to make the model's decisions more interpretable.
Fig 1 below describes the flowchart of the methodology employed in this paper.

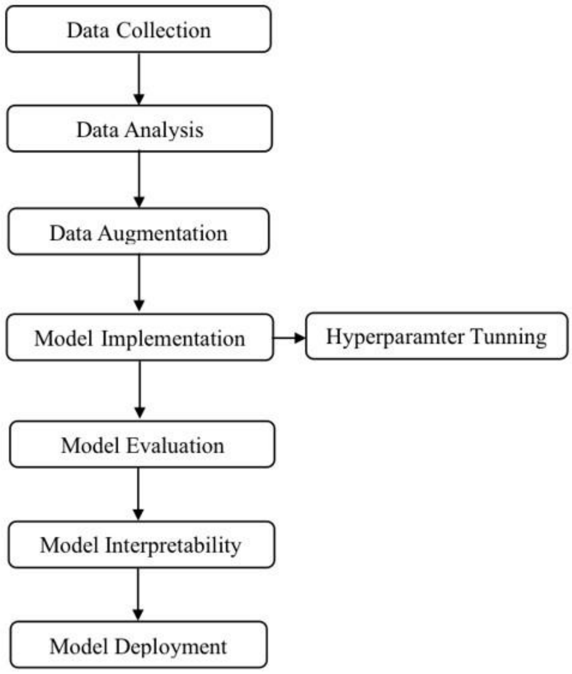


Fig 1. Methodology flowchart for the Serious Disc Bulge Detection model.

3.1. Data Collection

The dataset used for this study comprises MR composite images of 515 patients who reported symptomatic back pain. The images were sourced from the ground truth data provided by [13]. This study uses the axial view of the MRI, as shown in *Fig 2*, which is primarily obtained from the final three IVDs, including the one between the last vertebrae and the sacrum. The labelling of the figure was guided by [14]. Most of the slices have a 320x320 pixel image resolution, and all the pixels have 12-bit per pixel precision, which is greater than that of ordinary 8-bit greyscale images. All axial-view slices have a uniform thickness of 4 mm, with a centre-to-centre spacing of 4.4 mm. The axial-view slices have a constant horizontal and vertical pixel spacing of 0.6875 mm.

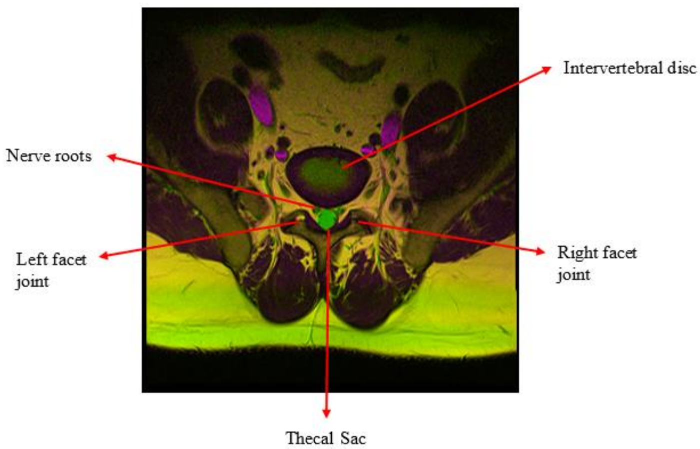


Fig. 2. A composite axial view of an MRI image of the L5-S1 intervertebral disc.

3.2. Data Analysis

Based on the radiologist note provided by [15], the data was grouped into *No Serious Bulge* and *Serious Bulge*. Conditions such as diffuse disc bulge compressing the thecal sac and exiting canals, disc protrusion compressing the thecal sac, and narrowing of the left neural foramen were classified as serious disc bulge cases. Cases of mild diffuse disc bulge, diffuse disc bulge mildly compressing the thecal sac, etc., were considered not to be serious cases. Hence, *no serious disc bulge*, like what has been reported by [16]. The count of all images for each class is presented in *Table 1* below.

Table 1: Summary of the different conditions present in the radiologist report

Disc condition	**Image Count**
No Serious Bulge	1074
Serious Bulge	366

3.3. Data Augmentation

Data augmentation was used to enrich model training data to improve the model's robustness. The data augmentation technique involves adding modest random stochastic modifications to the training image dataset and training the model with both the source and supplemented data. Data augmentation techniques are known to enhance model resilience and test accuracy [17]. In this study, the followingaugmentation parameters were considered using the *ImageDataGenerator* class from the *Keras* library.

ImageDataGenerator (width_shift_range=0.1; height_shift_range=0.1; horizontal_flip=True; fill_mode=’nearest’).

Table 2 below shows a summary of the data count before and after data augmentation performed in each section of this study.

Table 2: Summary of data augmentation done for serious disc bulge detection.

Disc condition	**Before Augmentation**	**Augmentation factor**	**No. of Augmented Images**	**Total**
No Serious Bulge	1074	1.33	1426	2500
Serious Bulge	366	5.83	2134	2500

3.4. Model Implementation using CNN

CNNs consist of neurones with learnable weights and several layers to extract features from images and then learn to classify them based on those features[18]. All the model's image data and labels were processed and resized to 150x150 pixels and then converted to *NumPy* arrays.

1) Model Architecture

CNN employs several layers to extract features from images. There are 4 layers in the architecture used in this study, as shown in *Fig 3* and highlighted below

- **Cropping layer**– This layer was added to get the region of interest (ROI) by removing 45 from the top and 25 from the bottom, left and right of the image.

- **Convolutional and pooling layers–**This layer is used to extract different features from the input images. This layer performs the mathematical operation of convolution between the input image and a filter of a specific size. There are four layers, and each layer has filters: the first layer has 64, the second 128, the third 256, and the fourth 512. As the convolutional layers deepen (from 64 to 512 filters), the model gains the ability to identify increasingly complex features within the image. Each layer uses the *ReLU* activation function to introduce non-linearity, enabling the model to capture complex patterns in the data. The same padding was used to retain spatial dimensions. They also include a *MaxPooling2D* layer with a pool size of (2, 2) to reduce spatial resolution and highlight key features, and a dropout rate of 0.2 was used to help prevent overfitting.

- **Flatten and Fully Connected layers** – Then three sequential *Dense* layers with 512, 256, and 128 neurones were added, each with a 'relu' activation function, to enable high-level feature interpretation and decision-making. The decreasing number of neurones across layers helps the model gradually make complex features into more compact representations, aiding efficient and robust classification/detection.

- **Output Layer** – this final layer is added with two neurones and a *softmax* activation function used to classify each image into *Serious Disc Bulge* and *NoSerious Disc Bulge*. The *RMSprop* optimizer was also used, which adapts the learning rate during training for efficient convergence. A *learning rate* of 0.001 was set to help change the weights and reduce losses.

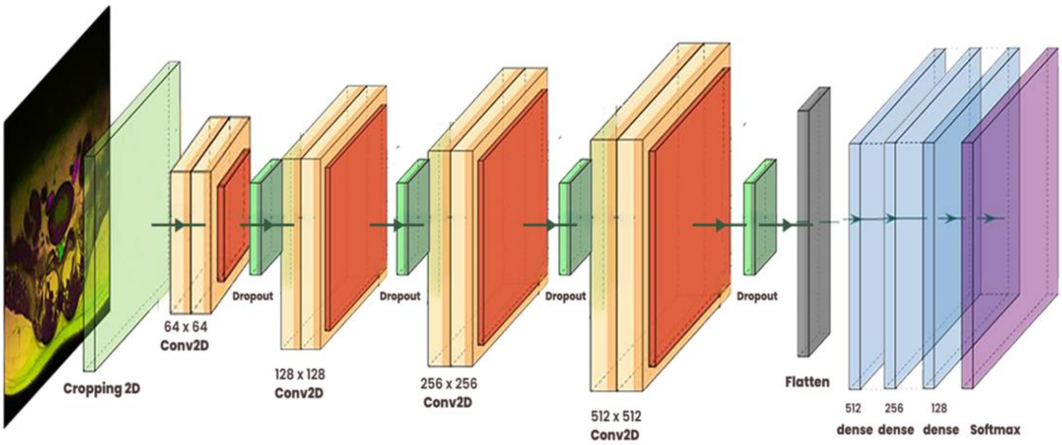


Fig 3. CNN's architecture

2) Model Training and Evaluation Metrics

The dataset was split into 70% for training, 20% for testing, and 10% for validation, following the recommendation by [19]. A batch size of 128 and 50 epochs were used to train the model. Callback parameters were implemented using *ModelCheckpoints* to monitor validation accuracies and save the best model. The model was evaluated by plotting the training and validation accuracies, a confusion matrix, and classification reports.

The following classification metrics were used to evaluate and compare the performance of the model:

- **Accuracy is** simply the percentage of correct label predictions against the sum of all the predictions, as shown below.

$$\text{Accuracy} = \frac{Correct\ label\ prediction}{Sum\ of\ all\ prediction}$$

- **Precision is** the ratio of true positives (TP) to the sum of false positives (FP) and TP, as shown below. It measures the capability of a model to not predict positive for a negative input.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall is** the ratio of TP to the sum of false negatives (FN) and TP. It measures the ability of a model to find all the positive samples.

$$\text{Recall}=\frac{TP}{TP + FN}$$

- **F1-scoreis** the harmonic mean of both precision and recall.

$$\text{F1}=\frac{2*(precision * recall)}{precision+recall}$$

3) Hyperparameter Tunning

Hyperparameters control the training and structure of AI models to influence the performance. Examples include learning rate, batch size, network depth, learning rate decay, and regularisation. Proper tuning of these parameters can significantly improve model outcomes.

3.5. Model Interpretability using LIME

A key reason for choosing LIME is its capacity to work with image data, which is often a challenge in model interpretability [20]. Although LIME can produce unstable explanations due to its reliance on perturbation-based sampling, its simplicity and ease of use make it a suitable option for this research, where interpretability is essential for understanding the influence of different features in a complex model. Moreover, LIME's ability to generate explanations that are intuitive for non-experts is an important consideration, as it facilitates understanding for users without a deep technical background [21]. While more advanced methods like SHAP offer global insights but demand higher computational resources, LIME’s instance-based explanations are well-suited for cases where understanding individual predictions is crucial. Therefore, LIME strikes a practical balance between clarity and efficiency, making it the best fit for this study.

Model interpretability allows for the explanation of complex AI models. It focuses on why the AI arrived at a particular decision, analysing its logical paradigms, as opposed to the inherent black box nature of AI [22]. They are called black boxes because it is hard to analyse and visualise how their inputs contribute to the predicted output [23]. LIME was employed to explain how different areas of the MRI contribute to the model’s decision. LIME creates an interpretable model by fitting a local linear model near the prediction point.

4. RESULT AND DISCUSSION

4.1. Model Evaluation

Five evaluation metrics were used to evaluate the performance of the model. Accuracy, recall, precision, f1-score, and confusion matrix. The model achieved 89% precision, recall, and accuracy, as shown in *Table 3*.

Table 3: Evaluation report for the serious disc bulge detection model

approx. values	**Accuracy**	**Recall**	**Precision**	**F1- score**
Serious Disc Bulge Detection	0.89	0.89	0.89	0.89

A confusion matrix plot was generated for the model to assess where errors were made. Fig. 4 below shows the confusion matrix plot of actual against predicted for the model. The confusion The matrix accurately classified 439 cases with no disc bulge and 450 cases with serious disc bulge.

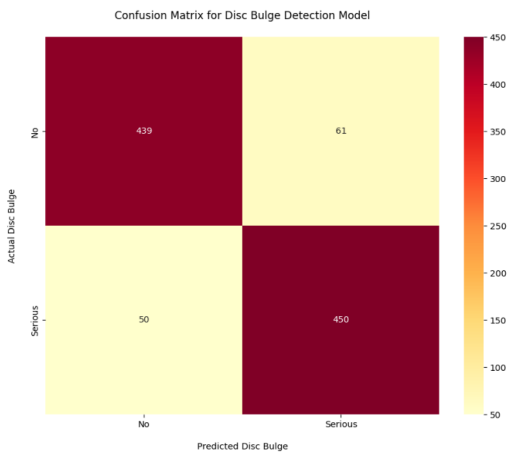


Fig 4: Confusion matrix for serious disc bulge detection mode

4.2. Model Interpretability Results

In this section, we provide the interpretability findings model. Model interpretability is critical for assuring the confidence, transparency, and comprehension of artificial intelligence (AI) models in medical imaging. By investigating how the models produce predictions, we get insights into their decision-making process, thereby increasing their clinical utility and promoting trust among healthcare practitioners.

Fig.5 shows the output of LIME reflecting the contribution of each region to the model's prediction. The heatmap on the right breaks down the level of contribution for each region.

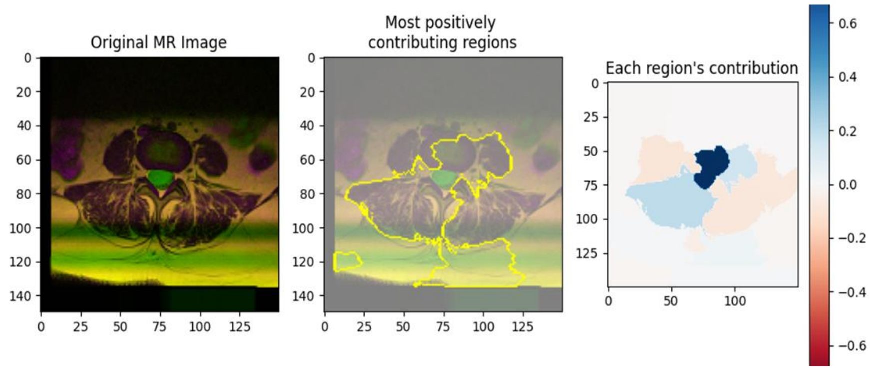


Fig 5: Contribution of each region to serious disc bulge detection

4.3. Benchmark Against Related Work

Despite the promising results achieved by models developed in similar studies [8], we didn't find any study that applied explainable AI techniques to disc bulge detection. Studies involving the development of deep learning models to date have not explored the potential of explainable AI to provide interpretability. This research gap was filled by this study, as LIME was employed to explain the model's decisions and thus eliminate the black box problem of this model. Table 4 below summarises the benchmarking of related work to this study.

Table 4: Benchmark against related work

Authors	Lin et al. (2024) [8]	Al Masarweh et al. (2024) [9]	Pan et al. (2021) [6]	Chukwurah et al., 2025 (This study)
Study	Development of a Machine Learning Algorithm to Correlate Lumbar Disc Height on X-rays with Disc Bulging or Herniation.	Automatic Detection of Lumbar Spine Disc Herniation	Automatically Diagnosing Disk Bulge and Disk Herniation with Lumbar Magnetic Resonance Images by Using Deep Convolutional Neural Networks: Method Development Study.	Serious Disc Bulge Detection in Axial MR Image of Lumbar Spine Using CNN and Image Explainer.
Methodology	Machine learning algorithms.	Computer vision and Artificial Intelligence	Deep convolutional neural networks, image analysis, classification.	Convolutional neural networks and image explainer LIME
Results	The average F1 score on the testing dataset for the LASSO, MARS, decision tree, random forest, and *XGBoost* models were 0.706, 0.778, 0.569, 0.729, and 0.706 respectively.	The study achieved 100% detection of intervertebral discs using Mask R-CNN, but classification accuracy was limited to 70% with the best-performing CNN model, likely due to a small training dataset.	Development of an automated technique for identifying disk bulge and herniation with excellent accuracy. Achieved an average accuracy of 88%.	Achieved an excellent result of 89% and use of LIME for model interpretability.Explaining the regions contributing to the model's decision.
Model Interpretability	Lack of explainable AI	Absent	Absent	Present

4.4. Clinical Implications and PACS Integration

The developed deep learning model for detecting serious disc bulge in axial lumbar spine MRI has significant clinical utility. Early and accurate identification of disc bulges can enhance diagnostic workflows, reduce radiologist workload, and support timely interventions. Its integration into hospital Picture Archiving and Communication Systems (PACS) could streamline radiology reporting by flagging suspected cases for priority review, ensuring consistency across readings. Moreover, automated triaging through PACS can support rural or understaffed settings by providing decision support to general practitioners and less-experienced clinicians. For real-world deployment, the model would require validation on diverse datasets and compliance with regulatory standards, but its potential to improve patient outcomes and diagnostic efficiency is clear.

5. CONCLUSION

This paper explored the development of a CNN model to detect serious disc bulge in lumbar spine MRI scans. The model achieved high accuracy, recall, precision, and F1-score of 89% in all the metrics. This paper demonstrates the potential use of AI in detecting spinal abnormalities which can provide support to radiologists and clinicians. It further emphasises the importance of AI interpretability in healthcare decision-making. The use of LIME to interpret the model has ensured that the model is not just a black box but offers insights into its decision-making process.

While the model developed demonstrated remarkable metrics, further improvements can be made to the AI model's performance and clinical application. Expanding the dataset to include a larger and more diverse range of lumbar spine MRI images, encompassing various pathologies and patient demographics, will enable the creation of a more robust and generalisable model capable of properly detecting serious disc bulges.

REFERENCES

[1] F. D'Antoni, F. Russo, L. Ambrosio, L. Bacco, L. Vollero, G. Vadalà, M. Merone, R. Papalia and V. Denaro, Artificial Intelligence and Computer Aided Diagnosis in Chronic Low Back Pain: A Systematic Review, vol. 19, 2022, p. 5971.

[2] The Bonati Institute, "Bulging disc - Causes, symptoms & treatment," [Online]. Available: https://www.bonati.com/conditions/bulging-disc/. [Accessed 2024].

[3] N. C. Lehnen, R. Haase, J. Faber, T. Rüber, H. Vatter, A. Radbruch and F. C. Schmeel, "Detection of Degenerative Changes on MR Images of the Lumbar Spine with a Convolutional Neural Network: A Feasibility Study," Diagnostics (Basel), vol. 11, no. 5, p. 902, 2021.

[4] A. S. Al-Kafri, S. Sudirman, A. Hussain, D. Al-Jumeily, F. Natalia and H. Meidia, "Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks," IEEE Access, vol. 7, pp. 43487-43501, 2019.

[5] X.-l. Ma, A new pathological classification of lumbar disc protrusion and its clinical significance., vol. 7(1), Orthopaedic Surgery, 2015, p. 1–12.

[6] Q. Pan, K. Zhang, L. He, Z. Dong, L. Zhang, X. Wu, Y. Wu and Y. Gao, "Automatically Diagnosing Disk Bulge and Disk Herniation with Lumbar Magnetic Resonance Images by Using Deep Convolutional Neural Networks," JMIR Med Inform, vol. 9, no. 5, p. 14755, 2021.

[7] T. Kim, Y.-G. Kim, S. Park, J.-K. Lee, C.-H. Lee, S.-J. Hyun, C. H. Kim, K.-J. Kim and C. K. Chung, "Diagnostic triage in patients with central lumbar spinal stenosis using a deep learning system of radiographs," J Neurosurg Spine, vol. 37, no. 1, pp. 104-111, 2022.

[8] P.-C. Lin, W.-S. Chang, K.-Y. Hsiao, H.-M. Liu, B.-C. Shia, M.-C. Chen, P.-Y. Hsieh, T.-W. Lai, F.-H. Lin and C.-C. Chang, "Development of a Machine Learning Algorithm to Correlate Lumbar Disc Height on X-rays with Disc Bulging or Herniation," Diagnostics, vol. 14, no. 2, p. 134, 2024.

[9] M. A. Masarweh, O. Oluseyi, A. Alkafri, H. Alsmadi and T. Alwadan, “Automatic Detection of Lumbar Spine Disc Herniation,” International Journal of Advanced Computer Science and Applications(ijacsa), vol. 15, no. 11, 2024.

[10] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari and U. R. Acharya, “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022),” Computer Methods and Programs in Biomedicine, vol. 226, p. 107161, 2022.

[11] J.-W. Kang, C. Park, D.-E. Lee, J.-H. Yoo and M. Kim, “Prediction of bone mineral density in CT using deep learning with explainability,” Front Physiol, vol. 13, 2023.

[12] Y. Otaki, A. Singh, P. Kavanagh, R. J. Miller and T. Parekh, “Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease.,” JACC: Cardiovascular Imaging, vol. 15, no. 6, p. 1091–1102, 2022.

[13] S. Sudirman, A. Al-Kafri, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi and M. Al-Jumaily, “Label Image Ground Truth Data for Lumbar Spine MRI Dataset,” Mendaley data, 2019.

[14] F. Natalia, H. Meidia, N. Afriliana, A. S. Al-Kafri, S. Sudirman and A. Simpson, “Development of Ground Truth Data for Automatic Lumbar Spine MRI Image Segmentation,” IEEE Xplore, pp. 1449-1454, 2018.

[15] S. Sudirman, A. Al-Kafri, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi and M. Al-Jumaily, “Radiologists Notes for Lumbar Spine MRI Dataset,” Mendeley Data, 2019.

[16] J. Zou, H. Yang, M. Miyazaki, Y. Morishita, F. Wei, S. McGovern and J. C. Wang, “Dynamic Bulging of Intervertebral Discs in the Degenerative Lumbar Spine,” Spine, vol. 34, no. 23, pp. 2545-2550, 2009.

[17] S. Shahinfar, P. Meek and G. Falzon, ““How many images do I need?” Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring,” Ecological Informatics, vol. 57, p. 101085, 2020.

[18] R. N. Shaw, A. Ghosh, V. E. Balas and M. Bianchini, Artificial Intelligence for Future Generation Robotics, Elsevier, 2021.

[19] D. A. Blackman, S. J. Johnson, F. Buick, D. E. Faifua, M. O'Donnell and M. Forsythe, “The 70:20:10 model for learning and development: an effective model for capability development,” Academy of Management Proceedings, vol. 1, p. 10745, 2016.

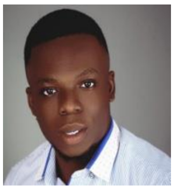
[20] M. T. Ribeiro, S. Singh and C. Guestrin, “"Why Should I Trust You?": Explaining the Predictions of Any Classifier,” KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, 2016.

[21] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions.," NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, vol. 30, pp. 4768 - 4777, 2017.

[22] V. Vishwarupe, P. M. Joshi, N. Mathias, S. Maheshwari, S. Mhaisalkar and V. Pawar, “Explainable AI and Interpretable Machine Learning: A Case Study in Perspective,” Procedia Computer Science, vol. 204, pp. 869-876, 2022.

[23] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions.,” NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, vol. 30, pp. 4768 - 4777, 2017.

AUTHORS



Paul Chukwurah is a skilled full-stack developer with over four years of experience building user-friendly, high-performance applications across sectors like finance, education, and enterprise. He holds a master's degree in artificial intelligence, where he developed a deep learning model to improve lumbar spine MRI diagnosis using large health datasets. Paul has worked on impactful projects, including the 9ijakids gaming platform, where he led frontend development, and fintech applications.

Daniel Chukwurah is a highly skilled front-end developer with over six years of experience delivering innovative web and mobile applications across the FinTech, Banking, Insurance, and Investment sectors. My expertise spans technologies like React.js, React Native, Vue.js, JavaScript, and TypeScript, coupled with a strong grasp of cloud technologies and user experience design. He has a proven track record of building scalable, high-performance applications, driving business outcomes, and collaborating with cross-functional teams to achieve impactful results. He loves to contribute to open-source projects and learn new skills.

Uthman Oyebanji is a skilled data&performance analyst with a master's degree in artificial intelligence.He specialises in using data-driven insights to improve performance, streamline processes, and support strategic decision-making, particularly within the renewable energy sector. At SSE Renewables, he has built data analysis frameworks and reporting systems that help track key metrics and optimise operations. Uthman combines strong analytical skills with clear communication, effectively bridging the gap between complex data and actionable business outcomes.

Dr AlaAlKafri is a highly experienced senior lecturer in data science&artificial intelligence at Teesside University. With over 16 years of teaching and academic expertise in computer science, he holds a PhD from Liverpool John Moores University focused on using machine learning such as SVMs and deep neural networks to automate the diagnosis of lumbar spine diseases from MRI data. His research extends into AI applications in healthcare and education, with recent projects in clinical video analysis for therapist training and AI-based diagnostic tools for spinal conditions.

Dr Mohammad Alkasasbeh is an experienced clinician at East Lancashire Hospitals NHS Trust, Burnley (UK), with years of experience across inpatient, outpatient and community services. Known for his calm, evidence-led approach, he has steered multiple quality-improvement projects within the Trust. Dr Alkasasbeh remains an active contributor to multidisciplinary research collaborations and regularly presents practical insights at regional clinical forums.