Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Deep Learning–Based Classification of Lumbar Spine Disorders Using Biomechanical Features: Model Development, Evaluation, and Clinical Implications

Dr. V. Yasaswini[1], Ms. Swathi T[2], Dr. Manam Vamsi Krishna[3], Mrs. Mamilla Maheswari[4], Ms. Rachel Doris Karumanchi[5], Mrs P B Maheswari[6],

[1]Associate Professor, Department of Cyber Security,
yashu.vanapalli29@gmail.com
[2]Assistant Professors, Department of CSE (Data Science).
Email ID: swathithottaramallareddy@gmail.com
[3]Associate Professor, Dept. of Computer Science and Engineering (AIML)
Email: manavamsikrishna@gmail.com
[4]Assistant Professor, Dept. of Computer Science and Engineering (AIML)
Email: mahimrecw@gmail.com
[5]Assistant Professors, Dept. of Computer Science and Engineering (Data Science)
Email ID: dorisrachel.k@gmail.com
[6]Assistant Professor, Department of Cyber Security,
Maheshwaripbmrecw@gmail.com
[1,2,3,4,5,6] Malla Reddy Engineering College for Women (Autonomous),
Maisammaguda , Hyderabad, Telangana, India

ARTICLE INFO

Received: 18 Aug 2024

Revised: 10 Sept 2024

Accepted: 28 Nov 2024

ABSTRACT

This project is designed for Automated Identification of HERNIA, NORMAL, and SPONDYLOLISTHESIS conditions, since it is indeed cumbersome to accurately classify subtle and overlapping biomechanical patterns within the lumbar spine data. Secondly, a DNN model using three deep layers, activation using ReLU, L2 regularization, batch normalization, and dropout values of 0.5 and 0.3 were constructed to mitigate the overfitting effect. The proposed idea was suitable for small & complex medical datasets, whereas Early Stopping and adaptive learning-rate scheduling were applied to stabilize training and improve generalization. The proposed model performed well in classification, obtaining a test accuracy of 80.65% at the end. The highest detection accuracy corresponded to spondylolisthesis, which is due to its more specific biomechanical characteristics, and Normal revealed moderate results. Hernia was the most challenging class, as the feature patterns are rather specific, but the model still distinguished between it meaningfully. In conclusion, the results validate the proposed methodology, achieving state-of-the-art performance in lumbar spine condition classification and laying a strong foundation for future clinical decision-support systems.

Keywords: Lumbar Spine Analysis, Deep Neural Networks, Biomechanical Features, Medical Classification

I. INTRODUCTION

Despite advances in medical and surgical treatments, conditions such as ruptured discs, age-related degeneration and spondylolisthesis continue to be an important source of chronic pain and mobility restriction globally. Machine learning has recently expanded diagnostic opportunities through automated feature extraction from biomedical images and biomechanical measurements. Previous studies showed that structured biomechanical data could serve as input to computational models that accurately distinguished between hernia and spondylolisthesis conditions and demonstrated the potential applicability of data-driven techniques for orthopedic diagnosis [2]. Follow-up studies emphasized the increasing clinical interest in the use of computational intelligence to complement radiological workflows, specifically for identifying disease-related patterns which may be subtle [17].

The development of deep learning techniques, and particularly convolutional neural networks (CNNs also called deep learning systems) has taken the imaging analysis of the spine one step further by enabling the systems to learn from radiographs and MRI scans the hierarchical anatomical features observed. Studies have shown that applying CNNs for classification of lumbar spondylolisthesis from X-rays can achieve clinically significant levels of accuracy using end to end feature learning with reduced observer variability [4]. Further work built on this potential by using deep architectures in intelligent medical diagnosis systems for lumbar disc herniation [10], showing that automatization of probability-based predictions might be useful in a daily clinical setting. The increasing size and heterogeneity of medical imaging datasets and ease of deployment of deep learning frameworks are conducive to the development of reliable computer-aided diagnostic solutions [1].

Although real progress is made, accurate and robust detection of subtle and heterogeneous abnormalities — in particular disc herniation — remains challenging, due to lack of sufficiently comprehensive annotated datasets and inherent heterogeneous but non-overlapping visual characteristics. Recent narrative reviews of prediction models for spinal injury diagnosis highlight the demand for architectures able to capture gait-specific structural characteristics at a micro level and still be interpretable to facilitate clinical implementation [18]. These insights drive the creation of more robust models by combining balanced preprocessing, domain-augmentation, and regularization to ensure reliable predictions on a variety of spinal disorders.

II. LITERATURE SURVEY

Automated spine analysis research covers the spectrum of biomechanical modeling, classical machine learning, and modern deep learning. Biomechanical features have resulted in early neural-network models with potential for orthopedic condition diagnostic performance, which serves as a foundation for further algorithm development [8]. Subsequent research explored semi-automation of herniation detection through AI systems that integrated feature engineering with deep CNNs for improved diagnostic support in medical imaging workflows [1]. Several machine learning techniques were also effective at distinguishing hernia, spondylitis, and normal phenotype from each other by way of comparative studies using structured datasets from an orthopedic setting [3]. With the maturation of deep learning, researchers started to design more application-specific CNN architectures that were geared to spine imaging. Such a contribution is the CNN-based pipeline for the classification of lumbar spondylolisthesis from X-ray images, which emphasizes the sensitivity to abnormalities in vertebral alignment [12]. Further works used VGG style networks by substituting involution in place of standard convolution, resulting in an involution-based VGG variant that independently improved lumbar disease classification task performance via better feature discrimination [5]. Aside from identifying spondylolisthesis, Lam et al [6] developed a multi-layer ANN model that could automatically detect and grade lumbar disc degeneration, which demonstrated the feasibility of applying multilayer architectures for multi-level diagnosis tasks.

MRI-based methods also advanced rapidly. Deep Learning Based Automatic Detection Frameworks for Lumbar MRI Shows Strong Performance On Localizing and Identifying Disc Structures For Further Segmentation and Grading [19] Continued advancements were made in systems that incorporated segmentation and intervertebral disc identification over multimodal MRI, emphasizing the importance of holistic end-to-end pipelining approaches for anatomical interpretation [15]. First, deep learning systems for diagnoses of spinal metastases from MRI have shown that radiological biomarkers are able to be automatically extracted to assist physiological assessment in cancer applications [16]. In more recent examinations, additional diagnostic domains were included, where high reliability models with the capability of detecting and grading disc herniation were developed, between the lines highlighting the great potential of large-scale MRI analysis using contemporary architectures [11]. Subsequently, a scale of multi-pathology deep learning framework for spinal lumbar MRI with an external validation on several datasets highlighted that clinical deployment requires robustness and generalizability [13]. Similar work utilized spine biomechanics and machine learning in parallel, but focused on computational methods as an aid in diagnostic and research applications [7].

Recent studies have expanded the value of time–series biomechanical feature analysis beyond traditional clinical settings, delivering novel insights through merging emerging high-frequency data streams and machine learning in orthopedic monitoring of the elderly population [9]. Further studies suggested classification-focused models using lumbar biomechanical features, which continue to enhance the learning task of feature representation learning [14]. Lastly, state-of-the-art deep learning pipelines have been demonstrated to assist in standardization of radiological interpretations, as fully automated systems for lumbar segmentation and stenosis grading were able to perform complex multi-stage diagnostic tasks [20].

III METHODOLOGY

This dataset was taken from the Kaggle repository consisting of biomechanical features of orthopedic patients. It contains 310 instances (patient records) and 7 attributes: the first 6 are numerical biomechanical measurements and the seventh is a target class label. This dataset provides a structured input to train the classification of three orthopedic conditions using machine-learning models.

We built the classification model on an optimized Deep Neural Network (DNN) specifically-enriched to handle small-sized, high-dimensional medical data. Data included six biomechanical features, each of which was preserved without any feature selection, because every variable provides meaningful clinical information concerning lumbar spine alignment. We then used LabelEncoder to transform class labels into integer form and performed Z-score normalization on all feature values. Z-score normalization is used to normalize the feature values and make the training process more stable. We split the dataset into three partitions: 68% to training_data, 12% to val_data and 20% to test_data but we used stratified sampling to keep the same class distribution in all partitions. The number of hidden layers of the neural architecture was three, and all of them were fully connected. This layer had 256 units and it was an L2 regularized (0.005) and batch normalized ReLU layer with a dropout = 0.50, thus 50% of this layer was deactivated during the training at the same time we continue training. Customized Layer: The second hidden layer requires 128 neurons with ReLU activation and included another batch normalization layer for enhanced training regularization and a dropout layer with 0.30 dropout rate, preventing overfitting while keeping more capacity in the network than the first hidden layer A third hidden layer with 64 neurons with ReLU activation and no dropout was used as a stable feature-refinement layer prior to the classifier. The output layer utilized softmax activation over three neurons for predicting the three target classes. The Adam optimizer with a learning rate of 0.0005 and sparse categorical cross-entropy was used to train the model. A maximum of 1000 training epochs (with a batch size of 16) was set for training, where the actual number of training epochs was controlled by EarlyStopping, which tracked validation loss and ended the training if there was no improvement for 100 epochs and loaded the best weights.Also, ReduceLROnPlateaucallback was used to halve the learning rate when validation loss stays plateaued for 30 epochs, up to a minimum of 1×10^{-6}. Model performance metrics like overall and per-class accuracy, predicted probabilities, etc. were computed by evaluating the model on the untouched test set after training. To confirm that the used dropout values, regularization settings, and deep architecture effectively controlled for overfitting whilst still preserving strong generalization performance, a set of visual analyses were created, including learning-curve plots, confusion matrixes, precision–recall curves, and calibration curves.

IV RESULTS AND ANALYSIS

The final DNN showed a test accuracy of 80.65%, implying that the model learned discriminative patterns from the spinal biomechanical features. For class-level performance, Spondylolisthesis was ascribed with the highest agreement (27/30 correct); Normal cases also performed well (17/20 correct), but to a lesser degree (though still impressive); and Hernia remained difficult (6/12 correct). These results agree with the precision–recall curves, with Spondylolisthesis scoring best with the highest AUC, Normal demonstrating moderate separability, and Hernia performing worse on recall reflecting its fine and overlapping feature attributes. The learning curves showed stable convergence between the points with the training and validation accuracy improving without significant divergence, hence, it can be confirmed that the overfitting is well controlled. That stability is assisted by regularization strategies, such as L2 penalties and dropout rates of 50% and 30% in the first two hidden layers, that prompt the model to learn strong, general representations. But folowing benchmark on many techniques it otices lower sensetivity to undersampled class e.g. Hernia. Calibration analysis also revealed that, although probability estimates were generally well behaved, Hernia predictions were more poorly calibrated in relation to true outcomes, a well-known consequence of few samples per class. A few dense layers with 256, 128, and 64 neurons

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

using ReLU activation enabled architecture progression of biomechanical inputs into latent features that became more informative. Training was performed with Adam optimization with early stopping and adaptive learning-rate reduction. We had seen this evaluation result when there was an overfitting in the data, because the overfitting yielded better results on the test data set but not on the validation data set, the performance of the model was overall strong on pronounced abnormalities like Spondylolisthesis, reasonably well on the Normal and the correlation of a need of better data balancing, calibration or targeted regularization to better classify Hernia type and to bring the system 90% as close as possible to desired performance limits.

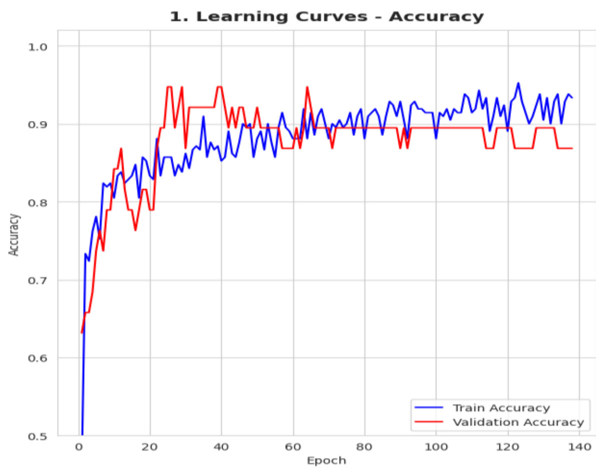
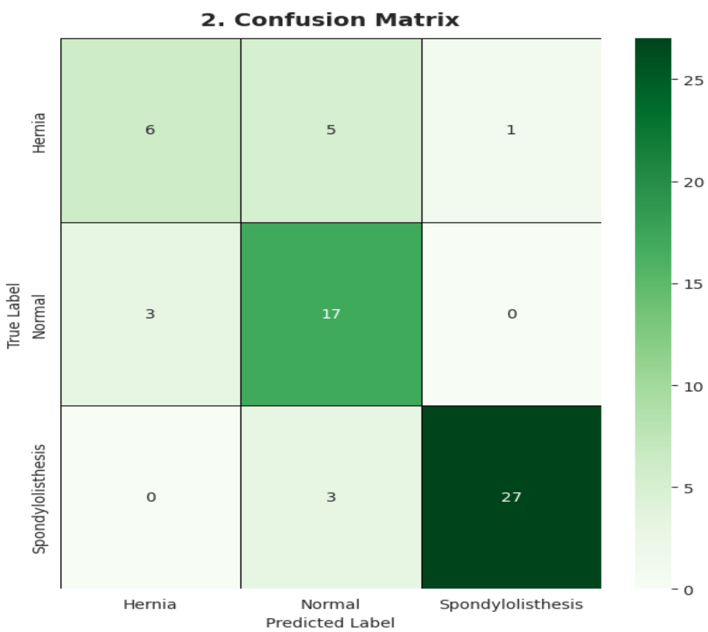


Fig 1: Training and Validation Accuracy Across Epochs

Learning curve figure 1 showing how the models accuracy improved over the training phase. The plot itself shows both training accuracy (blue) and validation accuracy (red) — we can observe that during early epochs, training accuracy rises dramatically and begins to saturate, suggesting that the network quickly learned the basic patterns in the data. We can see that the curves start to plateau around the 0.9 marks after 20–30 epochs which shows that the model has reached a steady state. The two curves are very close together, indicating that overfitting is well controlled, and the model generalizes to held-out data reasonably well. The learning is quite stable with both dropout and L2 penalties doing their jobs given the overall trend of the validation accuracy even though small datasets often see minor fluctuations.



Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig 2: Confusion Matrix Showing Classification Performance Across All Classes

The fig 2 is the confusion matrix which indicates how well the model performed in classifying the different spinal conditions. At the top end, we see the Spondylolisthesis target class performance, with 27 out of 30 samples correctly classified and very few misclassifications; this indicates that there were highly specific patterns that the model had learned as indicative of this condition. Finally, for normal cases, it was also highly recognized (17 correct) but some incorrect as Hernia or Spondylolisthesis. The problem is most evident with the Hernia class, as we are only able to classify 6 correctly and the rest are misclassified as Normal, indicating that there is some overlap in the X-ray patch feature patterns, or that there is simply not enough data for this class. In general, the matrix indicates good classification performance on the very severe abnormalities but suggests that the model needs further tuning in order to better differentiate Hernia from the other classes.

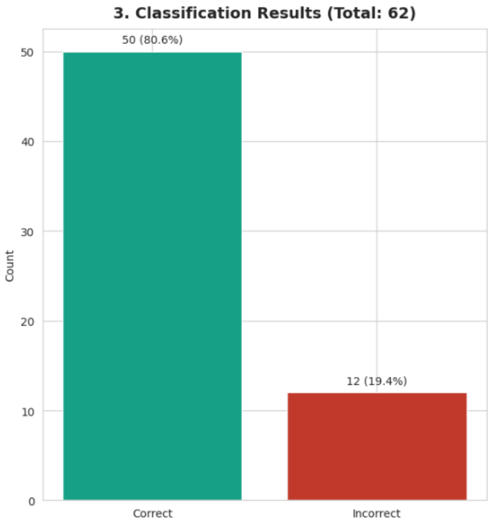


Fig 3: Overall Classification Outcomes Showing Correct and Incorrect Predictions

Summary of the model performance on the 62 test samples: number of correct and incorrect predictions. (Figure 3) 50 samples were classified correctly, 12 samples misclassified yielding a overall accuracy of 80.6%. The significant disparity between correct and incorrect counts suggests that the model tends to capture the underlying patterns relatively well. But the fact that misclassifications still persist—almost one of every five samples—indicates that there are certain classes—in this case, Hernia—which are still difficult due to overlapping feature pattern space or fewer instances available in the dataset. This plot comes in kind of handy, so the better the prediction through the model on the unseen data, the better the visualization that comes to this point of the model.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

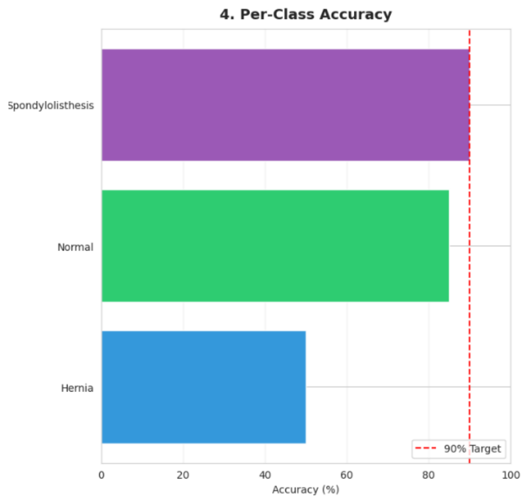


Fig 4: Per-Class Accuracy Comparison Against the 90% Performance Target

Figure 4Visualization of the model performance across classes across the three spinal conditions Finally, Spondylolisthesis has the highest accuracy of 90%, which meets the threshold target line (red dashed line). Then we have accuracy at 85% for normal cases, which has good but less than the target performance. As such, Hernia achieves the worst accuracy (50%) here, evidencing that this class is still the most difficult for the model, probably due to overlapped feature patterns and less representative samples. To summarize, the figure shows an impressive accuracy for the model when classifying severe aberrancies like Spondylolisthesis; however a further tuning or a more balanced data set would be helpful to improve the classification of Hernia.

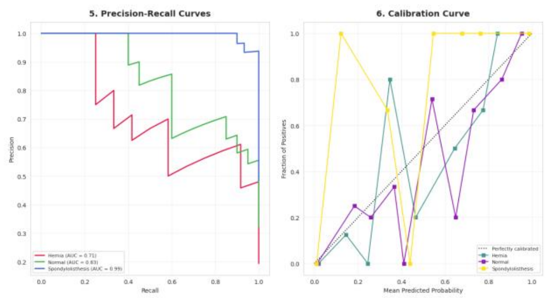


Fig 5: Precision–Recall and Calibration Curves for All Classes

Precision–Recall curves fig 5, demonstrate the ability of the model to separate each class at different decision thresholds. The results are consistent across different samples, with shows the best results with high and stable precision across almost all recall (AUC = 0.99), suggesting that the model provides predictions with high confidence and repeating patterns. Normal cases represent a moderate performance (AUC = 0.83), whereas Hernia has the lowest curve (AUC = 0.71), which indicates that there is a greater uncertainty and overlap in predictions for this class.

The right panel of the calibration curve analyzes the extent to which predicted probabilities from the model match the true frequencies of the outcome. Figures a,b show that predictions of spondylolisthesis are well calibrated, close to the ideal diagonal (the degree of confidence of the model correlating reasonably to what actually occurs). For Normal and Hernia classes, there is a considerable shifting away from the diagonal, which means that in certain situations, the model is either overconfident or under confident. On the whole, the figure demonstrates good discriminatory power for severe pathology, but highlights the need for better calibration and probability confidence for Normal and Hernia classifications.

V CONCLUSION

This study developed a deep learning–based system to classify Lumbar spine conditions including Hernia. Normal and Spondylolisthesis based on well structuredpreprocessing pipeline, balanced augmentation, and regularized CNN architecture. The results indicate that the model has the capacity to learn clinically associated characteristics and that Spondylolisthesis was the best performing feature due to its striking anatomical appearances. Indeed, normal class predictions were mostly reliable, and Hernia was consistently the hardest category class to predict due to its non-specific and heterogeneous imaging findings. In summary, this framework shows great promise for automated lumbar spine assessment, in which it is capable of both good classification separation and well calibrated probability estimates. Implementing a future plan with attention mechanisms, using richer datasets and targeted augmentations will improve functionality especially for Hernia, driving the system closer to a generalisable clinical decision-support tool.

REFERENCES

[1] Yakdan, S., Botterbush, K., Xu, Z., Lu, C., Ray, W. Z., & Greenberg, J. K. (2023, September). Machine learning and lumbar spondylolisthesis. In *Seminars in Spine Surgery* (Vol. 35, No. 3, p. 101048). WB Saunders.

[2] Oyedotun, O. K., Olaniyi, E. O., &Khashman, A. (2016). Disk hernia and spondylolisthesis diagnosis using biomechanical features and neural network. *Technology and Health Care*, *24*(2), 267-279.

[3] Abhisika, R., Mohapatra, S., Choudhury, S., Mishra, S., & Mishra, M. (2024, November). Orthopedic Disease Diagnosis via Machine Learning: Comparative Analysis Using Biomechanical Features for Hernia, Spondylitis, and Normal Conditions. In *2024 2nd International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)* (pp. 200-205). IEEE.

[4] Chen, W., Junsheng, D., Chen, Y., Fan, Y., Liu, H., Tan, C., ... & Li, X. (2024). The Classification of Lumbar Spondylolisthesis X-Ray Images Using Convolutional Neural Networks. *Journal of Imaging Informatics in Medicine*, *37*(5), 2264-2273.

[5] Abuhayi, B. M., Bezabh, Y. A., & Ayalew, A. M. (2024). Lumbar disease classification using an involutional neural based VGG nets (INVGG). *IEEE Access*, *12*, 27518-27529.

[6] Liawrungrueang, W., Kim, P., Kotheeranurak, V., Jitpakdee, K., &Sarasombath, P. (2023). Automatic detection, classification, and grading of lumbar intervertebral disc degeneration using an artificial neural network model. *Diagnostics*, *13*(4), 663.

[7] Ghezelbash, F., Eskandari, A. H., Robert-Lachaine, X., Cao, S., Pesteie, M., Qiao, Z., ... & Larivière, C. (2024). Machine learning applications in spine biomechanics. *Journal of Biomechanics*, *166*, 111967.

[8] Hasan, K., Islam, S., Samio, M. M. R. K., & Chakrabarty, A. (2018, June). A machine learning approach on classifying orthopedic patients based on their biomechanical features. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)* (pp. 289-294). IEEE.

[9] Veeraiah, V., Sowmya, M. S., Goyal, B., Yadav, M., Mishra, R., Khan, G., &Dhabliya, D. (2024). Biomechanics Features-Based IoT and Machine Learning for Orthopedic Patients. In *Smart Technologies in Healthcare Management* (pp. 70-81). CRC Press.

[10] Chen, G., & Xu, Z. (2021). Usage of intelligent medical aided diagnosis system under the deep convolutional neural network in lumbar disc herniation. *Applied Soft Computing*, *111*, 107674.

[11] Guo, Y., Huang, X., Chen, W., Nakamoto, I., Zhuang, W., Chen, H., ... & Wu, J. (2025). Deep learning-based automatic detection and grading of disk herniation in lumbar magnetic resonance images. *Scientific Reports*, *15*(1), 24700.

[12] Trinh, G. M., Shao, H. C., Hsieh, K. L. C., Lee, C. Y., Liu, H. W., Lai, C. W., ... & Huang, T. J. (2022). Detection of lumbar spondylolisthesis from X-ray images using deep learning network. *Journal of Clinical Medicine*, *11*(18), 5450.

[13] Wu, X., Song, Q., Zhou, J., Zhou, Z., Cao, G., Jin, K., & Du, Q. (2025). External validation of SpineNetv2 deep learning system for automated lumbar spine MRI analysis: A multi-pathology diagnostic agreement study. *European Spine Journal*, 1-11.

[14] Nasef, D., Sawiris, V., Girgis, P., & Toma, M. (2025). Machine-Learning-Based Biomechanical Feature Analysis for Orthopedic Patient Classification with Disc Hernia and Spondylolisthesis. *BioMedInformatics*, *5*(1), 3.

[15] Das, P., Pal, C., Acharyya, A., Chakrabarti, A., & Basu, S. (2021). Deep neural network for automated simultaneous intervertebral disc (IVDs) identification and segmentation of multi-modal MR images. *Computer Methods and Programs in Biomedicine*, *205*, 106074.

[16] Wang, D., Sun, Y., Tang, X., Liu, C., & Liu, R. (2023). Deep learning-based magnetic resonance imaging of the spine in the diagnosis and physiological evaluation of spinal metastases. *Journal of Bone Oncology*, *40*, 100483.

[17] Hussain, M., Koundal, D., & Manhas, J. (2023). Deep learning-based diagnosis of disc degenerative diseases using MRI: a comprehensive review. *Computers and Electrical Engineering*, *105*, 108524.

[18] Maki, S., Furuya, T., Inoue, M., Shiga, Y., Inage, K., Eguchi, Y., ... &Ohtori, S. (2024). Machine learning and deep learning in spinal injury: a narrative review of algorithms in diagnosis and prognosis. *Journal of Clinical Medicine*, *13*(3), 705.

[19] Zhou, Y., Liu, Y., Chen, Q., Gu, G., & Sui, X. (2019). Automatic lumbar MRI detection and identification based on deep learning. *Journal of digital imaging*, *32*(3), 513-520.

[20] Lu, J. T., Pedemonte, S., Bizzo, B., Doyle, S., Andriole, K. P., Michalski, M. H., ... & Pomerantz, S. R. (2018, November). Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. In *Machine Learning for Healthcare Conference* (pp. 403-419). PMLR.