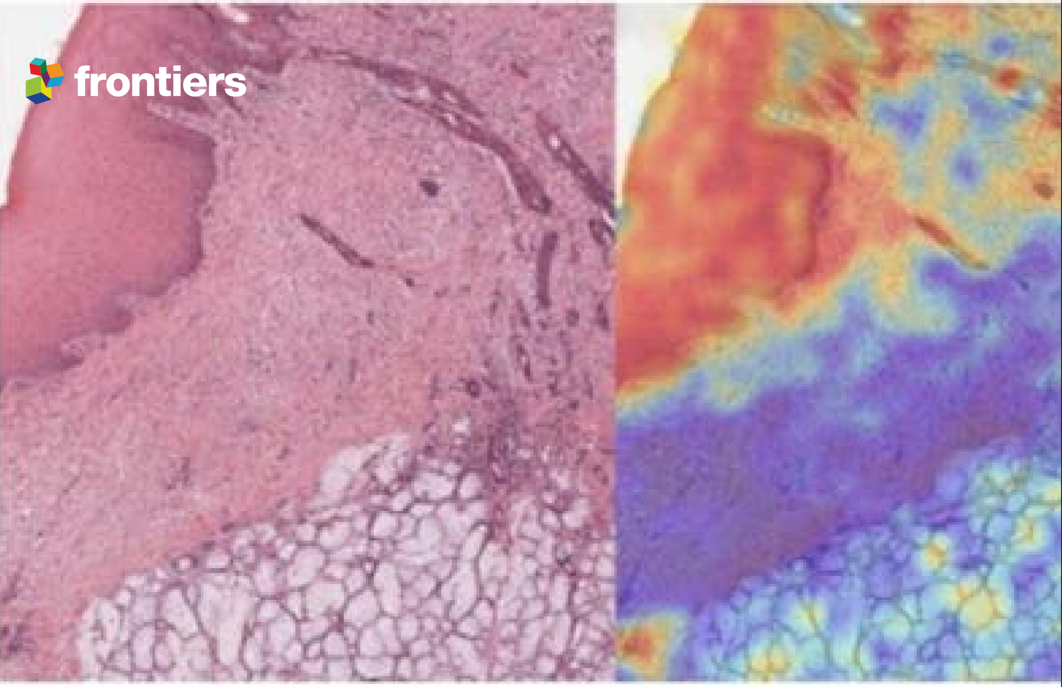
Multiple sclerosis diagnosis with deep learning and explainable AI

Nighat Bibi
Technological University Dublin, d22125041@mytudublin.ie

Jane Courtney
Technological University Dublin, Ireland

Kathleen M. Curran
University College Dublin, Ireland

Funder: Science Foundation Ireland Center for Research Training in Machine Learning



Medical Image Understanding and Analysis

Manchester, UK

Multiple sclerosis diagnosis with deep learning and explainable AI

Author
Nighat Bibi – School of Electrical and Electronic Engineering, Technological University Dublin, Dublin, Ireland
Jane Courtney – School of Electrical and Electronic Engineering, Technological University Dublin, Dublin, Ireland
Kathleen M. Curran – University College Dublin, School of Medicine, UCD Belfield, Dublin 4, Ireland

Citation
Bibi, N., Courtney, J., Curran, M.K. Multiple sclerosis diagnosis with deep learning and explainable AI.

Abstract
Diagnosing multiple sclerosis (MS) presents significant challenges due to its complex clinical presentation and the subjective interpretation of imaging findings. Machine learning (ML) and deep learning (DL) models, despite their potential, often exacerbate these challenges with their opaque decision-making processes, hindering clinical integration. This study addresses these limitations by employing eXplainable Artificial Intelligence (XAI) techniques, specifically integrating Grad-CAM within a Convolutional Neural Network (CNN) framework, EfficientNetB1, for the diagnosis of MS. The primary objective is to enhance the transparency and reliability of MS diagnosis by providing clear visual insights into the model's decision-making process, while also identifying and mitigating potential biases and irrelevant features. Using a dataset comprising FLAIR axial and sagittal MRI images of MS patients and healthy individuals, the CNN model is trained and integrated with Grad-CAM. Post-integration observations revealed

potential biases and irrelevant features, particularly in the erroneous highlighting of certain regions by the model. Subsequent adjustments and re-training using 10-fold cross-validation led to an improved model with accuracy rates of 99.82% for axial, 99.76% for sagittal images, and 99.36% overall. Furthermore, testing on a separate dataset confirmed the model's ability to generalize and perform well across various clinical contexts. In conclusion, this study underscores the critical role of transparent and interpretable models in medical diagnostics, demonstrating that the integration of XAI techniques can significantly enhance the reliability and clinical applicability of models.

Dataset

The dataset used in this study is obtained from a research article published in 2021, collected from Ozal University Medical Faculty [4]. It consists of FLAIR MRI images of the brains of 72 patients diagnosed with MS and 59 healthy individuals. The dataset is organized into four groups: MS-Axial (n = 650), MSSagittal (n = 761), Healthy-Axial (n = 1002), and Healthy-Sagittal (n = 1014). To evaluate the generalizability of the proposed model, an additional publicly available dataset by Muslim et al. [5] is used, which includes FLAIR MRI images from a different population.

Methods & Results

For the classification of MRI images to diagnose MS, a series of steps were taken. Initially, all images were resized to a fixed dimension of 240x240 pixels to ensure uniformity in data processing. The architecture chosen for the classification task is EfficientNetB1 [7], selected for its proven efficacy in image classification and computational efficiency. Using transfer learning, the model leveraged pre-trained weights from the ImageNet dataset to extract meaningful features from the MRI images. The model architecture consisted of a base EfficientNetB1 model followed by custom layers tailored for classification purposes. To assess the robustness of the model performance, a 10-fold cross-validation strategy is implemented. This involved training the model on nine folds of the data while validating the remaining fold for each iteration. Following cross-validation, the final model is trained on the entire dataset to maximize its learning potential for subsequent evaluation and testing on unseen data. During the initial 10-fold cross-validation process, the model achieved an accuracy of 98.13% for the overall dataset, combining both axial and sagittal images. Gradient-weighted Class Activation Mapping (Grad-CAM) [6] is integrated to provide visual explanations for the model's predictions. Grad-CAM highlighted regions of the input image that contributed the most to the model's decision-making process. Initially, certain regions on the left side of the images were incorrectly highlighted by the model. To address this issue, images were cropped to focus solely on relevant brain regions. Subsequently, the model is re-trained on the cropped images, leading to improved accuracy. Grad-CAM is then reapplied, effectively highlighting the relevant regions in the MRI images (shown in Figure 1).

To evaluate the generalizability of the model, a separate set of images collected from a different source [5] is used. The model tested on some random images and made accurate predictions on these images and Grad-

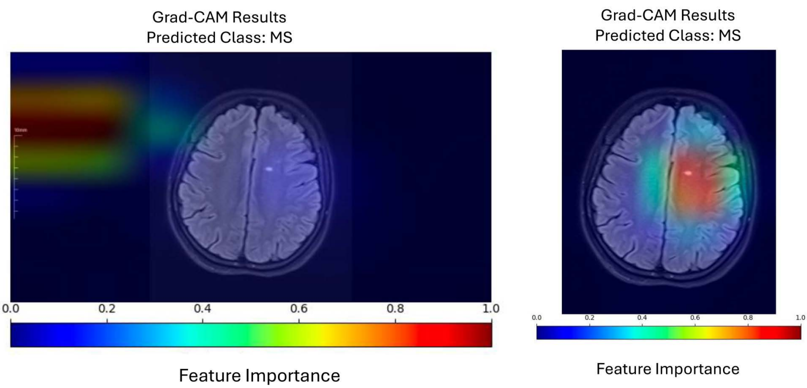


FIGURE 1
Grad-CAM visualisation for MS class before cropping and after cropping the images.

CAM effectively highlighted the regions that influence classification decisions (shown in Figure 2). In general, the integration of Grad-CAM improved the diagnostic capabilities of the model by accurately highlighting regions indicative of Multiple Sclerosis. The visual insights provided by Grad-CAM facilitated better understanding and trust in the model's decision-making process. Upon completion of the 10-fold cross-validation, the proposed CNN architecture achieved a final accuracy of 99.82% for axial images, 99.76% for sagittal images, and 99.36% overall. Furthermore, a comparative analysis demonstrated in Table 1 with existing studies showcased the competitive performance of the proposed model.

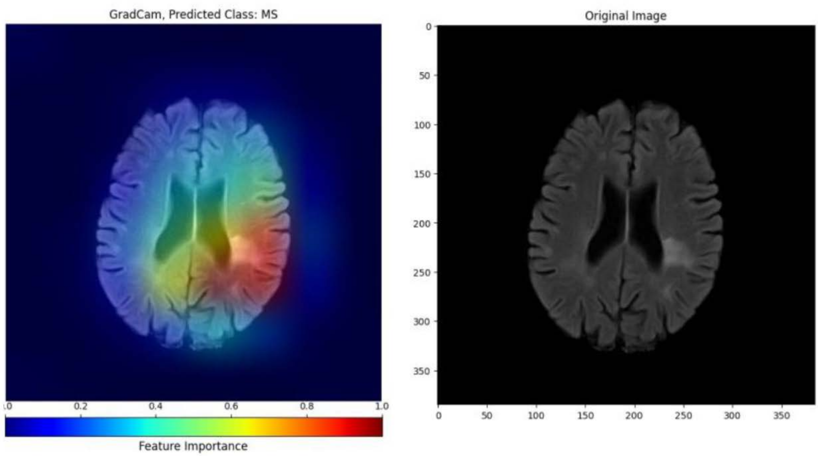


FIGURE 2
Grad-CAM visualisation for MS class on an image from Muslim et al. [5].

TABLE 1: Comparison of model performance with existing studies

Study	Dataset	Accuracy (%)	Methodology	XAI Method
Eitel et al. [1]	Overall dataset	87.04	CNN	LRP
Lopatina et al. [3]	Overall dataset	91 to 95	CNN	Attribution algorithms
	Axial plane images	98.37	(ExMPLPQ,	
Macin et al. [4]	Sagittal plane images	97.75	INCA,	No
	Overall dataset	98.22	KNN)	
	Axial plane images	99.76	(exemplar MobileNetV2,	
Ekmekyapar et al. [2]	Sagittal plane images	99.48	IMrMr,	No
	Overall dataset	98.02	KNN)	
	Axial plane images	**99.82**		
Our Work	**Sagittal plane images**	**99.76**	**EfficientNetB1**	**Grad-CAM**
	Overall dataset	**99.36**		

Conclusion

In conclusion, this work presents an approach to the diagnosis of multiple sclerosis using eXplainable Artificial Intelligence techniques to improve the explainability of the model. Achieving higher classification accuracy than existing work on the same dataset with Grad-CAM integration provides valuable insights into the decision-making process, highlighting potential biases and irrelevant features learned by the model. By addressing these issues, this research underscores the importance of transparent and interpretable models in medical diagnostics. Furthermore, it demonstrates robust performance on a separate unseen dataset and affirms its generalizability.

Looking ahead, future research will explore the incorporation of multiple modalities and advanced XAI techniques to further improve diagnostic accuracy and broaden the scope of our findings. Clinical user evaluation will also be performed in collaboration with clinicians to develop methods and metrics toward clinical interpretability.

Acknowledgements
This research is funded by the Science Foundation Ireland Center for Research Training in Machine Learning (18/CRT/6183).

References

[1] Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A.U., Ruprecht, K., Giess, R.M., Kuchling, J., Asseyer, S., Weygandt, M., Haynes, J.D., et al.: Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional mri using layer-wise relevance propagation. NeuroImage: Clinical 24, 102003 (2019)

[2] Ekmekyapar, T., Taşcı, B.: Exemplar mobilenetv2-based artificial intelligence for robust and accurate diagnosis of multiple sclerosis. Diagnostics 13(19), 3030 (2023)

[3] Lopatina, A., Ropele, S., Sibgatulin, R., Reichenbach, J.R., Güllmar, D.: Investigation of deep-learning-driven identification of multiple sclerosis patients based on susceptibility-weighted images using relevance analysis. Frontiers in neuroscience 14, 609468 (2020)

[4] Macin, G., Tasci, B., Tasci, I., Faust, O., Barua, P.D., Dogan, S., Tuncer, T., Tan, R.S., Acharya, U.R.: An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: Exmplpq. Applied Sciences 12(10), 4920 (2022)

[5] Muslim, A.M., Mashohor, S., Al Gawwam, G., Mahmud, R., binti Hanafi, M., Alnuaimi, O., Josephine, R., Almutairi, A.D.: Brain mri dataset of multiple sclerosis with consensus manual lesion segmentation and patient meta information. Data in Brief 42, 108139 (2022)

[6] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)

[7] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)