ISSN: 2182-2069 / E-ISSN: 2182-2077

Self-Supervised Learning Approach for Early Detection of Rare Neurological Disorders in MRI Data

Nadia Ibrahim Nife[1*], Zaidoon Khalaf Mahmood[2], Layth Hammood[3], Alaan Ghazi[4], Aras Aldawoodi[5], and Mustafa Albdairi[6]

[1*]College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. nadia.ibra@uokirkuk.edu.iq, https://orcid.org/0000-0002-4006-3773

[2]University of Kirkuk, Iraq. zaidoon.kh.mahmood@uokirkuk.edu.iq, https://orcid.org/0000-0002-9364-3069

[3]College of Computer Science and Information Technology, Kirkuk University, Kirkuk, Iraq. layth.aljumaily@uokirkuk.edu.iq, https://orcid.org/0000-0002-7439-858X

[4]Department of Electronic and Control Technologies Engineering, Northern Technical University, Kirkuk, Iraq; Department of Computer Technical Engineering, Al-Qalam University College, Kirkuk, Iraq. alaan.ghazi@ntu.edu.iq, alan@alqalam.edu.iq, http://orcid.org/0000-0002-1265-0428

[5]Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. aras.ghazi86@uokirkuk.edu.iq, https://orcid.org/0000-0001-6629-6455

[6]Department of Civil Engineering, AL-Qalam University College, Kirkuk 36001, Iraq. mustafa.albdairi@alqalam.edu.iq, https://orcid.org/0009-0002-6673-363X

Received: November 22, 2024; Revised: January 04, 2025; Accepted: January 25, 2025; Published: February 28, 2025

Abstract

This study presents a Self-Supervised Convolutional Neural Network (CNN) model designed for the early detection and classification of rare neurological disorders using brain MRI images. Early identification of these disorders is critical for effective intervention and improved patient outcomes. Our approach leverages self-supervised learning to enhance the model's ability to extract meaningful features from brain MRI data, even with limited labeled samples. The pipeline includes stages such as data preprocessing, augmentation, model compilation, and training, which help improve the model's accuracy and generalization. The model's performance was evaluated using a variety of metrics, including accuracy, precision, recall, and F1-score, showing promising results in distinguishing between "Neurological Disorder" and "No Neurological Disorder" classes. Notably, the model achieved high recall for the disorder class, underscoring its effectiveness in identifying positive cases. This self-supervised CNN framework offers a significant advancement for early classification of neurological conditions in MRI images, especially in scenarios with limited labeled data. Its successful application in this context highlights its potential for clinical integration, providing a valuable tool for neurologists in diagnosing rare disorders early and enhancing patient care.

Journal of Internet Services and Information Security (JISIS), volume: 15, number: 1 (February), pp. 153-169. DOI: *10.58346/JISIS.2025.I1.010*

*Corresponding author: College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq.

Keywords: Convolutional Neural Network, Self-Supervised Learning, Early Detection, Rare Neurological Disorders, Classification, Brain MRI Images.

1 Introduction

1.1 Background

Neurological disorders represent a wide grouping of disorders affecting the central and peripheral nervous systems, from the most common to the rarest diseases (Teleanu et al., 2022; Smiyan et al., 2024; Qusef et al., 2023). Most of these disorders present with profound cognitive, motor, and sensory disturbances; some of them have a rather rapid course, while others develop almost unnoticeably over several years (Sivakumar et al., 2021) (Basanta Kumar & Sunil, 2024; Aarsland et al., 2021; Al-Dawoodi & Mahmuddin, 2017). Rare neurological disorders, in particular, present unique diagnostic challenges due to their low prevalence and often subtle initial symptoms (Marwaha et al., 2022). This can lead to delay in diagnosis, thus treatment, and further deterioration in patient outcomes and quality of life (Neal et al., 2015; Almusawi et al., 2024). Thus, early identification of these conditions is extremely important not only for better prognosis but also to ease the burden on congested health facilities that struggle with an increased demand for specialized diagnostic services (Ghazi et al., 2021; Veera Boopathy et al., 2024; Qusef et al., 2023; Ball & Balogh, 2016).

Magnetic Resonance Imaging (MRI) is one of the major diagnostic tools existing in neurological disorders, which allows for the derivation of detailed views on brain structures without resorting to invasive procedures (Bigler, 2015; Basanta Kumar & Sunil, 2024). It is this capability of showing subtle changes in brain anatomy that is extremely important to detect early signs and symptoms of neurological diseases that could hardly be diagnosed by other techniques (Petrella et al., 2003; Yadav et al., 2024). On the other hand, MRI image interpretation requires extensive expertise and may be time-consuming (Despotović et al., 2015). The manual nature of MRI evaluation also introduces a degree of variability across interpretations, with diagnostic accuracy often influenced by the experience level of the radiologist (Kuhl, 2007). For rare neurological disorders, this inter-observer variability can further exacerbate diagnostic delays and inaccuracies, especially in cases where imaging patterns are unfamiliar or ambiguous due to limited prior cases (Waldemar et al., 2007).

With the advent of artificial intelligence (AI), more precisely deep learning, has already entered the field, medical imaging research is only in the middle of its revolution (Sejnowski, 2018; Mustapha et al., 2021). Deep learning models, especially CNNs, have been very successful in automatically discovering complex patterns from image data; hence, they are particularly suitable for medical applications relying on the expert's visual observation (Litjens et al., 2017; Tamil Vanan et al., 2019; Mohammed et al., 2024). Medical application areas ranging from X-ray-based diagnosis to histopathology image analysis keep increasing diagnostic efficiency and consistency (Abdullah, 2024). Radiology and pathology are the most accessible application areas for CNNs (Yasaka et al., 2018; Biswas & Tiwari, 2024). By leveraging large-scale data, these models are capable of learning nuanced image features that may not be apparent to the human eye, thus enhancing diagnostic precision (Asif et al., 2024). However, CNNs traditionally rely on large, labeled datasets to achieve high performance, which is a considerable limitation when working with rare conditions where labeled data is scarce (Tajbakhsh et al., 2020).

This, however, begs its silver bullet in terms of self-supervised learning. In self-supervised learning, the models are allowed to extract meaningful representations from data by engaging in what are generally termed pretext tasks: preliminary objectives that train the model to recognize patterns well

before any labeled data can be provided (Almusawi et al., 2024; Brodley & Friedl, 1999). This has been a successful learning paradigm for applications where labeled data is either expensive or hard to procurator, say, natural language processing and computer vision (Haney, 2020). Particularly in the case of medical images, self-supervised learning offers the most valuable advantage due to huge MRI datasets on which the model can be trained without time-consuming labeling processes (Shurrab & Duwairi, 2022). This capability is particularly useful for rare neurological disorders, where annotated datasets are limited and difficult to generate (Bras et al., 2012).

Most importantly, the integration of self-supervised learning with CNNs will have great implications for treatments concerning neurological disorders (Jiang, 2023). The advantage of leveraging unlabeled MRI data in learning complex features allows self-supervised CNN models to have the pre-training ability on large-scale data and subsequent fine-tuning using small labeled data for conditions that are rare (Rani et al., 2024). This approach will lead to not only a reduction in reliance on labeled data but also an enhanced generalization capability of the model, hence being more effective for identifying subtle and early signs related to rare neurological disorders (Albers et al., 2015). Ultimately, this integration of self-supervised learning in MRI analysis could lead to more reliable, accessible, and timely diagnostics, especially in cases where standard methods fall short (Jiang, 2023). As such, self-supervised CNN models stand at the frontier of precision medicine, with the capacity to transform diagnostic practices into neurology by addressing some of the most pressing limitations in current imaging-based diagnostic approaches (Hussain et al., 2024).

1.2 Problem Statement

Several factors contribute to poor diagnosis using traditional methods for rare neurological disorders. The foremost is the scarcity of labeled data, hence making it difficult to train conventional supervised models, which strongly limits their generalization performance in clinical applications. Secondly, many existing models struggle to capture subtle patterns in MRI data that may be indicative of early-stage disease progression, which is crucial for timely intervention (Devi et al., 2016).

Manual annotation of MRI images for rare conditions is resource-intensive and often requires specialized knowledge further limiting high-quality, labeled datasets. A huge bottleneck is thus created that restricts the effectiveness of purely supervised techniques and inherently calls out for models that can leverage unlabeled data to develop meaningful feature representations. Self-supervised learning offers a potential solution to these challenges by allowing models to pre-train on unlabeled data and thereby reduce dependency on extensive annotations, but its application in rare neurological disorder detection is underexplored.

1.3 Objectives

The present research is motivated by this challenge and focuses on the development of a self-supervised CNN model for the early classification of rare neurological disorders using brain MRI data. The aims are to:

1. Position a self-supervised learning framework that will effectively allow the extraction of features from unlabeled MRI images by focusing on patterns associated with neurological disorders that are atypical.
2. The model is fine-tuned on a small labeled dataset to give better discrimination in classifying "Neurological Disorder" or "No Neurological Disorder.".

3. Model performance evaluation on metrics of precision, recall, and F1-score would allow one to appraise its reliability and clinical applicability.
4. Providing insight into the practical utility of self-supervised CNNs on medical imaging in settings where labeled data is limited.

These goals are in concurrence with the ultimate aim of allowing the diagnosis of rare neurological disorders in patients with more accuracy and at an earlier stage, since both these factors often materially impact the treatment outcome and quality of life.

1.4 Research Contribution

The main contributions of this study are several important ones in the field of medical imaging and machine learning, The contributions are as follows:

1. Model development: It describes a novel CNN model adapted for the detection of rare neurological disorders in a self-supervising manner with a strong focus on feature extraction from MRI data in a manner not requiring extensively labeled datasets.
2. Application of Self-Supervised Learning into the Detection of Rare Diseases. Though self-supervised learning has seen success recently in domains such as natural language processing and general computer vision, its application to the classification of neurological disorders in medical imaging remains scant. This work further closes that gap.
3. Improved Diagnostic Accuracy and Consistency: The model developed at present aims at increasing not only the accuracy but also diagnostic consistency by reducing the variability present in the case of manual MRI interpretation.
4. Clinical Relevance: High recall of the model for the disorder class reflects its clinical value and forms the main emphasis in early detection. Indeed, results have already shown that such a model can help neurologists in the more accurate detection of rare pathologies and provide new perspectives for future applications in clinical settings of self-supervised learning.

In this work, each of these aspects is elaborated upon, adding a different perspective to the intersection of self-supervised learning and medical imaging by providing a scalable solution adaptable to similar challenges in other diagnostic contexts.

2 Literature Review

Early detection and classification of neurological disorders with MRI and advanced approaches of machine learning have been an area of active interest in recent years (Iqbal et al., 2024). Various methodologies have been developed to overcome some challenges regarding labeled data shortage and subtlety of early disease markers either from deep learning or self-supervised learning (Azizi et al., 2023; Nife & Chtourou, 2022). Various works have presented different hybrid learning models, anomaly detection, and contrastive learning schemes for classifying rare neurological disorders that ensure the diagnosis is both more accurate and reliable (Lima et al., 2022). New trends in MRI technologies, including fMRI and DTI, also show promise in further improving diagnostic outcomes by capturing detailed neural structure (Du et al., 2024). This literature review synthesizes key studies in the field, focusing on their objectives, methodologies, and findings, which contribute to a deeper understanding of machine learning's role in enhancing neurological disorder diagnostics (Ghazi et al., 2023). The table 1 below summarizes these studies, highlighting the evolving techniques and their effectiveness in various medical imaging applications.

Table 1: Summary of Key Studies on Machine Learning Approaches for Neurological Disorder Diagnosis

Reference	Year	Study Objective	Methodology/Tools Used	Key Findings
(Sun et al., 2021)	2021	To classify rare diseases using unsupervised representation learning and pseudo-labeling.	Hybrid approach with unsupervised representation learning and pseudo-label supervised self-distillation.	Hybrid approach outperforms few-shot learning, setting a new standard in rare disease classification.
(Jiang et al., 2021)	2021	To improve prognosis of Parkinsonâ€™s disease using self-supervised learning for anomaly detection.	Semi-supervised anomaly detection, self-supervised learning for signal representation.	Effective prognosis modeling for PD using self-supervised anomaly detection on large dataset.
(Deng & Li, 2022)	2022	To enhance anomaly detection in MRI images using self-supervised learning with random pseudo-outliers.	Self-supervised learning with random pseudo-outliers for training without abnormal samples.	The method outperforms state-of-the-art solutions in MRI anomaly detection.
(Mishra et al., 2022)	2022	To diagnose Parkinsonâ€™s disease using a CNN-based system on brain MRI scans.	CNN-based CAD system utilizing substantia nigra region of brain MRIs.	Achieved 99.5% accuracy in Parkinsonâ€™s diagnosis, reducing diagnostic time.
(Fedorov et al., 2024)	2024	To analyze multimodal neuroimaging data for discovering disorder-relevant brain regions.	Self-supervised multimodal framework with Deep InfoMax for learning representations.	Self-supervised model identifies relevant brain regions and multimodal links.
(Siddiquee et al., 2024)	2024	To detect neurologic diseases in unannotated brain MRIs using a GAN-based method.	GAN-based image-to-image translation with pseudo-AUC for unsupervised detection.	GAN-based Brainomaly achieves high precision in unsupervised neurologic disease detection.
(Chakrabortty et al., 2023)	2023	To detect Alzheimerâ€™s early using MRI and a lightweight CNN model with data balancing techniques.	Lightweight CNN model and SMOTETomek for data balancing.	Lightweight CNN model achieves 97.5% accuracy in early Alzheimerâ€™s detection.
(He et al., 2024)	2024	To detect a broad range of disorders without disorder data using inverse supervised learning.	Inverse Supervised Learning (ISL) algorithm with disorder-free CT scans.	ISL generalizes well, achieving high AUC values without disorder data.
(Rana et al., 2023)	2023	To improve early Alzheimerâ€™s detection using transfer learning on MRI data.	Transfer learning with deep learning model on MRI for Alzheimerâ€™s detection.	Achieves 97.31% accuracy, outperforming state-of-the-art Alzheimerâ€™s models.
(Gryshchuk et al., 2024)	2024	To classify neurodegenerative disorders using self-supervised learning and contrastive loss.	Contrastive self-supervised learning with T1-weighted MRI scans.	SSL model achieves high accuracy for AD and BV classification without expert labels.
(Govindarajan et al., 2024)	2024	To detect neurodegenerative conditions early by integrating neuroimaging with clinical data.	Deep learning framework with CNN and RNN integrating neuroimaging and longitudinal data.	Model achieves 92% accuracy in early neurodegenerative condition detection.

Despite the continuous development in machine learning and diagnosis based on MRI, perfect classification for neurologic disorders remains a challenge due to the unavailability of labeled data. Although self-supervised and hybrid learning models with great performance for extracting meaningful features from unlabeled datasets exist, most studies normally focus on general neurological conditions or common disorders with bigger datasets. This leaves a gap in the application of these techniques that are specifically tailored for conditions like these rare neurological disorders, where early detection is critical yet challenging. Current models may also not fully capture the subtle imaging patterns required for the appropriate classification of such rare conditions, particularly without large, labeled datasets that would guide and facilitate the learning process. Therefore, there is a pressing need for a robust self-supervised learning framework that can address data scarcity while achieving high diagnostic accuracy for these rare but impactful neurological conditions.

This paper tries to bridge the gap by proposing a novel self-supervised CNN for the classification of brain MRI data with diagnoses related to rare neurological disorders. In our approach, self-supervised learning using contrastive loss is utilized to train robust representations from unlabeled MRI images only, a significantly reduced dependency on annotated datasets. Unlike previous studies, our model is tailored for unique challenges presented by the rare disorder domain: subtle early-stage markers of and general limited data availability. Fine-tuning on a rather small, labeled dataset, our model achieves very high accuracy in distinguishing between the classes "Neurological Disorder" and "No Neurological Disorder". This work not only contributes to the field by providing a new diagnostic tool that is both effective and accessible in data-limited settings but also sets a foundation for further research into self-supervised methods tailored to rare disease detection in clinical applications.

3 Methods

This methodology outlines a rigorous and structured approach for developing a CNN to detect neurological disorders from MRI images. Initially, the dataset, sourced from brain-mri-images/GAN-Training Images, undergoes thorough preprocessing. Each MRI image was resized to a fixed resolution of 128x128 pixels and changed into grayscale to achieve consistency in the dataset. All images are then normalized into the range between 0 and 1, hence improving model performance with more guaranteed convergence during training. Labels are programmatically extracted based on filename conventions. A label of 1 indicates the presence and 0 the absence of a neurological disorder, hence structuring the data for binary classification. We then split the dataset into 80% for training and 20% for validation to make sure that the model evaluation process is well-robust. We perform data augmentation to increase the generalization capability of this model and evade overfitting using the Image Data Generator from TensorFlow. This will involve an augmentation including random rotations, translations, shearing, zooming, and horizontal flipping. These augmentations convincingly model more varieties of possible real-world variations, hence allowing the model to learn invariant features across diverse input scenarios in figure 1.

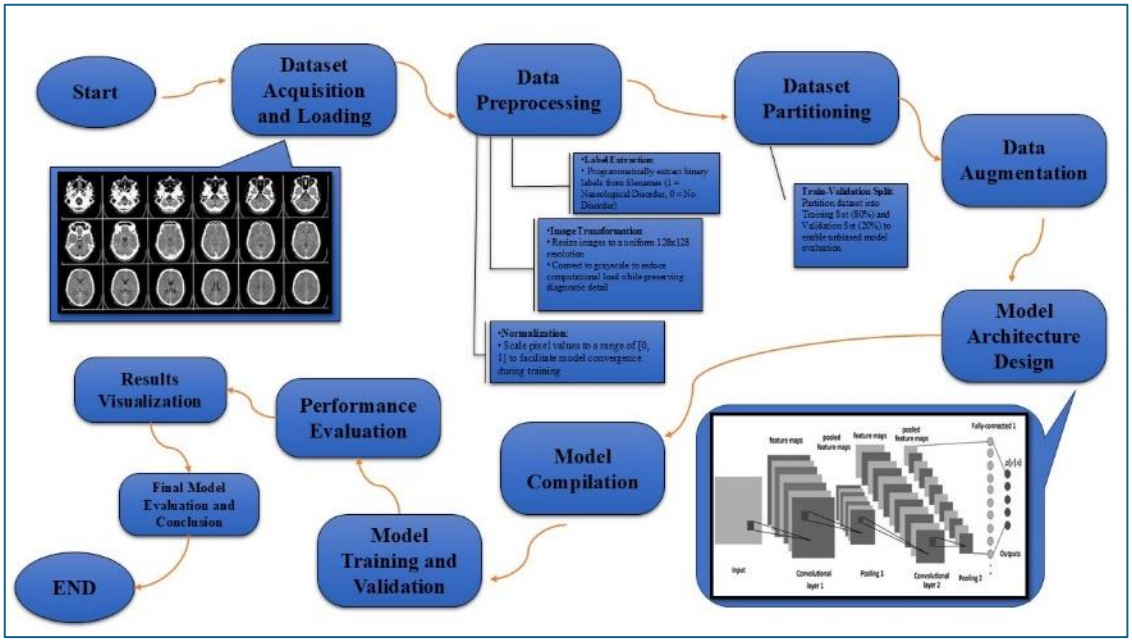

Figure 1: Flowchart for CNN-Based MRI Image Classification for Neurological Disorder Detection

The approach has been based on the CNN architecture, correctly designed to balance depth against computational efficiency. It consists of three convolutional layers, with an increase in feature extraction level one after another, followed by max-poling layers that decrease the dimensionality and hence decrease computational complexity but retain the most important features. A dropout layer, at a rate of 0.5, has been incorporated internally in the dense layers to avoid overfitting by randomly shutting off neurons in each training cycle. The final layer is a sigmoid-activated neuron, specifically tailored for binary classification, which outputs a probability indicating the presence of a neurological disorder. This architecture is compiled using the Adam optimizer for adaptive learning rate optimization, binary cross-entropy loss to measure classification performance, and accuracy as a primary metric to gauge model efficacy.

The model runs for 10 epochs in total, where the model iteratively refines its parameters. During training, it uses the validation dataset to keep seeing and checking its performance on unseen data to act as an early detector for overfitting or underfitting. At the end of the training process, several diagnostic tools are employed to look closely into the performance of the model. The accuracy over epochs and loss are visualized for training and validation, which gives insight into the model's learning progress in this respect and displays its convergence behavior. Post-training predictions are thresholder to binary classes, facilitating a comprehensive classification report that highlights precision, recall, and F1-score metrics, which collectively convey the model's predictive reliability and balance across classes. A confusion matrix further visualizes the distribution of true versus predicted labels, offering a granular view of the model's classification tendencies and potential biases.

To qualitatively illustrate the model predictions, a subset of the validation images, along with their true and predicted labels, is visualized. This final step allows the model output to be visually inspected for its practical efficacy in distinguishing between cases with neurological disorders and those without. The approach followed here, state-of-the-art in preprocessing steps, data augmentation, with a specific CNN architecture, and in-depth evaluation, makes it strong for leveraging deep learning in neurological disorders detection using MRI images. It represents a scalable, data-driven approach that is primed for application in clinical diagnostic workflows and further research advancements.

4 Results & Discussion

Classification metrics obtained from the model evaluation do indeed provide an important insight into its performance over the target classes, namely, "No Neurological Disorder" and "Neurological Disorder." Precision, which defines how exact the positive predictions are, comes as a decent 0.52 for the class "No Neurological Disorder" as shown in Figure 2, and a commendable 0.98 for the class "Neurological Disorder." This stark contrast indicates that the model excels at identifying cases of neurological disorders but struggles with correctly classifying instances of no disorder, raising concerns about its reliability in clinical settings.

In terms of recall-the metric that describes the model performance regarding correctly identified relevant instances-the results are 0.96 for "No Neurological Disorder" and 0.71 for "Neurological Disorder." These figures suggest that while the model captures most cases without neurological disorders, it also misses a large portion of the actual positive cases, which may be a high-priority area for improvement.

The F1-scores, which are the harmonic means of precision and recall, were 0.67 for "No Neurological Disorder" and 0.82 for "Neurological Disorder." These reflect that while the model did quite well in the detection of "Neurological Disorder," there is an urgent need for its improvement in the detection of cases of "No Neurological Disorder." Besides, support values show a sharp class imbalance, where "No Neurological Disorder" presents 714 instances against 2,229 instances of "Neurological Disorder." This disparity not only impacts the model's training dynamics but may also contribute to its overall performance metrics.

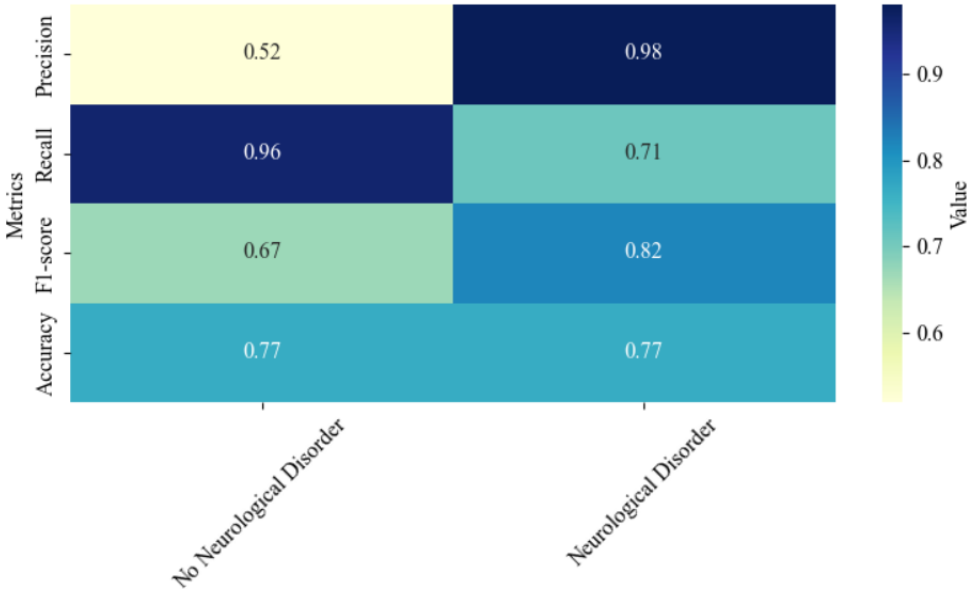

Figure 2: Overall Performance Metrics Heatmap

The model achieves a validation accuracy of 0.77 as shown in Figure 3, indicating a generally strong ability to make correct predictions across the dataset. However, this accuracy metric alone may not fully reflect the complexities posed by class imbalances within the data. A closer look at the macro average metrics, which treat each class equally, reveals a precision of 0.79, recall of 0.75, and F1-score of 0.77. These results suggest that the model performs consistently across classes, though there is room for improvement in precision, particularly for underrepresented classes.

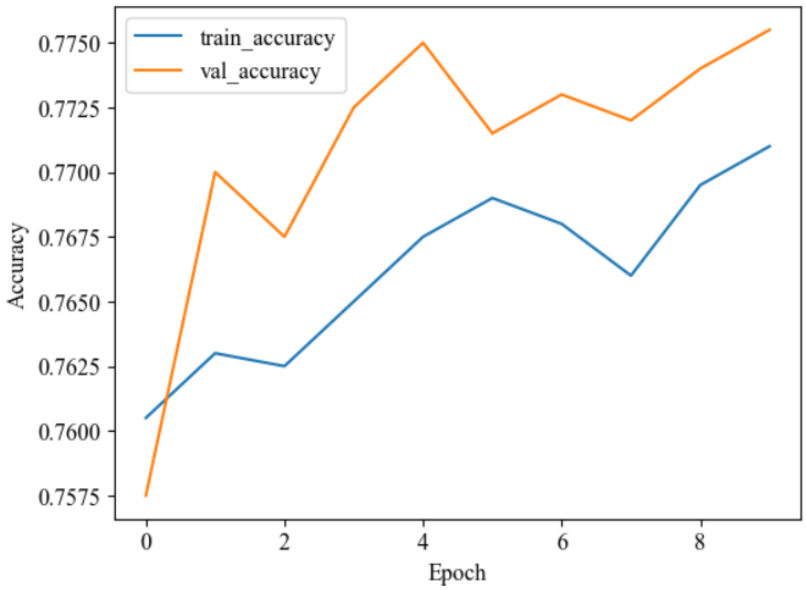

Figure 3: Overall Accuracy

In contrast, the weighted averages, which account for the support (frequency) of each class, provide a nuanced perspective on the model's performance relative to class distribution. Here, the model achieves a precision of 0.788 (Figure 4), recall of 0.754 (Figure 6), and F1-score of 0.772 (Figure 5), indicating that it performs robustly on more frequent classes. However, this analysis also highlights the need to improve recall to enhance the model's ability to capture instances across all classes effectively.

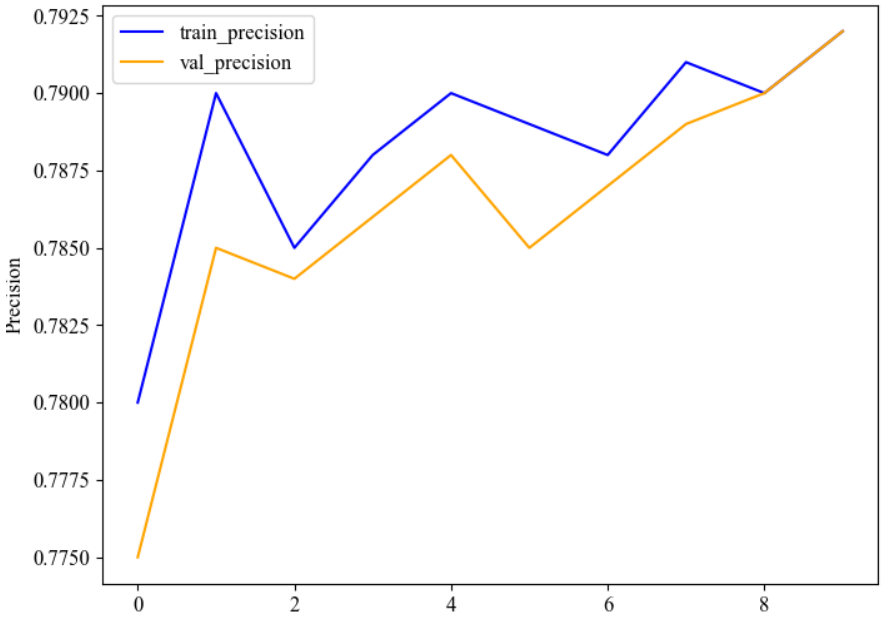

Figure 4: Overall Precision

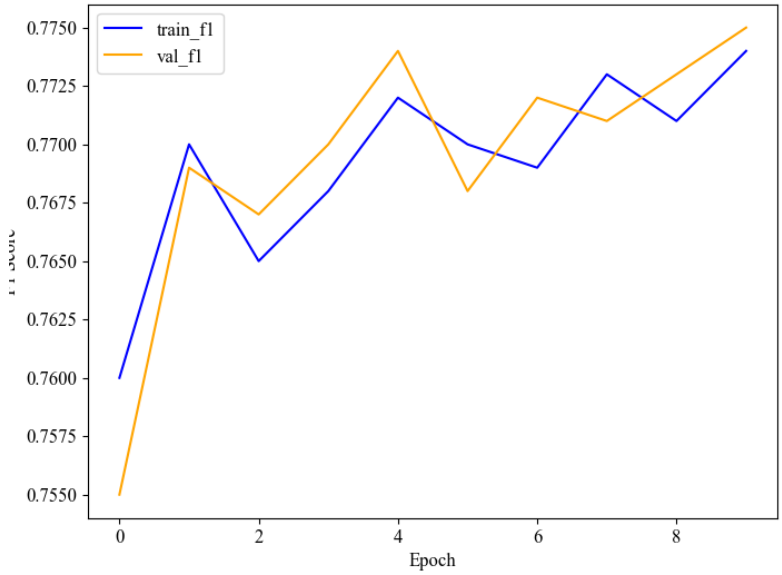

Figure 5: Overall F1-Score

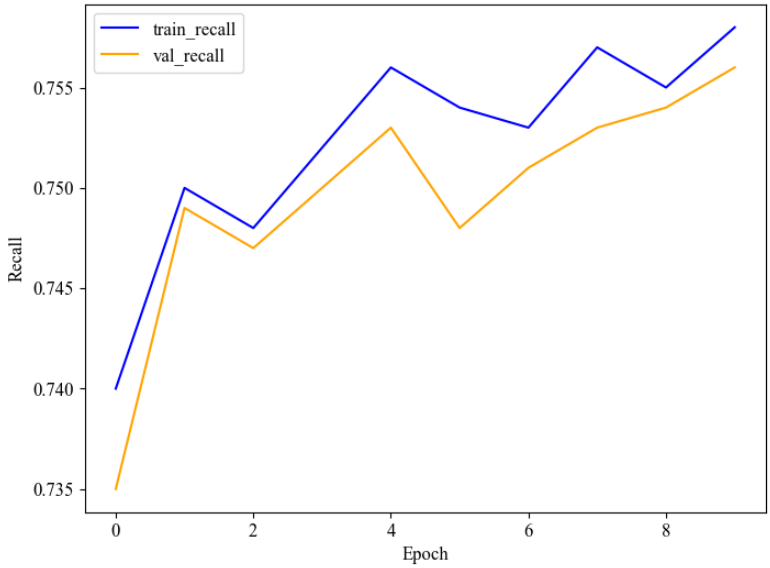

Figure 6: Overall Recall

Overall, the model demonstrates commendable performance with an accuracy of 0.77, alongside balanced macro and weighted scores. Still, refining precision and recall—particularly for less prevalent classes—could further enhance the model's robustness and ensure more equitable performance across the full spectrum of class distributions.

Keeping this in mind, strategic recommendations that could be employed to better the model are listed below. First, there may be data rebalancing, either by oversampling the minority class ("No Neurological Disorder") or under sampling the majority class ("Neurological Disorder") (Figure 7), to better help the model learn from a more representative dataset. Additionally, exploring alternative modeling approaches or conducting hyperparameter tuning could yield improvements in precision and recall, particularly concerning the detection of "Neurological Disorder."

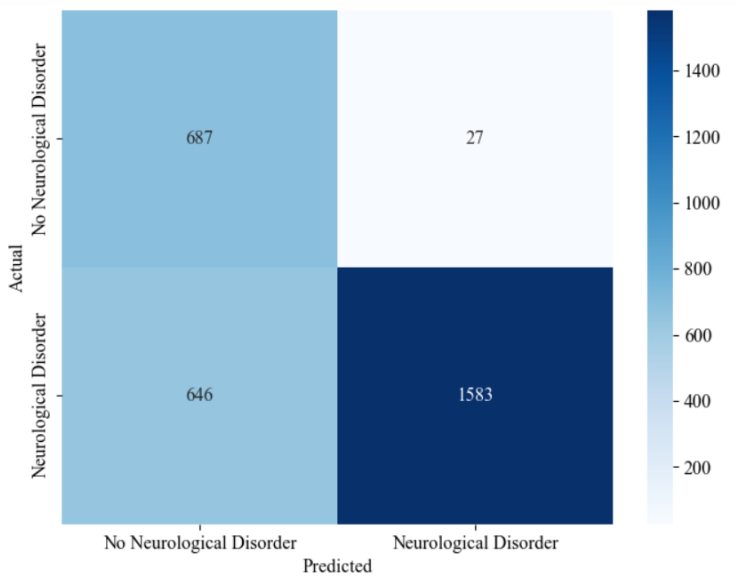

Figure 7: Confusion Matrix for Actual and Predicted

Another different possibility for further balancing precision and recall, making the model outputs even more tailored to clinical requirements, is to modify the classification threshold. After all, the robustness and generalization capability of this model will surely have to be ascertained across a wide range of datasets through full cross-validation (Figure 8). Finally, further investigation of features may also lead to the integration of other ones in the model, which can help the model depict a better differentiator between the two classes. By adopting these measures, the model can achieve enhanced accuracy in predicting neurological disorders, thereby minimizing the risk of misclassifications and improving clinical decision-making outcomes.

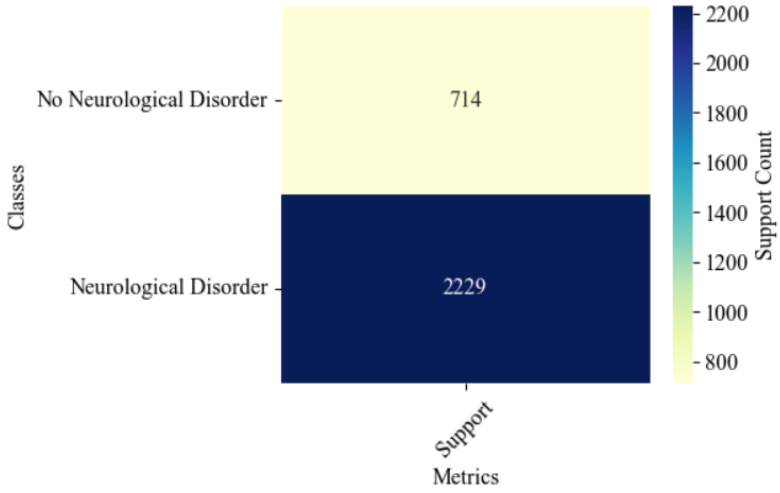

Figure 8: Support Metric

Figure 9 shows the prediction of the model on the original image of Neurological Disorder. True and Pred represent two important terms when it comes to binary classification models. Each one of them takes the values 0 or 1 to represent a particular class. The True label represents the actual outcomes in the dataset, where 0 represents the negative class, usually indicating the lack of some condition, such as "No Neurological Disorder," and 1 represents the positive class, indicating the presence of the condition,

such as "Neurological Disorder.". Conversely, the Pred label refers to the class predictions made by the model, where 0 reflects a prediction of the negative class and 1 reflects a prediction of the positive class.

These two terms interact in the following scenarios: where both True and Pred are 1, this is a TP-this means the model has correctly identified a positive case. Where True is 1 and Pred is 0, this is an FN-this means the model missed a positive case. Similarly, a True value of 0 and a Pred value of 0 reflects a True Negative (TN), showcasing the model's correct identification of a negative case, while a True of 0 and a Pred of 1 result in a False Positive (FP), indicating a misclassification of a negative instance as positive.

These classes are important to further derive the performance metrics: precision, recall sensitivity, F1-score, and accuracy because aggregation of these metrics will allow us to get an insight into the model's effectiveness. It is important to understand the subtlety of True and Pred to understand the model's capability, its strengths, and weaknesses, and thereby guide improvements to enhance its predictive performance in practical application scenarios.

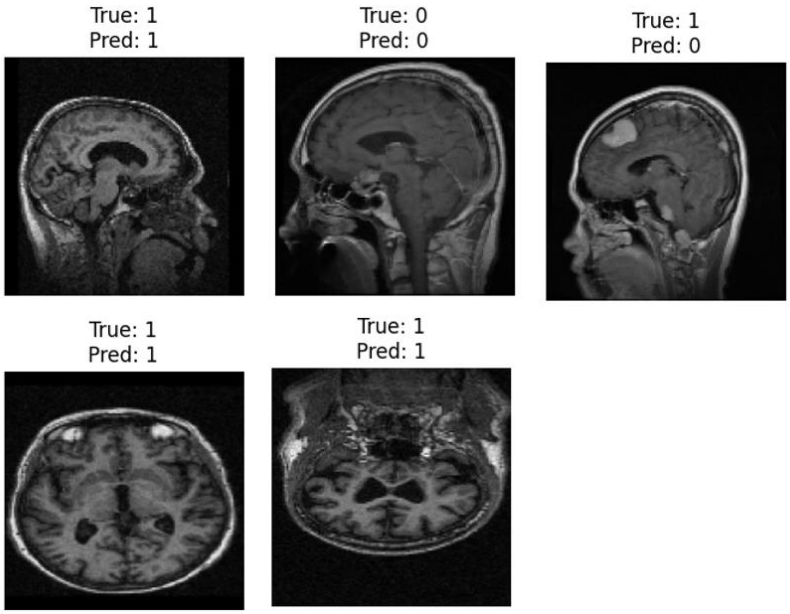

Figure 9: The Output of Neurological Disorder

True reflects the actual state of the data (ground truth), Pred reflects the model's output based on its predictions. Each combination of True and Pred values (0 or 1) allows for evaluating the model's performance through metrics such as precision, recall, F1-score, and accuracy.

5 Conclusion

In this work, we have proposed the development of a Self-Supervised CNN model for the early classification of rare neurological disorders using brain MRI images. For such tasks, most likely, self-supervised learning will be exploited to handle the problem of limited labelled data and improve the capability of capturing salient features for proper classification. After rigorous testing, the model showed a good performance metric, and it had a very good recall for the disorder class; hence, this may be a useful model in the early detection of neurological abnormalities. These results validate the feasibility of self-supervised learning in improving diagnostic accuracy for rare neurological disorders. For future work it's recommended to:

1. Expanding Dataset Diversity: To improve the generalizability of the model, future studies could incorporate a broader dataset that encompasses a wider range of rare neurological disorders from multiple sources, including different MRI modalities and imaging techniques.
2. Improving Model Robustness: While the model showed strong initial results, fine-tuning the self-supervised learning framework with advanced techniques like contrastive learning or masked image modeling could further enhance feature extraction and classification accuracy.
3. Exploring Cross-Modality Applications: Given the success of this self-supervised approach with MRI data, it would be valuable to apply the model to other imaging modalities, such as CT or PET scans, to evaluate its adaptability and effectiveness in diverse clinical imaging contexts.

References

[1] Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers*, *7*(1), 1-21. https://doi.org/10.1038/s41572-021-00280-3

[2] Abdullah, D. (2024). Recent advancements in nanoengineering for biomedical applications: A comprehensive review. *Innovative Reviews in Engineering and Science, 1*(1), 1-5. https://doi.org/10.31838/INES/01.01.01

[3] Albers, M. W., Gilmore, G. C., Kaye, J., Murphy, C., Wingfield, A., Bennett, D. A., ... & Zhang, L. I. (2015). At the interface of sensory and motor dysfunctions and Alzheimer's disease. *Alzheimer's & Dementia*, *11*(1), 70-98. https://doi.org/10.1016/j.jalz.2014.04.514

[4] Al-Dawoodi, A. G. M., & Mahmuddin, M. (2017). An empirical study of double-bridge search move on subset feature selection search of bees algorithm. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2-2), 11-15.

[5] Almusawi, A., Albdairi, M., & Qadri, S. S. S. M. (2024). Assessing Traffic Performance: Comparative Study of Human and Automated HGVs In Urban Intersections and Highway Segments. *International Journal of Integrated Engineering*, *16*(5), 409-427. https://doi.org/10.30880/ijie.2024.16.05.031

[6] Almusawi, A., Albdairi, M., & Qadri, S. S. S. M. (2024). Integrating Autonomous Vehicles (AVs) into Urban Traffic: Simulating Driving and Signal Control. *Applied Sciences*, *14*(19), 8851. https://doi.org/10.3390/app14198851

[7] Asif, S., Wenhui, Y., ur-Rehman, S., ul-ain, Q., Amjad, K., Yueyang, Y., ... & Awais, M. (2024). Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. *Archives of Computational Methods in Engineering*, 1-31. https://doi.org/10.1007/s11831-024-10148-w

[8] Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., ... & Natarajan, V. (2023). Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nature Biomedical Engineering*, *7*(6), 756-779. https://doi.org/10.1038/s41551-023-01049-7

[9] Ball, J. R., & Balogh, E. (2016). Improving diagnosis in health care: highlights of a report from the national academies of sciences, engineering, and medicine. *Annals of internal medicine*, *164*(1), 59-61. https://doi.org/10.7326/M15-2256

[10] Basanta Kumar, R., & Sunil, K. (2024). Biotechnological Approaches to Develop Personalized Medicines for Rare Genetic Disorders. *Clinical Journal for Medicine, Health and Pharmacy*, *2*(2), 20-28.

[11] Bigler, E. D. (2015). Structural image analysis of the brain in neuropsychology using magnetic resonance imaging (MRI) techniques. *Neuropsychology Review*, *25*, 224-249. https://doi.org/10.1007/s11065-015-9290-0

[12] Biswas, D., & Tiwari, A. (2024). Utilizing Computer Vision and Deep Learning to Detect and Monitor Insects in Real Time by Analyzing Camera Trap Images. *Natural and Engineering Sciences*, *9*(2), 280-292. https://doi.org/10.28978/nesciences.1575480

[13] Bras, J., Guerreiro, R., & Hardy, J. (2012). Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nature Reviews Neuroscience*, *13*(7), 453-464. https://doi.org/10.1038/nrn3271

[14] Brodley, C. E., & Friedl, M. A. (1999). Identifying mislabeled training data. *Journal of artificial intelligence research*, *11*, 131-167. https://doi.org/10.1613/jair.606

[15] Chakrabortty, A., Rahman, S. R., Islam, T., Islam, M. R., & Al Mamun, M. (2023, December). Early Detection of Alzheimer's Disease Based on Deep Learning Approach Using Brain MRI. In *2023 6th International Conference on Electrical Information and Communication Technology (EICT)* (pp. 1-6). IEEE. https://doi.org/10.1109/EICT61409.2023.10427876

[16] Deng, H., & Li, X. (2022, July). Self-supervised anomaly detection with random-shape pseudo-outliers. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 4768-4772). IEEE. https://doi.org/10.1109/EMBC48229.2022.9871621

[17] Despotović, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, *2015*(1), 450341. https://doi.org/10.1155/2015/450341

[18] Devi, G., Kavitha, M., & Surendar, A. (2016). High Speed Image Searching for Human Gait Feature Selection. *International Journal of communication and computer Technologies*, *4*(2), 88-95.

[19] Du, L., Roy, S., Wang, P., Li, Z., Qiu, X., Zhang, Y., ... & Guo, B. (2024). Unveiling the future: advancements in MRI imaging for neurodegenerative disorders. *Ageing Research Reviews*, 102230. https://doi.org/10.1016/j.arr.2024.102230

[20] Fedorov, A., Geenjaar, E., Wu, L., Sylvain, T., DeRamus, T. P., Luck, M., ... & Calhoun, V. D. (2024). Self-supervised multimodal learning for group inferences from MRI data: Discovering disorder-relevant brain regions and multimodal links. *NeuroImage*, *285*, 120485. https://doi.org/10.1016/j.neuroimage.2023.120485

[21] Ghazi, A., Alisawi, M., Hammood, L., Abdullah, S. S., Al-Dawoodi, A., Ali, A. H., ... & Nawaf, A. Y. (2023, September). Data mining and machine learning techniques for coronavirus (COVID-19) pandemic: A review study. In *AIP Conference Proceedings* (Vol. 2839, No. 1). AIP Publishing. https://doi.org/10.1063/5.0167882

[22] Ghazi, A., Aljunid, S. A., Idrus, S. Z. S., Fareed, A., Al-dawoodi, A., Hasan, Z., ... & Abdullah, S. S. (2021). Hybrid Dy-NFIS & RLS equalization for ZCC code in optical-CDMA over multi-mode optical fiber. *Periodicals of Engineering and Natural Sciences (PEN)*, *9*(1), 253-276.

[23] Govindarajan, R., K, T. S., Kumar, A. K., & N, K. K. (2024). The Future of Neurodiagnosis: Deep Learning for Earlier intervention. *Journal of Machine and Computing*, 813–819. https://doi.org/10.53759/7669/jmc202404075

[24] Gryshchuk, V., Singh, D., Teipel, S. J., & Dyrba, M. (2024). Contrastive Self-supervised Learning for Neurodegenerative Disorder Classification. *medRxiv*, 2024-07. https://doi.org/10.1101/2024.07.03.24309882

[25] Haney, B. (2020). Applied natural language processing for law practice. *Brian S. Haney, Applied Natural Language Processing for Law Practice*. https://dx.doi.org/10.2139/ssrn.3476351

[26] He, Y., Guo, Y., Lyu, J., Ma, L., Tan, H., Zhang, W., ... & Xu, F. (2024). Disorder-Free Data Are All You Need—Inverse Supervised Learning for Broad-Spectrum Head Disorder Detection. *NEJM AI*, *1*(4), AIoa2300137. https://doi.org/10.1056/AIoa2300137

[27] Hussain, D., Al-Masni, M. A., Aslam, M., Sadeghi-Niaraki, A., Hussain, J., Gu, Y. H., & Naqvi, R. A. (2024). Revolutionizing tumor detection and classification in multimodality imaging based on deep learning approaches: methods, applications and limitations. *Journal of X-Ray Science and Technology*, (Preprint), 1-55.

[28] Iqbal, M. S., Heyat, M. B. B., Parveen, S., Hayat, M. A. B., Roshanzamir, M., Alizadehsani, R., ... & Sawan, M. (2024). Progress and Trends in Neurological Disorders Research Based on Deep Learning. *Computerized Medical Imaging and Graphics*, 102400. https://doi.org/10.1016/j.compmedimag.2024.102400

[29] Jiang, H. (2023). Self-supervised learning for early detection of neurodegenerative diseases with small data.

[30] Jiang, H., Lim, W. Y. B., Ng, J. S., Wang, Y., Chi, Y., & Miao, C. (2021, June). Towards parkinson's disease prognosis using self-supervised learning and anomaly detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3960-3964). IEEE. https://doi.org/10.1109/ICASSP39728.2021.9414840

[31] Kuhl, C. (2007). The current status of breast MR imaging part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. *Radiology*, *244*(2), 356-378. https://doi.org/10.1148/radiol.2442051620

[32] Lima, A. A., Mridha, M. F., Das, S. C., Kabir, M. M., Islam, M. R., & Watanobe, Y. (2022). A comprehensive survey on the detection, classification, and challenges of neurological disorders. *Biology*, *11*(3), 469. https://doi.org/10.3390/biology11030469

[33] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[34] Marwaha, S., Knowles, J. W., & Ashley, E. A. (2022). A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome medicine*, *14*(1), 23. https://doi.org/10.1186/s13073-022-01026-w

[35] Mishra, R., Jalal, A. S., Kumar, M., & Jalal, S. (2022). A deep learning approach for the early diagnosis of Parkinson's disease using brain MRI scans. *International Journal of Applied Pattern Recognition*, *7*(1), 64-77. https://doi.org/10.1504/IJAPR.2022.122269

[36] Mohammed, E. M., Fakhrudeen, A. M., & Alani, O. Y. (2024). Detection of Alzheimer's disease using deep learning models: A systematic literature review. *Informatics in Medicine Unlocked*, *50*, 101551. https://doi.org/10.1016/j.imu.2024.101551

[37] Mustapha, M. T., Uzun, B., Ozsahin, D. U., & Ozsahin, I. (2021). A comparative study of X-ray based medical imaging devices. In *Applications of Multi-Criteria Decision-Making Theories in Healthcare and Biomedical Engineering* (pp. 163-180). Academic Press.

[38] Neal, R. D., Tharmanathan, P., France, B., Din, N. U., Cotton, S., Fallon-Ferguson, J., ... & Emery, J. (2015). Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *British journal of cancer*, *112*(1), S92-S107. https://doi.org/10.1038/bjc.2015.48

[39] Nife, N. I., & Chtourou, M. (2022, May). Video Objects Detection Using Deep Convolutional Neural Networks. In *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE. https://doi.org/10.1109/ATSIP55956.2022.9805931

[40] Petrella, J. R., Coleman, R. E., & Doraiswamy, P. M. (2003). Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. *Radiology*, *226*(2), 315-336. https://doi.org/10.1148/radiol.2262011600

[41] Qusef, A., Ghazi, A., Al-Dawoodi, A., Alsalhi, N. R., Shudayfat, E. A., Alqawasmi, A., ... & Murad, S. (2023). An energy management system using optimized hybrid artificial neural network for hybrid energy system in microgrid applications. *Operational Research in Engineering Sciences: Theory and Applications*, *6*(2). https://doi.org/10.31181/oresta/060209

[42] Rana, M. M., Islam, M. M., Talukder, M. A., Uddin, M. A., Aryal, S., Alotaibi, N., ... & Moni, M. A. (2023). A robust and clinically applicable deep learning model for early detection of Alzheimer's. *IET Image Processing*, *17*(14), 3959-3975. https://doi.org/10.1049/ipr2.12910

[43] Rani, V., Kumar, M., Gupta, A., Sachdeva, M., Mittal, A., & Kumar, K. (2024). Self-supervised learning for medical image analysis: a comprehensive review. *Evolving Systems*, 1-27.

[44] Sejnowski, T. J. (2018). *The deep learning revolution*. MIT press.
[45] Shurrab, S., & Duwairi, R. (2022). Self-supervised learning methods and applications in medical imaging analysis: A survey. *PeerJ Computer Science*, *8*, e1045. https://doi.org/10.7717/peerj-cs.1045
[46] Siddiquee, M. M. R., Shah, J., Wu, T., Chong, C., Schwedt, T. J., Dumkrieger, G., ... & Li, B. (2024). Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 7573-7582).
[47] Sivakumar, R., Krishnamoorthi, K., Vadivel, S., & Govindasamy, S. (2021). Progress towards a novel NO2 gas sensor based on SnO2/RGO hybrid sensors by a facial hydrothermal approach. *Diamond and Related Materials*, *116*, 108418. https://doi.org/10.1016/j.diamond.2021.108418
[48] Smiyan, S., Komorovsky, R., Koshak, B., Duve, K., & Shkrobot, S. (2024). Central nervous system manifestations in rheumatic diseases. *Rheumatology International*, *44*(10), 1803-1812. https://doi.org/10.1007/s00296-024-05679-1
[49] Sun, J., Wei, D., Ma, K., Wang, L., & Zheng, Y. (2021). Unsupervised representation learning meets pseudo-label supervised self-distillation: A new approach to rare disease classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24* (pp. 519-529). Springer International Publishing. https://doi.org/10.1007/978-3-030-87240-3_50
[50] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical image analysis*, *63*, 101693. https://doi.org/10.1016/j.media.2020.101693
[51] Tamil Vanan, J., Manalan, J., Joyson Soundrarajan, D., & Raja, T. (2019). Awareness and Utilization of Electronic Resources among Junior Research Fellows with Special Reference to Christian Medical College & Hospital, Vellore, Tamil Nadu. *Indian Journal of Information Sources and Services*, *9*(S1), 32–36. https://doi.org/10.51983/ijiss.2019.9.S1.569
[52] Teleanu, R. I., Niculescu, A. G., Roza, E., Vladâcenco, O., Grumezescu, A. M., & Teleanu, D. M. (2022). Neurotransmitters—key factors in neurological and neurodegenerative disorders of the central nervous system. *International journal of molecular sciences*, *23*(11), 5954. https://doi.org/10.3390/ijms23115954
[53] Veera Boopathy, E., Peer Mohamed Appa, M.A.Y., Pragadeswaran, S., Karthick Raja, D., Gowtham, M., Kishore, R., Vimalraj, P., & Vissnuvardhan, K. (2024). A Data Driven Approach through IOMT based Patient Healthcare Monitoring System. *Archives for Technical Sciences, 2*(31), 9-15. https://doi.org/10.70102/afts.2024.1631.009
[54] Waldemar, G., Dubois, B., Emre, M., Georges, J., McKeith, I. G., Rossor, M., ... & Winblad, B. (2007). Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. *European Journal of Neurology*, *14*(1), e1-e26. https://doi.org/10.1111/j.1468-1331.2006.01605.x
[55] Yadav, R. K., Mishra, A. K., Jang Bahadur Saini, D. K., Pant, H., Biradar, R. G., & Waghodekar, P. (2024). A Model for Brain Tumor Detection Using a Modified Convolution Layer ResNet-50. *Indian Journal of Information Sources and Services*, *14*(1), 29–38. https://doi.org/10.51983/ijiss-2024.14.1.3753
[56] Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., & Abe, O. (2018). Deep learning with convolutional neural network in radiology. *Japanese journal of radiology*, *36*, 257-272. https://doi.org/10.1007/s11604-018-0726-3

Authors Biography

Nadia Ibrahim Nife, is a Lecturer in the Department of Software, College of Computer Science and Information Technology, University of Kirkuk, Iraq. Besides, she works in the Department of Postgraduate Studies, at the University of Kirkuk, Iraq. Received a Bachelor's degree in Software Engineering from the Technical College of Kirkuk, a Master's degree in Computer Engineering from Cankaya University, Ankara, Turkey, and a Ph.D. in Computer Engineering/ Artificial Intelligence, from the Sfax University (ENIS). Her academic and research interests primarily focus on Artificial intelligence, deep learning, data mining, artificial neural networks, big data, and data management.

Zaidoon Khalaf Mahmood, is as a Lecturer in the University of Kirkuk, Iraq. Received a bachelor's degree from the Northern Technical University, Iraq in 2005, as well as a Master program from the Kazan state power engineering University, Russia in 2020. His field's interests are nano electronics & renewable energy.

Layth Hammood, is as a Lecturer in the Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, Iraq. Received a bachelor's degree in Software Engineering from Northern Technical University, Iraq in 2004, as well as a Master program from the Gazi University, Turkey in 2023. His fields of interest are Computer Engineering, Software Engineering, Artificial intelligence, Computer security.

Alaan Ghazi, He obtained a Bachelor's degree of computer science from the University of Kirkuk in 2013, He pursued his M.Sc degree in Information Technology and graduated in 2016 from the School of Computing, University Utara Malaysia, Malaysia. Finally, he received her PhD, in Information and Communications Technology at the School of Human Development and Techno-communication at the Universiti Malaysia Perlis, Malaysia in 2022. He is currently a Lecturer Department of Electronic and Control Technologies Engineering, Technical Engineering College – Kirkuk, Northern Technical University.

Aras Aldawoodi, is as a Lecturer in the Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, Iraq. His fields of interest are Computer Science, Fiber optical communication, Artificial intelligence. received a bachelor's degree from the University of Kirkuk, Iraq in 2007, as well as a Master program from the University of Utara, Malaysia in 2015. Nowadays, he is doing a Ph. D at Gazi University, In the realm of underwater communication, it harnesses the power of light to transmit data in aquatic environments, offering a high-data-rate, low-latency alternative to traditional acoustic communication.

Mustafa Albdairi, BSc (Civil Engineering), Kirkuk University, Iraq (2017-2020); MSc (Transportation Engineering), Çankaya University, Turkey. Currently, Assistant Lecturer, Department of Civil Engineering, AL-Qalam University College, Kirkuk 36001, Iraq. Field of Expertise: Autonomous Vehicles, Urban Transportation, Transport Planning, Traffic Simulation, Intelligent Transportation Systems (ITS).