Iraqi Journal for Computer Science and Mathematics

Volume 5 | Issue 4 Article 14

2024

Convolutional Neural Networks for Dementia Severity Classification: Ordinal Versus Regular Methods

Ambresh Bhadrashetty
Assistant Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi – 585105, Karnataka, India,
ambresh.bhadrashetty@gmail.com

P. Sandhya
Associate Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Centre for PG Studies, Mysuru - 570019, Karnataka, India,
sanjoshi17@yahoo.com

Recommended Citation
Bhadrashetty, Ambresh and Sandhya, P. (2024) "Convolutional Neural Networks for Dementia Severity Classification: Ordinal Versus Regular Methods," *Iraqi Journal for Computer Science and Mathematics*: Vol. 5: Iss. 4, Article 14.
DOI: https://doi.org/10.52866/2788-7421.1224
Available at: https://ijcsm.researchcommons.org/ijcsm/vol5/iss4/14

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2024;5:169–174

RESEARCH ARTICLE

Convolutional Neural Networks for Dementia Severity Classification: Ordinal Versus Regular Methods

Ambresh Bhadrashetty [a,*], P. Sandhya [b]

[a] Assistant Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi – 585105, Karnataka, India
[b] Associate Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Centre for PG Studies, Mysuru - 570019, Karnataka, India

ABSTRACT

Dementia, a chronic neurodegenerative disorder, progressively impairs cognitive functions such as memory, reasoning, learning, and recall, placing a significant burden on patients and healthcare systems. Early and accurate classification of dementia severity is crucial for personalized care and intervention. This study introduces a novel Convolutional Neural Network (CNN) designed to classify dementia into four ordinal severity levels (None, Very Mild, Mild, and Moderate) based on MRI brain scans. Utilizing the extensive Open Access Series of Imaging Studies (OASIS) dataset, which includes 86,437 MRI scans (67,222 'none,' 13,725 'very mild,' 5,002 'mild,' and 488 'moderate'), our model addresses severe class imbalance with a combination of the Synthetic Minority Over-sampling Technique (SMOTE) and advanced over- and under-sampling methods. By implementing ordinal classification, the model effectively captures the progressive nature of dementia, showing comparable or improved performance against current diagnostic benchmarks. This approach highlights the benefits of ordinal classification in medical imaging, paving the way for enhanced severity assessment and supporting better treatment planning.

Keywords: Dementia, CNN, MRI Brain Scans, OASIS, SMOTE

1. Introduction

Dementia is a complex, progressive neurodegenerative disorder affecting millions worldwide, with rates expected to rise as populations age. It profoundly impacts cognitive functions, including memory, reasoning, and learning abilities, leading to significant limitations in daily functioning and quality of life. Current diagnostic methods for dementia rely on clinical assessment, cognitive testing, and neuroimaging techniques, which, although effective, often face limitations in accurately classifying severity levels due to subjective interpretation and inter-rater variability. Early and accurate assessment of dementia severity is crucial, as it enables more effective intervention planning and symptom management, potentially delaying disease progression [1, 2].

Neurological disorders like dementia are one of the most serious health issues across the world. Apoptosis, a disease in which brain cells cease to function normally and eventually die, is one of the symptoms of dementia. Certain brain regions that control an individual's thinking, memory, mobility, behavior, and emotion are affected by dementia. Dementia has an early onset before the age of 65, but as the illness progresses, the condition gets worse [3]. Making a fast and correct diagnosis is a major obstacle in the detection of dementia. Dementia has no recognized treatment. But to identify the illness, laboratory testing or imaging methods like MRI are needed. To

Received 1 October 2024; accepted 16 November 2024.
Available online 25 November 2024

* Corresponding author.
E-mail addresses: ambresh.bhadrashetty@gmail.com (A. Bhadrashetty), sanjoshi17@yahoo.com (P. Sandhya).

https://doi.org/10.52866/2788-7421.1224

tackle this difficulty, significant progress has been achieved with computer-aided algorithms that use neuroimaging. The absence of a reliable and effective general-purpose algorithm makes dementia detection a difficult research problem even with significant advancements in this field [4]. Deep learning cutting-edge technology to detect mental health are discussed in [5–9].

The worldwide burden of dementia has been recognized by the World Health Organization (WHO) as a public health issue. The World Health Organization highlights the rising number of dementia-related deaths globally because of ageing populations, even if it does not forecast specific deaths. It is projected that 82 million individuals will have dementia by 2030, and that figure will climb to 152 million by 2050. In high-income nations, dementia is already one of the main causes of mortality [10, 11].

CNN may become biased toward the majority class as a result of this imbalance, making it difficult for them to accurately identify and categorize examples from the minority classes. Confronting this disparity in class is essential to developing a strong and objective model [12]. To lessen these difficulties and improve the model's performance on the under-represented classes, strategies like oversampling the minority class, under sampling the majority class, or utilizing sophisticated approaches like Synthetic Minority Over-sampling Technique (SMOTE) might be used. We experimented with a number of these techniques before deciding on a combination of over- and under sampling [13].

2. Literature review

Bron et al. [14] has conducted a study using Support Vector Machine (SVM) in which the author has reviewed more than 20 studies and concluded that SVM has performed better in diagnosing Dementia when SVM was combined with feature extraction techniques like Principal Component Analysis (PCA). Liu et al. [15] used Convolutional Neural Network (CNN) for feature extraction from Medical Resonance Imaging (MRI) data, accuracy of 89.5% was achieved in the detection of Alzheimer's. The Performance was boosted significantly on the limited datasets by using transfer learning.

Zhang et al. designed and developed a model which was hybrid using CNN-LSTM for the prediction of cognitive deterioration in MCI patients. The newly designed hybrid model was able to achieve 83% accuracy in estimating MCI converting into Alzheimer's [16]. Zhou et al. developed a model by considering Random Forest classifier and he combined MRI data with cognitive scores which resulted in achieving 86% accuracy in differentiating Alzheimer's from other dementias [17]. Gupta et al. applied an autoencoder to reduce MRI data dimensionality and K-means clustering for knowing the different stages of the dementia the patients were suffering from [18]. Qiu et al. has shown that the critical regions of the brain which are identified by the model aligned better with clinical expectations. To do this the authors integrated attention mechanisms in CNN to improve the explainability of detection of dementia [19].

With advancements in artificial intelligence, deep learning models—particularly CNNs—have demonstrated substantial promise in medical imaging applications. The CNNs have excelled in classifying, segmenting, and identifying structural patterns in MRI data, making them well-suited for neurodegenerative disease assessment [20]. In recent studies, CNNs have been applied to dementia classification tasks, yet most approaches treat severity as categorical, ignoring the ordinal nature of progression from mild to severe stages [21]. This can limit the models' clinical relevance, as the transition between dementia stages is inherently sequential rather than discrete.

To address this, our study introduces a CNN-based approach that leverages ordinal classification to better capture the progression of dementia severity. Using MRI brain scans from the Open Access Series of Imaging Studies (OASIS) dataset, we classify dementia into four severity levels: None, Very Mild, Mild, and Moderate. However, one of the significant challenges in this study is the imbalance in class distribution, with most scans categorized as "None" and fewer in more severe categories. Such imbalances can lead to biased predictions, reducing the model's utility in accurately assessing moderate cases [22]. To mitigate this, we apply the Synthetic Minority Over-sampling Technique (SMOTE) along with advanced resampling techniques, enhancing model performance across all severity levels. The advancement of ML and DL will be contingent upon the incorporation of multi-modal data and interpretability methodologies.

3. Methodology

This study utilizes EfficientNet, a state-of-the-art CNN model that employs a novel scaling method to balance model width, depth, and resolution. EfficientNet's variants—such as EfficientNetB0 and EfficientNetV2—are designed for high accuracy with optimized computational efficiency. EfficientNet models are ideal for handling complex tasks like MRI classification due to their depth and efficiency in extracting high-level features from input images.

The methodology of this work starts from the collection of data Open Access Series of Imaging Study (OASIS) and ends with the model evaluation. The step-by-step procedure is discussed in the next subsections.

3.1. Data collection

The dataset in this study comes from the Open Access Series of Imaging Study (OASIS) which consists of 86,437 Magnetic Resonance Imaging (MRI) brain scans. The dataset includes 67.222 scans in the 'non' category, 13.725 scans in 'very mild', 5002 in 'mild' and 488 in the 'moderate' category.

3.2. Data preprocessing

A test set of images for each category is separated before performing any preprocessing so that we could test the model on data that it had never seen previously and on which we had not adjusted the hyperparameters. This will provide a more realistic depiction of the real performance of our model. This work uses one hot encoding to transform categorical variables into numerical format. The dataset was then split into three parts, one for training the models, the other for testing and the last one is for validation.

3.3. Model selection

In this work, EfficientNetB0 is used and its other versions as follows: EfficientNetB0, Basic, EfficientNetV2B1, Basic + EfficientNetB0 and Basic + EfficientNetV2B1. The kind of classification challenge (ordinal versus regular categorical) when utilizing EfficientNetB0 can greatly affect the model's performance and the way the output is handled. Regular classification is the conventional arrangement in which there is no intrinsic ranking or order among the classes. Assigning an input to one of multiple possible categories is the only task, and all classes are handled equally.

Ordinal classification: Although there may not always be an equal difference between successive classes, the classes do have a natural order or ranking. Ordinal examples include rating scales (1–5) and severity levels (mild, moderate, severe).

3.4. Model training

The train images from every class were converted into the desired $128 \times 128 \times 3$ shape. To cut down on computing time, we go with 128 pixels. The 'data' object was created by combining the categories. The class labels were saved in the 'result' object after being converted into a one-hot encoded format.

3.5. Model evaluation

The precision of positive forecasts is measured. It emphasizes categories with higher rankings. Like sensitivity, recall quantifies the genuine positive rate. It emphasizes categories with lower rankings. Recall and precision make up the F1 score. False positives and false negatives are its main topics.

In statistics and machine learning, receiver operating characteristic (ROC) curves are used to assess how well binary classification algorithms perform.

Area Under the Curve, or AUC, is a scalar metric that represents the performance of a classification model at every threshold that could be used. The calculation of Scott's pi was done to gauge test performance.

4. Results and discussion

The dataset consists of 86,437 as shown in the Fig. 1. It clearly shows that there is an imbalance in the data.

To bring the data set in the balance, in this study, a combination of undersampling (of the majority classes) and oversampling (of the minority classes) are considered 5000 and which are balanced for training or testing the model.

The dataset was split into two categories, each category used for respective tasks, training, and testing. Test set of images for each category was split so that we can evaluate the model on data that it has never seen before. To complete our objective of dividing MRI images into four ordinal categories, we used CNN

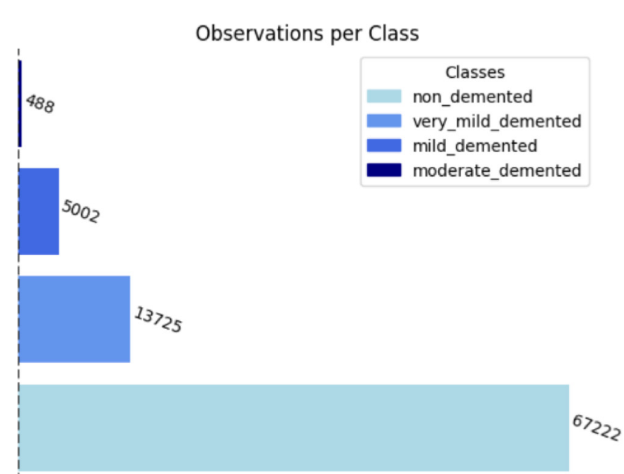


Fig. 1. Dataset with imbalanced classes.

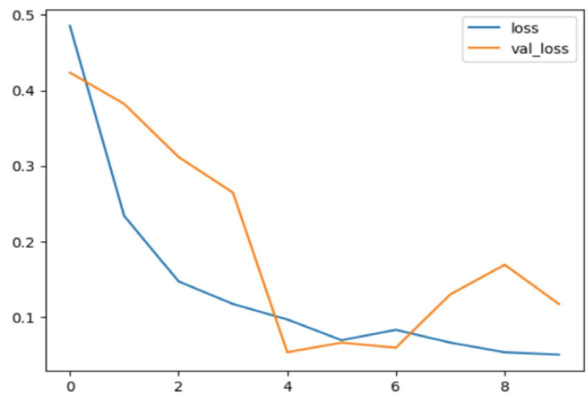


Fig. 2. Loss - Ordinal Classification.

that was created expressly for this purpose. Next, to determine whether an ordinal classifier CNN more accurately captures the ordinality in the data, we then compared this model to a CNN for “regular” classification.

In this work, the MRI images into the four groups using several models. We experimented with a variety of methods, including developing our architecture (a “basic” model), adding multiple layers, combining pre-trained bases and basic models, and more. These models include combinations of pre-trained bases and the basic model, the EfficientNetV2B1 model, and the basic model. Ultimately, the model that had a pre-trained basis of EfficientNetB0 outperformed the others.

4.1. Ordinal classification

The accuracy and loss scores of the model trained for ordinal classification is shown in Figs. 2 and 3. The scores for test loss were 0.0709 and test accuracy was 0.9782.

The obtained values of Precision, Recall and F1 Score in order to assess models’ performance as

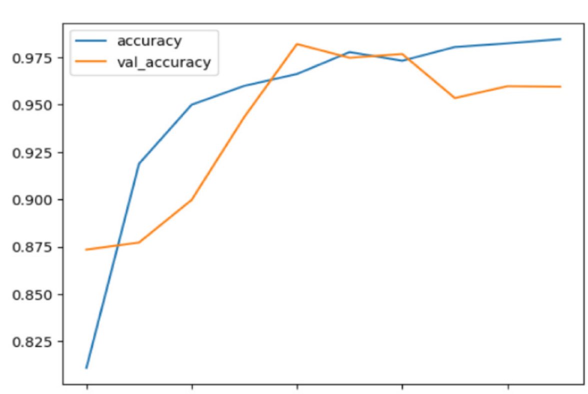
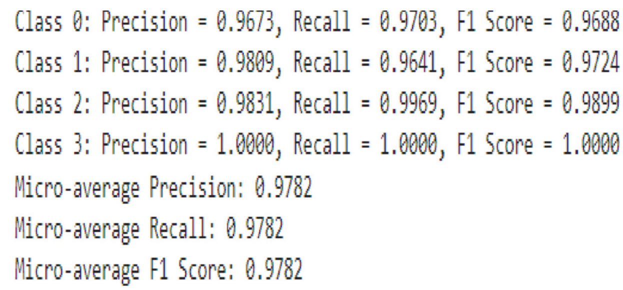


Fig. 3. Accuracy - Ordinal Classification.



```
Class 0: Precision = 0.9673, Recall = 0.9703, F1 Score = 0.9688
Class 1: Precision = 0.9809, Recall = 0.9641, F1 Score = 0.9724
Class 2: Precision = 0.9831, Recall = 0.9969, F1 Score = 0.9899
Class 3: Precision = 1.0000, Recall = 1.0000, F1 Score = 1.0000
Micro-average Precision: 0.9782
Micro-average Recall: 0.9782
Micro-average F1 Score: 0.9782
```

Fig. 4. A screenshot of Ordinal - Precision, Recall and F1 Score.

shown in Fig. 4 for ordinal classification. These measures account for both positive and negative precisions, making them more complex and accuracy.

The initial design of the F1 score, precision, and recall metrics was for a binarity environment; however, they can be readily extended to a multiclass setting and to ordinal classification. This is accomplished by taking into account each class separately first, and then calculating a micro-average, which is just the average of the results for each individual class. The problem is treated as a binary classification problem for each class by the micro-average method, which gives each class the same weight. This is the best approach for our assignment since, unlike weighted averages or macro-averages, which are better suited for unbalanced data, resampling has eliminated class imbalance. Furthermore, since overall performance is important to us, we wish to approach the issue.

The ROC-AUC score was computed as shown in Table 1 for ordinal classification. Although ROC-AUC is also intended for use in binary classification settings, it may be effortlessly adapted to multiclass (ordinal) settings by calculating the micro-average and calculating the score for each class. The model’s total performance is shown by the ROC-AUC, which takes into account the true positive and false positive rates.

A measure of test performance called Scott’s pi, which was obtained 0.9686681 for ordinal classification. This was done since the previous metrics might not adequately convey the ordinal relationship between the classes, even though they were modified for ordinal categorization by computing the micro-average and calculating scores per class. Scott’s pi is the most appropriate to employ for these kinds of

Table 1. Ordinal: ROC-AUC.

Class	ROC-AUC
Class 0	0.9966
Class 1	0.9977
Class 2	0.9997
Class 3	1.0000
Micro-average ROC AUC: 0.9988	

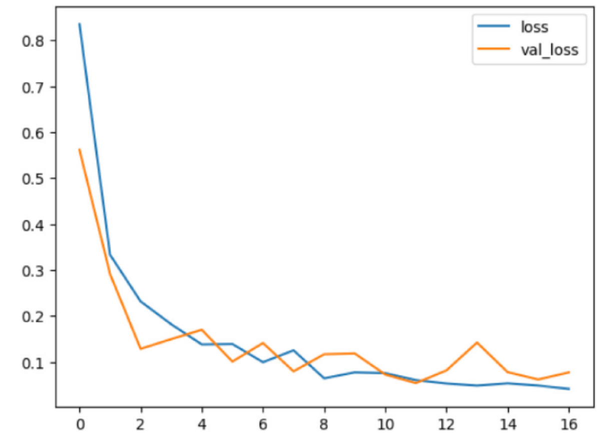


Fig. 5. Regular Classification: Loss.

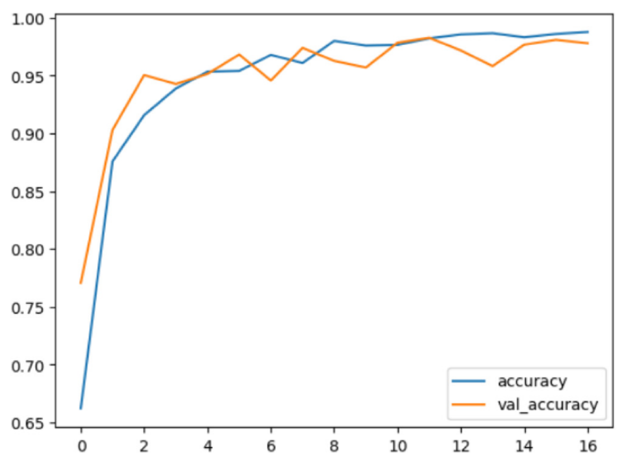
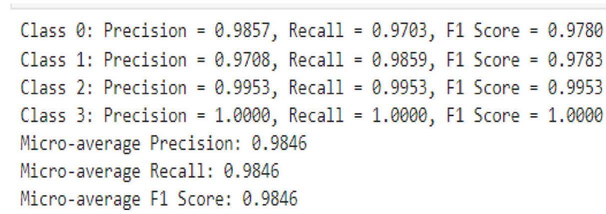


Fig. 6. Regular Classification: Accuracy.

activities, in which numerous performance metrics for ordinal classification were examined. Although it was designed as a measure of inter-rater reliability, its consideration of the ordinality of the data makes it ideal for tasks involving ordinals, such as ours. Similar to accuracy, Scott's pi is evaluated, with scores nearer 1 denoting strong performance.

4.2. Regular classification

The accuracy and loss scores of the model trained for ordinal classification is shown in Figs. 5 and 6. The scores for test loss were 0.0549 and test accuracy was 0.9846. A screenshot of obtained values of Precision, Recall and F1 Score in order to assess models' performance as shown in Fig. 7 for regular classification. It is obtained the Scott's pi 0.977925 for regular classification. The ROC-AUC score was computed as shown in Table 2 for regular classification.



```
Class 0: Precision = 0.9857, Recall = 0.9703, F1 Score = 0.9780
Class 1: Precision = 0.9708, Recall = 0.9859, F1 Score = 0.9783
Class 2: Precision = 0.9953, Recall = 0.9953, F1 Score = 0.9953
Class 3: Precision = 1.0000, Recall = 1.0000, F1 Score = 1.0000
Micro-average Precision: 0.9846
Micro-average Recall: 0.9846
Micro-average F1 Score: 0.9846
```

Fig. 7. Regular - Precision, Recall and F1 Score.

Table 2. Regular: ROC-AUC.

Class	ROC-AUC
Class 0	0.9971
Class 1	0.9977
Class 2	0.9999
Class 3	1.0000
Micro-average ROC AUC: 0.999	

Table 3. Comparison results considered models.

Model	Classification	Loss	Accuracy	Scott's Pi
EfficientNetB0	Ordinal	0.06	0.98	0.97
	Regular	0.11	0.96	0.95
Basic	Ordinal	0.07	0.97	0.96
	Regular	0.11	0.96	0.94
EfficientNetV2B1	Ordinal	0.12	0.96	0.94
	Regular	0.17	0.95	0.92
Basic+ EfficientNetB0	Ordinal	0.09	0.97	0.95
	Regular	0.18	0.95	0.93
Basic+ EfficientNetB1	Ordinal	0.06	0.98	0.97
	Regular	0.18	0.93	0.91

4.3. Performance comparision

The below is the Table 3 which shows the loss, accuracy, and Scott's pi for each of our models using test data, including the models found in the appendix. Since the numbers may change between runs, we manually enter the numbers from the same run.

5. Conclusion

This study presents a novel approach to dementia severity classification by leveraging EfficientNet and its variants, including EfficientNetB0, EfficientNetV2B1, Basic + EfficientNetB0, and Basic + EfficientNetV2B, on a large dataset of 86,437 MRI brain images from the Open Access Series of Imaging Studies (OASIS). Recognizing the substantial class imbalance in the dataset—with a predominance of "None" and fewer "Moderate" severity cases—we applied a combination of undersampling and Synthetic Minority Over-sampling Technique (SMOTE) to achieve a balanced dataset, ensuring robust model training across all classes. The classification models were fine-tuned using various hyperparameters and performance metrics, such as Precision, Recall, F1 Score, ROC-AUC, and Scott's Pi, to optimize both ordinal and regular classification methods. Results demonstrated that the ordinal classification approach significantly outperformed regular categorical classification, highlighting its effectiveness in capturing the progressive nature of dementia. This finding underscores the potential of ordinal methods in clinical applications, where nuanced gradations in disease severity are crucial for personalized treatment planning.

Acknowledgement

The authors are grateful to Visvesvaraya Technological University, Belagavi, for the assistance for providing the infrastructure and support.

Funding

This research was not funded by any external sources.

Conflicts of interest

The authors declare that they have no conflicts of interest to this work.

References

1. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. V. D. Laak, B. V. Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
2. C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman et al., "NIA-AA research framework: toward a biological definition of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 14, no. 4, pp. 535–562, 2018.
3. Z. Breijyeh and R. Karaman, "Comprehensive review on alzheimer's disease: Causes and treatment," *Journal of Molecules*, vol. 25, no. 24, p. 5789, December 2020. doi: https://doi.org/10.3390/molecules25245789.
4. Z. Arvanitakis, R. C. Shah, and D. A. Bennett, "Diagnosis and Management of Dementia: Review," *Journal of American Medical Asscociation*, vol. 322, no. 16, pp. 1589–1599, October 2019. doi: https://doi.org/10.1001/jama.2019.4782.
5. B. M. Khrisna, V. Ch Jhansi, P. S. Shama, A. Boya Leelambika, C. Prakash, and B. V. V. N. Manikanta, "Novel solution to improve mental health by integrating music and IoT with neural feedback," *J of Computl Inform Sys.*, vol. 15, no. 3, pp. 234–239, 2019.
6. S. K. Aazad, T. Saini, A. Ajad, K. Chaudhary, and E. E. Elsayed, "Deciphering Blood Cells-Method for Blood Cell Analysis using Microscopic Images," *Journal of Modern Technology*, pp. 9–18, 2024.
7. R. S. Gopi, R. Suganthi, J. Jasmine Hephzipah, G. Amirthayogam, P. N. Sundararajan, and T. Pushparaj, "Elderly People Health Care Monitoring System Using Internet of Things (IOT) For Exploratory Data Analysis," *Babylonian Journal of Artificial Intelligence*, vol. 2024, pp. 54–63, 2024.
8. T. Sakirin and S. Kusuma, "A survey of generative artificial intelligence techniques," *Babylonian Journal of Artificial Intelligence*, vol. 2023, pp. 10–14, 2023.
9. S. Y. Mohammed and M. Aljanabi, "Advancing Translation Quality Assessment: Integrating AI Models for Real-time Feedback," *EDRAAK*, vol. 2024, pp. 1–7, 2024. doi: https://doi.org/10.70470/EDRAAK/2024/001.
10. S. Liu, H. Zhang, L. Song, Z. Li, C. Tan, D. Xing, and D. Shen, "Deep learning based classification of Alzheimer's disease using combined structural and functional MRI features," *Brain Imaging and Behavior*, Vol. 12, no. 3, pp. 891–901, 2018.
11. World Health Organization (W.H.O.), "*The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*," World Health Organization, Geneva, 2004 (Re-print: 2021).
12. D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults," *Journal of Cognitive Neuroscience*, Vol. 22, no. 12, pp. 2677–2684, 2010. doi: https://doi.org/10.1162/jocn.2009.21407.
13. Z. Li, Y. Wang, Z. Jiang, Z. Luo, J. Wu, and T. T. Toe, "Ensemble of CNN Models for Identifying Stages of Alzheimer's Disease: An Approach Using MRI Scans and SMOTE Algorithm," *2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS)*, July 2023. doi: https://doi.org/10.1109/isctis58954.2023.10213182.
14. E. E. Bron, M. Smits, W. M. Van Der Flier, H. Vrenken, F. Barkhof, P. Scheltens, and S. Klein, "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CAD Dementia challenge," *Neuro Image*, vol. 111, pp. 562–579, 2015.
15. M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 17 no. 4, pp. 425–436, 2018.
16. L. Zhang, S. Wang, C. Liu, and J. Zhang, "Deep learning for Alzheimer's disease diagnosis based on magnetic resonance imaging: A survey," *International Journal of Imaging Systems and Technology*, vol. 30, no. 1, pp. 64–75, 2020.
17. T. Zhou, K. H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Human Brain Mapping*, vol. 38, no. 4, pp. 2047–2061, 2017.
18. A. Gupta, M. Ayhan, and A. S. Maida, "Natural image bases to represent neuroimaging data," *International Conference on Learning Representations*, 2020.
19. S. Qiu, Z. Lu, L. Zhang, Z. Zhong, and Y. Liu, "A novel convolutional attention model for explainable Alzheimer's disease classification," *Frontiers in Neuroscience*, vol. 15, p. 661957, 2021.
20. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
21. M. A. Myszczynska, P. N. Ojamies, A. M. B Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Reviews Neurology*, vol. 16, no. 8, pp. 440–456, 2020.
22. G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard et al., "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, no. 10113, pp. 2673–2734, 2017.