Automated Alzheimer's Disease Diagnosis Using Deep Learning and Neuroimaging Data

Nirajan Bhattarai*[1] and Mostafa M. Fouda*†§[2]
*Department of Biomedical and Pharmaceutical Sciences, Idaho State University, Pocatello, ID 83209, USA.
†Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA.
§Center for Advanced Energy Studies (CAES), Idaho Falls, ID 83401, USA.
Emails: [1]bhatnira@isu.edu, [2]mfouda@isu.edu

***Abstract*—Alzheimer's disease, one of the most prevalent neurodegenerative disorders, although has no definitive cure, requires early diagnosis for managing adverse effects of the disease and improving patient's quality of life. Artificial intelligence in combination with neuroimaging could lead into rapid and accurate diagnosis of the disease. The aim of this study is to develop a computationally efficient, accurate and interpretable deep neural network model for the robust, rapid and accurate identification of dementia/Alzheimer's cases through neuroimaging data, Magnetic Resonance Imaging(MRI). To achieve the goal, we utilized kaggle's four class MRI dataset on Alzheimer's detection, preprocessed it for binary classification tasks i.e. dementia vs non-dementia, where all different severity of dementia were regrouped into a single dementia class. Due to the limited availability of data, we developed lightweight models, basic Convolutional Neural Network(CNN) and its variant CNN with attention head. We also incorporated transfer learning, with DenseNet121, VGG16, ResNet50. The modeling and evaluation was carried out in triplicate experiments to ensure reliability, and model interpretation with attention map was incorporated to ensure transparency and explainability on model's performance. We found that CNN with attention mechanism outperformed all the other model's in terms of accuracy, robustness and interpretability. Area Under Curve(AUC) of Receiver Operating Characteristics(AUC ROC), which was the primary evaluation metric, was 0.99+/- 0.002 for CNN model with attention. Model interpretation was also possible by analyzing the attention mechanisms at different epochs of the model training, revealing distinct differences in how the model attended to dementia and non-dementia MRI images. Finally, the attention mechanism is less explored in the domain of Alzheimer's detection using MRI images. Therefore, we recommend further in-depth studies with larger datasets and different variants of attention mechanisms to achieve the best possible model outcomes.**

***Index Terms*—Alzheimer's, Magnetic Resonance Imaging(MRI),Convolutional Neural Network, attention, model interpretation, transfer learning.**

I. Introduction

Alzheimer's disease (AD), a neuro-degenerative disorder and the most common form of elderly dementia affecting individuals over 65 years, has impacted 6 million people in the US alone and 50 million worldwide. This number is projected to rise to 152 million by 2050 [1]. Patients suffer from neurocognitive disorders, incipient cognitive decline, memory loss, and psychological and physical trauma. The financial burden of the disease is immense, estimated in trillions of dollars, accounting for approximately 1.16% of the world's GDP [2]. There is no definitive cure available for Alzheimer's disease because the underlying molecular mechanisms of its pathologies are still not clearly understood. However, there are a few medications intended to slow its progression and provide symptomatic relief. The management of the disease relies heavily on the early diagnosis of dementia, which supports prompt intervention and helps patients maintain good mental health and improve their quality of life [3].

Diagnosis of AD is challenging, as there is no definitive diagnostic method other than autopsy. However, brain scanning provides insight into characteristic lesions that may be associated with the disease [4].While family practice physicians, neurologists, geriatric psychiatrists, and geriatricians are expected to have diagnostic competency, automation and artificial intelligence approaches could aid in accurate early-phase diagnosis of the disease [5]. The diagnostic process is complex due to the variability and intricacy of the human brain. A diagnostic framework using AI requires substantial training with neuroimaging images of appropriate size and pixel details for accurate disease characterization and classification. Modern deep learning classifier algorithms present a promising solution due to their ability to learn and extract features with minimal preprocessing, enabling diagnostic discrimination [5].

The study aims to develop deep learning techniques to create efficient classification models for the rapid identification of dementia from Magnetic Resonance Imaging (MRI). Model interpretation was incorporated to explain the attention mechanism in attention-based networks. A comparative analysis of different models based on the available dataset size will be presented, along with a comparison to previously developed models.

II. Related Work

There are several successful studies which applied deep learning approaches to develop diagnostic models for identifying Alzheimer's disease using neuroimaging data, particularly MRI. Korolev et al. developed VoxCNN, a standard convolutional neural network (CNN) with four convolution blocks, and ResNet, a 21-layer residual network incorporating VoxRes blocks, to perform six one-vs-one binary classification tasks. Both VoxCNN and ResNet achieved competitive accuracies of approximately 80% without requiring complex preprocessing steps [6].

Similarly, Aderghal et al. utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) MRI dataset, including 815 subjects: 188 with Alzheimer's disease (AD) and 228 normal controls (NC), and developed a fusion model employing deep convolutional neural networks (CNNs) using 2D+ projections for the binary classification task of distinguishing AD from NC. The "majority-vote late fusion" model outperformed individual projections and intermediate fusion methods (FuseMe), achieving outstanding accuracy of 94.1% using a majority-vote late fusion approach [7]. Cheng et al. developed a CNN-based multi-modal model that applied cascaded convolutional neural network, utilizing 3D-CNNs for feature extraction and a 2D-CNN for combining multimodal features. They used MRI and FDG-PET images from the Alzheimer's Disease Neuroimaging Initiative (ADNI), achieving accuracy of 89.64% for distinguishing Alzheimer's cases against normal control [8]. Liu et al. used MRI images from 311 subjects from the ADNI database, classified as normal control (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD), to develop a hybrid deep learning algorithm consisting of "stacked sparse auto-encoders" for feature extraction. The proposed method achieved an accuracy of 87.67% for the binary classification task distinguishing between AD and MCI and outperformed support vector machine(SVM) in multi-class classification [9]. Mehmood et al utilized a transfer learning approach, using pre-trained VGG architecture, using ADNI dataset consisting of 700 patients distributed over normal control, mild cognitive impairment(MCI),late mild cognitive impairment(LMCI) and Alzheimer's disease(AD) patients, and had a performance accuracy of 98.73%, and 83.72% accuracy for classification AD vs NC, EMCI vs LMCI [10].

III. Method

A. Dataset

The four-class Alzheimer's detection dataset is publicly available and can be accessed on Kaggle (www.kaggle.com), authored by Survesh Dubey. The MRI images are categorized into four classes: Mild Dementia, Moderate Dementia, Non-Dementia, and Very Mild Dementia, with sample sizes of 28 subjects, 2 subjects, 100 subjects, and 70 subjects, respectively [11]. Due to the imbalanced data distribution among these classes, we reformulated the data by combining Mild Dementia, Moderate Dementia, and Very Mild Dementia into a single Dementia class, while keeping Non-Dementia as a separate class. Essentially, we redefined the dementia identification task as a binary classification problem.

B. Models

1) Convolutional Neural Network(CNNs): The basic convolutional neural network (CNN) architecture takes an RGB image of size 100×100×3 as input. It consists of three convolutional layers. The first layer includes 64 filters with a kernel size of 3×3, ReLU activation, batch normalization, and max-pooling with a 2×2 window. The second and third layers have 128 and 256 filters, respectively, while maintaining the same structure as the first layer.The feature maps produced by the convolutional layers are flattened and passed to a fully connected layer with 512 neurons and ReLU activation. A dropout rate of 0.5 is applied for regularization to prevent overfitting. The final layer consists of a single neuron for the binary classification task, using binary cross-entropy as the loss function and Adam as the optimizer.

2) Transfer learning with DenseNet 121, VGG16, and Resnet50: We utilized three pre-trained transfer learning models—DenseNet121, VGG16, and ResNet50—due to their proven effectiveness in image classification tasks. DenseNet121 is a densely connected feed-forward deep neural network with 121 layers, enabling efficient feature reuse. VGG16, originally developed by the Visual Geometry Group at Oxford, consists of 16 weight layers, including 13 deep convolutional layers and 3 fully connected layers.Each model processes input images of size 100×100×3 through its respective deep convolutional architecture. This is followed by feature reduction using a Global Average Pooling (GAP) layer, which captures essential features while reducing dimensionality. The extracted features are then passed through a dense layer with 512 neurons, ReLU activation, and a dropout rate of 50% for regularization. The final layer comprises a single neuron with a sigmoid activation function for binary classification.During training, the models were optimized using the Adam optimizer, with binary cross-entropy loss serving as the objective function.

3) Attention CNN: Attention CNN is a deep convolutional neural network with an attention mechanism. The architecture consists of input later that takes an image of size 100 X 100 X 3, has three convolutional layers of 64, 128 and 256 filter sizes respectively. The attention mechanism enforces a convolution operation to generate the attention map, which helps to apply weight to the feature map. This mechanism allows the model to attend to more relevant regions. The output after enforcing the attention mechanism is flattened, and passed to fully connected layers of 512 neurons with ReLU activation, along with a drop rate of 0.5 for regularization.Again, the final output neuron consists of a single neuron with sigmoid activation function. Similarly to the architecture above, the model is compiled with Adam optimizer and binary-cross entropy loss function.

C. Evaluation Metrics

The Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve, a plot against the True Positive Rate (TPR) against the false positive rate (FPR) provides a balanced and elaborative evaluation of the performance of classification models across different thresholds. AUC of ROC was used as the primary evaluation criteria of the models. The True Positive Rate (TPR) represents the ratio of true positive instances to the sum of True Positive (TP) and False Negative (FN) instances, while the false positive rate is a ratio of false positive instances to the sum of False Positives (FP) and True Negatives (TN). Mathematically, additional evaluation metrics included Precision, Recall, F1 score, and Mathew Correlation Coefficient defined by the following equations:

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

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

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

$$F_1 \text{ Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{MCC} = \frac{(\text{TP} \cdot \text{TN}) - (\text{FP} \cdot \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

D. Interpretation of Attention Mechanism

Attention mechanism focuses on the most important region of the feature space or image, which acts a learned weight, to highlight important parts of features map having influence in the decision of the model. It enhances the decision making process and, most importantly, the interpretation of it.

Mathematically, it is expressed as follows [12]:

$$A = \sigma(w_a * f + b_a) \tag{2}$$

f is a feature map derived after applying the last convolutional operation and before applying the attention mechanism, with dimensions $H \times W \times C$, where H, W, and C are the height, width, and number of channels of the feature map, respectively. A is the attention map, which is a single-channel output of size $H \times W \times 1$. w_a and b_a are the weights and bias applied to the convolutional layers that generate the attention. σ is the sigmoid activation function used to normalize the attention values between $[0, 1]$.

$$f_{\text{att}} = f \cdot A \tag{3}$$

$f_{\text{attention}}$, an attention mechanism, is basically a re-weighted feature map obtained through dot product between features and the attention weight.

IV. Results

Since the replication of the model used in previous similar studies to identify dementia/Alzheimer's using MRI image dataset was not fully possible because of the unavailability of complete code, model parameters, or weights, direct comparison with previous studies was challenging. We developed various model architectures while keeping downstream structure consistent i.e.fully connected layers of 512 neurons with ReLU activation, drop out rate of 0.5 for regularization, the final output neuron consisting of a single neuron with sigmoid activation function for binary classification. Also, all the models were compiled with adam optimizer and binary cross entropy loss function. A particular emphasis on the transfer learning approach,DenseNet121,ResNet50 and VGG16, because the dataset was small.

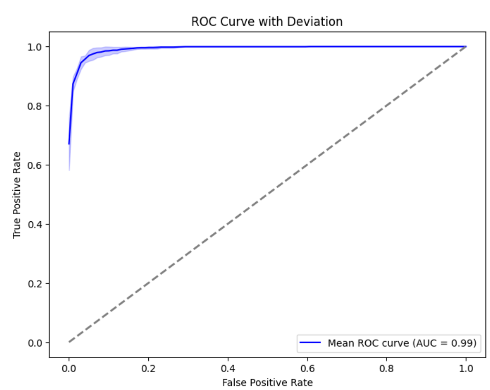


Fig. 1: **Receiver Operating Characteristic(ROC) plot showing performance of the best performing model: CNN with attention.** The Area Under the Curve (AUC) of the ROC is a numerical metric that indicates the classification performance of the model, as shown in the bottom-right corner of the plot. The ROC curve represents the mean, while the shaded area indicates the standard deviation from triplicate experiments.

The results demonstrate that the CNN architecture with attention mechanisms outperforms all other models, achieving superior performance across all evaluation

Model Name	Accuracy	ROC AUC	Precision	Recall	F1 Score	MCC
Simple CNN	0.91 ± 0.084	0.95 ± 0.059	0.92 ± 0.082	0.90 ± 0.089	0.91 ± 0.086	0.82 ± 0.17
DenseNet121	0.64 ± 0.096	0.77 ± 0.094	0.65 ± 0.14	0.85 ± 0.14	0.71 ± 0.024	0.35 ± 0.15
ResNet50	0.75 ± 0.18	0.79 ± 0.23	0.76 ± 0.19	0.91 ± 0.065	0.80 ± 0.10	0.50 ± 0.36
VGG16	0.50 ± 0.00	0.50 ± 0.00	0.33 ± 0.24	0.67 ± 0.47	0.44 ± 0.31	0.00 ± 0.00
CNN with Attention	**0.96 ± 0.009**	**0.99 ± 0.003**	**0.95 ± 0.006**	**0.96 ± 0.023**	**0.96 ± 0.009**	**0.92 ± 0.017**

TABLE I: **Performance metrics of different models evaluated using triplicate experiments with different test subsets.** The results are presented as the mean $\pm$ standard deviation, rounded to two significant figures. The area under the curve (AUC) of the receiver operating characteristic (ROC) is the primary evaluation metric.

metrics. The low variability across experiments suggests the model's robustness, as indicated by a ROC-AUC score of 0.9932 ± 0.0025, an accuracy of 0.9538 ± 0.0058 (Table 1) and as shown in ROC(Fiurge this). The performance of transfer learning for classification was poor, particularly for VGG16. The classification evaluation appeared very poor (ROC AUC = 0.5, precision = 0.5) and seemingly random, as reflected in the ROC curve, confusion matrix, and evaluation metrics.The simple CNN architecture demonstrated competitive performance compared to the architecture with an attention mechanism. However, the performance metrics were approximately 4–5% lower (as shown in the table), and the variability across experiments was higher compared to the CNN architecture with the attention mechanism.

One of the important features of attention-based model architectures is their ability to offer model interpretation, providing an explanation of the attention mechanism through attention map visualization for classification. Two examples of the attention mechanism applied to MRI images of the dementia and non-dementia classes are shown in Figure 2. The attention patterns are clearly distinguishable between dementia and non-dementia MRI images. The attention map MRI image associated with dementia may highlight key structural features affected in Alzheimer's disease. This could potentially capture underlying pathological features or changes associated with the disease.

V. Discussion

As we already discussed, effective management of Alzheimer's disease relies on early diagnosis and preventive interventions. One of the cost-effective, rapid, and accurate diagnostic approaches is the integration of machine learning and deep learning into a diagnostic pipeline. Jo et al. analyzed sixteen different studies focusing on traditional machine learning and deep learning approaches conducted between 2013 and 2018 [13] with limited data availability. Four studies applied a hybrid model combining deep learning and traditional machine learning, achieving up to 98.8% accuracy, which was with Stacked Autoencoder (SAE), in Alzheimer's disease classification [8]. Our study achieved approximately 96% accuracy and a 99% AUC-ROC score. Since our modeling used a limited amount of data, we built and analyzed the model in a triplicate experimental setup with different train and test sizes. Although the CNN model with an attention mechanism demonstrated superior and robust performance in our study, the attention mechanism had not been actively employed in the model architecture in previous studies using deep learning for MRI-based Alzheimer's classification. Not only does it provide more accurate predictions by focusing on important features, but it also offers model interpretability through attention maps. Our result is consistent with findings of Ranghavan et al. that ensemble deep learning models, particularly those enhanced with attention layers, significantly outperformed base CNN models and various transformer learning models for infrared breast cancer detection task [14].

Inspired by the functioning of human vision, attention is based on the cognitive mechanism of limiting information processed at a single time [15]. Based upon this mechanism, a model with selective attention was developed aiming to simulate focus on certain parts of the images or video. In the domain of computer vision, this attention mechanism allows the computer to attain important information in the images, often using a mask [16]. There have been numerous studies that applied attention as a key component of their model architecture for medical image classification.Ilse et al. utilized gated attention in histopathology images [17], while

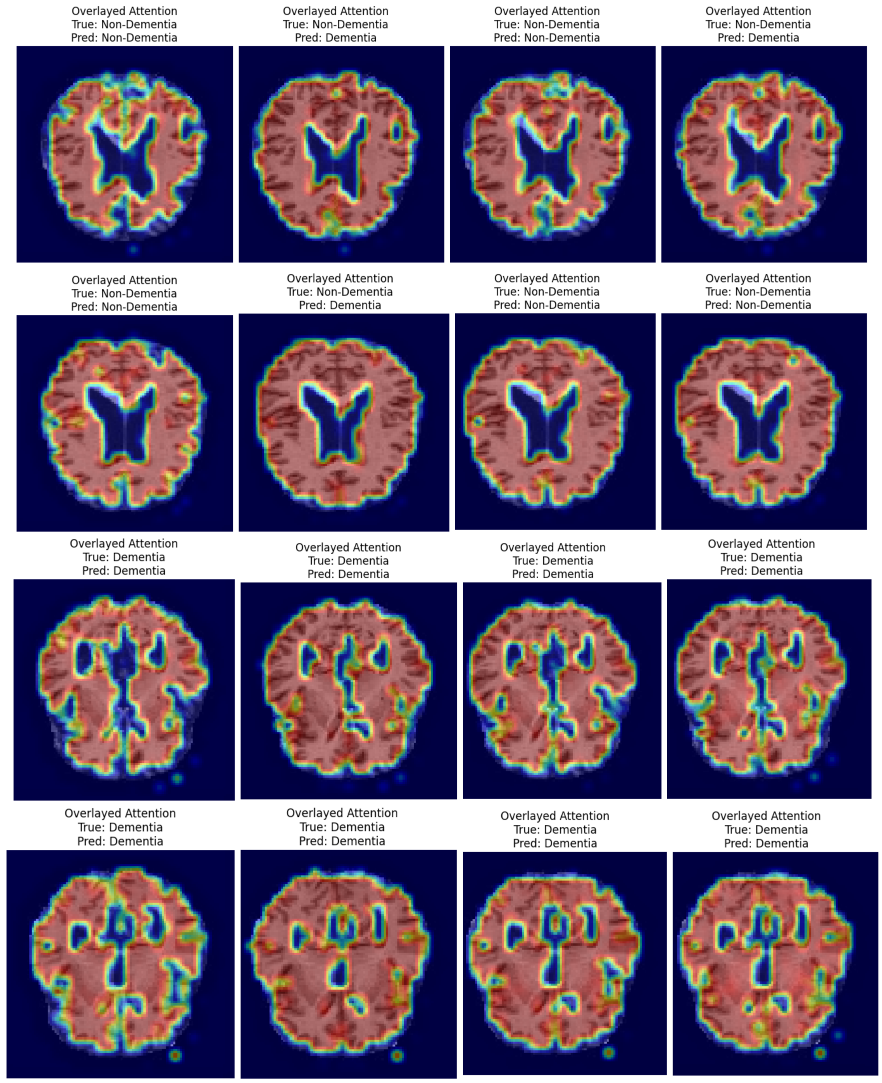
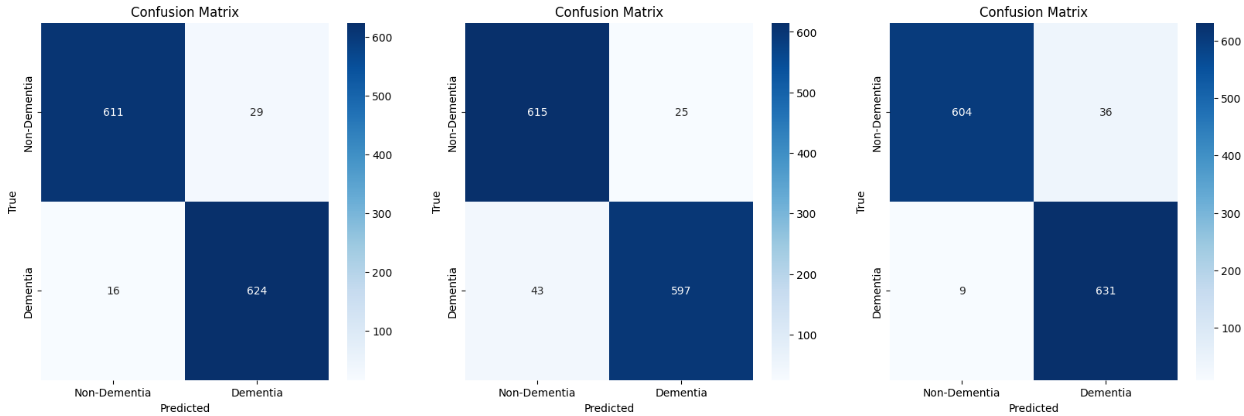


Fig. 2: **Attention map visualization.** The attention heat maps of two non-dementia cases (first and second rows) and two dementia MRI image classifications (last two rows) are shown. Each row consists of attention heat maps overlaid on the original MRI images, obtained at 10, 20, 30, and 50 epochs during model training. The images are labeled with their true labels and the predicted classification labels.

Gheflati and Rivaz leveraged transformer-based models in multi-database approaches [18] for classification of histopathology images into benign, malignant, or normal categories. Similarly, the attention mechanism has been incorporated while modeling automated detection of COVID-19 with CT and X-ray image data, including Vision Transformers, hybrid Transformer-CNN models, and federated learning, improving classification and facilitating collaborative learning [19].

A significant hurdle in the adoption of deep learning approaches in clinical decision-making is the lack of transparency and interoperability in the model's decision-making process, also known as explainable artificial intelligence (XAI). Incorporation of attention mechanisms enables the mapping of attention weights, as mentioned earlier, to highlight important regions or features in an image. These can be visualized for interpretation of model's functioning using various approaches, such as GRAD-CAM, CAM, or attention maps [14], [20]. Our study showed a characteristic difference in attention mapping of important features in MRI images between dementia and non-dementia, depicting key regions of focus by the model in a heatmap. These explanations can be corroborated with medical data if they are capturing relevant anatomical regions affected in Alzheimer's. In a study by Gu et al., spatial, channel,



(a) Experiment 1 (b) Experiment 2 (c) Experiment 3

Fig. 3: **Confusion matrices of the attention-based CNN model across three different experiments are presented from left to right.**

and scale attention techniques were utilized to analyze and understand how neural networks derive pixel-level predictions for image segmentation tasks [20].

VI. Conclusion

The addition of an attention head to the CNN architecture enhanced the model's performance, robustness, and interpretability for automated detection of Alzheimer's disease using MRI data. To ensure rigor and validity, modeling and evaluation were conducted in triplicate experiments. The most robust model was identified based on low variation in performance across experiments, with AUC-ROC used as the primary metric, as it evaluates the model's performance at different thresholds and provides a balanced assessment.While the basic CNN also demonstrated competitive performance, models employing transfer learning approaches, such as DenseNet121, ResNet50, and VGG16, exhibited lower predictive performance across all evaluation metrics. We attributed this discrepancy to the limited dataset size used in the modeling and evaluation process.For future studies, we recommend utilizing larger datasets, exploring different variants of attention mechanisms, and implementing strong control measures to develop more effective attention-based CNN models for Alzheimer's classification. These models are computationally efficient and provide interpretability, ensuring transparency in predictions.In our study, we observed a clear distinction in the attention maps between dementia and non-dementia MRI images, further supporting our recommendations. This highlights the potential of attention mechanisms, which remain underexplored in the field of Alzheimer's detection using MRI data.

VII. Disclaimer Statement

The AI assistant was used to improve the grammar and sentence structure, as well as to derive summaries from the research article. However, AI was not used to generate content or write any part of the paper.

References

[1] R. S. Turner, T. Stubbs, D. A. Davies, and B. C. Albensi, "Potential New Approaches for Diagnosis of Alzheimer's Disease and Related Dementias," *Frontiers in Neurology*, vol. 11, p. 496, Jun. 2020.

[2] K. Sharma, S. Pradhan, L. K. Duffy, S. Yeasmin, N. Bhattarai, and M. K. Schulte, "Role of Receptors in Relation to Plaques and Tangles in Alzheimer's Disease Pathology," *International Journal of Molecular Sciences*, vol. 22, no. 23, p. 12987, Nov. 2021.

[3] A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, and M. Nami, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," *Brain Informatics*, vol. 10, no. 1, p. 17, Dec. 2023.

[4] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain Imaging in Alzheimer Disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 4, pp. a006 213–a006 213, Apr. 2012.

[5] S. Basu, K. Wagstyl, A. Zandifar, L. Collins, A. Romero, and D. Precup, "Early Prediction of Alzheimer's Disease Progression Using Variational Autoencoders," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, vol. 11767, pp. 205–213, series Title: Lecture Notes in Computer Science.

[6] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. Melbourne, Australia: IEEE, Apr. 2017, pp. 835–838.

[7] K. Aderghal, J. Benois-Pineau, K. Afdel, and C. Gwenaëlle, "FuseMe: Classification of sMRI images by fusion of Deep CNNs in 2D+ projections," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*. Florence Italy: ACM, Jun. 2017, pp. 1–7.

[8] D. Cheng and M. Liu, “CNNs based multi-modality classification for AD diagnosis,” in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. Shanghai: IEEE, Oct. 2017, pp. 1–5.

[9] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, “Early diagnosis of Alzheimer’s disease with deep learning,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. Beijing, China: IEEE, Apr. 2014, pp. 1015–1018.

[10] A. Mehmood, S. Yang, Z. Feng, M. Wang, A. S. Ahmad, R. Khan, M. Maqsood, and M. Yaqub, “A Transfer Learning Approach for Early Diagnosis of Alzheimer’s Disease on MRI Images,” *Neuroscience*, vol. 460, pp. 43–52, Apr. 2021.

[11] Sarvesh Dubey, “Alzheimer MRI 4 classes dataset,” 2020. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[12] Z. Wang and F. A. Ghaleb, “An attention-based convolutional neural network for intrusion detection model,” *IEEE Access*, vol. 11, pp. 43 116–43 127, 2023.

[13] T. Jo, K. Nho, and A. J. Saykin, “Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data,” *Frontiers in Aging Neuroscience*, vol. 11, p. 220, Aug. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2019.00220/full

[14] K. Raghavan, S. Balasubramanian, and K. Veezhinathan, “Explainable artificial intelligence for medical imaging: Review and experiments with infrared breast images,” *Computational Intelligence*, vol. 40, no. 3, p. e12660, Jun. 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/coin.12660

[15] J. H. Maunsell, “Neuronal Mechanisms of Visual Attention,” *Annual Review of Vision Science*, vol. 1, no. 1, pp. 373–391, Nov. 2015. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-vision-082114-035431

[16] X. Yang, “An Overview of the Attention Mechanisms in Computer Vision,” *Journal of Physics: Conference Series*, vol. 1693, no. 1, p. 012173, Dec. 2020. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1693/1/012173

[17] M. Ilse, J. M. Tomczak, and M. Welling, “Attention-based Deep Multiple Instance Learning,” 2018, version Number: 4. [Online]. Available: https://arxiv.org/abs/1802.04712

[18] B. Gheflati and H. Rivaz, “Vision Transformer for Classification of Breast Ultrasound Images,” 2021, version Number: 2. [Online]. Available: https://arxiv.org/abs/2110.14731

[19] T. Goncalves, I. Rio-Torto, L. F. Teixeira, and J. S. Cardoso, “A Survey on Attention Mechanisms for Medical Applications: are we Moving Toward Better Algorithms?” *IEEE Access*, vol. 10, pp. 98 909–98 935, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9889720/

[20] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, “CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 699–711, Feb. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9246575/