EXPLORATION OF EXPLAINABLE DEEP LEARNING MODEL FOR CLASSIFICATION AND ANALYSIS OF ALZHEIMER’S DISEASE

Project Reference No.: *47S_BE_0090*

College ***:*** *Bapuji Institute of Engineering and Technology, Davanagere*
Branch ***:*** *Computer Science and Engineering*
Guide(s) ***:*** *Dr. Pradeep N and Prof. Bharath H C*
Student(S) ***:*** *Mr. Manoj Bhavihal*
Mr. Swaroop Aradhya N M
Ms. A Lasya Priya
Ms. Bhoomi Shree C P

Keywords:

Alzheimer’s Disease, Deep Learning, Convolutional Neural Network (CNN), eXplainable Deep Learning (XDL), eXplainable AI (XAI), Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP).

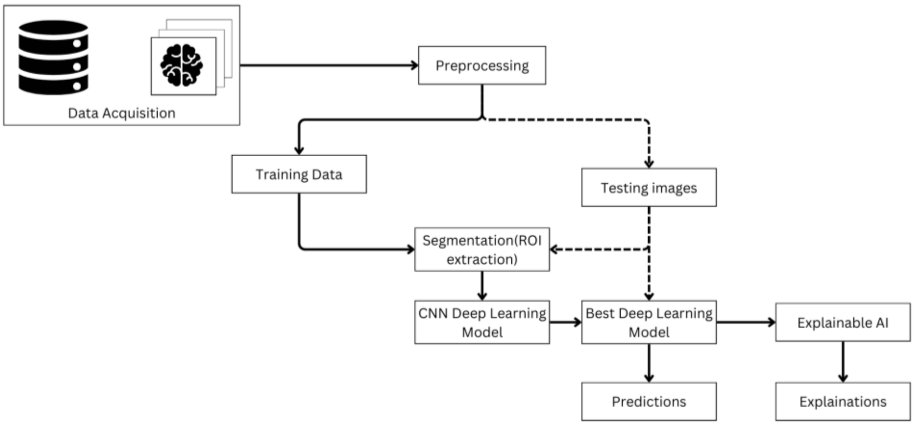
Introduction:

Alzheimer’s disease is a brain disorder that slowly destroys memory and thinking skills, and eventually, the ability to carry out the simplest tasks. It is the most common cause of dementia, accounting for 60-70% of cases. This study aims to use the power of deep learning techniques to enhance the understanding and diagnosis of AD. This project is built by incorporating explainable deep learning models such as LIME and SHAP. By combining the deep learning with the transparency of explainable AI techniques, the model developed not only achieves high classification accuracy but also provides insights into the features driving the classification decisions. Recent advancements in deep learning have shown promise in deep learning classification using complex data. However, the lack of interpretability in deep learning models hinders their adoption in medical domain. To address this, Integration of eXplainable Artificial Intelligence (XAI) techniques is explored to enhance interpretability in AD classification. This work delves into various XAI methods, such as feature importance ranking, providing transparent insights into the features used for AD classification. These understanding aids personalized treatment strategies and disease management by identifying critical features contributing to clinical diagnosis. After conducting a thorough literature review of studies by Alatarny et al., Grigas et al., and Elgammal et al., it becomes apparent that existing research in the field of Alzheimer's disease diagnosis using deep learning models suffers from a lack of transparency. These studies highlight the black box nature of deep learning models, wherein the reasoning behind their predictions remains opaque. To address this limitation, the present research endeavors to develop methods that enhance the interpretability of deep learning models for Alzheimer's disease diagnosis.

Objectives of the project:

- To acquire publicly available dataset and preprocess the dataset that will aid for the subsequent objectives.
- To train and test various CNN models and to identify best CNN trained model and its hyper parameters in the classification of the Alzheimer's disease.
- To investigate and integrate best XAI model to the best CNN model to derive the various insights/explanation for the classification made by the CNN model.
- To compare and analyze the performance of the best CNN model and XAI model with the state-of-the-art techniques.

Methodology:

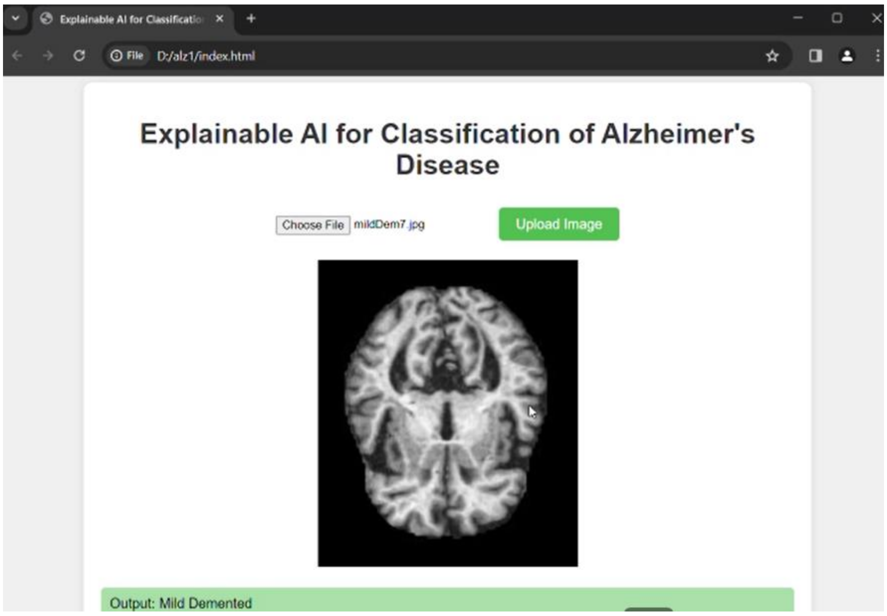


Project implementation starts with data acquisition, gathering relevant images for analysis, followed by preprocessing to refine and enhance the dataset's quality. Subsequently, the data is divided into training and testing sets, allowing a Convolutional Neural Network (CNN) to learn from the training data and segment regions of interest (ROIs) in the testing images.

Selecting the most effective deep learning model involves evaluating various architectures to pinpoint the optimal one for accurate segmentation. Post- segmentation, explainable AI techniques come into play, shedding light on CNN's decision-making process. These methods, like attention mechanisms or saliency maps, provide insights into how the model identifies and segments ROIs, enhancing transparency and interpretability.

This comprehensive approach, encompassing data preparation, model selection, and the integration of explainable AI, ensures a systematic pathway for accurate image segmentation while enabling a clear understanding of the CNN's segmentation decisions.

Results and Conclusion:



The accuracy of the implemented model is 96.3%, precision of the implemented model is 96.9%, the recall of the implemented model is 96.5% and F1-score of the implemented model is 96.7%.

eXplainable AI (XAI) has been integrated using the SHAP (SHapley Additive exPlanations) to explain the predictions of a Convolutional Neural Network (CNN) model trained for image classification and Local Interpretable Model-agnostic Explanations (LIME) technique to explain the predictions of a pre-trained InceptionV3 model for image classification.

Prioritizing the creation of explainable AI systems for Alzheimer's diagnosis stands as a critical pursuit. Such efforts enable early disease detection and informed medical decisions. This initiative gives technological advancement, it underscores a deep commitment to healthcare integrity and patient well-being. By clarifying diagnostic processes and enhancing their reliability, this work improves the outcomes for individuals affected by Alzheimer's and their caregivers. This endeavor embodies a concerted step toward advancing healthcare efficacy and fostering a more promising future for those grappling with this prevalent condition.

Innovation in the Project:

Using eXplainable Artificial Intelligence (XAI) for Alzheimer's disease classification is innovating because it illuminates the decision-making process of machine learning models. Instead of treating AI predictions as incomprehensible black boxes, XAI methods offer transparency and interpretability, allowing researchers to discern the rationale behind the model's conclusions. This transparency is particularly significant in Alzheimer's diagnosis, where precise insights are crucial. Furthermore, the interpretability afforded by XAI empowers clinicians to confirm the model's predictions, fostering confidence in its diagnostic utility. This collaboration between AI and human expertise improves diagnostic precision and enables tailored treatment strategies.

Scope for future work:

- Feature Importance Analysis: Continued exploration of feature importance methods to identify the most relevant biomarkers and clinical indicators of Alzheimer's disease, leading to more targeted diagnostic approaches.
- Integration of Multi-Modal Data: Incorporating diverse data sources such as neuroimaging, genetic markers, and clinical records to build more comprehensive models capable of capturing the complex nature of Alzheimer's progression.
- Addressing Ethical Considerations: Ethical considerations surrounding data privacy, consent, and bias must be addressed to ensure responsible deployment of AI-based diagnostic tools in healthcare.
- Robustness and Generalization: Enhancing the robustness and generalization capabilities of XAI models to perform reliably across diverse populations and healthcare.
- Interactive Visualization Tools: Developing user-friendly visualization tools that allow clinicians and researchers to interactively explore model predictions and insights, facilitating knowledge discovery and decision-making.