A Hybrid Approach to Alzheimer's Diagnosis: Predictive Modeling Using Diagnostic Data and MRI Images Integrated in a Web Interface

Megha Aby Alex [1], *Shyma* Sulthana [1], and *Swarnalatha* R [2]*
[1] Department of Biotechnology, Birla Institute of Technology and Science, Pilani, Dubai Campus, DIAC, Dubai, 345055, United Arab Emirates.
[2] Department of Electrical and Electronic Engineering, Birla Institute of Technology and Science, Pilani, Dubai Campus, DIAC, Dubai, 345055, United Arab Emirates.

Abstract Alzheimer's is a neurological disease that worsens with time and has no known cure. There is no turning back once the diagnosis is made; cognitive deterioration simply becomes worse. This research aims to reduce the time taken to visit the doctor for a diagnosis, which is only done once the patient or their family is suspicious because of any symptoms shown. This time is reduced by sifting through data when a patient goes to their normal checkups, this way, any abnormality that points at a chance of Alzheimer's can be caught and flagged by the ML (machine learning) model. Two ML models were trained to predict the percentage of chance of a patient having Alzheimer's. One model was trained by feeding it datasets of biomarkers and the other by using MRI brain scans of patients with and without Alzheimer's. These models were then integrated into a web interface for ease of use, where a combined prediction using probabilities calculated by both the models would be displayed. ***Keywords— Alzheimer's, machine learning, predictive modelling, MRI, diagnosis***

1 Introduction

Alzheimer's is a neurodegenerative disease which is progressive in nature, and is the most common cause of dementia. There is currently no known cure for Alzheimer's [1]. The disease is characterized by the accumulation of abnormal plaques and neurofibrillary tangles in the brain [2]. The aforementioned abnormalities are one of the reasons for the nerve cell death, while the beta-amyloid plaques, built of protein fragments, could be found in between the nerve cells. The neurofibrillary tangles are made of twisted fibers of a protein called Tau and these build up inside the cells, causing inflammation and eventually death of the cell [2].Though there are a few hypotheses for the cause of Alzheimer's Disease, there is no definite cause of disease known yet. Researchers have still not figured out if the pathological changes are what cause Alzheimer's Disease or if it's the disease itself that is causing the drastic pathological changes.

* Corresponding author: swarnalatha@dubai-bits-pilani.ac.in

Due to their many similarities, Alzheimer's disease pathology and aging pathology are extremely difficult to distinguish from one another, for instance, the spectrum of pathological changes in both cases, like the pathology of Braak stage VI disease [3], are marked as "definite" Alzheimer's Disease by CERAD (Consortium to Establish a Registry for Alzheimer's Disease) criteria, and marked "high likelihood" Alzheimer's Disease by the NIA-Reagan criteria, but this pathology has also been observed in the cognitively intact, that is, the patients who do not show the most common symptom of Alzheimer's Disease, loss of cognitive function. The Braak stages are a set of 6 stages that categorize the progression of the neurofibrillary tangles, topographically and these are highly used as the diagnostic criteria for Alzheimer's disease. Since Alzheimer's Disease is not considered a normal part of aging, it is important to diagnose if the dementia experienced by the patient can be classified as Alzheimer's Disease, aging or other conditions. This is also because dementia can be the symptom for many different diseases like Lewy body disease, Vascular disease, Frontotemporal degeneration etc. There are three main stages of Alzheimer's, they are Asymptomatic, Mild cognitive impairment (MCI) and Dementia. The third stage is further classified into three more depending on severity, these are namely Mild, Moderate and Severe Dementia. Machine learning techniques are used in this research for predictive modelling. Here, computers are fed vast amounts of data. The machines improve their accuracy with respect to their performance with exposure and experience to additional data. Performance and test results improve with larger training data sets. This research uses a combination of ANN (Artificial Neural Networks) and CNN (Convolutional Neural Networks) ML models. Artificial neural networks can be used to predict and classify anything from written data that is letters and numbers, images, speech like required in voice recognition and even robotics. ANNs have weighted connections that act like interconnected neurons, they mimic a biological nervous system. The model starts out with data collection, then the network structure is configured and all the hidden layers are put into place, after which the weight or bias is balanced by optimization. The model is trained till it reaches the threshold of accuracy required for the research.

Convoluted neutral networks are a type of machine learning model used for recognition of images. These models can be used to assess images in the healthcare field like images of various types of scans. These include but are not limited to X-Rays, MRIs (Magnetic Resonance Imaging), PETs (Positron Emission Tomography), SPECTs (Single photon emission computed tomography) CTs (Computed Tomography) etc. By analyzing these scans, the model can categorize or classify the images into the classifiers set by the user. CNN is an excellent model that could be used to identify patterns in images and learn various image attributes. However, CNNs have a drawback in that they need a lot of image datasets to train. There is also the disadvantage of it not being interpretable by the user, wherein once the results come, the user would not be able to interpret how the algorithm arrived at such a conclusion leaving them to blindly trust the model's working. Machine Learning can be used in many areas that come under healthcare, like imaging, treatment personalization, monitoring of patients, pandemic prediction etc. Notably, Machine learning can be used for early disease prediction, by having the model read and learn prediction patterns by feeding the model data. An ML model that can efficiently swift through large amounts of data can be created and this way it is possible to increase the accuracy with respect to prediction even with a huge number of variables involved.

What the research aims to do with ML is place these models in hospital systems such that when patients go for any biomarkers testing, the machine stores and sends this data to an ML model. The model can be built such that it can store patient history and can be updated as and when doctors give a diagnosis, including differentials. This way the data of patients who aren't aware they could possibly have Alzheimer's can also be fed through the system and can be flagged in case they show any pathological result that may hint at a possibility of Alzheimer's. This information can then be provided to the doctor so that they can interpret the flagged biomarkers as hints, pointing towards cognitive impairment and the like. Since there is no cure for the disease, current treatment plans focus on delaying the onset and further development of the disease. To delay its onset, there needs to be an increased accuracy in prediction of the disease in its early stages so treatment plans can be considered as early as possible to reduce symptoms before they occur. ML models can be used in this case as they can help with faster diagnosis by automating the process of reading patient data, and predicting possibilities. By using ML models in clinics and hospitals, we can flag any abnormalities or coincidences that may hint at a patient being in the asymptomatic stage of Alzheimer's. By identifying this, flagged patients can be sent to doctors specializing in Alzheimer's and similar conditions to confirm a diagnosis and start treatment management plans as soon as possible.

2 Literature Review

As of all diagnosis, doctors start by enquiring about the patient's medical and family history. After that, they examine every medicine the patient is taking in an attempt to find any that might be contributing to the patient's cognitive decline or make their cognitive abilities worse. Thirdly, a bedside cognitive exam is done, like the MMSE (Mini Mental State Examination) or MoCA (Montreal Cognitive Assessment) to test the level of the patients' cognitive functions [1]. They also do blood tests to make sure that the cause for the dementia is not something reversible.All possible reasons for the dementia are ruled out before scans and imaging tests are done. An MRI Scan of the brain is done, to look for atrophy of brain tissue. In recent days, volumetric MRI has become a much more preferred imaging technique as it can reveal the shrinkage of the brain tissues and hence the shrinkage in the brain volume better. Particularly, the hippocampus's shrinkage in the medial temporal lobe. The hippocampus shrinkage is a notable identifier of Alzheimer's Disease. Other than MRI's, functional brain imaging techniques like PET, F MRI (Functional MRI), SPECT are used. Once these scans hint at Alzheimer's, doctors perform a small invasive surgery known as a spinal tap, to confirm the diagnosis of Alzheimer's Disease. Even though it does not help with determining the severity of Alzheimer's Disease, this is especially helpful as a diagnostic tool in preclinical stages of the disease. Here, a hollow needle is injected into the lower back of the patient and cerebrospinal fluid (CSF) is withdrawn [4]. The CSF is then tested to check for the levels of beta-amyloid and tau proteins. Decrease in the levels of beta-amyloid and an increase in tau protein levels compared to normal levels are characteristic indicators of Alzheimer's Disease.

As there is no cure for Alzheimer's Disease, current treatments include symptom management and general patient care. Currently, two types of drugs have been approved to be administered to patients with Alzheimer's Disease. One of the types are cholinesterase inhibitors like donepezil, rivastigmine, and galantamine. The second types are partial N-methyl D-aspartate (NMDA) antagonists also referred to as Memantine. The first type works by increasing levels of acetylcholine, a type of neurotransmitter involved in increasing neurological function. It helps the patient by improving their memory, learning and other

cognitive functions. The second type works by slowing down the accumulation of Calcium intracellularly.

As for early Alzheimer's Disease, there are currently two FDA-approved drugs, namely Lecanemab and Donanemab. These drugs target the misfolded beta-amyloid proteins that have accumulated outside the nerve cells to form plaques; by doing so they reduce the amyloid plaques present in the brain. These drugs have proved to slow the rate of cognitive decline by reducing the amyloid plaques. Despite FDA approval, some patients, particularly those with the apolipoprotein E4 genotype, may experience potentially fatal adverse effects on consumption of these drugs, including fluid accumulation or brain bleeding. This is due to the body's immune response to the amyloid- targeting immunotherapies. Due to this, scans are ordered before any medication is prescribed for Alzheimer's Disease. Recent studies are being done on finding lesser invasive and more faster methods of diagnosis, as current methods require spinal taps, an invasive procedure and scans like PET, which are rather expensive. The methods that are being researched include potential correlations between certain blood biomarkers [5] and Alzheimer's. Researchers are looking at blood biomarkers that are biomarkers that can be tested through blood like cholesterol [6], diabetes along with other biomarkers that can be easily accessed/tested like saliva, olfactory mucus, lipids etc. Currently a study is being done on 19 blood biomarkers that are being considered for Alzheimer's disease detection. Due to the fast-paced advancements in computation and machine learning, disease diagnosis takes place faster than ever before. The major goal of this research is to identify a method that uses relatively simple model architectures and requires fewer resources without sacrificing prediction accuracy, despite it being known that machine learning accuracy improves with growing model architecture complexity.

3 Methodology

Figure 1 is a flowchart outlining the proposed methodology of this research. Initial stages include literature review and data collection as well as exploration. Next, we will select two models to train the obtained datasets. We will train both models separately on their respective datasets, and then create a web interface to enable user data input. Finally, the web interface will combine predictions from both models using a weighted average and display the results. The following paragraphs explain the details of the methodology. Two publicly available datasets that were sought on Kaggle were used to train the machine learning models. One such dataset consisting of diagnostic data was of particular interest to us, due to the presence of cholesterol levels. As previously mentioned in the literature review, cholesterol levels have been found to play a significant role in Alzheimer’s diagnosis.

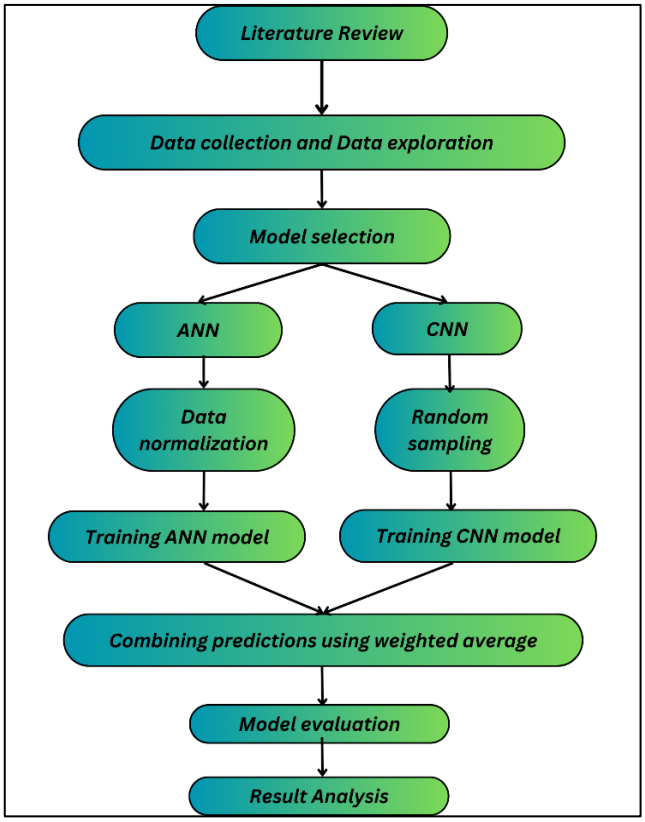

Figure 1. Flowchart representation of the proposed methodology

Patient ID, age, gender, ethnicity, education level, BMI, smoking, alcohol use, physical activity, food, family history of Alzheimer's, diabetes, heart disease, depression, head injury, high BP (blood pressure), systolic and diastolic BP, Cholesterol levels (Total, LDL, HDL, and Triglycerides), MMSE, Functional evaluation, memory issues, behavioral issues, ADL, personality changes, confusion, disorientation, forgetfulness, diagnosis, responsible physician and sleep quality were all included in the complete set of diagnostic information for each patient, which ranged in age from 60 to 90. The second dataset that was used to train our model consisted of pre-processed T-1 weighted, 2D MRI images of the brain. The photos in this dataset were all 128x128 pixels in size and in the.jpg format. The following sub-folders were included in each of the test and train folders:

1. No Impairment
2. Very Mild Impairment
3. Mild Impairment
4. Moderate Impairment

The images were obtained from 100, 70, 28, and 2 patients corresponding to the categories mentioned above. Each patient's brain was sliced into 32 MRI scans and also pre-processed to remove the skull. The original dataset contained heavily imbalanced classes in both test and train sub-folders; Under the train folder, No Impairment class contained 2560 images, Very Mild Impairment contained 1792 images, Mild Impairment had 717 images and Moderate

Impairment had merely 12 images. The dataset used by us was pre-processed using WGAN-GP Wasserstein Generative Adversarial Networks + Gradient Penalty) in such a way that synthetic MRI scans were generated to balance all the classes so that each sub-folder contained 2560 images for the train sub-folder. The test folder was pre-processed but remained untouched in terms of the addition of synthetic images.

Figure 2 is a flowchart that outlines the train and test folders along with the 4 sub-folders in the Alzheimer's dataset consisting of 2D MRI images.

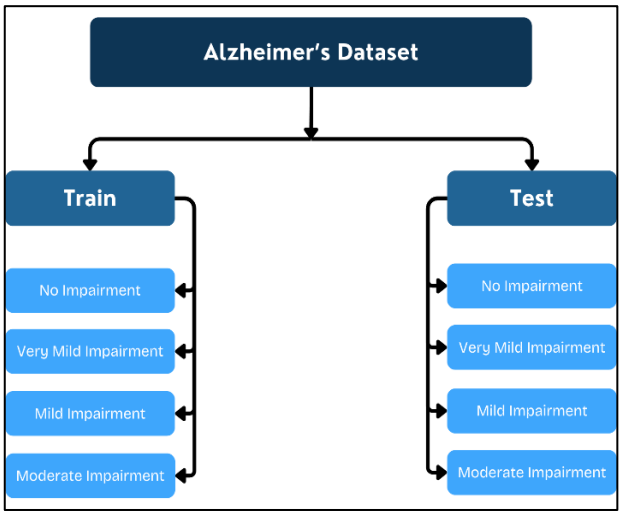

Figure 2. Flowchart representation of the folders in the dataset

In order to accurately depict the features pertinent to the suggested ANN model, a Pearson correlation graph first had to be created. In order to include all of the highly correlated features as well as a few less correlated ones that still affect the diagnosis, the input features were selected using the plot displayed in Figure 3. The following are the features that were chosen: Age, gender, cardiovascular disease, hypertension, HDL cholesterol, MMSE, memory complaints, behavioral issues, ADL, diagnosis, and family history of Alzheimer's disease. The dataset was then preprocessed using Standard Scaler. There was a class disparity between patients with and without Alzheimer's in the original dataset. In the training dataset, the number of samples falling under class 0 or No Alzheimer's were 1112, while the number of samples falling under class 1 or Alzheimer's were 607. Consequently, a method called SMOTE (Synthetic Minority Oversampling Technique) was used to balance both classes and enhance the model's capacity for generalization. After applying this technique, both the classes now contained 1112 samples each.

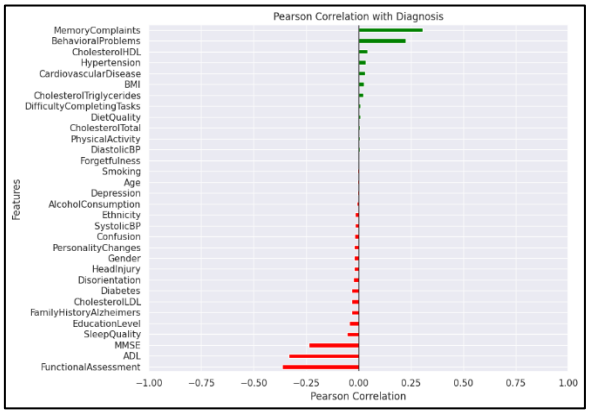

Figure 3. Pearson Correlation Plot

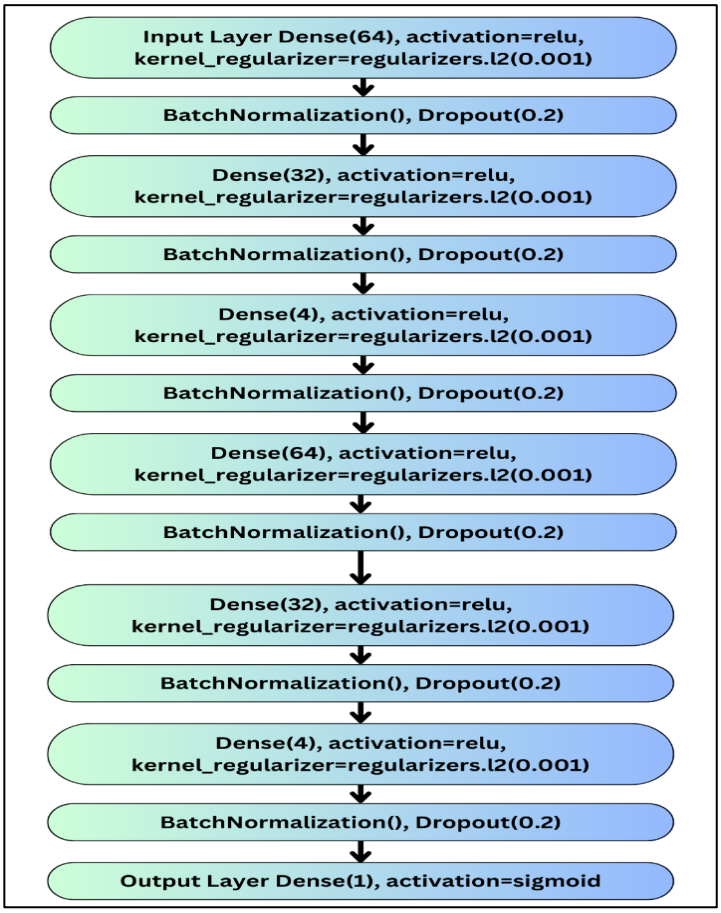

Figure 4. Flowchart of ANN model

TensorFlow's Sequential API was used to build the model [7-8]. 64 densely linked neurons with the activation function ReLU and input shape = 11 made up the input layer. A dropout layer that randomly removes 20% of the neurons from the preceding layer comes after a batch normalization layer that speeds up convergence. This is done in order to prevent overfitting. The activation function and 32 tightly linked neurons make up the following layer, and so forth. Finally, a single neuron and a sigmoid activation function that produces a probability value between 0 and 1 function as the output layer, which focuses on binary categorization. The Adam optimizer, a widely used optimizer, was used to compile the model. The loss function was calculated using the binary cross-entropy function. The SMOTE resampled data was used to train the model, which ran for 100 epochs. Additionally, callbacks like ReduceLROnPlateau and early stopping were employed; the former lowers the learning rate when the loss function does not drop considerably, while the latter halts training when the loss function stops falling. The ANN model architecture used to train the diagnostic dataset in this research is shown in Figure 4.

The original dataset needed to be changed to accommodate the binary output provided by the ANN model before the CNN model could be developed. The "No Impairment" folder remained unaltered, but the classes "Very Mild Impairment," "Mild Impairment," and "Moderate Impairment" were merged into a single folder named "Impairment." The reorganized Alzheimer's dataset, which now only has two subfolders rather than four, is displayed in Figure 5.

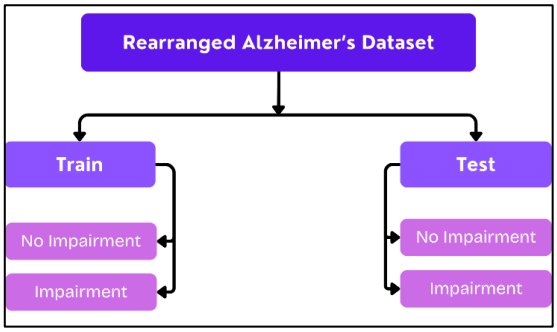

Figure 5. Flowchart of rearranged folders

Similar to the ANN model, the CNN mode was initiated using the Sequential API [9-8]. The first convolutional block applies 32 filters, each 3 by 3, to the input images. Here, non-linearity is introduced using the ReLU activation function. The input shape is the same size as the training and test datasets. The feature maps undergo a 2 by 2 down-sampling by the max pooling layer. With 64 filters, the subsequent convolutional block resembles the first layer. The flatten layer transforms the 2D feature maps from the earlier layers into a 1D vector. Lastly, a fully linked layer with 256 densely coupled neurons was applied to ensure a 50% dropout rate. The last output layer has a "softmax" activation function that converts the raw output scores into probabilities. The model's compilation makes use of the optimizer "Adamax," which modifies the learning rate for every parameter. The loss was calculated using the categorical entropy function. Twenty epochs were used to train the model, and the callback early stopping was also used. The CNN model architecture used to train the MRI dataset in this study is shown in Figure 6. Integrating the learned models into a web application that lets

users enter 2D MRI scans and pertinent diagnostic information to determine their likelihood of having Alzheimer's disease was the final step in this research. Streamlit was used to construct the web application.

Additionally, weights were constructed for subsequent use in integrating CNN and ANN model predictions. This is because the ANN model fared better than the CNN model. Consequently, the ANN model's forecast was given more weight than the CNN model's prediction. A prompt to upload a 2D MRI image was provided to the user because the CNN model requires this in order to function. The pre-processing features in the code precisely transform the input images into the same size that were used for training. The model computes and provides its prediction immediately when the image is uploaded. The same diagnostic markers that were utilized to train the ANN model were also employed as prompts to provide diagnostic data.

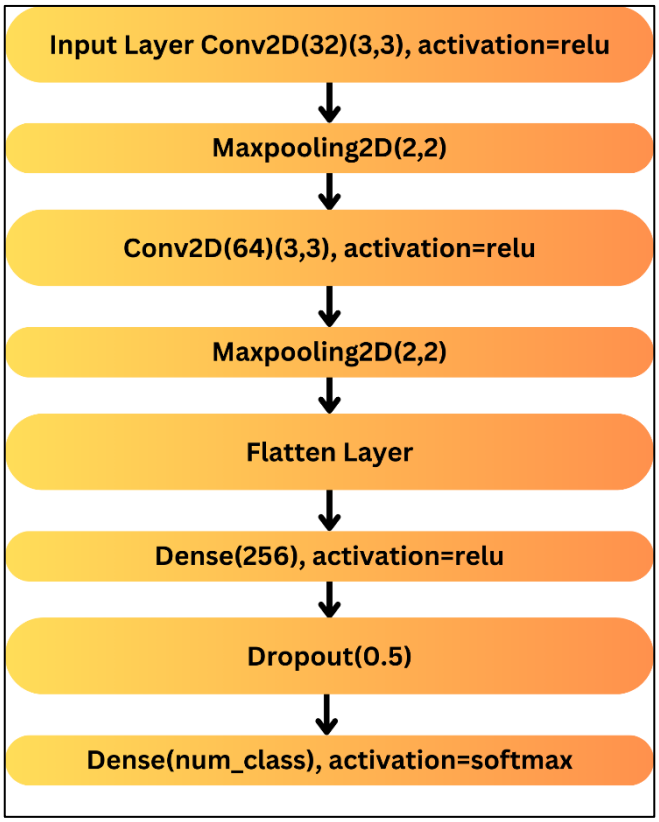

Figure 6. Flowchart of CNN model

To make calculations easier, the interface explicitly tells the user to enter 0 if the answer is no and 1 if the answer is yes. When it comes to gender, 0 denotes a man and 1 denotes a woman. Functional assessment and HDL cholesterol are examples of inputs that have been coded to accept integer values. Lastly, a button labeled "Get Prediction" was placed after the data input prompts. When this button is clicked, the ANN model is prompted to provide a

prediction. Additionally, this button uses the following formula to determine the combined probability (CP) of developing Alzheimer's disease.

4 Results and Discussion

An estimate of the ANN model's performance can be obtained by looking at the model's accuracy which came up to 90% as well as performance metrics plot, which includes F1, recall, and precision score. According to definitions, recall stands for the number of positives the model accurately predicted out of all potential positives, and precision is the number of positives that has been predicted correctly by the model. The aforementioned scores are harmonically aggregated to form the F1 score [10]. The below equations were used to calculate the metrics. Keep in mind that FP, FN, TN and TP mean false positives, false negatives, true negatives and true positives, respectively.

$$Precision = TP \div (TP + FP) \quad (1)$$

$$Recall = TP \div (TP + FN) \quad (2)$$

$$F1\ score = 2 \times \frac{(Recall \times Precision)}{Precision + Recall} \quad (3)$$

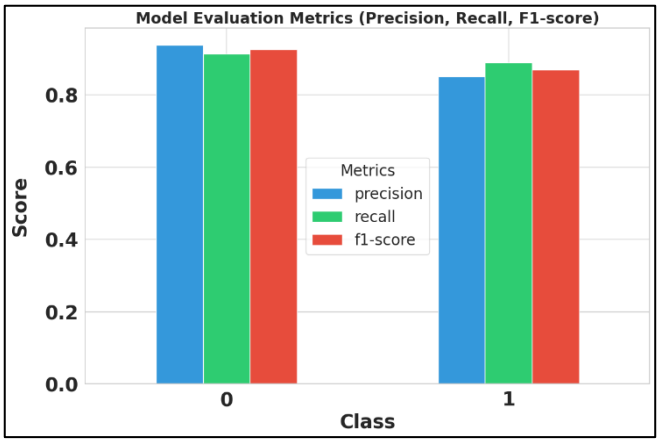

Figure 7. Plot of f1, recall and precision score of all classes ANN model

Evaluating the plot shown in figure 7, for class 0 (no Alzheimer's), the f1, recall and precision scores are as follows:

1. Precision: 94%
2. Recall: 91%
3. F1 score: 93%

For class 1 (Alzheimer's), the f1, recall and precision scores are as follows:

4. Precision: 85%
5. Recall: 89%
6. F1 score: 87%

Class 0's performance metrics were marginally superior to Class 1's. This could be explained by the fact that there was a class imbalance, which may have had an impact on the training process even though the SMOTE technique had greatly reduced it. Since SMOTE does not add diversity to the synthetic samples, the model may not have learnt the full disease complexity. Following this, to further assess the ANN model's performance, a confusion matrix was plotted. Out of all the positives in the sample i.e., 253 + 17 = 270, 253 samples were accurately recognized as positive, whereas the remaining 17 samples were mistakenly categorized as negative, according to the plot in Figure 8. 136 of the 160 negative samples in the negative class were correctly classified as negative, but 24 of them were mistakenly classified as positive.

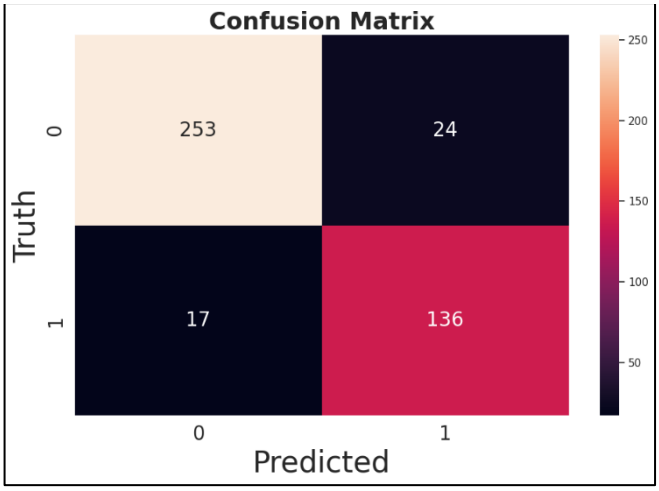

Figure 8. Confusion matrix for ANN model

Of 430 samples, 389 were correctly classified and 41 samples were incorrectly classified. This misclassification by the model could be due to the classes being ambiguous or unclear in pattern. The classes also had a significant overlap; through manual error analysis of the misclassified examples, the ADL values were noted to be higher than average for the cases incorrectly identified as ‘not Alzheimer’s’. Further analyzing ADL values for the correctly classified samples, it seemed to have no correlation with respect to the output being positive for Alzheimer’s or not. That is, cases having high ADL were sometimes classified as positive for Alzheimer’s and other times not. To further try and explain the misclassification, a permutation feature importance plot was generated as shown in Figure 9.

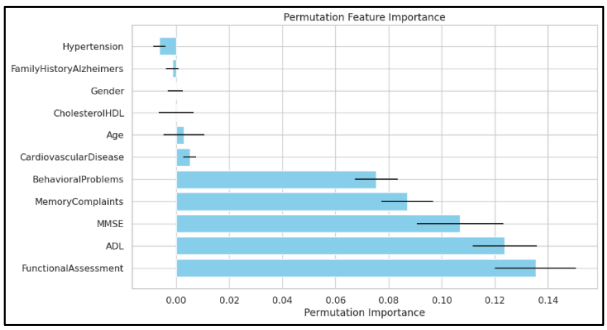

Figure 9. Permutation Feature Importance Plot

Using the permutation feature importance plot, it can be interpreted that although we had initially only picked relevant features for the prediction of Alzheimer's using diagnostic data, the model learnt that some features like Hypertension, Cholesterol HDL and Cardiovascular disease were irrelevant. This means that some relevant features have high correlation among each other, leading to its irrelevance in the prediction by the model.

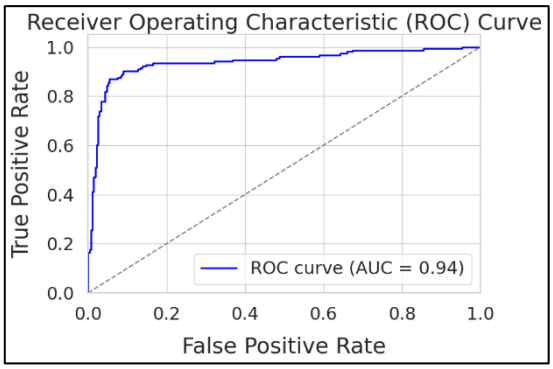

Figure 10. ROC curve for ANN model

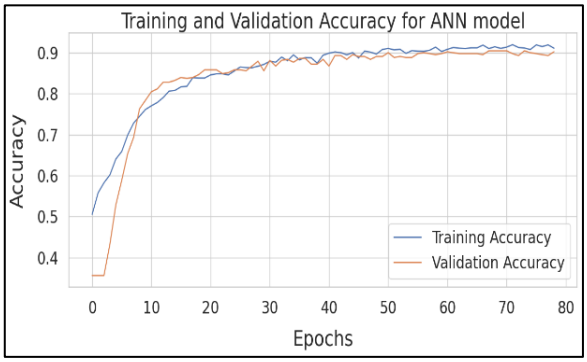

Figure 11. Plot comparing validation and training accuracy of the ANN model

We can infer with certainty that the ANN model can effectively distinguish between the classes based on an analysis of the ROC curve produced for the model, which is displayed in Figure 10. The value obtained for the area under the curve was 0.94 highlighting excellent performance.

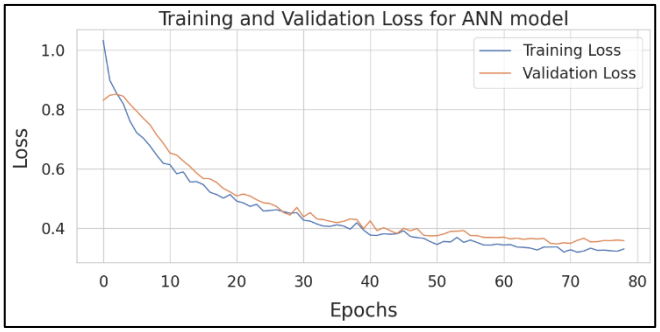

Figure 12. Plot comparing training and validation loss of the ANN model

From Figures 11 and 12 it can be inferred that the training and validation loss decreases over epochs whereas the accuracy increases over epochs, which is expected. To evaluate the CNN model, the following plot was generated as shown in figure 13, consisting of f1, recall and precision scores across all classes and the following inferences were made:

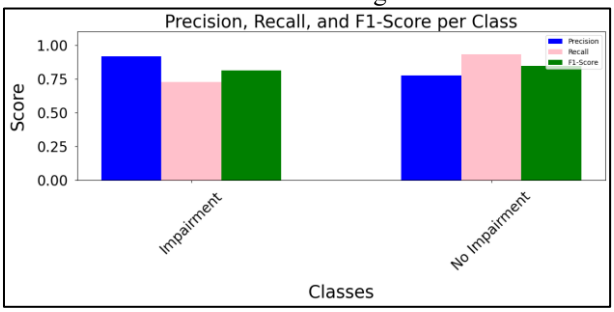

Figure 13. Plot of precision, recall and f1 scores of all classes for the CNN model

For class 0 (No Impairment), the f1, recall and precision scores are as follows:

1. Precision:78%
2. Recall: 93%
3. F1 score: 85%

For class 1 (Impairment), the the f1, recall and precision scores are as follow:

1. Precision: 92%
2. Recall: 73%
3. F1 score: 81%

Based on these findings, it is fair to say that the model performed almost as well in both classes. The accuracy of the model, which was 83%, provided more proof of its effectiveness.

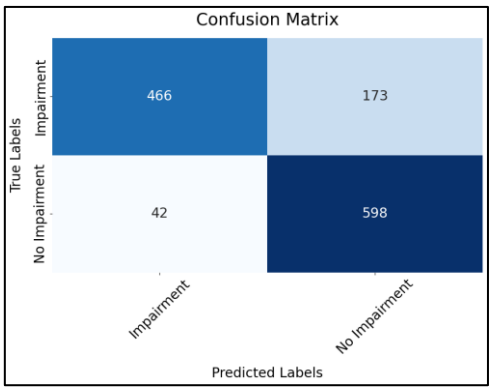

Figure 14. Confusion matrix for CNN model

A confusion matrix was plotted for the CNN model, as depicted in Figure 14, to understand how well the model distinguishes between the classes. Out of 1279 samples, 466 and 598 samples were correctly classified as impaired and not impaired respectively. On the other hand, 42 and 173 samples were misclassified as impaired and not impaired respectively.

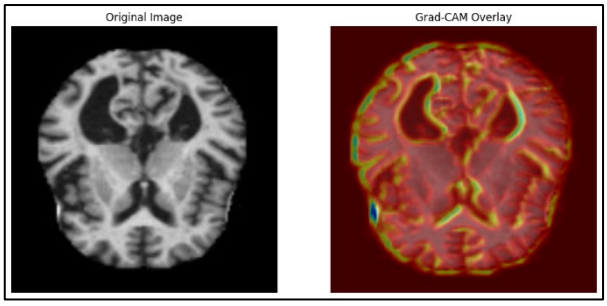

Figure 15. Class-activation map using grad-CAM

To possibly explain the misclassification, a class-activation map was generated using Grad-CAM as shown in figure 15. The red overlay highlights the areas that were important for prediction while the green and blue areas were less relevant. From the visualization it can be inferred that the entire area consisting of the brain scan was relevant for the prediction for the model, which can be explained by the reason that brain volume is a crucial deciding factor in whether a person is suffering from Alzheimer's or not. Although this was as expected, it would have been more beneficial if the model learnt to focus on key areas that indicate Alzheimer's such as the Hippocampus and temporal lobes. An ROC curve was generated for the CNN model as shown in figure 16 to analyze further its ability to distinguish between the classes. The curve shows that the model performs fairly well as the area of the curve is 0.89.

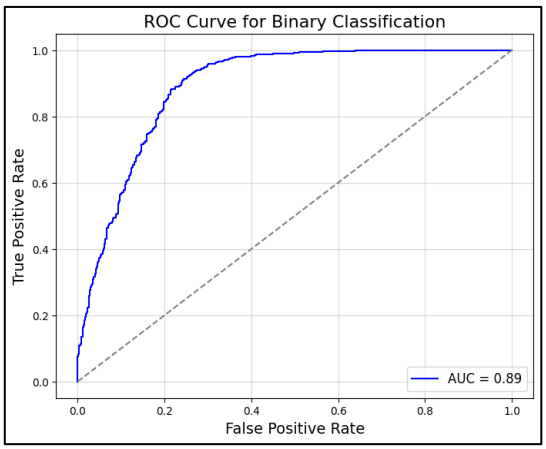

Figure 16. ROC curve for CNN model

Additionally, the CNN model was also evaluated on the basis of graphs shown in Figures 17 and 18. The former compares the accuracy of validation and training, whereas the latter compares validation and training loss.

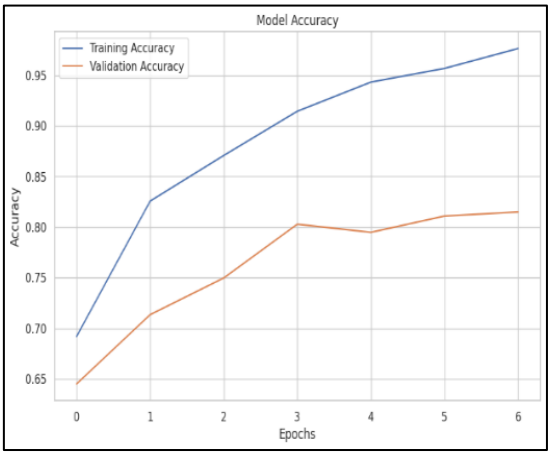

Figure 17. Plot comparing training and validation accuracy of the CNN model

Figure 16 suggests that the model is functioning properly since the training and validation accuracy increased over time. As seen in figure 17, a plot comparing training and validation loss was also generated.

It is clear from analyzing the plot in Figure 17 that the training loss converges as predicted. In contrast, the validation loss just marginally converges, then dips and begins to rise, suggesting that the model may be overfitting.

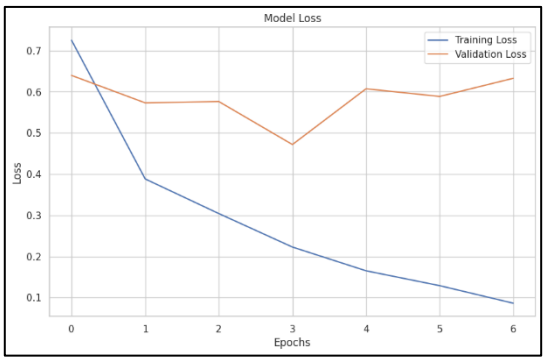

Figure 18. Plot comparing training and validation loss of the CNN model

Both the models were then loaded into a web interface from which users can enter diagnostic data and a 2D MRI image. A screen with a few prompts shows after the web interface code has been run. The online interface will then provide an output that includes a combined probability that takes into consideration both types of data. The formula for calculating combined prediction (CP) using weights is given in equation (4).

$$CP = (wCNN \times CNNpred) + (wANN \times ANNpred) \quad (4)$$

Here. wCNN and wANN are the weights, while CNNpred and ANNpred are the predicted values from the CNN and ANN model respectively. As mentioned earlier, more weight was given to the ANN model due to the fact that it performed better than the CNN model, as can be inferred from table 1. The weights were calculated using the following formula:

$$wCNN = \frac{CNN\ accuracy}{CNN\ accuracy + ANN\ accuracy} \quad (5)$$

$$wANN = \frac{ANN\ accuracy}{CNN\ accuracy + ANN\ accuracy} \quad (6)$$

From this, the assigned weights were calculated as 0.45 for the CNN model and 0.55 for the ANN model.

Table 1. *Performance summary of ANN and CNN models*

	ANN		**CNN**	
Class	Class 0	Class 1	Class 0	Class 1
Precision	94%	85%	78%	92%
Recall	91%	89%	93%	73%
F1 score	93%	87%	85%	81%
Accuracy	90%		83%	

The proposed methodology directly addresses the lack of accessible diagnosis for the general public who are in dire need of a diagnosis; by combining the predictions of two well-known and robust models, this approach enables a rather easy-to-understand as well as high accuracy diagnosis without the use of extremely complicated machine learning techniques. A significant drawback of current machine learning models is their utilization of heavy computational resources [11]. The simplicity of the methodology means that it can be employed in a relatively decent computational setup. Moreover, such an approach can be easily scaled up for its use in larger and complex environments such as hospitals, with minimal modifications to the architecture. Such systems can be encouraging for people with limited machine learning knowledge or experience to implement concepts of artificial intelligence in their work-flow.

5 Conclusion

Machine learning concepts can significantly improve the duration of days or even months taken to reach a patient's diagnosis and the effort that goes into putting together the puzzle, that is the various miniscule symptoms the patient shows, leading to the big picture, the diagnosis. It seems logical to presume that a machine will be able to conduct a more comprehensive analysis of a patient's profile than a human can, provided the correct filters/ layers have been included into the system. Keeping this in mind, there are many enhancements that can be added to the model to bring it to its best, most efficient version, like a multimodal system. This research has shown that it is possible to create and executive such a system and how useful it would be in its full-fledged state.

6 Limitations and Future scope

The model has not been verified by medical professionals which could impact the performance of such systems. Nevertheless, the suggested methodology's current state yields encouraging findings, which ought to motivate more research. Currently, the model lacks explainability features that could potentially make the model predictions more interpretable. This could be fixed by using 3D MRI scans rather than 2D ones in order to better preserve significant details and characteristics that 2D brain MRI images may overlook [12]. To add to this hindrance, there is also a lack of publicly available datasets that are documented properly. Preserving these models' capacity for diagnosis should be a task undertaken by medical practitioners, who must update and control them on a regular basis. The grave importance we may be placing on these artificially intelligent models that have the potential to determine your life's path in the future is empirically supported.

The model uses relatively simple architecture and datasets for training, which may hinder the process of automation in hospitals due to the requirement of pre-processing a variety of data. The models were also trained on relatively small datasets, due to lack of computational power. This can be improved by using pre-trained models such as ResNet, EfficientNet, VGG models etc., [13-14] and increasing the processing power and resources. The next step in enhancing the previously mentioned architecture might be longitudinal data, which consists of data points that could be utilized to forecast the course of Alzheimer's disease. A multi-modal system [15] that can process diagnostic data and visuals prior to the web interface would be ideal. To provide a more accurate and coherent representation, this system would attempt to link features from the various data sources rather than training on each dataset independently.

The high computational cost of such models and the lack of publicly accessible data may be deterrents. Therefore, further study in this field is needed, to better understand machine learning and Alzheimer's in general.

The authors would like to thank BPDC for the opportunity to work on this paper.
No funding was required for the paper.
The datasets used were publicly available datasets from Kaggle.
Author contributions:
Megha A: Conceptualization, developed the ML models and implemented the algorithms, Collection and curation of datasets, Creation of web interface, analysis of results, writing and reviewing manuscript.
Shyma S: Conceptualization, Introduction, and literature review of manuscript, analysis of results, writing, and review of manuscript.
R. Swarnalatha: Conceptualization, Supervision, analysis of results, and review of manuscript.

References

1. J. Fessel, The several ways to authentically cure Alzheimer's dementia. *Ageing Res. Rev.* **92**, 102093 (2023). https://doi.org/10.1016/j.arr.2023.102093
2. P. Goel, S. Chakrabarti, K. Goel, K. Bhutani, T. Chopra, S. Bali**,** Neuronal cell death mechanisms in Alzheimer's disease: an insight. *Front. Mol. Neurosci.* **15**, 937133 (2022). https://doi.org/10.3389/fnmol.2022.937133
3. J. Therriault, et al.**,** Biomarker modeling of Alzheimer's disease using PET-based Braak staging. *Nat. Aging* **2**, 526–535 (2022). https://doi.org/10.1038/s43587-022-00204-0
4. H. Hampel, et al.**,** The use of lumbar puncture and safety recommendations in Alzheimer's disease: a plain language summary. *Neurodegener. Dis. Manag.* **12**, 221–229 (2022). https://doi.org/10.2217/nmt-2022-0012
5. O. Hansson, et al.**,** The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimer's Dement.* **18**, 2669–2686 (2022). https://doi.org/10.1002/alz.12756
6. F.M. Feringa, R. Van Der Kant**,** Cholesterol and Alzheimer's disease: from risk genes to pathological effects. *Front. Aging Neurosci.* **13**, 690372 (2021). https://doi.org/10.3389/fnagi.2021.690372
7. S.-H. Han, K.W. Kim, S. Kim, Y.C. Youn**,** Artificial neural network: understanding the basic concepts without mathematics. *Dement. Neurocogn. Disord.* **17**, 83–89 (2018). https://doi.org/10.12779/dnd.2018.17.3.83
8. S. Ray, A quick review of machine learning algorithms. In: *2019 Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, 35–39 (2019). https://doi.org/10.1109/COMITCon.2019.8862451
9. L. Alzubaidi, et al.**,** Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8
10. C. Goutte, E. Gaussier**,** A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: *Adv. Inf. Retr.* **3408**, 345–359 (2005). https://doi.org/10.1007/978-3-540-31865-1_25
11. S.E. Sorour, A.A.A. El-Mageed, K.M. Albarrak, A.K. Alnaim, A.A. Wafa, E. El-Shafeiy**,** Classification of Alzheimer's disease using MRI data based on deep learning techniques. *J. King Saud Univ. Comput. Inf. Sci.* **36**, 101940 (2024). https://doi.org/10.1016/j.jksuci.2024.101940
12. A. Ebrahimi, S. Luo, Alzheimer's Disease Neuroimaging Initiative**,** Convolutional neural networks for Alzheimer's disease detection on MRI images. *J. Med. Imaging* **8**, 024503 (2021). https://doi.org/10.1117/1.JMI.8.2.024503
13. M. Heenaye-Mamode Khan, P. Reesaul, M.M. Auzine, A. Taylor**,** Detection of Alzheimer's disease using pre-trained deep learning models through transfer learning: a review. *Artif. Intell. Rev.* **57**, 275 (2024). https://doi.org/10.1007/s10462-024-10914-z

14. S. Savaş**,** Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. *Arab. J. Sci. Eng.* **47**, 2201–2218 (2022). https://doi.org/10.1007/s13369-021-06131-3
15. Z. Kong, M. Zhang, W. Zhu, Y. Yi, T. Wang, B. Zhang**,** Multi-modal data Alzheimer's disease detection based on 3D convolution. *Biomed. Signal Process. Control* **75**, 103565 (2022). https://doi.org/10.1016/j.bspc.2022.103565