Christy Hernandez
Livan Valdes
Pace University
CS 672 - Fall 2024

AI-Driven Dementia Detection:
A Multi-Model Approach Using Images and Patient Data

Abstract

Dementia, a progressive neurodegenerative disorder, affects millions worldwide, with early detection being critical for effective intervention. This project leverages deep learning and machine learning to enhance dementia diagnosis using both imaging and patient demographic data. A MobileNetV2-based convolutional neural network was trained on labeled medical images to classify individuals as "With Dementia" or "Non-Demented." The dataset was preprocessed and augmented to ensure robust training and testing, with the model achieving strong performance metrics while maintaining computational efficiency.

In parallel, patient demographic and medical information—including age, gender, ethnicity, BMI, alcohol consumption, and diagnosis—was analyzed using Decision Tree and Naive Bayes models. This dataset, comprising over 2,000 patient records, provided a complementary perspective to image-based classification, enabling deeper insights into dementia predictors. Together, these approaches highlight the potential of integrating deep learning and traditional machine learning in developing comprehensive, AI-driven diagnostic tools for dementia screening. This report details the methodologies, model architectures, and results of this dual-dataset strategy.

Data Description

For this project 2 different datasets from Kaggle were used, each dataset was used with the intention of training different Machine Learning models and a Deep Learning Model within a Python Notebook.

The main dataset that was used to train and test the deep learning MobileNetV2 model for the project, consists of 4 folders containing brain scan images categorized into 'Non Demented', 'Very Mild Dementia', 'Mild Dementia', and 'Moderate Dementia', the images are labeled and merged into a single directory ('All Labeled Images') that contained 2 folders labeled

‘Dementia’ or 1 and ‘No Dementia’ or 0, this creates a binary classification problem for the image data, even though the original dataset had multiple dementia severity levels, after that they were resized and preprocessed for deep learning. The original folders in the dataset are distributed the following way:

All Labeled Images: 86437

- With Dementia: 19215
  - Mild Dementia: 5002
  - Moderate Dementia: 488
  - Very mild Dementia: 13725
- Non Demented: 67222

The second dataset used for this project consists of a CSV file that contains patient demographics like age, gender, and ethnicity, medical information like BMI, and alcohol consumption, and dementia diagnosis. This dataset is considered as a helping hand for the main dataset as this dataset was used to train 2 different Machine Learning models called the Decision Tree and the Naive Bayes models. The CSV file contains the information of more than 2000 patients.

Description of Chosen Algorithms

The choice of Random Forest and Naive Bayes for analyzing the demographic data was guided by their balanced accuracy and interpretability. Random Forest was selected for its ability to handle diverse feature sets effectively while minimizing overfitting, making it well-suited for the structured data in this project. Naive Bayes, on the other hand, was chosen for its efficiency with smaller datasets and categorical variables, enabling rapid insights. Together, these models provided complementary strengths, allowing for a robust exploration of dementia predictors within the CSV dataset and enhancing the overall analytical framework.

MobileNetV2 is an advanced deep learning architecture specifically designed for computational efficiency and scalability in image classification tasks. Developed by Google Research, this architecture is part of the MobileNet family, which is optimized for mobile and edge devices. Its lightweight nature and ability to achieve high accuracy with fewer parameters make it a popular choice for resource-constrained environments while still maintaining strong performance on complex datasets, such as medical imaging data. For this project, MobileNetV2 was selected due to its balance between computational efficiency and predictive power, which is essential when working with large-scale image datasets like the dementia classification task.

At its core, MobileNetV2 introduces two key innovations: depthwise separable convolutions and inverted residuals with linear bottlenecks. Depthwise separable convolutions significantly reduce the number of computations required during training and inference by separating the spatial and channel-wise operations in the convolution process. Meanwhile, the inverted residual structure allows the network to pass features through narrow bottlenecks, enabling efficient information flow while minimizing resource usage. This design ensures that MobileNetV2 achieves a high degree of accuracy without the need for heavy computational resources, making it ideal for training on large medical datasets while remaining accessible to researchers with limited hardware.

Additionally, MobileNetV2 is equipped to handle transfer learning, which was leveraged in this project to fine-tune the model on the dementia image dataset. Transfer learning involves initializing the model with pre-trained weights from ImageNet, a large-scale image recognition database, and then adapting the model to the specific classification task. This approach significantly reduces training time and improves model performance, especially when the dataset size is limited. In the context of this project, transfer learning allowed the model to quickly learn to differentiate between dementia-related images (labeled as 1) and non-dementia-related images (labeled as 0), capitalizing on the features already learned from ImageNet.

The flexibility and robustness of MobileNetV2 make it a fitting choice for this project's goal of dementia detection through image analysis. By combining its computational efficiency with the power of transfer learning, MobileNetV2 was instrumental in achieving reliable and accurate classification results. Its performance in this project underscores its suitability for medical imaging tasks where precision and efficiency are paramount.

Simplified Review of the Approach and Process

1. Data Loading and Preprocessing:

The CSV data was imported from Kaggle and then loaded to the notebook using pandas, after loading the data we dropped the columns that were not relevant for the model like PatientID, DoctorInCharge, etc. We also used Label Encoding to encode the categorical columns of the dataset, some examples are gender and ethnicity, we also had to normalize the numeric features of the data using a MinMaxScaler. Finally we splitted the data into training and testing sets using stratified sampling.

For the images dataset we had to reorganize the images in the folders based on the condition of the patients, the images were labeled in folders as a “0” or “Non Demented” or as “1” for all the other categories of Dementia, this was to create a binary classification task. Next we re-sized the images and preprocessed them using an ImageDataGenerator, we also included data augmentation for the training set. Finally for the images, we had to split them into training, testing and validation sets to have an outcome as we expected.

2. Model Building and Training:

For the CSV dataset, we decided to train two different Machine Learning Models to see which one had a higher accuracy for our dataset. We decided that the best options were a Decision Tree and a Naive Bayes model since they explored different approaches to making decisions regarding the data since the Decision Tree model is rule-based, while Naive Bayes is probabilistic.

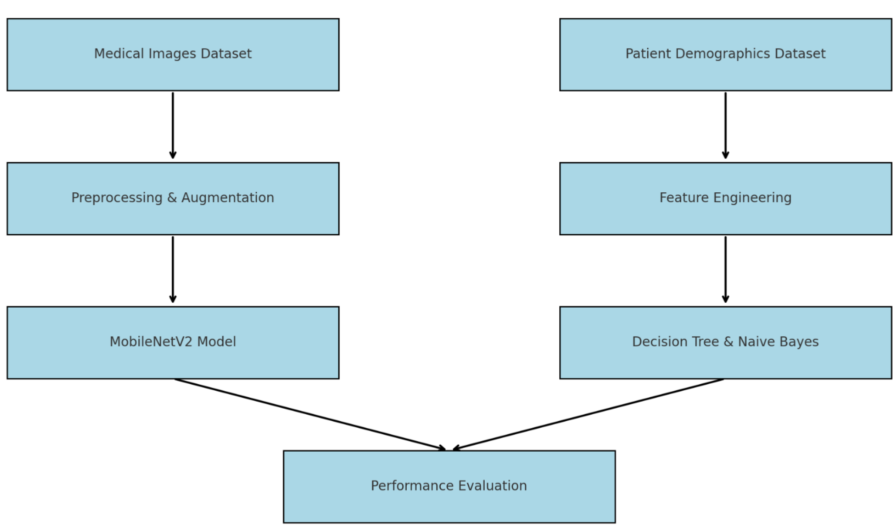
For the images dataset we decided to use the MobileNetV2 model as a base for feature extraction, we added some custom layer for the classification of the data, the modes is then trained using transfer learning, initially with a frozen base model and then we fine-tuned it with the base model unfrozen.

3. Model Evaluation:

To evaluate the models performance with the CSV data, we decided to use metrics like accuracy and classification reports for both models, we compared the performance of the Machine Learning Models and then we printed some example predictions to compare them.

For the images dataset, the model performance is evaluated using accuracy parameter, confusion matrix, ROC curve and a classification report, we also printed tome test image predictions to visualize the examples.

<u>High level diagram</u>



Review and Validation of the Output

Hypothesis 1: Using a Convolutional Neural Network (MobileNetV2), the classification model will be able to accurately distinguish between brain scans of individuals with and without dementia with a classification accuracy of at least 80%.

Hypothesis 2: The models trained on patient CSV data will correctly predict the diagnosis of dementia (Demented or Non-Demented) with an accuracy of at least 80%.

Hypothesis 3: The Decision Tree model will achieve a higher accuracy in predicting dementia diagnosis compared to the Naive Bayes model when trained and evaluated on the patient CSV data.

When evaluating the performance of the MobilNetV2 model, we get a test accuracy of 88.99% with a test loss of 0.2448, with these parameters we can say that the first hypothesis is confirmed since the original estimation was a classification accuracy of at least 80%.

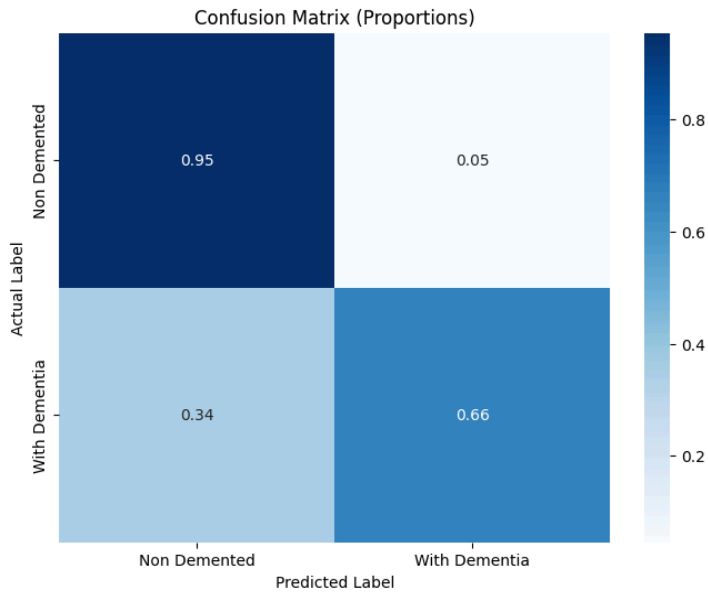
For the evaluation of both Machine Learning models, we got an accuracy of 90% for the Decision Tree model and of 77% for the Naive Bayes, disconfirming the second hypothesis as the Naive Bayes model did not achieve the desired accuracy during the testing.

Lastly, we can confirm the third hypothesis by comparing both Machine Learning models. As previously mentioned, the Decision Tree model achieves an accuracy of 90% which is significantly higher than the Naive Bayes 77% accuracy.

While evaluating the confusion matrix of the MobileNetV2 model, we can see that although the accuracy of the model can be considered high, if the dataset is big enough, having a 90% accuracy sometimes means that maybe thousands of patients can get misclassified. In the medical field, getting the correct classification everytime means being able to save lives, therefore, the importance of aiming for higher accuracy and reducing misclassifications as much as possible is emphasized. Even a seemingly small percentage of error can have significant consequences in real-world medical applications.

Conclusion

The results of this project highlight the potential of combining deep learning and traditional machine learning approaches for dementia detection. The MobileNetV2 model demonstrated strong performance, achieving a test accuracy of 88.99% and a weighted F1-score of 0.89. While it showed excellent precision and recall for non-demented cases, its ability to detect cases with dementia, as indicated by a lower recall of 0.69, suggests room for improvement in handling minority class data. The fine-tuning process further enhanced the model's predictive capabilities, achieving a perfect validation accuracy of 100% in some epochs, albeit with some variability in test set performance.



In parallel, the Decision Tree and Naive Bayes models, trained on demographic and medical data, provided complementary insights. The Decision Tree achieved an accuracy of 90%, outperforming the Naive Bayes model's 77%, with stronger precision, recall, and F1-scores. These traditional models confirmed the utility of structured patient data in identifying patterns relevant to dementia detection. Together, these findings underscore the effectiveness of leveraging diverse data types and models to gain a comprehensive understanding of dementia prediction tasks, albeit with distinct strengths and limitations for each approach.

References

Aithal, N. (2023, June 18). *Oasis Alzheimer's detection*. Kaggle. https://www.kaggle.com/datasets/ninadaithal/imagesoasis/data

GeeksforGeeks. (2024, September 16). *Decision Tree vs. naive Bayes classifier*. https://www.geeksforgeeks.org/decision-tree-vs-naive-bayes-classifier/#

Kharoua, R. E. (2024, June 11). *Alzheimer's disease dataset* . Kaggle. https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset

Liu, S., Masurkar, A.V., Rusinek, H. et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. Sci Rep 12, 17106 (2022). https://doi.org/10.1038/s41598-022-20674-x

Future Work

Given more time, resources, and expertise, several avenues for improvement and extension of this project are evident. First, transitioning from a multi-model approach to a multimodal framework could offer substantial benefits. By integrating image data and patient demographics into a single unified model, the architecture could potentially capture richer, cross-correlated features, enhancing its predictive power. This shift would require significant expertise in multimodal deep learning techniques and additional computational resources.

Another key area of improvement involves data acquisition. Increasing the diversity and size of both imaging and structured datasets would address class imbalance and improve the model's generalizability. Gathering higher-quality, balanced datasets could also mitigate the underperformance in detecting dementia cases. Furthermore, exploring a broader range of deep learning architectures, such as ResNet or EfficientNet, could uncover alternative models better suited to the dataset's characteristics.

Additionally, continued training and hyperparameter tuning for the MobileNetV2 model might yield improved results. Employing advanced techniques like learning rate scheduling, early stopping, or ensemble methods could further refine the model's accuracy and robustness. If time and expertise had permitted, upfront training on GPU utilization, cloud platforms like AWS or Google Cloud, and containerized frameworks like TensorFlow Serving could have streamlined model deployment and performance evaluation.

Lastly, enhancing our programming toolkit by leveraging more advanced Python libraries and functions—such as TensorFlow's tf.data API for efficient data pipelines or better visualization tools for interpretability—would have enriched the analysis. Integrating cutting-edge explainability tools like LIME or SHAP to analyze the feature importance of demographic and medical data could also enhance trust in the model's decisions. These enhancements, alongside comprehensive training in deep learning technology, could substantially elevate the project's impact and scalability.