Exploring the Use of Synthetic Data from AI Chatbots for Predicting Alzheimer's Disease: Methods for Validation and Barriers to Real-World Implementation

Siddhi Ananya

1. Introduction

1.1 Advancements of AI in Healthcare

Artificial intelligence has rapidly advanced in the last decade for applications in healthcare from diagnostic tools to predictive analytics. Machine learning, a subset of the wide range of AI tools undergoing implementation, uses its sophisticated abilities to recognize patterns in subsequent data sets to help enable the early detection of chronic illnesses like Alzheimer's disease.

1.2 Alzheimer's Disease and the Need for Early Diagnosis

Alzeihemer's, "a type of dementia that affects memory, thinking and behavior" (Alzheimer's Association, 2025), affects millions worldwide and remains a significant challenge in public health. Early diagnosis is vital to allowing opportunities for intervention in the early growth activity of the disease, which could slow disease progression and improve patient outcomes. However, a tenacious challenge persists in the reliance of real world data, when developing machine learning models, raising ethical concerns over privacy, security, and bias.

1.3 The Role of Synthetic Data and Chatbots

Synthetic data offers a potential solution by mimicking real-world medical datasets while avoiding privacy issues. Traditional approaches rely on statistical analytics and simulations to generate viable data sets, unlike this research, which will be proposing a new source of data generation: ChatGPT.

Chatbots such as ChatGPT utilize conversational AI techniques like natural language processing systems to understand user questions and simulate responses to them (IBM, 2025). This can provide the edge of the implementations of conversational, contextual, and behavioral aspects onto the data generated, making the model increasingly authentic. This study investigates the viable use of AI generated synthetic data in effectively training machine learning models to predict the future onset of Alzheimer's disease. It specifically seeks to answer if synthetic data from AI chatbots can produce accurate and reliable predictive models with further investigation into methods that can validate these models' effectiveness and barriers that might prevent their adoption in real-world healthcare settings.

1.4 Research Significance and Future Implications

The desire of advancing this research topic originates from a fascination with AI's potential in predictive analytics within healthcare. Since the public debut of ChatGPT, AI has been implemented into various domains with a wide breadth of exploration of its applications, including health care. However, chatbots are generally utilized in website management, its true potential remaining unexamined. This interest is personally further driven by an aspiration to merge computer science and healthcare, through observing the challenges and opportunities within the healthcare field through familial connections.

This research dives into uncharted territory, with the exploration of the intersection of natural language processing and predictive machine learning models, leveraging synthetic data generated by AI chatbots to train machine learning models for predicting the onset of Alzheimer's disease. This new approach seeks to plug the existing gaps in healthcare technology with emphasis on current ethical dilemmas on the usage of patient data. Exploring AI-generated synthetic data shows promise for a future in enhanced early diagnostic capabilities, and if successful, goes far beyond Alzheimer's. My research shows the possibilities of AI in healthcare developments, where it is further magnified for a maximized psychosocial

impact, creating a future for technology in which early intervention and better patient care becomes a priority.

2. Literature Review

2.1 Machine Learning in Healthcare

Machine learning is the process of identifying various characteristics and attributes to identify complex patterns within data, which can be applied to carry out deeper analysis and gain valuable insights (Kaul et al., 2020). This can be greatly useful in areas such as health care with its applicability to optimize diagnostic processes, with the potential use of data tailored to look very similar to real world data.

2.2 Definition and Importance of Synthetic Data

Specifically, synthetic data refers to artificially created datasets that retain statistical patterns similar to real data which can participate in model training without endangering patient privacy in this study. Other studies have similarly defined it with specifications of being created through algorithms, generative models, or simulations (Lui et al., 2024). However, this study will focus on its ability to train a machine learning model if it is generated from AI chatbots like ChatGPT.

2.3 AI Chatbots as Data Sources

AI chatbots can be defined in this study as conversational AI systems designed to simulate human interaction, which is how it is referred to in most publications, with a focus on the abilities of ChatGPT. The synthetic data produced from these chatbots will be used to produce responses based on medical data patterns, better known as predictive modeling. These responses are targeted to forecast health outcomes for chronic illnesses. One such illness that will be focused on is Alzheimer's disease, which is a disorder that impairs memory and function, with an inclusion of damage to neurons (Ding et al., 2024). Research generally agrees with synthetic data's potential in safely training models and addresses its potential in early disease detection, however there are limited sources addressing its feasibility in healthcare as well as the potential of synthetic data produced through chatbots rather than mathematical techniques.

2.4 Historical Development of AI in Healthcare

The application of Artificial Intelligence in healthcare dates back several decades, with the development of AI being very sparse throughout the early years of technology. To understand the development of artificial intelligence, however, it is important to understand how it can be defined as well as the differences between artificial thinking and natural thinking as described by Fetzer (1990). Artificial thinking is the more complex process, which is generally undergone by artifacts known as machines. Building on this idea is the application of machines in healthcare which is discussed by Kaul in his literature review on the development of AI in medicine with a focus on the major applications of AI in gastroenterology and endoscopy (2008). He discussed the concepts of the increase in diagnostic accuracy following integration of AI into clinical trials, as well as improved efficiency in provider workflow.

2.5 Existing Methodologies Using Synthetic Data

Several studies have employed the same methodology as the one I will be using, with a specific focus on using synthetic data to train machine learning models. In the 2024 study, Mills et al. explore the integration of machine learning techniques in predicting rheumatic disorders through models trained with datasets containing over 10,000 records, made through binary classification. With the same end goal of disease prediction, Garza-Frias et al. explored this methodology in their 2024 study, by using real radiographs rather that synthetic data, passing it

through an AI model that obtained information on factors such as cardiac silhouette with the goal of then being able to train it to predict which patients were developing heart failure.

2.6 Bias Considerations in AI-Generated Data and Addressing Gaps

As AI continues to grow in the field of healthcare, with a specification in disease prediction, there are growing concerns regarding biases and ethical implications, especially with the newfound application of authentic data. Hao et al discuss this topic in their paper bringing in points of legal constraints into the argument to support that of ethicalities. Because synthetic data is often very closely related to real world data to get accurate results, it can oftentimes mimic biases that are inherent in real-world data sources, which generally leads to gender or racial biases. Besides the implications around synthetic data, there are also many chatbots with concerns arising because of their extreme impact on the job market (Følstad et al., 2021). To target these issues many organizations such as Microsoft's FATE (Fairness, Accountability, Transparency, and Ethics in AI) have taken initiatives to mitigate these issues. Without being addressed, these issues could lead to models that are less accurate for certain populations, particularly for chronic illnesses like Alzheimer's, where the disease's manifestation can vary significantly across different socio-economic groups. These concerns highlight the need for careful consideration when developing AI models for healthcare, especially when synthetic data is utilized.

Substantial studies have focused on research on utilizing synthetic data to train models for early diagnosis health care practices, but the synthetic data is generated using statistical techniques rather than AI chatbots. This approach is greatly effective in creating synthetic data sets that mimic data patterns in real data sets, making the model more accurate and increasing its adaptability rates in real time clinical trials. However, this mathematical approach lacks the nuanced conversational and contextual data that AI chatbots can provide to the data sets, adding more variability. Because chatbots mimic human interaction, it can add layers of behavioral and contextual richness to the data set. Training using these data sets can give the model more variability and the ability to account for more situations than just those that occur frequently, which is what will be primarily targeted through statistically generated data sets. This idea is further supported by Følstad et al. 's literature review in 2021, in which knowledge based advancement in fields like management analytics, marketing, communication science, etc can be underscored to the implementation of AI chatbots.

In summary, the existing research on using synthetic data to train predictive models for chronic illnesses highlights challenges and positive breakthroughs. Despite the significant positive implications many studies have seen, the significant challenge of privacy concerns and limited data availability remains. In addition, challenges of a lack of comprehensiveness in synthetic data also prove issues in providing variability to predictive models. This review, thus, identifies the gaps around predictive model accuracy if trained on synthetic data and its applicability in real clinical settings. Addressing these gaps can help put a lot of these models into practice and increase efficiency, advancing their role in chronic illness prediction.

3. Methodology

3.1 Research Problem and Data Collection

Beginning the project after analyzing the research problem was challenging, particularly due to the broad scope of the study. To address this, a mixed-methods approach was implemented, starting with an analysis of datasets from similar studies and refining the data to enhance model accuracy. Data collection was a crucial first step, requiring an examination of real-world data and comparable research to define the expected characteristics of the synthetic

dataset. To ensure reliability, extensive research was conducted to evaluate the credibility of institutions providing relevant datasets. After reviewing multiple sources, the University of Southern California's Image and Data Archive was identified as a trusted resource. A specific dataset from this archive was selected as the foundation for the testing dataset, ensuring alignment with the study's objectives. This dataset included various distribution methods, depicting data on patient visits and age. While patient visits were less relevant to this research, as they indirectly affect the development of Alzheimer's, the age-related data played a crucial role in model refinement.

This graph (as shown below) depicts the most common age of active development centering around the range of 70-79, enhancing the correlation between old age and the disease. There are also various research groups present in the study, so that they can be compared against each other, so that causation can be implied. The primary to be noted include EMCI, LMCI, MCI, and CN. EMCI (Early Mild Cognitive Impairment) is an earlier stage of MCI with milder symptoms. Late Mild Cognitive Impairment (LMCI) is a more progressed stage of MCI, with more pronounced cognitive decline but not yet reaching full dementia. EMCI (326) and LMCI (182) combined make up a significant portion of MCI (mild cognitive impairment) cases, reflecting a progression model for cognitive decline. CN (cognitively normal) is the group of healthy individuals without any diagnosed cognitive impairment, likely used as a control group.

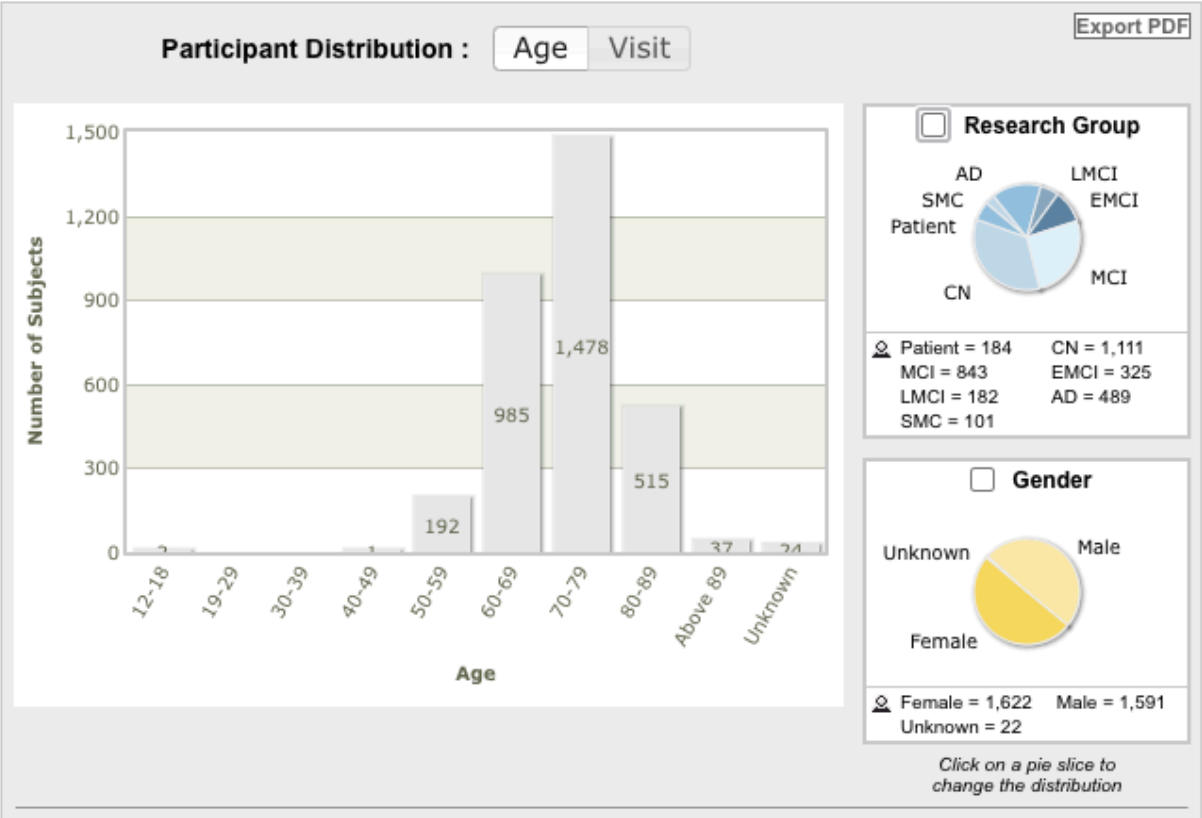

Figure 1: Participant distribution by age and gender in the study. Data is categorized into research groups, including AD (Alzheimer's Disease), MCI (Mild Cognitive Impairment), and CN

(Cognitively Normal). Source: Alzheimer's Disease Neuroimaging Initiative, USC Stevens Neuroimaging and Informatics Institute.

3.2 Synthetic Data Generation

The next step was the formation of the synthetic data, a key aspect that set my research apart. Unlike traditional methods, the synthetic data generation was conducted using ChatGPT rather than relying on standard statistical modeling or pre-existing synthetic datasets. This approach allowed for the creation of diverse and dynamic data points that closely mirrored real-world patterns. By fine-tuning the prompts and iterating through multiple generations, the dataset was refined to ensure it aligned with the characteristics observed in real patient data while minimizing biases and inconsistencies. Characteristics specifically focused on included adjusting the age distribution to match the patterns of the data set observed, to ensure normality and make the data more realistic. Participants were increased in the 70-79 (800 cases) and 80-89 (500 cases) age ranges since they were the most represented and those under 60 were reduced as they were rare in the original graph. The gender balance was also adjusted as the original generated data weighed females over males greatly with at least a 3:1 ratio. Whereas now there are nearly equal numbers of males and females, preventing gender bias, with an exponential increase in the number of participants, with a total of 1600 subjects.

The distribution of diagnoses has also been adjusted to better align with observed trends in Alzheimer's progression (as depicted below). The number of mild cognitive impairment (MCI) cases, including both early and late MCI, was increased to reflect real-world patterns, while Cognitively normal (CN) individuals still make up the largest group.

MMSE scores now better correspond to diagnosis, with lower scores (10-21) for AD, mid-range scores (22-27) for MCI, and higher scores (24-30) for CN individuals. Hippocampal volume was modified accordingly, with AD patients having the lowest volume (2.5-4.5 mL), MCI cases showing moderate volume (4.5-5.5 mL), and CN individuals exhibiting the highest volume (5.5-6.5 mL). These refinements ensure that the dataset more accurately reflects real-world patterns in Alzheimer's progression and risk factors.

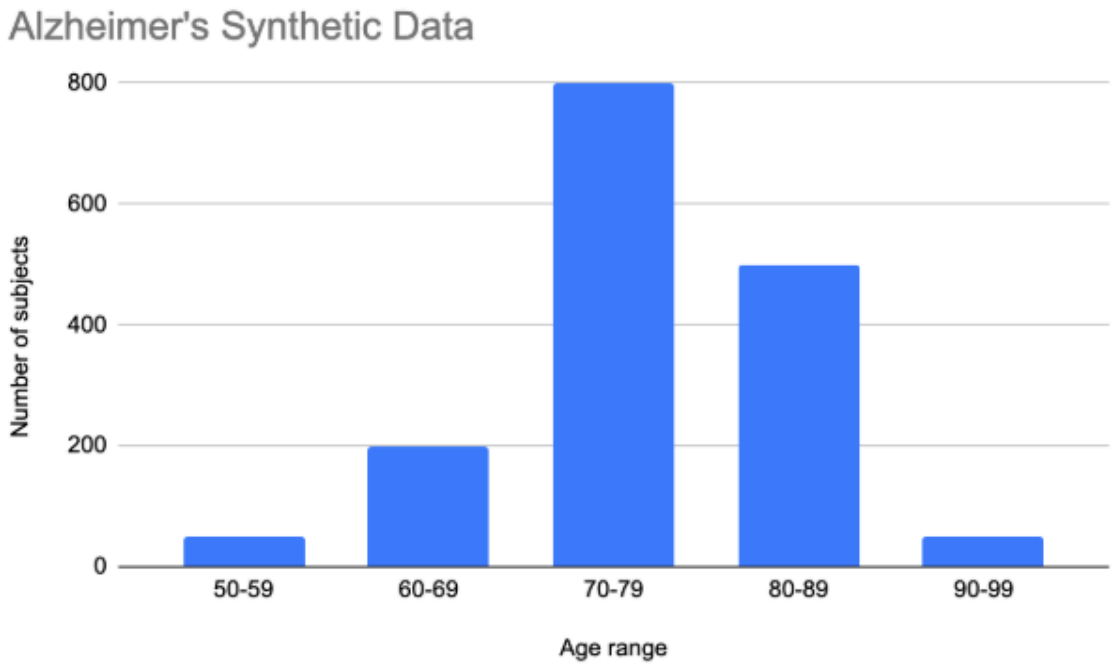

Figure 2: Distribution of synthetic data generated for Alzheimer's study. The graph represents the number of subjects across different age ranges, with the highest concentration in the 70–79 age group. This synthetic data was created to model participant demographics in Alzheimer's research.

3.3 Final Dataset variables

The final data set used in this study contains patient information compromising the following features as previously mentioned. To provide a clear understanding of its variables and the values in the data set, the following table outlines the key attributes recorded for each participant. These variables capture essential demographic, cognitive, and genetic factors relevant to Alzheimer's research:

Variable	Description	Type
Patient_ID	Unique identifier for each patient	Qualitative
Age	Patient's age in years	Quantitative
Gender	Patient's biological sex (Male/Female)	Qualitative
MMSE_Score	Mini-Mental State Examination score (0-30)	Quantitative
Hippocampal_Volume	Brain hippocampus volume measured via MRI (mm^3)	Quantitative
Family_History	Whether the patient has a family history of Alzheimer's (Yes/No)	Qualitative
APOE4_Present	Presence of the APOE4 gene variant (Yes/No)	Qualitative
Alzheimers_Diagnosis	Diagnosis outcome (Positive/Negative)	Qualitative

3.4 Machine Learning Approach

The machine learning approach utilized for predictive analysis in this study is Supervised Learning, a category of machine learning where the model can learn from labeled training refined data to make predictions on unseen data, enhancing flexibility. In this study's context, the labeled outputs, such as that of the Diagnosis Outcome (yes/no), can allow the model to recognize patterns and correlations among patient characteristics. Utilizing this method, the model can determine what variables have the most effect towards the development of Alzheimer's for the individuals in the data set provided.

To implement this approach, the Random Forest Classifier machine learning algorithm was selected due to its effectiveness in handling nonlinearity, reduction of overfitting, as well as the improvement of generalization to a broader population outside this study. As seen by the

data set presented and its many variables, Alzheimer's has many stages and interdependent factors that influence its development which can be captured accurately by Random Forest. When making machine learning models, single decision trees can be prone to overfitting, which means they perform well on the training data set but can find issues adapting to new data. Decision trees in machine learning models are a type of supervised algorithm that splits data into multiple branches based on specific conditions, making predictions at each step until it reaches a final outcome. Each split point is called a node and represents a decision based on a feature. The tree henceforth continues branching until it reaches terminal nodes, or leaves, which hold the final classification. Random Forest is beneficial in this sense because it has the ability to mitigate potential overfitting by averaging predictions across multiple decision trees before coming to a final decision, enhancing model stability and reliability.

3.5 Model Implementation

3.5.1 Data Preprocessing

To begin implementing the Random Forest Classifier, the dataset was first preprocessed to ensure all variables were formatted correctly for the model. This included handling categorical variables and normalizing numerical features where necessary. The dataset was loaded onto Google colab using the pandas library and the feature variables (X) and target variable (y) were defined. Some variables, however, remained qualitative, which was converted into numerical values using one-hot coding.

3.5.2 One-hot coding and splitting the data

One-hot coding is a method of encoding categorical data into numerical values by assigning it the values 1 or 0. Typically the value of 1 is assigned if a specific category is present and 0 otherwise. The dataset was then split into 70% training data and 30% testing data to evaluate the model's performance on unseen data.

3.6 Random Forest Model Configuration

In beginning training for the model, the random forest classifier was initialized with the following hyperparameters:

Hyperparameter	Description	Value
N_decisiontrees	Number of decision trees	100
Random_state	Ensures reproducibility	42
Max_depth	No restriction on tree depth	None
split_limit	Minimum samples that are needed to split a node	2
minimum_leaf_samples	Minimum samples per leaf node	1
Bootstrap	Random subsets of data for training	True

The initialized model was trained on the training data set (X_train, y_train), 70% of the data set shown, and further tested on the unseen data set (X_test), which was the leftover 30% of the original data set.

3.7 Model Evaluation

3.7.1 Performance metrics

Once the random classifier model was trained, its performance was accessed using performance metrics. The primary metrics used in this study included precision, recall, F1-score, accuracy and a confusion matrix. These metrics provided insight into how well the model classified Alzheimer's diagnoses and how effectively it generalized to unseen data. Accuracy in this study is measured to be:

number of correctly classified cases/all the cases in the input into the model. A crucial area of Alzheimer's detection is classification of the number of cognitively normal (CN) cases versus fully developed Alzheimer's Disease (AD) cases, with the most significant information being how much more prevalent CN cases are than Alzheimer's cases. This is why precision, recall, and the F1-score had to also be calculated and evaluated with accuracy.

Precision measured the percentage of correctly predicted Alzheimer's cases out of all predicted Alzheimer's cases. A high precision value indicates that the model minimizes false positives, which is crucial in preventing misdiagnosis. A low precision value therefore indicates possible presence of false positives and leads into potential cases of encouraged misdiagnosis. Recall assessed how well the model identified actual Alzheimer's cases, ensuring that true positives were captured effectively. F1-score is the mean of precision and recall, and balances them both to ensure that the model had high success rates, whilst also maintaining generalizability.

3.7.2 Confusion Matrix

The confusion matrix depicts how often the model accurately or inaccurately predicts a scenario, in this paper's case, how often the model accurately predicts the presence of Alzheimer's compared to the training data it was trained on. In this study the confusion matrix was generated to examine the classification results in more detail. It displayed the number of true positives (correctly classified Alzheimer's cases), true negatives (correctly classified non-Alzheimer's cases), false positives (healthy individuals incorrectly classified as having Alzheimer's), and false negatives (Alzheimer's patients incorrectly classified as healthy subjects). Analyzing this matrix helped determine if the model had any bias toward a particular class and whether adjustments were necessary.

3.7.3 Feature Importance Analysis

After the random forest model was completely trained, feature importance values were calculated and extracted to provide an understanding of what features contributed most to the model's predictions. This also plays a part in determining whether the model tailored itself to look at a specific feature over others based on high feature importance values from the training data. Because this model is disease prediction based, it was predicted that age would hold one of the largest scores, as it is the strongest known predictor of Alzheimer's onset. Another high importance value predicted before running the model was that of Cognitive Test Scores, for example MMSE (Mini-Mental State Examination) or MoCA (Montreal Cognitive Assessment), as they can be key indicators of cognitive decline. These feature importance values provided critical insights into the biological and clinical factors driving model predictions.

3.8 User input

To allow real time functionality and usability, a predict_alzheimer's function was built in which all the user is prompted to enter all the key clinical data included in the training data set. The user's input is then processed and prepared to match the format used in training the model. Categorical variables such as gender are converted into binary values and any missing features are handled by setting default values, ensuring the input data aligns with the model's expected input.

Once the data has been processed, it uses the trained Random forest model to make a prediction, outputting whether the individual is likely to have Alzheimer's disease.

4. Discussion/Analysis:

4.1 Data Analysis

The model's validity was evaluated using the metrics of accuracy, classification report, confusion matrix, and feature importance as previously mentioned.

4.1.1 Accuracy

The model received an accuracy score of 1.00, indicating that there is 100% accuracy and that it correctly diagnosed each patient in the data set based on the designated variable values. While this perfect accuracy is an impressive result, it suggests that the model may be overfitting, meaning it could perform well on the current dataset but may not generalize to unseen data. Overfitting generally occurs when the model learns beyond just the patterns of the training data set, but also its noise and details specific to only that data set. In the case of this study, this could mean that the dataset is not diverse enough or is too small. This also does indicate that there may be biases present in the training data set. Bias could have been introduced if the data set wasn't representative of the broader population, making it hard to generalize the results.

This high accuracy rate could also be because of the simplicity of the data set. As the MMSE Score and Hippocampal Volume are the values most closely related to Alzheimer's disease prediction, they might dominate the prediction, leading the model to primarily look at those values when performing future analyses.

4.1.2 Classification Report

The classification report provides details about the model's performance through descriptions of precision, recall, and f1-score for both classes (0 = no Alzheimer's, 1 = Alzheimer's). Both classes have values of 1.00 for each statistical output, indicating that there were no false positives in precision, recall, or f1-score, which averages precision and recall.

4.1.3 Confusion Matrix

The confusion matrix further depicts the model accuracy and its perfect performance, once again encouraging signs of overfitting. It indicates there are 214 true negatives (TN) and 266 true positives (TP). There are no false positives (FP) or false negatives (FN). The high true negatives and true positive values compared to no false positives or negatives introduce the possibility of the model having memorized the data and not learned generalizable patterns, which is a type of overfitting and would mean there were would be high performance rates, but low accuracy rates when introduced to new data sets.

4.1.4 Feature Importances

The feature importance values provide insight on what variables held the most effect on the model's outputs and predictions. This model's feature importance values are represented below:

Feature	Feature Importance value
MMSE_Score	0.500641
Hippocampal_Volume	0.437367
Patient_ID	0.025364
Age	0.022071
APOE4_Present	0.013380
Family_History	0.000626
Gender	0.000552

MMSE Score and Hippocampal Volume have the highest scores indicating the most impact towards the model predictions. This suggests that cognitive decline, as measured by the MMSE score, and hippocampal volume, which is known to be affected by Alzheimer's, were key indicators in predicting Alzheimer's diagnosis. Other variables contain lower scores such as age and gender, indicating that they also had a contribution, but not nearly as much as the variables with higher scores.

The Patient_ID feature is a unique identification number differentiating patents from one another, and should hold no effect on model performance. However, it has a higher score than age, which is likely an anomaly, indicating further issues with the training data set. This issue likely arised, because when preprocessing the data, the Patient_ID variable was set as a feature, letting the model memorize the patterns and create correlations between them and diagnoses.

4.2 Future Scope

4.2.1 Evaluation Metrics for future testing

A Receiver Operating Characteristic (ROC) curve relative to this study would be a mathematical technique to assess how well your model discriminates between the two classes: Alzheimer's diagnosis (positive) and no Alzheimer's diagnosis (negative). The area under the curve (AUC) value is a quantitative measure of distinction. If the ROC curve shows that the model has a very high AUC (close to 1), it means the model is doing an excellent job at distinguishing between the two classes at various thresholds, and future testers can undergo model refinement until the AUC is a higher value than before. This model would likely have a lower AUC value as there are many signs of overfitting.

K Fold cross validation techniques could also be employed. Cross validation is used to assess how well your model generalizes to unseen data, especially beneficial in cases like this study with such high accuracy scores. To incorporate this technique, the data would likely be split into several subsets for which the model would be trained on different combinations of. If the performance is still high, the data set likely lacks variability. However, if the accuracy rate decreases, there may still be overfitting in the data set.

4.2.2 Future of Alzheimer’s and machine learning disease prediction

Advancements in machine learning indicate a strong presence of it in predictive diagnosing in the future of Alzheimer's. Machine learning models could identify individuals at higher risk for developing Alzheimer's, enabling early intervention and more personalized treatment options. This integration of multi-modal data could revolutionize how we diagnose and manage the disease, with the potential for tailored treatments and better patient outcomes.

Along with early detection, machine learning allows for acceleration of drug discovery and clinical trials, encouraging efficiency and precision However, for these advancements to be effective, it's essential to ensure that AI models are built with fairness, transparency, and ethical considerations in mind, ensuring that technology enhances human decision-making rather than replacing it.

5. Conclusion

The use of machine learning in Alzheimer's disease prediction holds great promise for the future of early diagnosis. Through applications of machine learning models, such as the random forest classifier, and evaluation metrics, diagnostic accuracy can undergo great improvement and encourage early identification of individuals at risk.

However, challenges like overfitting, biases in training data, and the need for robust evaluation metrics remain, highlighting the importance of refining models and ensuring they generalize well. With continued advancements and ethical considerations, AI has the potential to transform Alzheimer's care, from early detection to personalized interventions and improved patient outcomes.

References

Abidi, Y. (2024, April 23). The 5 Best Open-Source AI Image Generators. MUO; MakeUseOf. https://www.makeuseof.com/best-open-source-ai-image-generators/

Adel, S. M., Bichu, Y. M., Pandian, S. M., Sabouni, W., Shah, C., & Vaiid, N. (2024). Clinical audit of an artificial intelligence (AI) empowered smile simulation system: a prospective clinical trial. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-69314-6

Al-Antari, M. A. (2023). Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology! Diagnostics, 13(4), 688. PubMed Central. https://doi.org/10.3390/diagnostics13040688

AI Could Help Predict Alzheimer's Disease Early Using Language. (2020). Psychology Today. https://www.psychologytoday.com/us/blog/the-future-brain/202010/ai-could-help-predict-alzheimer-s-disease-early-using-language

Alice, M., Felipe, Fernando, Madeiro, F., & Lima, J. B. (2024). Machine Learning and Graph Signal Processing Applied to Healthcare: A Review. Bioengineering, 11(7), 671–671. https://doi.org/10.3390/bioengineering11070671

Applying artificial intelligence for early risk forecasting of Alzheimer's disease. (n.d.). ScienceDaily. https://www.sciencedaily.com/releases/2023/06/230607124033.htm

Are AI and Talking Cars the Future of Driving? (2024). Psychology Today. https://www.psychologytoday.com/au/blog/the-future-brain/202410/are-ai-and-talking-cars-the-future-of-driving

Artificial intelligence outperforms clinical tests at predicting progress of Alzheimer's disease. (2024). ScienceDaily. https://www.sciencedaily.com/releases/2024/07/240713121220.htm

Best Practices and Lessons Learned on Synthetic Data. (2022). Arxiv.org. https://arxiv.org/html/2404.07503

Davenport, T., & Kalakota, R. (2019). The Potential for Artificial Intelligence in Healthcare. Future Healthcare Journal, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

D'Hondt, E., Ashby, T. J., Chakroun, I., Koninckx, T., & Wuyts, R. (2022). Identifying and evaluating barriers for the implementation of machine learning in the intensive care unit. Communications Medicine, 2(1). https://doi.org/10.1038/s43856-022-00225-1

Ding, K., Chetty, M., Noori Hoshyar, A., Bhattacharya, T., & Klein, B. (2024). Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. Artificial Intelligence Review, 57(12). https://doi.org/10.1007/s10462-024-10961-6

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making, 19(1). https://doi.org/10.1186/s12911-019-0918-5

Emiliano Garza-Frias, Kaviani, P., Karout, L., Roshan Fahimi, Hosseini, S., Preetham Putha, Manoj Tadepalli, Kiran, S., Arora, C., Robert, D., Bizzo, B., Dreyer, K. J., Kalra, M. K., & Digumarthy, S. R. (2024). Early Detection of Heart Failure with Autonomous AI-Based Model Using Chest Radiographs: A Multicenter Study. Diagnostics, 14(15), 1635–1635. https://doi.org/10.3390/diagnostics14151635

Fetzer, J. H. (1990). What is Artificial Intelligence? Artificial Intelligence: Its Scope and Limits, 4(1), 3–27. https://doi.org/10.1007/978-94-009-1900-6_1

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary

research agenda. Computing, 103(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical research and machine learning in health care. The Lancet Digital Health, 2(9), e489–e492. https://doi.org/10.1016/s2589-7500(20)30186-2

Google Health. (n.d.). Google Health. Health.google. https://health.google/health-research/imaging-and-diagnostics/

Grueso, S., & Viejo-Sobera, R. (2021). Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. Alzheimer's Research & Therapy, 13(1). https://doi.org/10.1186/s13195-021-00900-w

Gugerty, L. (2006). Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 880–884. https://doi.org/10.1177/154193120605000904

Guzmán-Quezada, E., Mancilla-Jiménez, C., Rosas-Agraz, F., Romo-Vázquez, R., & Vélez-Pérez, H. (2024). Embedded Machine Learning System for Muscle Patterns Detection in a Patient with Shoulder Disarticulation. Sensors, 24(11), 3264. https://doi.org/10.3390/s24113264

Hao, S., Han, W., Jiang, T., Li, Y., Wu, H., Zhong, C., Zhou, Z., & Tang, H. (2024). Synthetic Data in AI: Challenges, Applications, and Ethical Implications. https://arxiv.org/pdf/2401.01629

Irving, S. J., Kocksch, L., & Munk, A. K. (2024). Synthetic Interlocutors. Experiments with Generative AI to Prolong Ethnographic Encounters. ArXiv.org. https://arxiv.org/abs/2410.11395

Jahan, S., Kazi Abu Taher, M. Shamim Kaiser, Mahmud, M., Md. Sazzadur Rahman, A. S. M. Sanwar Hosen, & Ra, I.-H. (2023). Explainable AI-based Alzheimer's prediction and management using multimodal data. PloS One, 18(11), e0294253–e0294253. https://doi.org/10.1371/journal.pone.0294253

Jog, C. (2024, October 12). The "strawberry" problem: How to overcome AI's limitations. VentureBeat. https://venturebeat.com/ai/the-strawberrry-problem-how-to-overcome-ais-limitations/

Kalota, F. (2024). A Primer on Generative Artificial Intelligence. Education Sciences, 14(2), 172. https://doi.org/10.3390/educsci14020172

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. Gastrointestinal Endoscopy, 92(4), 807–812. https://doi.org/10.1016/j.gie.2020.06.040

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key Challenges for Delivering Clinical Impact with Artificial Intelligence. BMC Medicine, 17(1). BMC. https://doi.org/10.1186/s12916-019-1426-2

Kuzucu, S., Cheong, J., Gunes, H., & Kalkan, S. (2024). Uncertainty as a Fairness Measure. Journal of Artificial Intelligence Research, 81, 307–335. https://doi.org/10.1613/jair.1.16041

Lawrence Livermore National Laboratory. (2024). The Birth of Artificial Intelligence (AI) Research | Science and Technology. St.llnl.gov. https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research

Malik, P., Pathania, M., & Rathaur, V. (2019). Overview of artificial intelligence in medicine. Journal of Family Medicine and Primary Care, 8(7), 2328–2331. https://doi.org/10.4103/jfmpc.jfmpc_440_19

Manjur Kolhar, Nazir, R., Mohapatra, H., & Al, A. M. (2024). AI-Driven Real-Time Classification of ECG Signals for Cardiac Monitoring Using i-AlexNet Architecture. Diagnostics, 14(13), 1344–1344. https://doi.org/10.3390/diagnostics14131344

Mearian, L. (2024, February 7). What are LLMs, and how are they used in generative AI? Computerworld. https://www.computerworld.com/article/1627101/what-are-large-language-models-and-how-are-they-used-in-generative-ai.html

Mills, G. A., Dey, D., Kassim, M., Yiwere, A., & Broni, K. (2024). Diagnostic Tool for Early Detection of Rheumatic Disorders Using Machine Learning Algorithm and Predictive Models. BioMedInformatics, 4(2), 1174–1201. https://doi.org/10.3390/biomedinformatics4020065

Moghaddam, M. T., Yones Jahani, Zahra Arefzadeh, Dehghan, A., Mohsen Khaleghi, Sharafi, M., & Ghasem Nikfar. (2024). Predicting diabetes in adults: identifying important features in unbalanced data over a 5-year cohort study using machine learning algorithms. BMC Medical Research Methodology, 24(1). https://doi.org/10.1186/s12874-024-02341-z

New AI Program Could Predict Likelihood of Alzheimer's | College of Engineering. (2025). Bu.edu. https://www.bu.edu/eng/2024/06/25/new-ai-program-could-predict-likelihood-of-alzheimers/

Nield, D. (2024, August). New AI Tool Predicts Alzheimer's With Higher Accuracy Than Clinical Tests. ScienceAlert. https://www.sciencealert.com/new-ai-tool-predicts-alzheimers-with-higher-accuracy-than-clinical-tests

Nuñez, M. (2024, October). Nvidia just dropped a bombshell: Its new AI model is open, massive, and ready to rival GPT-4. VentureBeat. https://venturebeat.com/ai/nvidia-just-dropped-a-bombshell-its-new-ai-model-is-open-massive-and-ready-to-rival-gpt-4/https://venturebeat.com/ai/nvidia-just-dropped-a-bombshell-its-new-ai-model-is-open-massive-and-ready-to-rival-gpt-4/

Orland, K. (2024, October 15). Apple Engineers Show How Flimsy AI "Reasoning" Can Be. WIRED. https://www.wired.com/story/apple-ai-llm-reasoning-research/

Osorio, P., Jimenez-Perez, G., Montalt-Tordera, J., Hooge, J., Guillem Duran-Ballester, Singh, S., Moritz Radbruch, Bach, U., Schroeder, S., Siudak, K., Vienenkoetter, J., Lawrenz, B., & Mohammadi, S. (2024). Latent Diffusion Models with Image-Derived Annotations for Enhanced AI-Assisted Cancer Diagnosis in Histopathology. Diagnostics, 14(13), 1442–1442. https://doi.org/10.3390/diagnostics14131442

Riem, L., DuCharme, O., Cousins, M., Feng, X., Kenney, A., Morris, J., Tapscott, S. J., Tawil, R., Statland, J., Shaw, D., Wang, L., Walker, M., Lewis, L., Jacobs, M. A., Leung, D. G., Friedman, S. D., & Blemker, S. S. (2024). AI driven analysis of MRI to measure health and disease progression in FSHD. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-65802-x

Rowe, T. W., Katzourou, I. K., Stevenson-Hoare, J. O., Bracher-Smith, M. R., Ivanov, D. K., & Escott-Price, V. (2021). Machine learning for the life-time risk prediction of Alzheimer's disease: a systematic review. Brain Communications, 3(4). https://doi.org/10.1093/braincomms/fcab24

Sandeep Singh Sengar, Affan Bin Hasan, Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-20016-1

Sharma, A. (2023, September 25). 15 Must-Try AI Social Media Content Creation Tools to Save Time. Buffer Resources. https://buffer.com/resources/ai-social-media-content-creation/

Siddiqui, I. A., Littlefield, N., Carlson, L. A., Gong, M., Chhabra, A., Menezes, Z., Mastorakos, G. M., Sakshi Mehul Thakar, Mehrnaz Abedian, Lohse, I., Weiss, K. R., Plate, J. F., Moradi, H., Amirian, S., & Tafti, A. P. (2024). Fair AI-powered orthopedic image segmentation: addressing bias and promoting equitable healthcare. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-66873-6

Stanford University. (2021). SQ2. What are the most important advances in AI? | One Hundred Year Study on Artificial Intelligence (AI100). Ai100.Stanford.edu. https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-1/sq2

tina. (2024, January 18). Guide to Open Source Large Language Models: Complete LLM Resource 2024. HYPEStudio - AI Automations, API Integrations, WordPress Website Development. https://hypestudio.org/blog/guide-to-open-source-large-language-models/

Zisis, K., Pavi, E., Geitona, M., & Athanasakis, K. (2024). Real-world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process. Journal of Pharmacy & Pharmaceutical Sciences, 27, 12302. https://doi.org/10.3389/jpps.2024.12302