Edge Hill University

Machine Learning Approach to Dementia Prediction

Student Name: **Chinonso Uche**

Student ID: **25830651**

Supervisor: **Prof. Ella Pereira**

5 July 2024

'This Report is submitted in partial fulfilment of the requirements for the MSc Data Science and Artificial Intelligence Degree at Edge Hill University.'

ACKNOWLEDGMENTS (Optional)

KEYWORDS

TABLE OF CONTENT

1 INTRODUCTION

Dementia is a broad term that describes a group of symptoms caused by changes in brain function that affect memory, thinking, and behaviour (Arvanitakis & Bennett, 2019). It is not a normal part of aging, although it is more common among older adults. Various underlying diseases or conditions can cause dementia, such as Alzheimer's disease, vascular dementia, Lewy body dementia, and frontotemporal dementia (Quinn et al., 2021). Alzheimer's disease, the most common type of dementia, accounts for 60-70% of cases (World Health Organization, 2023). It is a progressive neurological disorder characterized by the deposition of amyloid plaques, neurofibrillary tangles, and synaptic loss, leading to cognitive decline and memory impairment (Sheppard & Coleman, 2020).

The complex nature of Alzheimer's disease and other forms of dementia presents significant challenges in early detection and accurate diagnosis. Recent advancements in artificial intelligence, particularly in the field of machine learning, have shown promising results in addressing these challenges (Li et al., 2021). Machine learning techniques can analyse vast amounts of diverse data types, including neuroimaging, genetic information, cognitive test scores, and clinical records, to identify subtle patterns and biomarkers associated with dementia (Zhu et al., 2020). These sophisticated approaches enable the development of models for early detection, diagnosis, and prognosis of dementia, particularly Alzheimer's disease. Such models have shown promise in identifying individuals at risk of developing Alzheimer's before symptoms become apparent, potentially improving early detection, and enhancing diagnostic accuracy. Furthermore, these techniques can predict disease progression, offering the potential for more timely interventions, better management of the condition, and the development of personalized treatment strategies. By leveraging machine learning in this way, researchers and clinicians aim to significantly improve patient outcomes through earlier and more targeted interventions in the field of dementia care.

An example of how advanced machine learning techniques have improved Alzheimer's prediction can be seen in the work of Park et al. (2020). They developed a machine learning model using large-scale administrative health data to predict the incidence of Alzheimer's disease. Their model, which incorporated various health-related features and utilized advanced algorithms, demonstrated high accuracy in predicting Alzheimer's disease up to five years before clinical diagnosis. This approach showcases how machine learning can leverage complex, multi-dimensional data to provide early warnings of Alzheimer's disease, potentially allowing for earlier intervention and better patient outcomes. Given the promising applications of machine learning in dementia research, it's important to understand what machine learning is and how it's being specifically applied to Alzheimer's detection and prediction. Machine learning is a subfield of artificial intelligence that involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed (Mitchell, 1997). A machine learning model is a mathematical representation of a system or process that is trained on data to make predictions or classifications. Machine learning models have been widely used in various applications, including image and speech recognition, natural language processing, and predictive analytics. In the context of dementia prediction, machine learning models can be trained on large datasets of clinical, imaging, and biomarker features to identify patterns and predict the likelihood of developing dementia (Li et al., 2021). Figure 1 below shows a flowchart describing the process of Machine Learning Operation. The goal of dementia prediction is to identify individuals at high risk of developing dementia, allowing for early intervention, and potentially delaying or preventing disease progression.

Commented [EP1]: find a better connection between these two paragraphs. introduce the concept of using ML approaches to Alzheimer's detection, how advanced in ML can help with prediction and diagnosis then move on the ML

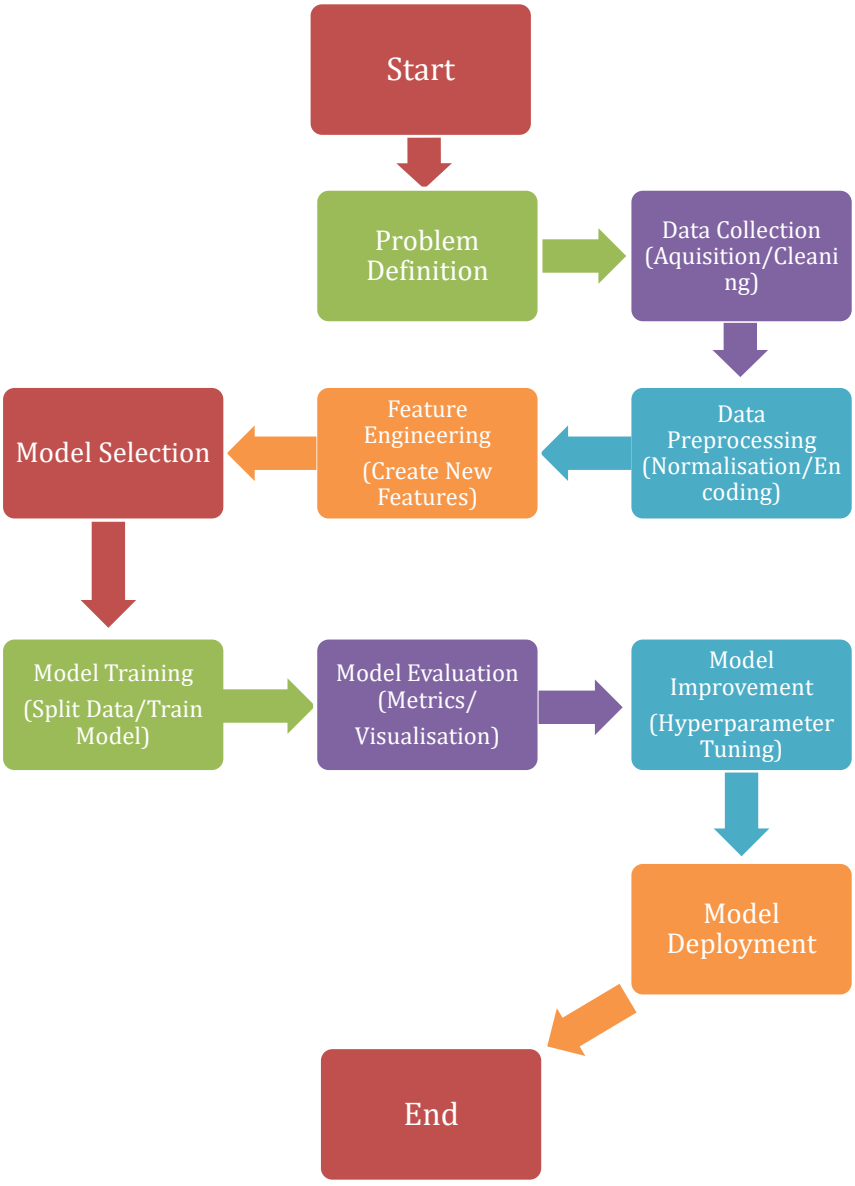

Figure 1: Flowchart of Machine Learning workflow

A recent systematic review by Javeed et al. (2023) highlighted the potential of machine learning for dementia prediction, but also identified several limitations and future

research directions. The review examined various studies that utilized different types of data for dementia prediction, including:

i. Neuroimaging data, (such as MRI and PET scans).
ii. Clinical-Variable Data which consists of medical tests and patient information such as age, sex, and cognitive assessments.
iii. Voice Data which involves speech analysis to detect neurodegenerative disorders affecting language processing.

While these diverse data types have shown promise in machine learning models for dementia prediction, the review emphasized several limitations:

1. Single Data Modality Focus: Previous systematic literature reviews (SLRs) focused on a single type of data modality for dementia detection, limiting the comprehensiveness of the evaluation.
2. Data Quality Issues: Poor quality of data and imbalance in dataset classes can lead to biased results from machine learning (ML) models.
3. Model Selection: Inappropriate selection of ML models and the complexity of training models can affect the performance of automated diagnostic systems.
4. Supervised Learning Constraints: Supervised ML techniques have inherent limitations, which can impact the effectiveness of automated diagnostic methods for dementia prediction.

The review emphasized the need for multimodal approach that combine image, clinical and voice data, larger datasets, and voice data improvement. This project therefore aims to develop a machine learning model for predicting Alzheimer's disease using a larger and more diverse dataset. The objectives are:

- To review existing literature on dementia, Alzheimer's disease, and machine learning models for dementia prediction.
- To combine multiple datasets, such as OASIS and ADNI dataset to create a larger dataset for model development.
- To develop a machine learning model for predicting Alzheimer's disease.
- To evaluate the performance of the developed model and compare it with existing models.

Commented [EP2]: add something on what type of data has been used so far than highlight the limitations

Commented [EP3]: be more specific

2 LITERATURE REVIEW

2.1 Background and Significance of Demetia

Dementia is a complex and multifactorial condition that affects millions of people worldwide, causing cognitive decline, memory loss, and changes in behaviour and mood (Dr Raina Loh, 2023). It is a significant public health concern, with an estimated 55 million people worldwide living with the condition, and this number is expected to increase to 78 million by 2030 and 139 million by 2050. The economic burden of dementia is significant, with estimated annual costs of over $1.3 trillion US Dollars in 2019 and may rise to $2.8 trillion by 2030 (Shin, 2022).

Commented [EP4]: organise your lit review into themes, sections clearly numbered and titled. start with background discussions and move on to challenges,

Early diagnosis or prediction is crucial for timely and treatment. Early intervention is critical for improving patient outcomes and reducing costs (Brookmeyer et al., 2017). However, dementia diagnosis is often delayed or inaccurate, leading to inadequate care and support for patients and caregivers (Pais et al., 2020).

2.2 Current State of Machine Learning in Dementia Prediction

The use of machine learning models for dementia prediction has shown promise in recent studies, with some models achieving high accuracy and generalizability across different populations (Javeed et al., 2023). Machine learning models have been increasingly used for dementia prediction, leveraging various features, and achieving promising performance (Park et al. 2020).

Researchers are showing interest in this field, with different people employing different data mode and mechanisms to produce better accuracy. Li et al. (2021) in their study employed logistic regression, decision trees, and random forests to predict dementia using cognitive tests and neuroimaging features, achieving an accuracy of 85.7%. Similarly, Park et al. (2020) used different machine learning algorithms and a combination of cognitive tests, neuroimaging, and biomarkers variables to predict Alzheimer's disease with high accuracy of 92.5%.

Several machine learning models have been developed for dementia prediction, but they have limited accuracy. A systematic review of machine learning models for dementia prediction found that the models had a median accuracy of 85% (range 70-100%) (Bari Antor et al., 2021). Another review found that the models had a median area under the receiver operating characteristic curve (AUC-ROC) of 0.85 (range 0.70-0.95) (Grueso and Viejo-Sobera, 2021).

Table 1:

Summary of Machine Learning Studies for Dementia and Alzheimer's Disease Prediction

Study	**ML Approaches Used**	**Types of Data Used**	**Specific Task**	**Accuracy/ Performance**
Li et al. (2021)	Logistic regression, Decision trees, Random forests	Cognitive tests, Neuroimaging	Dementia prediction	85.7% accuracy
Park et al. (2020)	Not specified (general "machine learning algorithms")	Cognitive tests, Neuroimaging, Biomarkers	Alzheimer's disease prediction	92.5% accuracy
Bari Antor et al. (2021)	Systematic review of various ML models	Not specified	Alzheimer's disease prediction	Median accuracy of 85% (range 70-100%)
Grueso and Viejo-Sobera (2021)	Systematic review of various ML models	Not specified	Predicting progression from mild cognitive impairment to Alzheimer's disease	Median AUC-ROC of 0.85 (range 0.70-0.95)

Musto et al. (2021)	Not specified	Not specified	Dementia prediction	83.6% accuracy
Battineni et al. (2020)	Not specified (discussed challenges in ML models)	Not specified	Not specified	Not reported

2.3 Applying Machine Learning to Dementia Prediction

Machine learning models have been applied to various datasets to predict the likelihood of developing dementia. For example, a study by Park et al. (2020) used large-scale administrative health data to predict the incidence of Alzheimer's disease with high accuracy. Similarly, a study by Musto et al. (2021) developed a machine learning approach to predict deterioration in Alzheimer's disease.

These models can be trained on various features, including clinical, imaging, and biomarker data. For instance, a study by Li et al. (2021) applied machine learning to omics, imaging, and clinical data to identify patterns and predict Alzheimer's disease. Another study by Battineni et al. (2020) used machine learning predictive models for chronic disease diagnosis, including dementia.

The use of machine learning models for dementia prediction has shown promise, with some models achieving high accuracy and generalizability across different populations. For example, a study by Kim and Lim (2021) developed a deep neural network-based method for predicting dementia using big data, achieving high accuracy and F1-score. Similarly, a study by Bucholc et al. (2023) developed a hybrid machine learning approach to predict conversion from mild cognitive impairment to dementia, achieving high accuracy and area under the receiver operating characteristic curve.

2.4 Challenges and Limitations

The results of several studies that have used machine learning algorithms to predict dementia have been inconsistent (Li et al., 2021). For example, Park et al. (2020) developed a machine learning model that predicted Alzheimer's disease with an accuracy of 85.7%, while Musto et al. (2021) developed a model that predicted dementia with an accuracy of 83.6%.

Existing studies have limitations. For example, Li et al. (2021) relied on a relatively small sample size (n=150) and did not consider important biomarkers like APOE genotyping. Park et al. (2020) used a larger dataset (n=1200) but did not report feature importance or model interpretability. The field faces challenges in terms of data quality, feature selection, and model generalizability (Battineni et al. 2020). Machine learning models for dementia prediction often suffer from overfitting and lack robustness due to small sample sizes and noisy data.

2.5 Future Direction

Other limitations of existing models highlight the need for larger and more diverse datasets to improve their performance (Bari Antor et al., 2021). Additionally, there is a need for more critical evaluation of the models, including their generalisability and clinical utility (Kumar et al., 2021). To address these challenges, future studies should prioritize data harmonization, feature engineering, and ensemble learning models. Additionally, incorporating domain knowledge and expert feedback can improve model interpretability and generalisability.

The literature review highlights the need for more research on machine learning models for dementia prediction, with a focus on larger and more diverse datasets, and more critical evaluation of the models.

3 ETHICAL IMPLICATIONS

The project utilises anonymised datasets from Kaggle and OASIS. These are reputable open-access repositories, ensuring the protection of personal information and maintaining ethical standards.

3.1 Individuals Affected:

The project benefits from the contributions of individuals who have shared their data through open-access repositories, promoting research and innovation.

3.2 Ethical Considerations:

- Acknowledgement: The project acknowledges the original creators and sources of the datasets, ensuring proper citation and credit.
- Compliance: The project adheres to the terms and conditions of the repositories and dataset licenses, respecting the intentions of the data providers.
- Data Security: Robust security measures are implemented to protect the dataset during storage, processing, and analysis, maintaining confidentiality and integrity.
- Transparency: Clear documentation and publication of the dataset's origin, limitations, and usage ensure accountability and transparency.

By leveraging anonymised datasets from Kaggle and OASIS open-access repositories and addressing these ethical considerations, the project promotes responsible innovation and respects the contributions of the data providers.

Commented [EP5]: name the repositories

4 PROJECT MANAGEMENT & PROGRESS REVIEW

Project management is crucial for achieving the objectives of this machine learning project for dementia prediction. A project plan has been developed, outlining the phases, tasks, and activities with estimated timelines.

4.1 Project Plan

- Topic Selection (Completed)
  - Duration: 5 weeks
  - Tasks: Find Research and Select Suitable Topic for Dissertation
- Literature Review (Completed)
  - Duration: 2 weeks
  - Tasks: Research and analyse existing studies on machine learning for dementia prediction
- Data Collection and Preprocessing (Ongoing)
  - Duration: 3 weeks
  - Tasks: Collect datasets from open-access repositories, preprocess and anonymize data
- Model Development and Training (Upcoming)
  - Duration: 6 weeks
  - Tasks: Develop and train machine learning models for dementia prediction
- Model Evaluation and Testing (Upcoming)
  - Duration: 1 weeks
  - Tasks: Evaluate and test the performance of the developed models
- Project Report and Documentation (Ongoing)
  - Duration: 1 weeks
  - Tasks: Document project progress, results, and conclusions

4.2 Progress Review:

- Topic Selection completed.
- Literature review completed, providing a solid foundation for the project.

- Data collection and preprocessing underway, with a focus on anonymised datasets from Kaggle and OASIS open access repositories.
- Interim Report submission completed.
- Model development and training scheduled to commence upon completion of data preprocessing.
- Model evaluation and testing scheduled to commence upon completion of model development.
- Project report and documentation ongoing, ensuring transparent progress tracking.
- Final Project submission process to commence upon completion of project report.

4.3 Critical Reflective Analysis:

- The project plan has helped maintain focus and direction, with the literature review providing a strong foundation.
- Time management has been effective, with tasks completed within estimated timelines.
- Collaboration with open-access repositories has ensured access to valuable datasets while maintaining ethical standards.
- Areas for improvement include more frequent progress updates and enhanced documentation.

4.4 Gantt Chart:

Agile Gantt Chart has been designed in Microsoft Excel 365 to manage and track the progress of this project. Click the link in this Dissertation - Gantt Chart to open the Gantt Chart designed in Microsoft Excel.

This project plan and progress review demonstrate the systematic approach adopted to achieve the project objectives. The critical reflective analysis highlights areas of strength and improvement, ensuring the project remains on track to deliver a machine learning model for dementia prediction.

5 REFERENCES

ARVANITAKIS, Z. and BENNETT, D.A., 2019. What Is Dementia? *JAMA* [online]. 322 (17), p. 1728. Available from: https://jamanetwork.com/journals/jama/fullarticle/2753900 [Accessed 5 Jul 2024].

AYODELE, T., ROGAEVA, E., KURUP, J.T., BEECHAM, G., and REITZ, C., 2021. Early-Onset Alzheimer's Disease: What Is Missing in Research? *Current Neurology and Neuroscience Reports* [online]. 21 (2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7815616/ [Accessed 5 Jul 2024].

BARI ANTOR, M., JAMIL, A.H.M.S., MAMTAZ, M., MONIRUJJAMAN KHAN, M., ALJAHDALI, S., KAUR, M., SINGH, P., and MASUD, M., 2021. A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. *Journal of Healthcare Engineering* [online]. 2021, p. e9917919. Available from: https://www.hindawi.com/journals/jhe/2021/9917919/ [Accessed 5 Jul 2024].

BATTINENI, G., SAGARO, G.G., CHINATALAPUDI, N., and AMENTA, F., 2020. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *Journal of Personalized Medicine* [online]. 10 (2). Available from: https://pubmed.ncbi.nlm.nih.gov/32244292/ [Accessed 5 Jul 2024].

BROOKMEYER, R., ABDALLA, N., KAWAS, C.H., and CORRADA, M.M., 2017. Forecasting the Prevalence of Preclinical and Clinical Alzheimer's Disease in the United States. *Alzheimer's & Dementia* [online]. 14 (2), pp. 121–129. Available from: https://www.sciencedirect.com/science/article/abs/pii/S155252601733813X [Accessed 5 Jul 2024].

BUCHOLC, M., TITARENKO, S., DING, X., CANAVAN, C., and CHEN, T., 2023. A Hybrid Machine Learning Approach for Prediction of Conversion from Mild Cognitive Impairment to Dementia. *Expert Systems with Applications*. 217, p. 119541.

DR RAINA LOH, 2023. Dementia. *Keystone Clinic & Surgery* [online]. Available from: https://keystonemedical.com.sg/dementia/ [Accessed 5 Jul 2024].

GRUESO, S. and 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).

JAVEED, A., DALLORA, A.L., BERGLUND, J.S., ALI, A., ALI, L., and ANDERBERG, P., 2023. Machine Learning for Dementia Prediction: a Systematic Review and Future Research Directions. *Journal of Medical Systems*. 47 (1).

KIM, J. and LIM, J., 2021. A Deep Neural Network-Based Method for Prediction of Dementia Using Big Data. *International Journal of Environmental Research and Public Health* [online]. 18 (10), p. 5386. Available from: https://pubmed.ncbi.nlm.nih.gov/34070100/#:~:text=A%20Deep%20Neural%20Network-Based%20Method%20for%20Prediction%20of.

KUMAR, S., OH, I., SCHINDLER, S., LAI, A.M., PAYNE, P.R.O., and GUPTA, A., 2021. Machine Learning for Modeling the Progression of Alzheimer Disease Dementia Using Clinical data: a Systematic Literature Review. *JAMIA Open*. 4 (3).

LI, Z., JIANG, X., WANG, Y., and KIM, Y., 2021. Applied Machine Learning in Alzheimer's Disease research: omics, imaging, and Clinical Data. *Emerging Topics in Life Sciences*. 5 (6), pp. 765–777.

MITCHELL, T.M., 1997. *Machine Learning*. New York: Mcgraw-Hill.

MUSTO, H., STAMATE, D., PU, I., and STAHL, D., 2021. A Machine Learning Approach for Predicting Deterioration in Alzheimer's Disease. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*.

PAIS, M., MARTINEZ, L., RIBEIRO, O., LOUREIRO, J., FERNANDEZ, R., VALIENGO, L., CANINEU, P., STELLA, F., TALIB, L., RADANOVIC, M., and FORLENZA, O.V., 2020. Early Diagnosis and Treatment of Alzheimer's disease: New Definitions and Challenges. *Brazilian Journal of Psychiatry*. 42 (4).

PARK, J.H., CHO, H.E., KIM, J.H., WALL, M.M., STERN, Y., LIM, H., YOO, S., KIM, H.S., and CHA, J., 2020. Machine Learning Prediction of Incidence of Alzheimer's Disease Using large-scale Administrative Health Data. *npj Digital Medicine* [online]. 3 (1), pp. 1–7. Available from: http://www.nature.com/articles/s41746-020-0256-0 [Accessed 5 Jul 2024].

QUINN, C., PICKETT, J.A., LITHERLAND, R., MORRIS, R.G., MARTYR, A., and CLARE, L., 2021. Living Well with dementia: What Is Possible and How to Promote It. *International Journal of Geriatric Psychiatry* [online]. 37 (1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9292841/ [Accessed 5 Jul 2024].

SHEPPARD, O. and COLEMAN, M., 2020. Alzheimer's Disease: Etiology, Neuropathology and Pathogenesis. *Exon Publications* [online]. pp. 1–21. Available from: https://exonpublications.com/index.php/exon/article/view/252/473#figures [Accessed 5 Jul 2024].

SHIN, J.-H., 2022. Dementia Epidemiology Fact Sheet 2022. *Annals of Rehabilitation Medicine* [online]. 46 (2), pp. 53–59. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9081392/ [Accessed 4 Jul 2024].

WORLD HEALTH ORGANIZATION, 2023. Dementia. *World Health Organization* [online]. Available from: https://www.who.int/news-room/fact-sheets/detail/dementia [Accessed 4 Jul 2024].

ZHU, F., LI, X., TANG, H., HE, Z., ZHANG, C., HUNG, G.-U., CHIU, P.-Y., and ZHOU, W., 2020. Machine Learning for the Preliminary Diagnosis of Dementia. *Scientific Programming*. 2020, pp. 1–10.

APPENDIX A SAMPLE HEADING (Heading A1)