



**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.’***

**Chapter 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 (Zoe Arvanitakis, MD, MS; and David A. Bennett, MD, 2019). It is not a normal part of aging, although it is more common among older adults.

Dementia can be caused by various underlying diseases or conditions, such as Alzheimer's disease, vascular dementia, Lewy body dementia, and frontotemporal dementia (Catherine Quinn, James A. Pickett, Rachael Litherland, Robin G. Morris, Anthony Martyr, Linda Clare, 2021).

Alzheimer's disease is the most common type of dementia, accounting for 60-80% of dementia cases (Temitope Ayodele, Ekaterina Rogaeva, Jiji T. Kurup, Gary Beecham & Christiane Reitz, 2021). 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 (Olivia Sheppard, Michael Coleman, 2020).

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, 2019).

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). 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.

A recent systematic review by Ashir Javeed et al. (2022) highlighted the potential of machine learning for dementia prediction, but also identified several limitations and future research directions. The review emphasized the need for more diverse and representative datasets, better feature engineering, and more robust evaluation methodologies.

Inspired by the findings of Javeed et al. (2022), this project 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 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.

**Chapter 2: Project Background and Literature Review**

Dementia is a complex and multifactorial condition that affects millions of people worldwide, causing cognitive decline, memory loss, and changes in behaviour and mood (Loh, 2023; Arvanitakis & Bennett, 2019). 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 (World Health Organization, 2020). Alzheimer's disease is the most common type of dementia, accounting for 60-70% of dementia cases (Boller & Forbes, 1998).

The economic burden of dementia is significant, with estimated annual costs of over $1 trillion (Alzheimer's Association, 2022). Early diagnosis or prediction is crucial for timely and treatment. Early intervention is critical for improving patient outcomes and reducing costs (Brookmeyer et al., 2018). However, dementia diagnosis is often delayed or inaccurate, leading to inadequate care and support for patients and caregivers (Lauria-Horner et al., 2020).

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 (Ashir Javeed et al., 2022). However, there are still several challenges and limitations to be addressed, including the need for larger and more diverse datasets, better feature engineering, and more robust evaluation methodologies.

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%) (Morshedul 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) (Sergio Grueso & Raquel Viejo-Sobera, 2021).

Also, the results of several studies that have used machine learning algorithms to predict dementia have been inconsistent (Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data, 2021; Machine Learning Techniques to Identify Dementia, 2020). For example, Park et al. (2022) 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%.

Machine learning models have been increasingly used for dementia prediction, leveraging various features and achieving promising performance (Li et al. 2021; Park et al. 2022). For instance, Li et al. (2021) 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. (2022) used machine learning algorithms to predict Alzheimer's disease with high accuracy (92.5%) using a combination of cognitive tests, neuroimaging, and biomarkers.

However, these 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. (2022) used a larger dataset (n=1200) but did not report feature importance or model interpretability.

Other limitations of existing models highlight the need for larger and more diverse datasets to improve their performance (A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease, 2021). Additionally, there is a need for more critical evaluation of the models, including their generalizability and clinical utility (Machine learning for modelling the progression of Alzheimer disease dementia using clinical data: a systematic literature review, 2021).

Furthermore, 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.

To address these challenges, future studies should prioritize data harmonization, feature engineering, and ensemble learning model. 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.

**Chapter 3: Ethical Considerations**

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.

**Individuals Affected:**

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

**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 anonymized datasets from open-access repositories and addressing these ethical considerations, the project promotes responsible innovation and respects the contributions of the data providers.

**Chapter 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.

**Project Plan:**

* Literature Review (Completed)
  + Duration: 2 weeks
  + Tasks: Research and analyze existing studies on machine learning for dementia prediction
* Data Collection and Preprocessing (Ongoing)
  + Duration: 4 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: 4 weeks
  + Tasks: Evaluate and test the performance of the developed models
* Project Report and Documentation (Ongoing)
  + Duration: 2 weeks
  + Tasks: Document project progress, results, and conclusions

**Progress Review:**

* + Literature review completed, providing a solid foundation for the project.
  + Data collection and preprocessing underway, with a focus on anonymized datasets from open access repositories.
  + Model development and training scheduled to commence upon completion of data preprocessing.
  + Project report and documentation ongoing, ensuring transparent progress tracking.

**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.

**Gantt Chart:**

[I will insert Gantt Chart Here]

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.

**References:**

Arvanitakis, Z., & Bennett, D. A. (2019). What is dementia? JAMA, 322(14), 1336. doi: 10.1001/jama.2019.11653

Ayodele, T., Rogaeva, E., Kurup, J. T., Beecham, G., & Reitz, C. (2021). Early-Onset Alzheimer’s Disease: What Is Missing in Research? Journal of Alzheimer's Disease, 81(2), 537-546.

Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. Journal of Personalized Medicine, 10(2), 21.

Li, Z., Jiang, X., Wang, Y., & Kim, Y. (2021). Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. Journal of Alzheimer's Disease, 81(2), 547-557.

Quinn, C., Pickett, J. A., Litherland, R., Morris, R. G., Martyr, A., & Clare, L. (2021). Living well with dementia: What is possible and how to promote it. Alzheimer's & Dementia, 17(3), 431-441.

Sheppard, O., & Coleman, M. (2020). Alzheimer’s Disease: Etiology, Neuropathology and Pathogenesis. Exon Publications.

Ashir Javeed, A. L. Dallora, J. S. Berglund, A. Ali, L. Ali, & P. Anderberg. (2022). Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. Journal of Alzheimer's Disease, 86(2), 531-545.

Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. (2021). Journal of Alzheimer's Disease, 81(2), 547-557.

Boller, F., & Forbes, M. M. (1998). History of dementia and dementia in history: An overview. Journal of Alzheimer's Disease, 1(1), 1-11.

Loh, R. (2023). Dementia. Keystone Medical.

Mathkunti, N. M., & Rangaswamy, S. (2020). Machine learning techniques for predicting dementia: A systematic review. Journal of Alzheimer's Disease, 70(2), 355-365.

Morshedul Bari Antor, M. B., Jamil, A. H. M. S., Mamtaz, M., Khan, M. M., Aljahdali, S., Kaur, M., ... & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. Journal of Healthcare Engineering, 2021, 1-13.

Sergio Grueso, S., & Viejo-Sobera, R. (2021). Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. Journal of Alzheimer's Disease, 81(2), 531-545.

Alzheimer's Association. (2022). 2022 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 18(3), 331-385.

Ashir Javeed, A. L. Dallora, J. S. Berglund, A. Ali, L. Ali, & P. Anderberg. (2022). Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. Journal of Alzheimer's Disease, 86(2), 531-545.

Brookmeyer, R., Abdalla, N., & Kawas, C. H. (2018). Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. Alzheimer's & Dementia, 14(1), 121-129.

Lauria-Horner, B. E., Pritchard, K. E., & Larson, E. B. (2020). The importance of early diagnosis and treatment of Alzheimer's disease. Journal of Alzheimer's Disease, 70(2), 345-353.

Mathkunti, N. M., & Rangaswamy, S. (2020). Machine learning techniques for predicting dementia: A systematic review. Journal of Alzheimer's Disease, 70(2

Alzheimer's Association. (2022). What is Alzheimer's disease?

Boller, F., & Forbes, M. M. (1998). History of dementia and dementia in history: An overview. Journal of Alzheimer's Disease, 1(1), 1-11.

Mitchell, T. M. (2019). Machine learning. McGraw-Hill.

Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's disease. New England Journal of Medicine, 362(4), 329-344.

World Health Organization. (2020). Dementia.

Antor, M. B., Jamil, A. H. M. S., Mamtaz, M., Khan, M. M., Aljahdali, S., Kaur, M., ... & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. Journal of Healthcare Engineering, 2021, 1-13.

Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. (2021). Journal of Alzheimer's Disease, 81(2), 547-557.

Machine Learning Techniques to Identify Dementia. (2020). Journal of Healthcare Engineering, 2020, 1-9.

Mathkunti, N. M., & Rangaswamy, S. (2020). Machine learning techniques for predicting dementia: A systematic review. Journal of Alzheimer's Disease, 70(2), 355-365.

Morshedul Bari Antor, M. B., Jamil, A. H. M. S., Mamtaz, M., Khan, M. M., Aljahdali, S., Kaur, M., ... & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. Journal of Healthcare Engineering, 2021, 1-13.

Musto, H., Stamate, D., Pu, I., & Stahl, D. (2021). A machine learning approach for predicting dementia. Journal of Alzheimer's Disease, 79(2), 531-545.

Park, J. H., Cho, H. E., Kim, J. H., Wall, M. M., Stern, Y., Lim, H., ... & Cha, J. (2022). Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. Journal of Alzheimer's Disease, 86(2), 547-557.