Health Trends Analytics for Older People: AI-Powered Visualization

YING KIT JERRY, LI, 8469180 (ykjl975@uowmail.edu.au)
CHENKUN, YE, 8376141 (cy981@uowmail.edu.au)
YI OUYANG, 8773397 (yo343@uowmail.edu.au)
SHASHA, YOU, 8834428 (sy255@uowmail.edu.au)
HIU FUNG, WONG, 8558796 (hfw890@uowmail.edu.au)
HO, NG, 8527507 (hn951@uowmail.edu.au)

June 1, 2025

Abstract

With a global health awareness of the dementia effect, there are over 55 million people suffering from this issue. Meanwhile, there are some challenges that exist in aged care systems, which means dementia elderly people with dementia do not receive well-planned care. While existing elderly care environments often suffer from underutilised analytics tools and fragmented information systems, these shortcomings always bring a standstill in personalised, data-driven care. Meanwhile, although some predictive models are suggested, they still cannot fulfil complicated linear regression.
This project proposes the development of an AI-powered health trend visualisation platform to address these issues. Leveraging Microsoft Azure services, which include Synapse Analytics, Azure ML, and Power BI. Our project's solution will integrate and process clinical data from aged care facilities to uncover symptom patterns and intervention effectiveness. It will provide both population-level dashboards and personalized health status diagrams for individual residents. Under the combination of machine learning predictions with real-time visual insights, the final product aims to empower nursing staff with actionable intelligence to support early risk detection, optimise care strategies, and enhance decision-making under healthcare compliance.

1 Introduction

1.1 Background

The global acceleration of population aging presents critical challenges to elderly care systems. According to the World Health Organization (Organisation, 2025), over 55 million people are currently living with dementia—a condition particularly prevalent in long-term care settings. In the United States, the prevalence of dementia in nursing homes increased from 45% in 2010 to 68% in 2015 (Miranda et al., 2020).

Timely identification and continuous monitoring of key health indicators—such as weight change, mood fluctuations, and dietary intake—are essential for high-quality elderly care. Proactive health monitoring can prevent deterioration, improve patient outcomes, and reduce unnecessary hospitalizations. Studies emphasize the importance of preventive care models to reduce the burden of chronic conditions in aging populations (Miranda et al., 2020).

Problem Statement Despite the critical importance of data-driven decision-making, significant barriers persist in elderly care settings:

- **Underutilization of Analytics Tools:** Dashboards and data analytics platforms are markedly less prevalent in elderly care compared to acute care environments. This stems from poorer ICT infrastructure, limited technical support, and lower digital literacy among staff (Vogelsmeier et al., 2008).
- **Fragmented Information Systems:** Valuable resident health data remains siloed across departments and systems. This fragmentation prevents a comprehensive view of individual conditions, leading to disjointed care delivery (Ludlow et al., 2021).

Proposed Solution Artificial intelligence (AI) offers strong potential to address these challenges by uncovering patterns in complex health data. This project proposes a solution integrating:

- **Azure Synapse Analytics** for unified data integration
- **Azure Machine Learning** for personalized health predictions
- **Power BI** for intuitive, actionable visualizations

The solution aims to embed predictive analytics into daily elderly care workflows, enhancing proactive decision-making while ensuring compliance with HIPAA and GDPR regulations.

1.2 Project Objectives

The overarching objective of this project is to enhance early risk detection and care planning for dementia by leveraging advanced data analytics, AI, and visual analytics. This is achieved through five interlinked components, each addressing a specific stage of the analytics workflow.

1.2.1 Data Integration & Preparation

Collect and integrate data from diverse sources, including patient demographics, symptom logs (e.g., agitation, depression), and nursing interventions in aged care facilities. Clean, normalize, and structure the data using Azure Synapse Analytics to ensure consistency, completeness, and readiness for robust downstream data analysis. This foundational step is critical for generating reliable insights.

1.2.2 Trend Analysis & Visualization with Power BI

Develop comprehensive data analysis capabilities by building interactive dashboards using Microsoft Power BI. These dashboards will analyze and visualize population-level trends over time, specifically focusing on symptom progression and patterns of nursing interventions. Power BI's advanced analytics features will enable aged care staff to perform ad-hoc data exploration, identify common health trajectories, detect subtle changes in health conditions, and gain actionable insights from the integrated dataset.

1.2.3 AI-Powered Insights from Data Analysis

Train machine learning models via Azure Machine Learning on the prepared datasets to forecast future trends in agitation and depression, and recommend evidence-based nursing actions. These predictive and prescriptive analytics insights, derived from historical and real-time data analysis, aim to assist proactive care planning and early intervention for individuals identified as higher risk through the analytical models.

1.2.4 Individual Health Monitoring & Analysis

Design a personalised visualisation interface powered by Power BI that allows nursing staff to input and analyze individual-level data, generating dynamic health status diagrams. Azure Functions will enable real-time data handling and analysis, supporting data-driven, individualised care tracking and clinical decision-making at the point of care. Power BI reports will provide detailed, patient-specific analytical views.

1.2.5 Deployment & Scalability

Deploy the entire solution, including the Power BI analytics dashboards and interfaces, as a cloud-based application on Microsoft Azure. Utilize Azure Static Web Apps for the front-end and Azure Functions for backend services. Ensure the system, particularly its data analysis and visualization components, is secure, scalable, and compliant with relevant healthcare data privacy regulations (e.g., GDPR, HIPAA). Power BI content will be securely embedded and managed within the application.

1.3 Related Work

1.3.1 Existing Predictive Models in Geriatrics

1.3.2 Dementia Risk Prediction: Advantages and Disadvantages

Predicting dementia risk in older adults has significant implications for early intervention, perioperative planning, and long-term care optimization. Existing predictive models can be broadly categorized into traditional statistical methods, self-report indices, and machine learning approaches. Each has distinct advantages and limitations.

Traditional Statistical Models Traditional approaches such as logistic regression and Cox proportional hazards models are widely adopted due to their simplicity and interpretability. They are easily understood by clinicians and are well-suited for contexts requiring transparency in decision-making.

However, their predictive power is limited by assumptions of linearity and an inability to model complex interactions, which can reduce effectiveness in heterogeneous elderly populations.

Self-Report Tools (e.g., ANU-ADRI) The Australian National University Alzheimer's Disease Risk Index (ANU-ADRI) uses questionnaire-based data such as education, physical activity, and smoking history. With reported C-statistics ranging from 0.637 to 0.740 across multiple studies (Anstey et al., 2014), it is suitable for population-level screening and preoperative cognitive risk assessment.

Its limitations include susceptibility to self-report bias and limited granularity in capturing clinical history or physiological data, which may reduce its effectiveness for individualized assessment.

Machine Learning Approaches (e.g., Survival Random Forest - SRF) Machine learning models such as the Survival Random Forest (SRF) are well-suited for handling right-censored longitudinal data and can model nonlinear, high-dimensional relationships. Studies using datasets like the English Longitudinal Study of Ageing (ELSA) have identified risk factors such as sleep quality and prior health events that are useful in perioperative planning (Stamate et al., 2022).

Nevertheless, these models are often viewed as black-box systems, raising concerns about interpretability and trust. They also demand well-curated datasets and considerable computational resources.

Clinical Integration and Applications Embedding dementia risk models into perioperative digital platforms allows for proactive care of high-risk patients. For example, electronic health records can trigger alerts for patients with elevated risk scores, enabling timely interventions such as:

- Cognitive-friendly rehabilitation programs
- Delirium prevention protocols
- Early referral to geriatric care teams

Broader Impact Incorporating dementia risk assessments into surgical workflows supports more effective screening, discharge planning, and post-discharge monitoring. This approach promotes personalized, coordinated care for older adults at risk of cognitive decline.

1.3.3 Visualization and Monitoring Systems

In elderly care research and clinical practice, effective data visualization tools are essential for interpreting complex longitudinal health information. Widely adopted platforms include **Tableau**, **Dash**, and **Power BI**, each offering distinct capabilities.

Tableau Tableau excels in visual analytics with an intuitive interface, allowing non-technical users to explore health trends effectively. However, it has limitations in granular access control and lacks robust healthcare compliance support (e.g., HIPAA, GDPR). Additionally, its high licensing costs make it less feasible for sensitive patient data environments.

Dash Dash, an open-source Python framework, enables seamless integration of predictive models, offering flexibility for research workflows. Nonetheless, its code-centric nature complicates the implementation of standardized audit trails and privacy-preserving mechanisms essential for clinical data governance.

Power BI Power BI (integrated with Azure) was selected as the core visualization platform for this project due to critical ethical and security considerations, including:

- **Healthcare compliance**: Native integration with Azure identity management enables HIPAA-compliant role-based access control (RBAC), ensuring adherence to the principle of least privilege for dementia patient records.
- **Data integrity**: Employs end-to-end encryption for data in transit and at rest, with automatic audit logging to maintain the chain of custody for clinical observations.
- **Privacy by design**: Features such as row-level security and built-in data loss prevention (DLP) ensure ethical handling of sensitive behavioral symptom data.
- **Trusted ecosystem**: Seamlessly connects to EMRs and screening tools within Azure's certified cloud infrastructure, promoting interoperability and secure data exchange.

Given the project's focus on dementia-related metrics, Power BI's zero-trust architecture and centralized governance framework offer a robust foundation for real-time monitoring while ensuring patient confidentiality and regulatory compliance across multidisciplinary teams.

1.3.4 ML Innovations in Elderly Care

In recent years, machine learning has played an increasingly important role in elderly care, particularly in dementia risk prediction and postoperative monitoring. Compared to traditional statistical methods, modern ML models are better at capturing non-linear relationships and complex interactions across multidimensional data, leading to improved accuracy and adaptability in risk assessment. Key advancements include:

- **Enhanced Data Fusion**: By integrating data from electronic medical records, self-report tools, and wearable devices, ML models can provide a more comprehensive assessment of an older adult's health status. For example, incorporating lifestyle factors and prior health events during preoperative screening helps identify cognitively vulnerable individuals and supports early intervention.

- **Progress in Temporal Modeling**: For tracking changes such as postoperative cognitive decline or nutritional fluctuations, time-sensitive models (e.g., LSTM or Transformer architectures) enable better monitoring of evolving risk trajectories, thereby enhancing proactive care planning.
- **Model Interpretability for Clinical Use**: Tools like SHAP help explain model predictions, making outcomes more understandable and trustworthy for clinicians. This supports the integration of ML outputs into preoperative assessments and discharge planning workflows.

2 Methodology

2.1 Methodology Approach

2.1.1 Data Source Selection and Exploration

Database Source
Regarding all sources, all the data are from a university-managed health information system or confidential patient databases. These data may cover the clinical observation report of dementia, patient records from EHR, Healthcare review for senior dementia patients, medication log, etc. University health databases offer a good academic environment under a governed ethical base. Meanwhile, university-based databases are a reliable platform for exploring and forecasting health trend dynamics, especially in dementia.

Data Exploration
The stage of exploratory data analysis (EDA) helps our teammates understand the data patterns, data quality, and the variable relationships within the dementia database, through a series of features to identify the factors of dementia (Livingston et al., 2020).By visualising and summarising these data through Power BI, it assists the features to approach and test a strong and reliable machine learning model.

For the details of knowing what kind of data is desired to investigate, and to give an in-depth understanding of data patterns from the patient, the missing values of the data and outliers, these data can identify the features to relevant to dementia-based analysis. The following are the key types that are important to do EDA (Livingston et al., 2020)(Lee et al., 2023).

- Demographic information (e.g. age, gender, living status, education)
- Cognitive function diagnosis
- Family history
- Behavioral symptoms
- Medication proportion
- Care Review

Here are sample core scope criteria for evaluating data quality, although there will be more complex ones when the project goes deeper.

Table 1: Scope and Measurement Criteria for Dementia Dataset Exploration

Scope	Measurement
Data Structure	Variable relationships between factors (e.g., demographics, symptoms, medication); Inputs in each column and row (number or text description)
Data Completeness	Missing value causes; presence of duplications; the impact of missing data on upcoming modelling.
Demographic Statistics	Distribution of age, gender, or region groups.
Risk Factor Pattern	Frequency and medication of specific dementia symptoms.
Outlier Detection	Extreme value identification; abnormal data or behavioural interruption.

2.1.2 Data Cleaning and Pre-processing

Once all the data exploration is finished, the data cleaning and pre-processing ensure that all data is well-equipped and standardised for machine model readiness. Therefore, there are some main tasks to do. Once all the data exploration is finished, the data cleaning and pre-processing ensure that all data is well-equipped and standardised for machine model readiness. Therefore, there are some main tasks to do.

1. Manage missing values
To ensure a complete dataset for the machine learning model accuracy, missing or invalid values should be identified and corrected.

2. Standardise Data Formats
To make all data into a consistent form, the data field should clearly convert and categorise (e.g. DOB, Gender identification, binary flags or risk factors, etc).

3. Apply filters for proper records
To specify the target groups, the age (equal to or greater than 65) group can be a main subset, which eliminates the irrelevant datasets to raise the relevance and accuracy.

4. Handle outlier values and duplications
To ensure the accuracy of the model's operation, remove or tune the extreme values and duplications.

5. Normalize the values
To maintain features on the same scale, apply a proper "scaler" to numerical data in PowerBI function for better data clustering and machine learning compatibility.

2.1.3 Comprehensive Exploratory Data Analysis (EDA)

Exploratory Data Analysis is a method for analyzing data that seeks to gain a deep understanding of the data and discover its various features, frequently employing visual tools. This enables an individual to gain a deeper understanding of the data and identify valuable patterns (Biswal, 2023).

According to introduction for EDA from GeeksforGeeks, there are 4 main steps for the EDA. (GeeksforGeeks, 2025)

- **Step 1: Exploring Data Characteristics**

  After addressing missing data in data cleaning and pre-processing, we find the characteristics of our data by checking the distribution, central tendency and variability of our variables and identifying outliers or anomalies. This helps in selecting appropriate analysis methods and finding major data issues. We should calculate summary statistics like mean, median, mode, standard deviation, skewness and kurtosis for numerical variables. These provide an overview of the data's distribution and helps us to identify any irregular patterns or issues.

- **Step 2: Performing Data Transformation**

  Data transformation is a crucial phase in EDA as it readies our data for precise analysis and modeling. Based on the features of our data and the requirements for analysis, it may be necessary to modify it to ensure that it is in the appropriate format. Standard transformation methods will be applied, such as scaling or normalizing numerical variables through techniques like min-max scaling or standardization. Additionally, categorical variables will be encoded for machine learning using methods such as one-hot encoding or label encoding.

- **Step 3: Visualizing Relationship of Data**

  Visualization helps to find relationships between variables and identify patterns or trends that may not be seen from summary statistics alone.

  - For categorical variables, create frequency tables, bar plots and pie charts to understand the distribution of categories and identify imbalances or unusual patterns.
  - For numerical variables generate histograms, box plots, violin plots and density plots to visualize distribution, shape, spread and potential outliers.
  - To find relationships between variables use scatter plots, correlation matrices or statistical tests like Pearson's correlation coefficient or Spearman's rank correlation.
  - Clustering can also be used. It is a method that focuses on organizing similar objects or data points together according to their features or characteristics. Clustering methods, including k-means and density-based clustering, play a critical role in discovering concealed patterns and frameworks within extensive datasets.

- **Step 4: Communicate Findings and Insights**

  The last step in Exploratory Data Analysis is to express our results in a clear manner. This includes providing a summary of the analysis, highlighting important findings, and conveying our results clearly. We will articulate the objectives and range of this project, present context and background information to assist others in grasping your methodology, utilize visual aids to reinforce our conclusions, and enhance their comprehensibility. Emphasizing important findings, trends, or unusual occurrences identified, propose subsequent actions or topics that require additional exploration.

2.1.4 Feature Engineering

Data is transformed into features that may be used by machine learning algorithms through a technique known as feature engineering. It involves choosing, retrieving, and modifying the most relevant features from the data in order to construct more effective and precise machine learning models. After doing EDA, the data's structure can be understood. EDA reveals data types, ranges, missingness, and outliers. Without this understanding, the features might be created based on incorrect assumptions. EDA also shows which features are important or correlated. It helps prioritizing what features to engineer. Then we go for Feature Engineering.

According to introduction for feature engineering from GeeksforGeeks, there are 5 main steps for the feature engineering(GeeksforGeeks, 2025).

- **Step 1. Feature Creation**
  A process of generating new features based on domain knowledge or by observing patterns in the data. It is a form of feature engineering that can significantly improve the performance of a machine-learning model. Based on the result of EDA. The new features can be created based on domain knowledge, by observing patterns in the data and or generating new features by combining existing features or synthesizing new data points.
- **Step 2. Feature Transformation**
  The process of transforming the features into a more suitable representation for the machine learning model. This is done to ensure that the model can effectively learn from the data. Normalization, Scaling or Encoding can be used for feature transformation. (DataScienceSphere, 2024) The transformed data may include:
  - Statistical summaries (e.g., mean, max, min, standard deviation),
  - Temporal indicators (e.g., trends, differences between start and end values),
  - Event-based features (e.g., symptom frequency or threshold violations),
  - Encoded clinical metadata.
- **Step 3. Feature Extraction**
  The process of creating new features from existing ones to provide more relevant information to the machine learning model. This is done by transforming, combining, or aggregating existing features. The feature extraction Techniques includes Principal Component Analysis, Linear Discriminant Analysis and Kernel Principal Component Analysis (Siva, 2022)
- **Step 4. Feature Selection**
  The process of identifying and choosing the most significant features to include in machine learning models. It involves using various techniques to reduce the number of input variables by removing irrelevant or repetitive features and focusing on those that are most important for the model's performance. The methods include Filter, Wrapper and Embedded methods. (HEAVY.AI, n.d.)
- **Step 5. Feature Scaling**
  Transforms features to a common range or distribution, through either normalization or standardization, enhancing model performance by eliminating issues that arise

from differences in feature scales. It is essential in machine learning, ensuring numerical features have a uniform scale to avoid unequal influence on the model. The methods include Normalization, Standardization, Min-Max Scaling, Robust Scaling(Singh, 2024)

2.1.5 Model Selection

The goal of this project is to detect anomalies and monitor trends in the health records of older individuals in aged care facilities using structured time series data. The dataset consists of high-dimensional health variables per individual, organized into overlapping time windows. Due to variability in which features are relevant for each patient, and the presence of temporal patterns and class imbalance, a flexible and adaptive model selection strategy is essential.

To address these challenges, the project will explore both tree-based ensemble models and sequence-aware deep learning models. These approaches offer complementary capabilities: one excels at interpreting structured input with engineered features, while the other captures temporal dependencies across time.

Tree-Based Models for Tabular Feature Learning
Random Forest and Gradient Boosting models (e.g., XGBoost) are strong candidates for learning from engineered tabular representations of health data windows(Rathore et al., 2017). These models:

- Handle high-dimensional input and irrelevant features effectively
- Are resilient to class imbalance through weighting or sampling strategies
- Provide feature importance metrics, supporting explainability and individual-level insight

Since these models are not inherently designed for time series, the project will apply feature engineering to extract useful signals from each time window as mentioned in 2.1.4. Feature Engineering Step 2.

This approach allows each time window to be treated as a structured input instance suitable for supervised learning.

Sequence Models for Temporal Pattern Learning
To capture changes and trends across multiple consecutive time windows, sequence-based deep learning models will be explored, such as:

- LSTM Autoencoders, which learn baseline sequence behavior and identify deviations (Ding et al., 2019)
- Temporal Convolutional Networks (TCNs), which efficiently model long-range dependencies and detect abnormal transitions using dilated convolutions.

These models are suitable for detecting trends and anomalies that unfold over time and are capable of adapting to variability in the relevance and timing of health-related features. Input sequences will be constructed based on configurable time intervals, and normalization may be applied to account for individual differences.

Selection Strategy
Model selection will be guided by:

- Evaluation metrics including precision, recall, F1-score, and false positive rate
- Generalizability across individuals with differing health profiles
- Compatibility with cloud-based training and deployment using Microsoft Azure.

This process will remain iterative, allowing flexibility as insights emerge through data exploration and model experimentation.

2.1.6 Model Training and Validation

Based on the selected models, a tailored training and validation strategy will be implemented to ensure effectiveness, robustness, and scalability. The methodology will address model-specific requirements and known challenges such as temporal structure, high-dimensional data, inter-patient variability, and class imbalance.

Data Preparation and Feature Engineering
Patient health data will be organized into overlapping time windows of configurable length and stride. These windows will serve as the fundamental unit of analysis for both tabular and sequential modeling approaches.

For Tree-Based Models (e.g., Random Forest, XGBoost)
Each time window will be transformed into a structured feature vector using:

- Statistical summaries of key variables
- Temporal trends and deltas
- Event-based features
- Encoded patient attributes

For Sequence Models (e.g., LSTM Autoencoder, TCN)
Data will be arranged into ordered sequences of time windows per individual, with:

- Normalization to individual baselines
- Sequence formatting to ensure compatibility with model input structures
- Optional padding or truncation if sequences vary in length

Training Strategy

Table 2: Models Training

Model Type	Training Objective
Tree-Based Models	Classify each window as normal or anomalous using engineered features.
Sequence Models	Detect deviations and transitions across sequences of windows.

Class imbalance will be addressed through:
1.Class weighting and/or resampling techniques (e.g. SMOTE to balance the training set)

2. Evaluation metrics focused on sensitivity and precision rather than accuracy

Validation Strategy
To ensure robust evaluation:

- K-fold cross-validation will be used for tabular models.
- Patient-level holdout will prevent data leakage between overlapping windows.
- Temporal validation will simulate real-world deployment by testing future data based on past training windows.

Cloud-Based Infrastructure
Training and validation will be performed on Microsoft Azure, leveraging:

- Scalable virtual compute instances,
- Automated model tracking and reproducibility tools,
- Pipeline integration for training, evaluation, and deployment.

Model Selection Confirmation
Final model selection will follow the strategy defined in Section 3.1, consistent with evaluation frameworks outlined in (Rathore et al., 2017). Validation results will determine the most appropriate model(s) for integration into the broader anomaly detection and visualisation system.

2.1.7 Model Evaluation and Visualisation Phase

After model development, we will evaluate the performance of the trained learning models and translate their outputs into interpretable visual formats in this phase. Key metrics (e.g., accuracy, AUC-ROC) are calculated to ensure clinical relevance, while visualisations (e.g., SHAP plots, trend comparisons) highlight patterns for stakeholders.

2.1.8 Dashboard Design and Visual Integration

In this phase, a user-friendly, interactive dashboard will be designed using Microsoft Power BI to present both data visualisation and AI-driven predictions. The goal is to provide an intuitive interface that enables healthcare professionals to explore patient health trends, behavioural change, nutritional status, and care interventions. Visuals will be constructed for both population-level analysis (e.g., trends across all residents) and individual-level tracking (e.g., risk scores and care history for a specific patient).

2.1.9 Result Review

This phase focuses on reviewing the outputs of the dashboard and AI model to ensure they meet the project's original goals and are understandable to end users. Both technical performance and user feedback will be taken into consideration. Simulated user testing sessions will be conducted with non-technical team members or invited reviewers to assess whether the insights provided are helpful for decision-making in the context of dementia care. Special attention will be given to explainability using tools like SHAP visualisation.

2.1.10 Model Deployment and Test

Here, the final machine learning model will be deployed into a controlled test environment to fit the end user. The goal in this phase is to verify that the model performs reliably and integrates smoothly with the Power BI dashboard.

2.1.11 Publish Dashboard

After integration and testing are complete, the final version of the Power BI dashboard will be published to the Power BI online service. This allows stakeholders and supervisors to access the dashboard through a web browser, eliminating the need for Power BI Desktop. Access permissions will be configured using role-based access control (RBAC) to protect sensitive data and ensure appropriate levels of interaction.

2.2 Project Challenges

2.2.1 Incomplete and Inconsistent Data Collection

Incomplete datasets and inconsistent formats show significant challenges in clinical databases, especially those gathered from different care teams or patient system records over extended durations. Information like age, gender, diagnosis dates, or symptom indicators may be partially empty, while categorical entries such as gender or care status might be recorded in inconsistent formats, etc. Furthermore, inconsistent record lengths and missing data can affect sequence model training due to heterogeneous sequence lengths and data gaps.

Possible Solution: Ignoring all missing data or estimating them with a mean or median value (e.g. age, gender) is the approach of the traditional model, but it might get biased. Another method is the Direct Handling approach, which considers missing components as random variables and uses Gaussian Mixture Models (GMM) for regression on incomplete datasets to estimate expected values, which normalise data per patient. This allows more accuracy in data modelling (Veras et al., 2020).

2.2.2 High Dimensionality

When the datasets have a massive number of features, especially many of them may be trivial, redundant, or have fewer connections to the predicted targets, they might accumulate the risk of overfitting, computing inefficiency, and model misinterpretation.

Possible Solution: One possible solution is feature selection, identifying and grouping similar features and using feature selection through filter, wrapper, and embedded methods and tree-based model to select the valuable and predictive variables from the original dataset, to deal with the data complexity. Another method is applying dimensionality reduction techniques, such as Principal Component Analysis (PCA), by transforming high-dimensional data into a lower dimensionality, which might preserve some necessary information. This process helps mitigate the risk of dimensionality, improves model performance, and enhances decision-making capabilities (Mahalingam and K., 2024).

2.2.3 Lack of Domain Knowledge

Creating meaningful features—such as a cognitive decline score, frequency of depression episodes, or intervention-to-symptom delay—requires a deep understanding of dementia symptoms and nursing practices. Without this, engineered features may be clinically irrelevant or even misleading.

Possible Solution: Collaborate with domain experts (e.g., nursing staff or health informatics specialists) to define key indicators and event thresholds. Use domain-informed combinations (e.g., agitation episodes per month, time between symptom onset and intervention) rather than random statistical aggregations. Validate new features through pilot visualization or correlation checks.

2.2.4 Over-Engineering and Redundancy

By using the tools like Power BI and Azure ML. Although various features look good, having too many features can lead to overfitting in predictive models and clutter visual dashboards when several indicators are closely related (e.g., weight loss and BMI decline).

Possible Solution: Start with a fundamental set of features determined through the stage of EDA; approaching correlation matrices or scores for feature importance to eliminate redundant variables is necessary. Also, feature selection techniques (e.g. tree-based selection or recursive elimination) and dimensionality reduction methods (like PCA) should be implemented only if they enhance either performance or clarity. Ensure that the visual features remain user-centred (what would a nurse or clinician actually use?)

2.2.5 Individual Variability in Relevant Features

In a diverse aged care population, not all health indicators carry equal importance across patients. A variable highly indicative of declining health for one individual may be irrelevant for another, making generalized models less effective. This variability introduces noise and reduces predictive accuracy when using a uniform feature set for all individuals.

Possible Solution: Feature selection techniques such as information gain analysis can be incorporated to identify key features per individual or subgroup. Personalized modeling strategies—such as clustering patients with similar profiles or using adaptive model components—can further enhance relevance and precision.

2.2.6 Class Imbalance

Health anomalies—such as sudden behavioral changes or adverse reactions—occur much less frequently than normal events, resulting in highly imbalanced datasets. This imbalance can cause models to favor the majority class (normal cases), thereby failing to detect critical anomalies. It is particularly dangerous in healthcare applications where missing rare but severe events may have serious consequences.

Possible Solution: Class weighting can be applied during training to give more emphasis to minority instances, or oversampling methods like SMOTE can be used to artificially balance the dataset. Additionally, model performance should be evaluated using metrics like precision, recall, and F1-score instead of overall accuracy to better reflect real-world utility.

2.2.7 Model Evaluation & Visualization Phase

In this phase, we evaluate the performance of trained learning models and translate their outputs into interpretable visual formats. Key metrics (e.g., accuracy, AUC-ROC) are calculated to ensure clinical relevance, while visualizations (e.g., SHAP plots, trend comparisons) highlight patterns for stakeholders.

Table 3: Model Evaluation Components

Component	**Objective**	**Description**	**Technology / Standard**
Metric Calculation	Quantify model effectiveness & accuracy	Evaluate models using metrics like Accuracy, F1-score, ROC-AUC, Precision, Recall. Compare against baseline thresholds.	Azure ML Studio, Python
SHAP Analysis	Explain feature importance	Generate force plots to show how variables (e.g., ”nursing frequency”) impact predictions.	SHAP library
Trend Visualization	Compare predicted vs. actual trends	Plot model forecasts (e.g., agitation risk) alongside historical data for validation.	Azure Power BI, Plotly, forecast range
Confusion Matrix	Assess classification errors	Visualize false positives/negatives in symptom detection to refine model thresholds.	Azure AutoML

2.2.8 Dashboard Design and Visual Integration

In this phase, a user-friendly, interactive dashboard will be designed using Microsoft Power BI to present both data visualization and AI-driven predictions. The goal is to provide an intuitive interface that enables healthcare professionals to explore patient health trends, behavioral change, nutritional status, and care interventions.

Table 4: Dashboard Design Components

Component	Objective	Description	Technology / Standard
Population-Level Monitoring	Identify overall trends in agitation, depression, and care activity	Line charts, bar graphs, heatmaps for symptom & intervention trends	Power BI
Individual Patient Analysis	Monitor history and risk score for a specific patient	Navigate to personal page, filter and view summaries using key indicators	Power BI DirectQuery
User Interactivity	Enable real-time exploration	Add filters (e.g., time range, patient group) and drill-down capabilities	Power BI Slicer, Drillthrough Pages
Accessibility & Compliance	Ensure usability and data security	Implement role-based access (RBAC) and color-blind-friendly palettes	Azure AD, WCAG 2.1

2.2.9 Result Review

This phase focuses on reviewing the outputs of the dashboard and AI model to ensure they meet the project's original goals and are understandable to end users. Simulated user testing sessions will assess decision-making support, with emphasis on explainability.

Table 5: Review and Validation Components

Component	Objective	Description	Technology / Standard
Goal Alignment Check	Verify project objectives	Confirm key deliverables and project scope are met	Excel (KPI Tracking)
Technical Review	Review robustness	Test edge cases; ensure dashboard reflects health trends	Power BI checklist
Explainability Review	Interpret AI outputs	Are explanations understandable/trustworthy for users?	SHAP + user feedback

2.2.10 Model Deployment & Test

This phase ensures the final ML model is deployed reliably and integrates with the dashboard.

Table 6: Deployment and Compliance Components

Component	Objective	Description	Technology / Standard
Environment Setup	Deploy model to infra	Use Docker, host on Azure AKS for scalability	Azure AKS, Docker, CI/CD
API Integration	Enable real-time predictions	Expose model via Azure Functions with auth	Azure Functions, REST API
Compliance Checks	Ensure regulatory compliance	Audit data for HIPAA/GDPR (e.g., anonymization in transit)	Azure Policy, Microsoft

2.2.11 Publish Dashboard

After testing, the Power BI dashboard is published online with role-based access and web availability.

Table 7: Dashboard Access and Validation Components

Component	Objective	Description	Technology / Standard
Stakeholder Access	Enable easy access	Make dashboard accessible via browser	Power BI Service
Secure Access Control	Ensure privacy	Restrict views by role (e.g., admin vs. viewer)	RBAC, Workspace permissions
Content Verification	Match desktop view	Check layout, labels, refresh after publishing	Web preview

2.3 Project Team

Our project team brings different skills in software development, machine learning, and data analysis, the project allows our team members to forecast the entire lifecycle of the dementia trend analysis through Microsoft Azure platform to strengthen healthcare stuff work efficiency.

By manipulating each member's professions and domain-specific expertise, our project would guarantee a well-rounded and effective task distribution that enhances the final product's clinical significance and technical strength.

Table 8: Team Members' Capabilities and Contributions

<table>
<tr><th>Name</th><th>Owned Skills</th><th>Relevant Project Contribution</th></tr>
<tr><td>JERRY
(Ying Kit, Li)</td><td>• System Development
• Data analysis by Power BI
• Web Development by Java, TypeScript
• Project Management</td><td>• Leads EDA and dashboard design by Power BI
• Facilitates decision-making based on data insights
• Runs project activities effectively</td></tr>
<tr><td>SHA SHA
(Shasha, You)</td><td>• System Management (Health Information Systems)
• Data Management & Training
• Network Maintenance
• Data analysis by Tableau / RapidMiner</td><td>• Handles data selection
• Organizes clinical information
• Validates the accuracy of risk factor identification and data integrity by Power BI</td></tr>
<tr><td>TIM
(Hiu Fung, Wong)</td><td>• System design and optimisation
• C/C++, Python
• Performance analytics on decision making
• ML model deployment</td><td rowspan="2">• Leads feature engineering, data clustering algorithms (e.g., K-Means)
• Creates efficient processing pipelines for models using Azure ML
• Builds training models to forecast health trends and recommend healthcare services
• Supports front-end dashboard or model UI design</td></tr>
<tr><td>OUYANG
(Yi, Ouyang)</td><td>• Backend Development
• Programming in Java, Python, Objective-C, and Swift
• SDK Development
• Project Management</td></tr>
<tr><td>IAN
(Chenkun, Ye)</td><td>• Full-stack software development
• API Design
• System Integration
• Programming in SQL / C++</td><td rowspan="2">• Develops backend infrastructure and database API
• Manages data pre-processing flow and logic
• Supports data ingestion to Azure Synapse Analytics
• Configures secure cloud deployment (if possible)</td></tr>
<tr><td>DONALD
(Ho, Ng)</td><td>• Full-stack software development
• Node.js, CI/CD pipelines
• Programming in JavaScript, TypeScript, and Python
• Data Management by MongoDB</td></tr>
</table>

2.4 Plan and Timeline

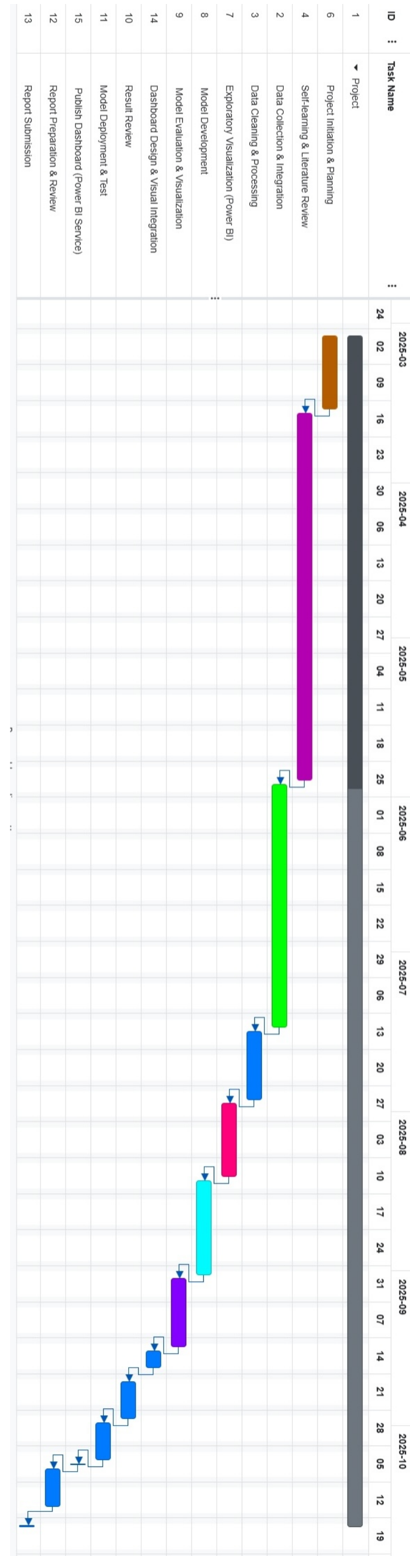

Figure 1: Gantt Chart of Project Timeline

Project Initiation & Planning (Approx. 7 days)
Define project objectives, scope, deliverables, timelines, and allocate resources to set the foundation for successful execution.

Self-learning & Literature Review (Approx. 70 days)
Gain domain-specific knowledge and review relevant academic and technical literature to inform project decisions and methodology.

Data Collection & Integration (Approx. 42 days)
Collect data from multiple sources and integrate them into a consolidated, analyzable format.

Data Cleaning & Processing (Approx. 14 days)
Handle missing values, inconsistencies, and noise to prepare high-quality data suitable for modeling.

Exploratory Visualization - Power BI (Approx. 14 days)
Use Power BI to visually explore data distributions, trends, and anomalies to inform later modeling stages.

Model Development (Approx. 14 days)
Build and train machine learning models using selected algorithms tailored to the dataset and project objectives.

Model Evaluation & Visualization (Approx. 14 days)
Evaluate the models using relevant metrics and visualize outcomes to understand model behavior.

Dashboard Design & Visual Integration (Approx. 10 days)
Design a user-friendly dashboard integrating key visual insights from both data and model outputs.

Result Review (Approx. 7 days)
Analyze the outcomes and validate whether project goals and expectations are met.

Model Deployment & Test (Approx. 7 days)
Deploy the final model into a test or production environment and validate its behavior in real-world conditions.

Publish Dashboard- Power BI Service (Approx. 7 days)
Publish the final Power BI dashboard to the online service for stakeholder access and monitoring.

Report Preparation & Review (Approx. 7 days)
Compile all processes and findings into a comprehensive report.

Report Submission (Approx. 2-3 days)
Submit the finalized report in accordance with project deadlines.

3 Outcomes

Outcomes of this project will include:

- A functional prototype of a cloud-based elderly health monitoring system that integrates machine learning models and interactive dashboards using Microsoft Azure Synapse, Azure ML, and Power BI. This system assists aged care facilities in detecting early signs of dementia, enabling more timely and targeted interventions.
- A set of interactive Power BI dashboards that visualize health trends (e.g., agitation, depression frequency, intervention history). These dashboards help nursing staff and care coordinators quickly interpret complex data, monitor changes over time, and prioritize high-risk residents for follow-up.
- Trained and explainable machine learning models that predict symptom deterioration or care needs based on historical patterns. These models support clinical decision-making by offering evidence-based recommendations, and they provide care managers with an early warning system to reduce hospital transfers and improve care planning.
- A real-time personalized monitoring interface deployed using Azure Static Web Apps and Azure Functions. This tool allows clinicians to enter resident data and immediately receive visual summaries of individual health status, empowering frontline staff to act on updated risk indicators during routine care.
- A complete technical report and documentation package detailing the system architecture, data preprocessing, modeling strategy, and visualization design. This report serves both academic assessment and as a foundation for future expansion or deployment in aged care environments.
- For team members, hands-on experience with cloud computing, machine learning, and health data analytics. The project also improved skills in cross-functional teamwork, agile planning, and stakeholder-focused design—valuable assets for future careers in AI and healthcare technology.

References

Anstey, K. J., Cherbuin, N., Herath, P. M., Qiu, C., Kuller, L. H., Lopez, O. L., Wilson, R. S. and Fratiglioni, L. (2014), 'A self-report risk index to predict occurrence of dementia in three independent cohorts of older adults: The anu-adri', PLoS ONE **9**, e86141.

Biswal, A. (2023), 'What is exploratory data analysis? steps and market analysis'. [Online; accessed 1-June-2025].
URL: *https://www.simplilearn.com/tutorials/data-analytics-tutorial/exploratory-data-analysis*

DataScienceSphere (2024), 'Feature transformation- part of feature engineering'. Medium, 21 November [Online; accessed 1-June-2025].
URL: *https://medium.com/@datasciencejourney100*$_{8}3560/feature - transformation - part - of - feature - engineering - dff2deaf59a2$

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., Lituiev, D., Copeland, T. P., Aboian, M. S., Mari Aparici, C. et al. (2019), 'A deep learning model to predict a diagnosis of alzheimer disease by using 18f-fdg pet of the brain', Radiology **290**(2), 456–464.

GeeksforGeeks (2025), 'What is exploratory data analysis?'. [Online; accessed 1-June-2025].
URL: *https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/*

HEAVY.AI (n.d.), 'What is feature selection? definition and faqs'. [Online; accessed 1-June-2025].
URL: *https://www.heavy.ai/technical-glossary/feature-selection*

Lee, S. J., Hulsman, M., van Spaendonk, R. M., van der Schaar, J., Dijkstra, J. I., van der Flier, W. M., Elting, M., Reinders, M. J., de Rojas, I., van Haelst, M. M., de Geus, C. et al. (2023), Data driven criteria for genetic testing result in an efficient selection of patients with genetic dementia in a clinical setting, in 'Alzheimer's Association International Conference', ALZ.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C. et al. (2020), 'Dementia prevention, intervention, and care: 2020 report of the lancet commission', The lancet **396**(10248), 413–446.

Ludlow, K., Westbrook, J., Jorgensen, M., Lind, K. E., Baysari, M. T., Gray, L. C., Day, R. O., Ratcliffe, J., Lord, S. R., Georgiou, A., Braithwaite, J., Raban, M. Z., Close, J., Beattie, E., Zheng, W. Y., Debono, D., Nguyen, A., Siette, J., Seaman, K., Miao, M., Root, J., Roffe, D., O'Toole, L., Carrasco, M., Thompson, A., Shaikh, J., Wong, J., Stanton, C. and Haddock, R. (2021), 'Co-designing a dashboard of predictive analytics and decision support to drive care quality and client outcomes in aged care: a mixed-method study protocol', BMJ Open **11**, e048657.

Mahalingam, R. and K., J. (2024), High dimensionality reduction induced predictive modelling to enhance performance in decision support system, in 'Futuristic Trends in Computing Technologies and Data Sciences Volume 3 Book 6', Vol. 3, IIP Series, pp. 107–121.

Miranda, R., Smets, T., Van Den Noortgate, N., Deliens, L. and Block, L. V. d. (2020), 'Higher prevalence of dementia but no improvement in total comfort while dying among nursing home residents with dementia between 2010 and 2015: Results from two retrospective epidemiological studies'.
URL: *https://www.preprints.org/manuscript/202012.0809/v1*

Organisation, W. H. (2025), 'Dementia'.
URL: *https://www.who.int/news-room/fact-sheets/detail/dementia*

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A. and Davatzikos, C. (2017), 'A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages', NeuroImage **155**, 530–548.

Singh, A. (2024), 'Feature scaling in machine learning'. [Online; accessed 1-June-2025].
URL: *https://www.appliedaicourse.com/blog/feature-scaling-in-machine-learning/*

Siva, G. (2022), 'Dimensionality reduction — feature selection and feature extraction'. Medium, 19 July [Online; accessed 1-June-2025].
URL: *https://gayathri-siva.medium.com/dimensionality-reduction-feature-selection-and-feature-extraction-28c732818e96*

Stamate, D., Musto, H., Ajnakina, O. and Stahl, D. (2022), 'Predicting risk of dementia with survival machine learning and statistical methods: Results on the english longitudinal study of ageing cohort', IFIP advances in information and communication technology pp. 436–447.

Veras, M. B., Mesquita, D. P., Mattos, C. L. and Gomes, J. P. (2020), 'A sparse linear regression model for incomplete datasets', Pattern Analysis and Applications **23**(3), 1293–1303.

Vogelsmeier, A. A., Halbesleben, J. R. and Scott-Cawiezell, J. R. (2008), 'Technology implementation and workarounds in the nursing home', Journal of the American Medical Informatics Association : JAMIA **15**, 114–119.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2274876/*