ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 4 No.2 (2024) 223-232
https://publisher.uthm.edu.my/periodicals/index.php/ekst

LASSO and Elastic Net in the Prediction of Dementia

Lim Phei Gee[1], Khuneswari Gopal Pillay[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA*

*Corresponding Author: khuneswari@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2024.04.02.026

Article Info

Received: 27 December 2023
Accepted: 11 January 2024
Available online: 12 December 2024

Keywords

Dementia, Major Neurocognitive Disorder, LASSO, Elastic Net, Logistic Regression, AICc, BIC

Abstract

Dementia, referred to in medical terms as 'major neurocognitive disorder,' is a prevalent condition impacting a substantial number of individuals in the United States, where approximately 5.5 million people currently grapple with the ailment. Despite its widespread occurrence and the profound consequences, it inflicts on individuals and their families, dementia remains underdiagnosed and poorly recognized. This study aims to identify the correlation between demographic factors and dementia, employing cross-tabulation and the chi-square test. Additionally, it aimed to compare the predictive capabilities of LASSO and Elastic Net regression models in forecasting dementia occurrences using AICc and BIC. Finally, the third objective was to identify the primary factor influencing dementia utilizing the most effective model. The data set implemented in this study is longitudinal dementia data obtained from MIT Press Direct web. The study's outcomes underscore the superior predictive performance of logistic LASSO regression in the context of dementia. This conclusion is substantiated by notably smaller AICc and BIC values, specifically -290.5912 and -243.7626, respectively, as compared to the corresponding metrics derived from Elastic Net regression. Furthermore, the research delineated that significant factors influencing dementia encompass gender, age, years of education, and socioeconomic status. These findings provide valuable insights for healthcare authorities and governments in both the United States and Malaysia, enabling them to allocate adequate medical resources and formulate guidelines tailored to the ageing population in specific districts of the country.

1. Introduction

Dementia is not a single disease that will develop in a person's body, it is an acquired syndrome brought on by several disorders. According to [1], dementia usually comes together with several commonly known neuro-related diseases such as Alzheimer's disease, Parkinson's disease, and vascular dementia. Significant memory loss is one of the dementia symptoms that affect people the most frequently, nonetheless. One may have trouble recalling recent activities and experience confusion with memories including their family members and own living skills. Most of the time, they require the assistance of their family member in completing everyday tasks or sequential activities and eventually lose the ability of decision-making causing them to fail to live on their own.

Presently, the World Health Organization (WHO) estimated that there are more than 55 million people worldwide are reported to have impacted by dementia, with over 60% of those people living in middle- or low-income countries. It is projected that 10 million new cases are reported annually [2]. Furthermore, it was predicted that this number would rise sharply to an estimated 131.5 million people by 2050 [3].

Nonetheless, dementia is underdiagnosed and unaware, despite its high prevalence and the severe impact it has on individuals and their families [4]. This may be due to people not knowing that demographic factors could cause dementia and the proper preventive action for getting dementia without a clear policy provided by the government and doctors. As of 2015, the estimated global expenditure for dementia was estimated to reach US$ 818 billion, equivalent to approximately 1.09% of the global GDP during that period [5]. This phenomenon suggests that medical expenditures are likely to increase substantially due to the growing demand for healthcare services. It is, therefore, critical for governments to prepare adequate funds to support the heavy medical expenses associated with dementia in the United States. Furthermore, statistical approaches more suited to medical data have yet to be extensively embraced, and too many researchers continue to utilize simple survival analysis techniques that are unsuitable for analysing or predicting complex hospital data with time-varying factors and numerous outcomes.

This study aimed to identify the correlation between demographic factors and dementia, employing cross-tabulation and the chi-square test. Additionally, it aimed to compare the predictive capabilities of LASSO and Elastic Net regression models in forecasting dementia occurrences using AICc and BIC. Finally, the third objective was to identify the primary factor influencing dementia utilizing the most effective model.

2. Materials

2.1 Data sources and data set

The dataset employed in this research was obtained from MIT Press Direct web (https://doi.org/10.1162/jocn.2009.21407). The dataset was published on December 01, 2010, and was freely accessible to the public [6]. The dataset includes both demented and non-demented individuals with right-handed (R) ages between 60 and 96. The dataset consists of 150 subjects, including both males and females, who underwent scanning sessions on at least two visits, with a minimum time interval of one year between sessions. Overall, there were 373 MRI imaging sessions. Table 1 shows details and descriptions of the data variables, which are generally demographics of the dementia patient.

Table 1 List of Variables of the Information of Dementia Patient

Variable	Variable Name	Description
y	Group	Nondemented and demented
x_1	MRI ID	Magnetic Resonance Imaging ID of the patient
x_2	Subject ID	Patient ID
x_3	Visit	Number of visits
x_4	MR Delay	Delay of the visit by a subject since the last visit (Number of days)
x_5	M/F	Gender (M= male, F= female)
x_6	Hand	Patient is right-handed
x_7	Age	Age at the time of image acquisition (years)
x_8	EDUC	Years of education
x_9	SES	Socioeconomic status (from 1 =highest status to 5 =lowest status)
x_{10}	MMSE	Mini-Mental State Examination score (from 0 = worst to 30 = best)
x_{11}	CDR	Clinical Dementia Rating (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD)
x_{12}	eTIV	Estimated total intracranial volume (cm^3)
x_{13}	nWBV	Normalized whole-brain volume (percentage)
x_{14}	ASF	Atlas scaling factor (unitless).

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 223-232

3. Methodology

3.1 Cross Tabulation

Understanding data patterns and making wise judgments require statistical analysis of relationships between variables. A fundamental method for analysing categorical data and finding connections between variables is cross-tabulation, also known as contingency table analysis. Cross-tabulation enables researchers to investigate trends, proportions, and associations within and across categories by arranging data into contingency tables. It is generally accepted to do cross-tabulation, specifically adopting a contingency table, to investigate the relationship between two binary variables [7].

3.2 Chi-square Test

The chi-square test is a non-parametric test, also known as the free distribution test [7]. By generating the test statistics based on the observed and estimated data and comparing them to the chi-square value, the chi-square test is a trustworthy and efficient technique for analysing categorical data and identifying significant associations between variables. According to [8], the chi-square test provides the evaluation of the existence of a relationship between the rows and columns inside a contingency table. So, before a chi-square distribution, cross-tabulation is usually done. The formula for the chi-square test is given in equation 1, whereby χ^2 represents the test statistic, E stands for expected value, and O stands for observed value.

$$\chi^2 = \frac{\sum (O-E)^2}{E} \tag{1}$$

3.3 *K*-Fold Cross-Validation

Accurately evaluating and selecting models is essential in statistical analysis to ensure reliable and robust results. One of the most popular techniques employed for achieving this is *k*-fold cross-validation (CV), which overcomes the overfitting of the data [9]. The training set is divided into *k* subsets of comparable dimensions during *k*-fold cross-validation. Cases from the training set are randomly chosen to create these subsets. The model is trained using *k*-1 subsets as the training set, while the remaining subset is utilized as the validation set. The performance of the model is then evaluated on the validation set [10]. In both LASSO and elastic net regression, *k*-fold cross-validation helps us find the best value for lambda. The model is trained *k* times once the data has been divided into k sets. Every time, *k*-1 parts are used for training, and the remaining parts are used for validation [10]. In this case, 10-fold cross-validation was used. Then, the model's performance can be accessed by binomial deviance.

3.4 Logistic Regression

A logistic regression model is a relationship built using the logit link function between the likelihood of a given event of interest and a linear combination of independent variables. The logit link function is defined as the natural logarithm of the odds ratio, whereby the odds ratio is the ratio of the probability of an event of interest occurring to the probability of it not occurring. The use of the logit link function in binary logistic regression has improved the interpretability of the results. This is due to its simpler nature compared to other link functions, allowing for a clearer comprehension of the relationship between the predictors and the outcome variable's log odds [11].

3.5 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO regression, known as the Least Absolute Shrinkage and Selection Operator, is a robust statistical method extensively employed in high-dimensional data analysis for variable selection and regularization. It aimed to minimize the sum of squared residuals while adding a restriction to the model parameters that "shrinks" the regression coefficients towards zero (L1 norm or L1 penalty) and subjects the total of the absolute values of the coefficients to a fixed constant. The nonzero variables will then be selected to remain in the model, and the rest will be eliminated [12][13]. The formula for LASSO is defined as equation 2.

$$\hat{\beta} = \arg\min \sum_{i=1}^{n} (y_i - \sum_{i=1} x_{ij}\beta_i)^2 + \lambda \sum_{j=1}^{p} |\beta_i| \tag{2}$$

where n stands for the number of observations, *p* stands for predictor variables, *t* stands for the turning parameter and $\lambda\sum_{j=1}^{p}|\beta_i|$ stands for the sum of the absolute coefficient penalty. In this case, the logistic LASSO regression "glmnet" package is implemented as it is one of the generalized linear models via penalized maximum likelihood [14].

3.6 Elastic Net

Although LASSO has outperformed in various situations, it still has limitations. Elastic net regression was proposed to further strengthen the statistical power of LASSO. It is known as the L1 penalty or L1 norm when the shrinkage is equal to zero. On the other hand, it is called the L2 penalty when the shrinkage is toward zero. Unlike LASSO, which only employs the L1 penalty, elastic net uses both penalties, combining the benefits and solving their limitations. Elastic Net performs automatic variable selection and continuous shrinkage simultaneously. However, what sets the elastic net apart is its additional capability of "grouped selection" [15]. An elastic net regression can be defined as equation 3.

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n}(y_i - \sum_{i=1} x_{ij}\beta_i)^2 + \lambda_1\sum_{j=1}^{p}|\beta_i| + \lambda_2\sum_{j=1}^{p}\beta_j^{\,2} \tag{3}$$

Similar to LASSO regression, *n* in elastic net stands for the number of observations, *p* stands for predictor variables, stands for turning parameter, $\lambda\sum_{j=1}^{p}|\beta_i|$ stands for the sum of the absolute coefficients (L1 penalty), and $\sum_{j=1}^{p}\beta_j^{\,2}$ stands for the sum of the square of the coefficient (L2 penalty).

3.7 Model Selection Criteria

3.7.1 Corrected Akaike's Information Criterion (AICc)

Although AIC is a helpful method in most cases, it has some obvious drawbacks. AIC does not explicitly account for sample size while the Akaike Information Criterion Corrected (AICc) proposed by [16] imposes stronger penalties than the original Akaike Information Criterion (AIC) for smaller sample sizes. So, instead of AIC, AICc is employed in our case for the model selection criterion. The equation of AICc can be written as equation 4.

$$AICc = -2\log L(\hat{\theta}) + 2k + \frac{2k(k+1)}{n-k-1} \tag{4}$$

where n refers to sample size, $\hat{\theta}$ denotes the maximum likelihood estimates of parameters, and the number of estimated parameters (regression coefficient) in the model is denoted by *k* in the database. Note that $-2\log(\theta)+2k$ is equivalent to AIC. The penalty term $\frac{2k(k+1)}{n-k-1}$ in AICc's equation adjusts the limited data and helps decrease any potential bias during the model selection process. The AICc provides a more reliable model selection criterion by integrating the modification of AIC. The fundamental principle underlying the AICc is that models with the smallest value indicate a better-fitting model [17]. Note that the difference between AIC and AICc lies in how they handle the penalty whereby AICc consider several parameters and sample size while AIC only considers several parameters.

3.7.2 Bayesian Information Criterion (BIC)

To ensure consistent comparisons with the AICc, another frequently utilized criterion called the Bayesian Information Criterion (BIC) is employed in this study for selecting the best model. The BIC, similar to the AICc, considers both the sample size and the number of parameters during the model selection process. Schwarz established the BIC, often known as the Schwarz criteria, in 1978 based on Bayesian theory. With given available data, the BIC seeks to maximize the posterior probability of a model. The equation of BIC can be written in equation 5.

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 223-232

$$BIC = -2\log L(\hat{\theta}) + k\log n \tag{5}$$

where the log-likelihood of the model given the data is denoted as $-2\log L\left(\hat{\theta}\right)$, the number of estimated parameters in the model is denoted by *k*, and n denotes the number of entries in the database. Like AICc, BIC serves as a cost function that must be minimized. In essence, the best model will be the one with the lowest BIC value [18].

3.7.3 Goodness of Fit Test

The goodness-of-fit test serves the purpose of evaluating the extent to which the selected regression model aligns with the observed data. It aids in determining whether the model effectively accounts for the variability in the dependent variable or if disparities exist between the model's predictions and the observed data. In summary, it provides a comprehensive assessment of the regression model's performance by comparing actual values to the predicted ones.

3.7.4 Deviance Residual Plot

Deviance residual plots aid in the detection of misfit patterns, outliers, and violations of model assumptions. Researchers can acquire insights about the model's fit by evaluating the distribution and trends of the deviance residuals. In the ideal scenario, well-fitted models would have deviance residuals that are randomly distributed around zero, indicating that the observed and projected responses are well aligned. Systematic patterns or trends in the residuals, on the other hand, indicate model misspecification or shortcomings. The deviance residual plot facilitates the detection of such patterns as well as the evaluation of the model's performance. Deviance residuals are calculated using the deviance, which quantifies the difference between the saturated model's log-likelihood and the fitted model's log-likelihood. The formula of deviance residuals is given in equation 6.

$$D(y,\hat{\mu}) = 2\sum_{i=1}^{n}\left\{\log[\,p(y_i\left|\hat{\theta}_s\right|] - \log[\,p(y_i[\hat{\theta}]\right\} \tag{6}$$

where represents the log-likelihood function for the saturated model, $\hat{\theta}_s$ denotes the set of the parameter estimates for the saturated model, represents the log-likelihood function of the fitted model and $\hat{\theta}$ indicates the set of parameter estimates for the fitted model. A saturated model, by definition, provides the best fit to the data and has the highest log-likelihood of any model [19].

3.7.5 Deviance P-value

Setting the degree of freedom of the null model to one signifies that only a single parameter, specifically the intercept, was incorporated into the null model. This simplification makes the null model the simplest representation of the null hypothesis model. In contrast, the best LASSO model obtained includes more than five predictor variables, rendering it a complex model. By comparing the best LASSO model with the null model, the evaluation of whether the additional parameters (predictor variables) in the LASSO models significantly enhance the model's fit compared to the null model can be accessed [20]. Also, the lowest the deviance, the better the model fits. The deviance p-value is obtained from the comparison of these two models. A deviance *p*-value that is lower than 0.05 indicates that the predictor variables in the model indeed improve the model's fit.

3.8 Results and Discussion

3.8.1 Data Preparation

Data preparation is a systematic process aimed at enhancing a dataset by cleaning, refining, and transforming the raw data into a more refined and immediately usable format, thereby facilitating subsequent data analysis. Within my dataset, which initially comprised 15 variables, the primary step involved thorough data cleaning. During this phase, variables such as Subject ID, MRI ID, number of visits, and hand were excluded from the dataset due to their lack of relevance to the specific analysis. As a result, only 11 variables remained in the dataset.

Then, the demographic factors such as age and education years were categorized into several groups to ease the process during analysis. Continuous age data was grouped into 8 groups and education years were divided into 3 main groups. Subsequently, a comprehensive missing data detection procedure was conducted, revealing that only the variables SES and MMSE contained missing values, totalling 19 and 2 instances, respectively. To

address this issue, an imputation process was applied. Given the nature of the variables, SES, an ordinal data type, was imputed using the mode of SES, while MMSE, characterized as continuous data, was imputed using the mean of MMSE.

3.9 Cross Tabulation and Chi-Square Test

To explore the relationship between demographic variables and dementia, cross-tabulation and chi-square tests were conducted with the following factors: age group, gender, educational level (EDUC), and Social Economic Status (SES). The resulting table generated by R under the *sj*plot package provided insights into the association between demographic factors and dementia. A *p*-value of each combination was calculated and those with *p*-value less than 0.05 were considered significant to the model. Table 2 shows the *p*-value of each variable. The result indicates that demographic variables like "Gender" are statistically significant in affecting dementia while the rest are not.

Table 2 The p-value obtained by the Chi-square Test.

Variable	*p*-value
Gender	0.016
Age Group	0.682
Socioeconomic Status	0.862
Education Years	0.085

3.10 *K*-Fold Cross-Validation

K-fold cross-validation is a widely adopted approach in the realm of statistical modelling to assess how well a predictive model performs and its capacity to generalize. This method involves iteratively partitioning a dataset into distinct training and testing subsets, which allows for a comprehensive evaluation of the model's performance. In this specific analysis, *k*-fold cross-validation was chosen to perform this process ten times, emphasizing its significance in the evaluation, which mean this is a 10-fold cross-validation analysis. Furthermore, it was also applied in the context of logistic LASSO and Elastic Net regression to gauge the models' capability to extend their predictive power to unseen data. Additionally, this approach aids in the determination of optimal hyperparameters, including the critical regularization strength parameter (lambda).

3.11 Regression Model

Logistic LASSO regression aimed to minimize the sum of squared residuals while adding a restriction to the model parameters that "shrinks" the regression coefficients towards zero (L1 norm or L1 penalty) and subjects the total of the absolute values of the coefficients to a fixed constant. The nonzero variables will then be selected to remain in the model, and the rest will be eliminated.

Fig.1 (a) depicts the LASSO regression regularization path, with the LASSO regression coefficients shown against various lambda values. Notice how the numbers on the figure change as the lambda value increases; these numbers represent the number of predictors considered for each unique lambda value. The graph indicates that as lambda values increase, the model retains fewer variables. This is because LASSO regression effectively shrinks irrelevant predictors to zero, causing them to be excluded from the model. In this instance, the optimal lambda was identified as 0.00707349. It is anticipated that this lambda value will yield the most accurate model.

Elastic Net simplifies the selection of variable groups while simultaneously minimizing dependency and redundancy by steering their coefficients toward zero. Elastic Net's regularization technique has a special significance in the automated identification and selection of variable groups, hence building a robust framework for linear regression. Elastic Net's innovation lies in the introduction of a combined L1 and L2 penalty strategy in the realm of linear regression, which enhances its versatility and applicability.

The process of the regularization path was visualised in Fig. 1 (b) for better understanding. Each curve in the plot represents a separate variable, and the *y*-axis shows the value of the coefficient for that variable. Particularly, in the context of Elastic Net regression, the optimal lambda value for meeting this condition is 0.004523921. Furthermore, determining the appropriate alpha value, which was determined at 0.512, is an equally vital component of this investigation. Combining these optimal alpha and lambda values, the regression model is expected to provide the highest level of predictive accuracy.

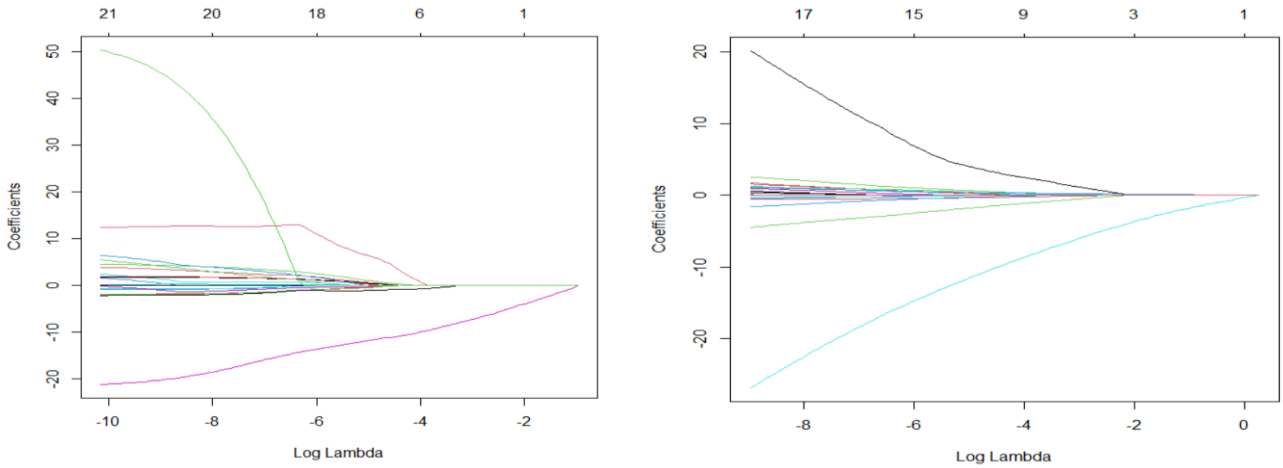
Upon a thorough examination of LASSO and Elastic Net regressions, the coefficient of each variable was recorded in Table 3. Both the LASSO and Elastic Net regression has diligently performed variable selection. The variables marked with an asterisk ('-') have undergone a rigorous shrinkage process and, as a result, have been considered eliminated or excluded from the final model. This elimination signifies that these variables did not contribute significantly to the model's predictive capacity and were therefore reduced to negligible coefficients.

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 223-232

Table 3 Regression's Coefficient of LASSO and Elastic Net

Predictor	LASSO	Elastic Net
Intercept	5.6458	2.5649
Magnetic Resonance Delay (MR. Delay)	0.0008	0.0009629
Gender (Male/Female)	-0.2307	-0.5136
Age Group (<70)	-0.3879	-0.6367
Age Group (<90)	-0.9988	-1.08321
Age Group (<95)	0.6342	1.1067
Education Level (<20 years)	-	0.2032
Education Level (20 years and above)	-0.1212	-0.2549
Socioeconomic Status (SES -2)	0.4948	0.8585
Socioeconomic Status (SES -3)	0.2977	0.7012
Socioeconomic Status (SES -4)	1.0448	1.7318
Socioeconomic Status (SES -5)	-	0.7171
MMSE	0.09199	0.1582
CDR	-11.6678	-12.4588
eTIV	0.002313	0.003482
nWBV	6.1940	7.9012

In the aftermath of this meticulous shrinkage process, a selected group of variables has emerged as the key contributors to the model's predictive power. Two of the regression model retained almost the same variables which include "Magnetic Resonance Delay (MR. Delay)", "Gender", patients aged below 70, patients aged below 90, individuals with an education level of 20 years and above, "socioeconomic status rank 2 to 4", "MMSE", "CDR", as well as the variables "eTIV" and "nWBV". Note that the Elastic Net regression model retain the variable "Socioeconomic Status (SES-5)" and "Education Level (<20 years)" in the model as well. These retained variables are integral to the final model and have proven to be pivotal in explaining the observed outcomes, emphasizing their substantial role in the regression framework. On contrary, those excluded variables like "Socioeconomic Status (SES -1)", "Age Group (<75)", "Age Group (<80)", "Age Group (<85)", "Age Group (95 and above)" and "Education Level (<15 years)" had a noble effect on developing dementia.



(a) **(b)**

Fig. 1 (a) Regularization path of LASSO Regression; (b). Regularization path of Elastic Net Regression

The final model of logistic LASSO and Elastic Net regression were represented in Eq. 7 and Eq. 8 respectively.

$$P(Group\_Demented) = \frac{\exp\left(\begin{array}{l}5.6428-0.0008MR.Delay-0.2307Gender-0.3879AgeGroup(<70)-0.9988AgeGroup(<90)+\\0.6342AgeGroup(<95)-0.1212EDUC(\geq 20)+0.4948SES(2)+0.2977SES(3)+1.0448SES(4)+\\0.09199MMSE-11.6678CDR+0.002313eTIV+6.1940nWBV\end{array}\right)}{1+\exp\left(\begin{array}{l}5.6428-0.0008MR.Delay-0.2307Gender-0.3879AgeGroup(<70)-0.9988AgeGroup(<90)+\\0.6342AgeGroup(<95)-0.1212EDUC(\geq 20)+0.4948SES(2)+0.2977SES(3)+1.0448SES(4)+\\0.09199MMSE-11.6678CDR+0.002313eTIV+6.1940nWBV\end{array}\right)} \quad (7)$$

(8)

$$P(Group\_Demented) = \frac{\exp\left(\begin{array}{l}2.5649-0.0009629MR.Delay-0.5136Gender-0.6367AgeGroup(<70)-01.08321AgeGroup(<90)+\\1.1067AgeGroup(<95)+0.2032EDUC(<20)-0.2549EDUC(\geq 20)+0.8585SES(2)+0.7012SES(3)+\\1.7318SES(4)+0.7171SES(5)+0.1582MMSE-12.4588CDR+0.003482eTIV+7.9012nWBV\end{array}\right)}{1+\exp\left(\begin{array}{l}2.5649-0.0009629MR.Delay-0.5136Gender-0.6367AgeGroup(<70)-01.08321AgeGroup(<90)+\\1.1067AgeGroup(<95)+0.2032EDUC(<20)-0.2549EDUC(\geq 20)+0.8585SES(2)+0.7012SES(3)+\\1.7318SES(4)+0.7171SES(5)+0.1582MMSE-12.4588CDR+0.003482eTIV+7.9012nWBV\end{array}\right)}$$

3.12 Comparison of Logistic LASSO Regression and Logistic Elastic Net Regression

A comparison of each best model was conducted to select the best-fit model that can represent dementia by comparing both models in terms of AICc and BIC. A model with the lowest value of AICc and BIC will be chosen as the best model. According to Table 4, it is obvious that LASSO regression achieved a lower AICc value of -290.5912, surpassing Elastic Net, which obtained -44.8801. Similarly, in the context of BIC, LASSO also outperformed Elastic Net with a smaller value of -243.7686, in contrast to Elastic Net's -33.8601. This suggests that LASSO regression may be a more suitable model in this study, as it demonstrates superior goodness-of-fit based on the AICc and BIC criteria.

Table 4 Comparison of LASSO and Elastic Net in terms of AICc and BIC

Regression Models	AICc	BIC
LASSO	−290.5912	−243.7626
Elastic Net	−44.8801	−33.8601

3.13 Better Model

After conducting a comprehensive analysis and comparing the best models between LASSO regression and Elastic Net regression, the most suitable model for representing dementia has been selected and shown in equation 7.

The coefficient's sign provides insight into the direction of the effect, with a positive sign indicating a positive correlation and a negative sign indicating a negative correlation. In addition, the larger the magnitude of the coefficient, the stronger the relationship it holds toward dementia.

The better model suggests that an uptick in Magnetic Resonance Delay (MR. Delay) is linked to a marginal rise in dementia likelihood, specifically an increase of 0.0008. In addition, being male (Gender-0) is associated with a reduction in dementia probability when compared to being female (Gender-1) by 0.2307. Individuals in age groups below 70 (-0.3879) and 90 (-0.9988) exhibit a negative association with dementia, implying that younger people face a lower risk of developing dementia. People aged between 65 to 69 and 85 to 89 will experience a decrease in their dementia risk by 0.3879 and 0.9988, respectively. However, the age group below 95 (0.6342) demonstrates a positive coefficient, suggesting an elevated risk of dementia, indicating that this age group may be more vulnerable to dementia.

Moreover, having an educational background of 20 years or more is linked to higher dementia risk, with an increase of 0.1212. This could imply a correlation between increased years of education and an elevated risk of dementia. Additionally, dementia can be influenced by various levels of socioeconomic status (SES). Individuals in the lower level of socioeconomic status (SES-4) are associated with a 1.0448 unit increase in the odds of developing dementia, followed by those in the moderate level of socioeconomic status (SES-3), who have a 0.2977 unit increase in dementia risk. Higher socioeconomic status (SES-2) is also linked to a unit rise in dementia risk by 0.4948.

Furthermore, a 0.09199 increase in the Mini-Mental State Examination (MMSE) score is tied to a one-unit increase in the risk of developing dementia. In contrast to the MMSE, an increase in the Clinical Dementia Rating (CDR) results in a substantial reduction in dementia odds, with a decrease of 11.6678. Estimated total intracranial volume (eTIV) has a slight impact on dementia, where a 0.002313 increase in eTIV corresponds to a unit increase in dementia risk. Similarly, normalized whole brain volume (nWBV) has a strong positive influence on dementia risk.

3.14 Goodness of Fit Test

3.14.1 Deviance Residual Plot

A deviance residual plot is a scatterplot where each data point represents a deviance residual for a particular observation. These residuals are typically plotted against the predicted probabilities. Based on Fig. 2, it was shown that the two lines were plotted similarly with zero slope and intercept. Meaning that, there was no significant model inadequacy and presence of outliers within the range of predictor variables. Therefore, it can be concluded that the best LASSO model obtained is a good fit model.

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 223-232

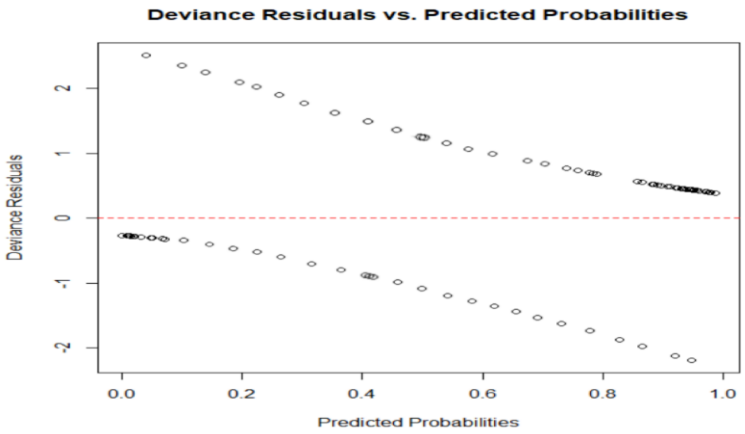


Fig. 2 Deviance Residual Plot of Better Model

3.14.2 Deviance P-value

As mentioned in the previous section, the lowest the deviance, the better the model fits. Table 5 illustrates that the residual deviance outperformed null deviance with a coefficient of 96.55 to 102.5317, showing the LASSO model as having a better fit than the null model. This result can be supported by the deviance *p*-value, obtained from the comparison of these two models, which is 0.03305, which is lower than the significance level of 0.05. Consequently, this indicated that the predictor variables retained in the LASSO regression model are reliable in explaining dementia occurrence in this study and indeed improve the model's fit.

Table 5 Comparison of Deviance and Null Deviance Residuals

Deviance Model	Value of Residuals
Residual Deviance	96.55
Null Deviance	102.5317

4. Conclusion

Overall, the attainment of the first objective occurred through the execution of cross-tabulation and chi-square testing. This analysis unveiled a robust correlation between gender and the incidence of dementia, while other demographic variables were determined to have an insignificant relationship with the occurrence of dementia. Moreover, the second objective focused on comparing the performance of logistic LASSO and Elastic Net regression, was accomplished through the implementation of AICc and BIC. The results unequivocally demonstrated that logistic LASSO regression outperformed Elastic Net regression in predicting dementia. This may be attributed to the dataset's low multicollinearity, a characteristic effectively managed by LASSO regression. Finally, the third objective was achieved by constructing a logistic LASSO model, revealing that demographic factors such as gender, age, years of education, and socioeconomic status significantly contribute to the likelihood of developing dementia. Notably, females, individuals aged 95 and above, those with 20 or more years of education, and those with a socioeconomic status of 2 to 4 were identified as having a higher chance of developing dementia.

Despite the geographical disparity between the location of data collection and the target region (Malaysia), this study can serve as a valuable reference for the Malaysian government and healthcare authorities. While recognizing potential variations in insights, the study can inform related authorities about the allocation of sufficient medical resources and guidelines for the potential ageing population in specific districts of Malaysia. This underscores the study's potential utility in healthcare planning, even with the acknowledged difference in the geographic origin of the data. Another limitation is that the sample size of this research was considered small. Future research may conduct dementia prediction analysis in a larger dataset to increase the liability and robustness of the analysis. In addition, this study only compared two methods which are logistic LASSO and Elastic Net regression for the prediction of dementia. Interested researchers may conduct comparative studies with other advanced machine learning techniques to benchmark the performance of LASSO and Elastic Net regression models.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author Contribution

The authors confirm their contribution to the paper as follows: ***study conception and design****: Lim Phei Gee, Khuneswari A/P Gopal Pillay;* ***data collection****: Lim Phei Gee;* ***analysis and interpretation of results****: Lim Phei Gee, Khuneswari A/P Gopal Pillay;* ***draft manuscript preparation****: Lim Phei Gee, Khuneswari A/P Gopal Pillay. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Gale, S. A., Acar, D., & Daffner, K. R. (2018). Dementia. The American Journal of Medicine, 131(10), 1161–1169.
[2] World Health Organization. (2023). Dementia. World Health Organization. Retrieved on 10th April, 2023 https://www.who.int/news-room/fact-sheets/detail/dementia
[3] World Health Organization. (2021). World failing to address dementia challenge. Retrieved on 10th April, 2023 https://www.who.int/news/item/02-09-2021-world-failing-to-address-dementia-challenge
[4] Amjad, H., Roth, D. L., Sheehan, O. C., Lyketsos, C. G., Wolff, J. L., & Samus, Q. M. (2018). Underdiagnosis of Dementia: An Observational Study of Patterns in Diagnosis and Awareness in US Older Adults. Journal of General Internal Medicine, 33(7), 1131–1138.
[5] Martin James Prince. (2015). World Alzheimer Report 2015.
[6] Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2010). Open Access Series of Imaging Studies (OASIS): Longitudinal MRI Data in Nondemented and Demented Older Adults. Journal of Cognitive Neuroscience, 22(12), 2677–2684.
[7] Momeni, A., Pincus, M., & Libien, J. (2018). Introduction to Statistical Methods in Pathology. Springer.
[8] Singhal, R., & Rana, R. (2015). Chi-square Test and Its Application in Hypothesis Testing. *Journal of the Practice of Cardiovascular Sciences, 1*(1), 69.
[9] Jung, Y. (2017). Multiple Predicting k-Fold Cross-Validation for Model Selection. *Journal of Nonparametric Statistics, 30*(1), 197–215.
[10] Berrar, D. (2019). Cross-Validation. Encyclopedia of Bioinformatics and Computational Biology, 1, 542–545.
[11] Srimaneekarn, N., Hayter, A., Liu, W., & Tantipoj, C. (2022). Binary Response Analysis Using Logistic Regression in Dentistry. *International Journal of Dentistry, 2022*, 1–7.
[12] Ranstam, J., & Cook, J. A. (2018). LASSO Regression. British Journal of Surgery, 105(10), 1348–1348.
[13] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267–288. Gebken, B., Bieker, K., & Peitz, S. (2022). On the structure of regularization paths for piecewise differentiable regularization terms. Journal of Global Optimization, 85(3), 709–741.
[14] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements Of Statistical Learning, Second Edition: Data Mining, Inference, And Prediction (2nd ed.). Springer.
[15] Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via The Elastic Net. Journal of The Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.
[16] Hurvich, C. M., & Tsai, C.-L. (1993). A Corrected Akaike Information Criterion For Vector Autoregressive Model Selection. Journal of Time Series Analysis, 14(3), 271–279.
[17] Wong, C. K. (2019). Minimum message length inference with application to genome-wide association studies data. Minerva-Access.unimelb.edu.au. Retrieved on 6th May, 2023 from http://hdl.handle.net/11343/223009
[18] Rossi, R., Murari, A., Gaudio, P., & Gelfusa, M. (2020). Upgrading Model Selection Criteria with Goodness of Fit Tests for Practical Applications. Entropy, 22(4), 447.
[19] Li, X., & Guo, Y. (2020). Model Diagnostics for Generalized Linear Models with Applications to Longitudinal Data. Journal of Applied Statistics, 47(9), 1657-1673.
[20] Tupper, L. (2021). 7.5 Deviance and Residuals | Stat 340 Notes: Fall 2021. In bookdown.org. Retrieved on 3 November 2023 https://bookdown.org/ltupper/340f21_notes/deviance-and-residuals.html