# Socio-demographic, Lifestyle and Neuropsychological Risk Factors on Alzheimer's Disease

# Socio-demographic, Lifestyle and Neuropsychological Risk Factors on Alzheimer's Disease

Tahera Ahmed
*Bond Business School*
*Bond University*
Gold Coast, Australia
tahera.ahmed@student.bond.edu.au

Ping Zhang*
*Menzies Health Institute Queensland*
*Griffith University*
Gold Coast, Australia
p.zhang@griffith.edu.au

Kuldeep Kumar
*Bond Business School*
*Bond University*
Gold Coast, Australia
kkumar@bond.edu.au

***Abstract*—Background: Alzheimer's is a common neurodegenerative disease that predominantly affects the elderly. However, the leading causes for the development of Alzheimer's disease (AD) are yet to be identified, and early detection and disease progression intervention are studied to help slow down its deterioration. Recognizing the possible threat of AD in individuals before neurodegenerative changes start to take effect may contribute to the discovery of preventive actions. This study aims to identify socio-demographic, lifestyle, and neuropsychological factors that may contribute to AD development.**

**Results: This study performed the Pearson's chi-squared test for categorical and one-way ANOVA for continuous variables to analyse the association of risk factors for developing AD from two separate cohorts. The patients in the first cohort have no record of the AD development period. In the second cohort, all the AD patients developed dementia within 36 months. Marital status, occupation, APOE4 genotype, Geriatric Depression Scale, and Functional Questionnaire Assessment score were significantly different between normal control and AD patients from both cohorts. The prediction models developed with either cohort showed high performance in ROC (receiver operating characteristic) measurements.**

**Conclusion: This study investigated the effects of socio-demographic and neuropsychological risk factors in AD development from two different cohorts. The significant risk factors from both cohorts can contribute to developing an Alzheimer's risk app that can be potentially reliable and easily accessible.**

***Keywords— Alzheimer's disease, Risk Factors, Logistic Regression, ADNI Database***

## I. INTRODUCTION

Dementia is perceived as a fatal age-related disorder that invariably results in a loss of independence and declining quality of life [1]. The ever-increasing number of dementia-affected people globally is expected to significantly impact the healthcare system and society as a burden in the near future [2]. The most widely recognized cause of dementia is Alzheimer's disease (AD). Researchers have indicated a 50-80 percent prevalence of AD in all dementia victims and sufferers and a high mortality rate in AD patients[3]. It is estimated that 1.6 million or roughly 43 percent of older adult deaths will occur from AD by 2050 [4].

Alzheimer's is usually diagnosed very late in the disease's deterioration process. By the time AD diagnosis is made, substantial brain tissue damage has already occurred, and there is a noticeable cognitive deficit and other symptoms as a result [5].

By identifying the people at the greatest risk of memory decline, there is potential for holding clinical trials of AD in an effort to make treatments more cost-effective and valid. A better- selection of subjects is important as treatments are likely to have the most significant impact when provided at the earliest possible stage of the disease process [6]. With new emerging treatment approaches, it is becoming increasingly important to develop ways to identify subjects at a real risk of developing AD later on [7].

Like other common chronic diseases, AD develops due to multiple factors rather than a single cause. It is of great relevance to AD prevention that an estimated 54% of late-onset cases are attributed to modifiable risk factors [8]. The most significant risk factors for late-onset of the disease are age (being 65 or older), having a family history of AD, and carrying the APOE e4 gene. A 25% reduction in just seven modifiable risk factors of AD, namely mid-life hypertension, obesity, diabetes, smoking, depression, low educational status, and a sedentary lifestyle, could reduce the global prevalence by 3 million cases [8][9].

Although there have been many studies on Alzheimer's risk factor identification, there is still a shortage of comprehensive studies and investigations on modifiable risk factors and how they may affect the development of Alzheimer's in individuals diagnosed with AD during a specific period. Besides, not enough checking has been done on whether the same risk factors had a similar effect on the cohort suffering from Alzheimer's for a long time and had no recorded specific timeframe of AD development.

The primary purpose of this study is to investigate the socio-demographic, lifestyle, and neuropsychological risk factors associated with the development of AD from two different cohorts and determine if certain variables have a more significant effect on the elderly population and make them more likely to suffer from AD. While in one cohort, the progression period to AD development is known, the actual period of AD development in the other is not known.

## II. METHOD

### *A. Study design and Sample*

* Corresponding Author

Data for this study has been sourced from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is a multi-cohort longitudinal study where cognitively healthy volunteers or those with various degrees of cognitive impairment have been tracked and followed since 2005. The relevant data were collected in four phases - ADNI 1, ADNI Go, ADNI 2, and ADNI 3 [10].

The participants enrolled for ADNI belonged to the 55-90 age group. Recruitment was made from 57 sites in the USA and Canada at baseline and multiple following up time points. Each participant was diagnosed with AD, mild cognitive impairment (MCI), or normal control (NC) at the corresponding time points.[11].

For this study, two separate cohorts were extracted from the ANDI datasets for analysis. One cohort (Cohort1) was selected from the baseline where the participants were either healthy or had been clinically diagnosed with Alzheimer's before or at the baseline data collection time. However, the actual period of how long their disease has been developed was unknown. In another cohort (Cohort2), only NC and the participants who had developed AD within the last 36-month (diagnosed as normal or MCI 36 months before) were included. The first cohort included 484 NC and 341 AD participants, and the second cohort consisted of 220 NC and 171 AD cases. In this article, we have used 0 to represent NC and 1 for AD.

### *B. Variable selection*

Participants' demographic information (age, sex, marital status, education, occupation), vital signs (body weight, systolic blood pressure, diastolic blood pressure, pulse rate), neuropsychological Functional Questionnaire Assessment score (FAQTOTAL), and Geriatric Depression Scale (GDTOTAL) were analyzed and used for building multivariate models. The values of these variables used for analysis and modelling are described below:

**Age category**: Participants' age was classified into four groups in a ten-year interval. The age groups were 55-64 years, 65-74 years, 75 to 84 years, and 85 years or more.

**Sex:** The sex of the participants was Male or Female only

**Participant's Education Level:** In both study datasets, years (year 1 to year 20) of each case study's education was given as a continuous variable. Subsequently, the education category was formed for the groups according to the Australian Qualifications Framework (AQF) as Primary (Year 1-6), Secondary (Year 7-12), Bachelor or Diploma (Year 13-16), and Masters or Higher Degree.

**Marital Status:** The participant's marital status was divided into four categories - married, widowed, divorced, and never married.

**Occupation Category:** In the ADNI study, participants' occupation history had an open-ended response. Therefore, the analysis followed the 'Australian Standard Classification of Occupations' [12] and classified the relevant primary occupation response. For the analysis purpose, the occupation groups were divided into four (04) categories – Managers or Administrators, Professionals, skilled occupations, and unskilled occupations. Participants who worked as associates or clerical, sales, or technician jobs, were classified as skilled occupations. Homemakers, laborers, and transport workers were classified as unskilled occupations.

**APOE4:** Presence of APOE4 gene in the individual's body where 0 means zero APOE4 allele, 1 means presence of one APOE4 allele, and 2 means presence of two APOE4 alleles.

**Body Weight:** This variable contained the bodyweight of the participants in kilogram (kg).

Systolic Blood Pressure: This variable means systolic blood pressure reading per mmHg

**Diastolic Blood Pressure:** This is the diastolic blood pressure reading of the study population per mmHg.

**Pulse rate:** This variable contains the per minutes pulse rate of the individuals.

**Geriatric Depression Scale (GDTOTAL):** It is a self-report measure of depression in older adults, ranging from 0 to 15. Total scores of 0-5 are considered normal, and 6-15 are considered a depressed state [13].

**Functional Questionnaire Assessment score (FAQTOTAL):** A daily functional activities total score ranges from 0-30. A 0 score is regarded as 'no impairment' and 30 as 'severely impaired [14]

It should be noted that for Cohort1, all the participant records are from the baseline time point. For Cohort2, sex, education level, marital status, and primary occupation were collected at baseline, and the rest (such as weight and blood pressure) were recorded at both baseline and at the 36-month follow-up time point.

### *C. Statistical analysis*

Continuous variables (body weight, systolic and diastolic blood pressure, pulse rate, GDTOTAL, FAQTOTAL) were analysed using one-way variance analysis (ANOVA). These measures were expressed as mean ± standard deviation. A chi-squared test was used to compare the difference between NC and AD groups. Differences in measures were considered significant if the p-value was less than 0.05.

Binary logistic regression was used with each cohort to evaluate the association between AD development and each risk factor. Results were presented as unadjusted and adjusted odds ratios (OR) and 95% CI. As age, sex, and APOE4 genotype have been commonly considered and frequently reported in the earlier literatures[15] are associated with AD but not modifiable, they were used as the covariates for adjustment.

The performance of the multivariate logistic regression models for predicting AD or predicting AD development within the next three years was assessed using ROC (receiver operating characteristic) analysis. To evaluate the predictability of each regression model with each cohort, 5-fold cross-validations were performed.

The entire statistical analyses were performed using R-studio software for windows.

## III. RESULTS:

The characteristics of NC and AD patients from two different cohorts are reported on TABLE I.

The number of NC and AD distributions across each age group in the collected cohorts are slightly different and not significant for Cohort2. Sex, marital status, occupation category, APOE4 genotype, GDTOTAL score, and FAQTOTAL score were found significantly different between NC and AD groups from both cohorts at 95% confidence level. Age, Education level, and participants' weight were found different from only Cohort1. Among the vital signs, only pulse rate was significantly different between AD and NC from cohort2.

TABLE Ⅱ shows the results from logistic regression models built for each variable with and without adjusting variables. Odds ratios between the AD and NC adjusted for age, sex, and APOE4 genotype are also reported in Table II.

Participants who completed Secondary, Bachelor's or Diploma, and Master's or Higher Degree levels of education showed a slightly lower risk of being AD in both cohorts, but not statistically significant. Widowed and never married showed a lower AD ratio than the married population only in Cohort1 before and after adjustment on age, sex, and APOE4 genotype. No significant difference across marital status was found from Cohort2. No significant difference was shown between divorced and married groups. Both skilled and unskilled occupations showed higher AD ratios in Cohort 1. The skilled occupation had a higher ratio of AD than the managers or administrators group in only Cohort2. There is no significant difference between the professionals' group and the managers or administrators group from either cohort before and after adjustment.

The neuropsychological assessments were highly associated with AD as expected from both Cohort1 and Cohort2. The OR of GDTOTAL was 1.91 (95% CI: 1.65 to 2.20) and 1.43 (95% CI: 1.26 to 1.63) respectively from Cohort1 and Cohort2, after adjustment for the control variables. These OR numbers of FAQTOTAL are 5.24 (95% CI: 3.47 to 7.92) and 2.22 (95% CI, 1.82 to 2.73). Vital signs and body weight showed a mild effect on their association with AD

TABLE Ⅲ reports the results from multivariate logistic regression models built with the data from Cohort1 or Cohort2. In the model1, the multivariate logistic regression model was built with all the listed variables in this article. Only significant variables from TABLE I were selected in model2 and model3, the variables were selected by a stepwise logistic regression by taking all the socio-demographic, health, and neuropsychological variables mentioned in TABLE-I. Coefficients of the variables from each model were determined from the logistic regressions, also known as log(OR).

A ROC curve method was used for measuring the models' overall performance. Area Under the Curve (AUC) was commonly used to compare the performance of prediction models. The values of AUC = 0.5 indicate random guessing, AUC = 1 indicates perfect prediction, AUC = 0 is also perfect prediction but with class label switched. Furthermore, the practical implications for the majority of classification systems are that values of AUC>0.9 indicate excellent prediction models, AUC>0.8 are good predictions, while AUC<0.7 indicates poor predictions. To test the predictability of the models built with different sets of variables using the data from Cohort1 and Cohort2. Five-fold cross-validation was performed on models built with each set of variables.

The models' performances were surprisingly good with the AUC values produced over 0.98 from all the models. As we can see, the neuropsychological test scores, GDTOTAL, and FAQTOAL contribute to all the models significantly. It might indicate that the neuropsychological tests had contributed to the diagnosis in the sample collection. We then built the models without GDTOTAL and FAQTOTAL included. The result is shown at the bottom of the table (TABLE III), with cross-validated AUC produced ranging from 0.77 to 0.84.

## IV. DISCUSSION

This study aims to identify characteristics that may help determine Alzheimer's disease risk. Researchers are yet to pinpoint Alzheimer's risk factors. In our research, the dependent variable is a binary variable indicative of the study participants' status as normal control or having Alzheimer's disease. The independent variables are picked based on the earlier studies to see if statistically significant variables could be helpful to get some idea on disease development.

For this study, the datasets were taken from two cohorts. Cohort1 included AD patients with how long the AD had been developed unknown, and Cohort2 included only the AD patients who had developed AD within the last 3 years. The main reason for having two cohorts is to check if it is possible to distinguish the risk factors that contribute to AD development in the short term from the ones that are associated with a long history of AD. It was assumed that in Cohort1 some of the AD patients had been diagnosed with AD long before the data collection

Initially, the Pearson chi-squared test and one-way ANOVA showed marital status, occupation category, APOE4 genotype, FAQTOTAL, and GDTOTAL are associated with AD.

TABLE I. CHARACTERISTICS OF SOCIO-DEMOGRAPHIC, VITAL SIGNS, AND NEUROPSYCHOLOGICAL RISK FACTORS TO DEVELOP ALZHEIMER'S

| Variables | Cohort1( Baseline) | | | Cohort2 (36-month follow-up period) | | |
|---|---|---|---|---|---|---|
| | *NC (n=484)* | *Alzheimer's (n=341)* | *P-Value* | *NC (n=220)* | *Alzheimer's (n=171)* | *P-Value* |
| **Age in years (n%)** | | | <.001 | | | 0.335 |
| 55-64 | 4.3% | 13.8% | | 3.20% | 7.00% | |
| 65-74 | 54.30% | 33.10% | | 33.20% | 29.20% | |
| 75-84 | 38.00% | 43.40% | | 53.60% | 53.80% | |
| 85+ | 3.30% | 9.70% | | 10.00% | 9.90% | |
| **Sex (n%)** | | | 0.011 | | | 0.016 |
| Male | 47.90% | 56.90% | | 46.80% | 59.10% | |
| Female | 52.10% | 43.10% | | 53.20% | 40.90% | |
| **Education Level (n%)** | | | <.001 | | | 0.155 |
| Primary | 0.20% | 0.60% | | 0.50% | 0.60% | |
| Secondary | 9.30% | 22.90% | | 8.60% | 15.80% | |
| Bachelor or Diploma | 43.80% | 46.90% | | 45.00% | 44.40% | |
| Masters or Higher Degree | 46.70% | 29.60% | | 45.90% | 39.20% | |
| **Marital status (n%)** | | | <.001 | | | 0.021 |
| Married | 69.40% | 83.00% | | 69.50% | 81.30% | |
| Widowed | 14.90% | 10.90% | | 15.00% | 9.90% | |
| Divorced | 10.10% | 4.10% | | 9.50% | 7.60% | |
| Never married | 5.60% | 2.10% | | 5.90% | 1.20% | |
| **Occupation Category (n%)** | | | <.001 | | | 0.003 |
| Managers or Administrators | 14.90% | 9.40% | | 15.50% | 14.60% | |
| Professionals | 56.00% | 46.90% | | 60.00% | 46.20% | |
| Skilled Occupants | 19.40% | 25.80% | | 14.10% | 29.20% | |
| Unskilled Occupants | 9.70% | 17.90% | | 10.50% | 9.90% | |
| **APOE4 Genotype (n%)** | | | <.001 | | | <.001 |
| 0 apoe4 allele | 72.70% | 33.70% | | 69.50% | 35.70% | |
| 1 Apoe4 allele | 24.40% | 46.90% | | 28.20% | 44.40% | |
| 2 Apoe4 alleles | 2.90% | 19.40% | | 2.30% | 19.90% | |
| **Vital signs (Mean ± SD)** | | | | | | |
| Body Weight | 76.53±16.39 | 73.59±14.46 | 0.008 | 76.18±16.03 | 75.83±15.99 | 0.830 |
| Systolic Blood Pressure | 132.97±16.1 | 133.43±17.04 | 0.697 | 131.1±15.62 | 133.68±18.35 | 0.134 |
| Diastolic Blood Pressure | 73.747±9.82 | 73.59±9.29 | 0.822 | 72.57±9.59 | 73.11±9.85 | 0.591 |
| Pulse Rate | 65.77±10.67 | 64.45±9.86 | 0.071 | 65.88±10.93 | 63.73±10.97 | 0.044 |
| **Neuropsychological test score (Mean ± SD)** | | | | | | |
| GDTOTAL | 0.7±1.074 | 1.69±1.464 | <.001 | 1.15±1.15 | 2.4±2.2 | <.001 |
| FAQTOTAL | 0.13±0.57 | 13.37±6.99 | <.001 | 0.38±1.189 | 15.87±7.733 | <.001 |

TABLE Ⅱ: ODDS RATIOS OF EACH FACTOR BETWEEN AD AND NC, AND THE RESULT ADJUSTED FOR AGE, SEX AND APOE4 GENOTYPE

| Variable | Cohort1( Baseline) (univariate) | | Cohort1( Baseline) (adjusted for age, sex, and APOE4 genotype) | | Cohort2(36-month follow-up period) (univariate) | | Cohort2(36-month follow-up period) (adjusted for age sex and APOE4 genotype) | |
|---|---|---|---|---|---|---|---|---|
| | ***OR*** | ***95% CI*** | ***OR*** | ***95% CI*** | ***OR*** | ***95% CI*** | ***OR*** | ***95% CI*** |
| **Education Level** | | | | | | | | |
| Primary | Reference | | Reference | | Reference | | Reference | |
| Secondary | 0.87 | [0.08, 9.83] | 1.47 | [0.13, 17.04] | 1.42 | [0.08, 24.16] | 1.49 | [0.08, 26.51] |
| Bachelor or Diploma | 0.38 | [0.03,4.20] | 0.57 | [0.05, 6.49] | 0.77 | [0.05, 12.47] | 0.68 | [0.04, 11.50] |
| Masters or Higher Degree | 0.22 | [0.02, 2.49] | 0.34 | [0.03, 3.86] | 0.66 | [0.04, 10.79] | 0.63 | [0.04, 10.70] |
| **Marital Status** | | | | | | | | |
| Married | Reference | | Reference | | Reference | | Reference | |
| Widowed | 0.61 | [0.40, 0.94] | 0.53 | [0.32, 0.89] | 0.57 | [0.30, 1.06] | 0.72 | [0.35, 1.49] |
| Divorced | 0.34 | [0.18, 0.63] | 0.52 | [0.26, 1.05] | 0.68 | [0.33, 1.41] | 0.92 | [0.40, 2.10] |
| Never married | 0.31 | [0.13, 0.72] | 0.34 | [0.13, 0.88] | 0.17 | [0.04, 0.76] | 0.29 | [0.06, 1.35] |
| **Occupation Category** | | | | | | | | |
| Managers & Administrators | Reference | | Reference | | Reference | | Reference | |
| Professionals | 1.33 | [0.84, 2.10] | 1.388 | [0.81, 2.34] | 0.81 | [0.45, 1.46] | 0.757 | [0.40, 1.45] |
| Skilled occupants | 2.12 | [1.27, 3.50] | 2.404 | [1.33, 4.35] | 2.19 | [1.11, 4.35] | 2.848 | [1.32, 6.13] |
| Unskilled occupants | 2.92 | [1.66, 5.13] | 3.628 | [1.89, 6.97] | 1 | [0.45, 2.27] | 1.107 | [0.43, 2.80] |
| **Body Weight** | 0.988 | [0.98, 1.0] | 0.98 | [0.97, 0.99] | 1 | [0.99, 1.01] | 0.99 | [0.97, 1.00] |
| **Systolic Blood Pressure** | 1 | [0.99, 1.01] | 1 | [0.99, 1.01] | 1.01 | [0.99, 1.02] | 1.01 | [0.99, 1.03] |
| **Diastolic Blood Pressure** | 1 | [0.98, 1.01] | 1 | [0.98, 1.02] | 1.01 | [0.98, 1.03] | 1 | [0.98, 1.03] |
| **Pulse Rate** | 0.99 | [0.97, 1.00] | 1.01 | [0.99, 1.02] | 0.98 | [0.96, 1] | 0.98 | [0.96, 1.00] |
| **GDTOTAL** | 1.86 | [1.64, 2.11] | 1.91 | [1.65, 2.20] | 1.43 | [1.26, 1.63] | 1.43 | [1.25, 1.63] |
| **FAQTOTAL** | 4.279 | [3.18, 5.75] | 5.24 | [3.47, 7.92] | 2.23 | [1.82, 2.73] | 2.22 | [1.82, 2.73] |

TABLE Ⅲ: PREDICTION MODELS BUILT WITH LOGISTIC REGRESSION USING DIFFERENT SETS OF VARIABLES

| Dependent Variables | Cohort1 (Baseline) | | | Cohort2(36-month follow-up period) | | |
|---|---|---|---|---|---|---|
| | *Model1* | *Model2* | *Model 3* | *Model1* | *Model2* | *Model 3* |
| **Age in years** | | | | | | |
| 55-64 | Reference | | | | | |
| 65-74 | -2.089 | -1.843 | -1.849 | -2.449 | | |
| 75-84 | -0.605 | -0.519 | -0.468 | -3.646 | | |
| 85+ | 2.53 | 2.222 | 1.597 | -3.304 | | |
| **Sex** | | | | | | |
| Male | Reference | | | | | |
| Female | 0.212 | 0.255 | | -1.196 | -0.65 | |
| **APOE4 Genotype** | | | | | | |
| No apoe4 allele | Reference | | | | | |
| 1 Apoe4 allele | 3.982 | 3.734 | 3.509 | 0.557 | 0.352 | 0.249 |
| 2 Apoe4 allele | 4.776 | 4.551 | 4.234 | 3.502 | 3.273 | 2.712 |
| **Education Level** | | | | | | |
| Primary (Year 1-6) | Reference | | | | | |
| Secondary (Year7-12) | 0.087 | 0.464 | | 7.399 | | |
| Bachelor or Diploma (Year13-16) | -0.401 | 0.191 | | 8.311 | | |
| Masters or Higher Degree | -1.26 | -0.608 | | 8.348 | | |
| **Marital Status** | | | | | | |
| Married | Reference | | | | | |
| Widowed | -2.122 | -2.048 | | 0.865 | 0.068 | |
| Divorced | -3.1 | -2.599 | | 0.457 | 0.53 | |
| Never married | 0.371 | 0.345 | | 2.523 | 1.589 | |
| **Occupation Category** | | | | | | |
| Manager & Administration | Reference | | | | | |
| Professionals | 0.538 | 0.941 | | 1.32 | 1.118 | 0.943 |
| Skilled Occupants | 0.461 | 0.941 | | 4.07 | 3.203 | 2.916 |
| Unskilled Occupants | 1.277 | 1.544 | | 4.055 | 3.121 | 2.293 |
| **Body Weight in Kg** | -0.007 | 0.003 | | -0.014 | | |
| **Systolic blood pressure, mm Hg** | 0.014 | | | 0.022 | | |
| **Diastolic blood pressure, mm Hg** | 0.041 | | | -0.03 | | |
| **Pulse rate/minute** | 0.006 | | | -0.025 | -0.034 | |
| **GDTOTAL** | 0.729 | 0.658 | 0.54 | 0.23 | 0.197 | |
| **FAQTOTAL** | 1.7 | 1.607 | 1.566 | 0.899 | 0.849 | 0.867 |
| **AUC** | 0.997 | 0.997 | 0.996 | 0.989 | 0.989 | 0.987 |
| **AUC_1*** | 0.824 | 0.823 | 0.823 | 0.792 | 0.768 | 0.784 |
| **Cross Validated AUC** | 0.988 | 0.988 | 0.996 | 0.976 | 0.989 | 0.987 |
| **Cross Validated AUC_1 *** | 0.803 | 0.806 | 0.823 | 0.746 | 0.742 | 0.784 |

* The models built without neuropsychological test scores included, used only demographic, social factors, and vital signs. AUC (and AUC_1): AUC values from the overall models built with the whole dataset. Cross Validated AUC (and AUC_1): AUC values from 5-fold cross-validated.

Aging is known as the leading risk factor for AD development. The positive diagnoses numbers and prevalence of AD-related disorders increase exponentially after 65 years and almost doubles every subsequent five years until the age of 90 [16].

APOE4 genotype is considered the most substantial genetic risk factor for late-onset AD [17]. One study stated that a statistical model could be developed in which APOE4 carrier status, demographic variables, and baseline cognitive status can be used to identify individuals at risk of clinical progression for being included in trials [18]. From our analyses, it should be noted that the presence of the APOE4 gene in the human body is significant whether the development period is known or unknown.

A study discovered that people in low-status jobs had a higher risk of developing dementia even when education was controlled [19]. In this study, the Chi-square association test found that both cohorts have a significant association between AD development and occupation. People in the skilled and unskilled occupation groups tend to have a higher risk of developing AD than those with managerial or administrative jobs.

Some studies have indicated that women are at a higher risk of dementia, especially of the AD variety, while other studies have shown no noticeable difference in sex [20].

Participants' weight, blood pressure, pulse rate, and this study have been taken as lifestyle risk factors. Even people's obesity-related to their weight [21]. High blood pressure also relates to lifestyle, and it also affects developing cardiovascular disease [22]. Moreover, in a study, it has been shown that traditional risk factors for cardiovascular and cerebrovascular disease (hypertension, diabetes mellitus, smoking, physical inactivity, obesity, and heart disease, including coronary artery disease and atrial fibrillation) are associated with risk of mild cognitive impairment (MCI) and progression to AD [23].

It has been reported that depression increases dementia and AD risk in older adults. [16]. The Geriatric Depression Scale (GDS) is a proven tool for screening for depression in the elderly with MCI. Its reliability, however, decreases with increasing cognitive impairment or dementia [24]. Another psychological factor, the FAQ, interests researchers for its potential to help discriminate people with cognitive impairment from unimpaired individuals, and it has been used in several studies for evaluating the occurrence of dementia [25]. While checking the association with AD development, both Neuropsychological tests FAQTOTAL and GDTOTAL were significant for cohort1 and cohort2.

There have been numerous studies on AD risk factors. Those studies have pointed to different lifestyle factors, such as alcohol habit, smoking habit, diabetes, and cardiovascular disease history, physical activities and cognitive test variables (Mini-Mental State Exam (MMSE), Delayed total recall (LDELTOTAL), and Rey Auditory Verbal Learning Test (RAVLT) ) as having significant effects on predicting the risk of developing AD in the elderly population [26][27][28].

In this study, the variables from ADNI were taken considering that they are relatively easy to collect in real life. With these easy-to-collect variables, it is possible to develop an app to assess the risk of Alzheimer's disease development and alert people who may be at high risk before any clinical diagnosis.

The prediction models built with FAQTOTAL and GDTOTAL scores included in the models have shown high performance for both cohorts. Even the models built with only demographic variables, APOE4 genotype, and social factors performed reasonably well without FAQTOTAL and GDTOTAL. It indicates that the set of variables are evident in association with the development of AD irrespective of whether the disease development period is short or long. These factors may be considered in developing an app for Alzheimer's development risk prediction, especially in third-world countries or developing countries. The study can also help form an idea on how the effects of risk factors may work if the AD develops within a certain period.

## V. CONCLUSION

This study showed the effects of socio-demographic, lifestyle, and neuropsychological risk factors on developing risk factors and determined the model accuracy from two different cohorts. A clearer picture has emerged through logistic regression regarding the effects of the risk factors, and the AUC from ROC analysis showed that the models performances of the two cohorts are good. The significant variables from the regression models may be used to develop a risk app for Alzheimer’s disease risk prediction.

## ACKNOWLEDGMENT

The authors wish to acknowledge The Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which is considered as the largest free access database on Alzheimer’s.

## REFERENCES:

[1] D. G. Blazer, K. Yaffe, and J. Karlawish, “Cognitive aging: A report from the Institute of Medicine,” JAMA - J. Am. Med. Assoc., vol. 313, no. 21, pp. 2121–2122, 2015, doi: 10.1001/jama.2015.4380.

[2] K. Deckers, A. Nooyens, M. van Boxtel, F. Verhey, M. Verschuren, and S. Köhler, “Gender and Educational Differences in the Association between Lifestyle and Cognitive Decline over 10 Years: The Doetinchem Cohort Study,” J. Alzheimer’s Dis., vol. 70, pp. 1–11, 2018, doi: 10.3233/jad-180492.

[3] D. Jha, S. Alam, J.-Y. Pyun, K. H. Lee, and G.-R. Kwon, “Alzheimer’s Disease Detection Using Extreme Learning Machine, Complex Dual Tree Wavelet Principal Coefficients and Linear Discriminant Analysis,” J. Med. Imaging Heal. Informatics, vol. 8, no. 5, pp. 881–890, 2018, doi: 10.1166/jmihi.2018.2381.

[4] M. A. McAdams-Demarco, M. Daubresse, S. Bae, A. L. Gross, M. C. Carlson, and D. L. Segev, “Dementia, Alzheimer’s disease, and mortality after hemodialysis initiation,” Clin. J. Am. Soc. Nephrol., vol. 13, no. 9, pp. 1339–1347, 2018, doi: 10.2215/CJN.10150917.

[5] M. Belathur Suresh, B. Fischl, and D. H. Salat, “Factors influencing accuracy of cortical thickness in the diagnosis of Alzheimer’s disease,” Hum. Brain Mapp., vol. 39, no. 4, pp. 1500–1515, 2018, doi: 10.1002/hbm.23922.

[6] D. A. Loewenstein, R. E. Curiel, R. Duara, and H. Buschke, “Novel Cognitive Paradigms for the Detection of Memory Impairment in Preclinical Alzheimer’s Disease,” Assessment, vol. 25, no. 3, pp. 348–359, 2018, doi: 10.1177/1073191117691608.

[7] M. Grassi, G. Perna, D. Caldirola, K. Schruers, R. Duara, and D. A. Loewenstein, “A clinically-translatable machine learning algorithm for the prediction of Alzheimer’s disease conversion in individuals with mild and premild cognitive impairment,” J. Alzheimer’s Dis., vol. 61, no. 4, pp. 1555–1573, 2018, doi: 10.3233/JAD-170547.

[8] M. Walters, K. Hackett, E. Caesar, R. Isaacson, and L. Mosconi, “Role of Nutrition to Promote Healthy Brain Aging and Reduce Risk of Alzheimer’s Disease,” Curr. Nutr. Rep., vol. 6, no. 2, pp. 63–71, 2017, doi: 10.1007/s13668-017-0199-5.

[9] M. Samadi, S. Moradi, M. Moradinazar, R. Mostafai, and Y. Pasdar, “Dietary pattern in relation to the risk of Alzheimer’s disease: a systematic review,” Neurol. Sci., vol. 40, no. 10, pp. 2031–2043, 2019, doi: 10.1007/s10072-019-03976-3.

[10] M. W. Weiner et al., “The Alzheimer’s disease neuroimaging initiative: A review of papers published since its inception,” Alzheimer’s Dement., vol. 8, no. 1 SUPPL., pp. S1–S68, 2012, doi: 10.1016/j.jalz.2011.09.172.

[11] S. Supplies, “ADNI General Procedures Manual,” vol. 208, no. 5, p. 2004, 2003, doi: 10.1002/ejoc.201200111.

[12] I. Castles, “Australian standard classification of occupations: Statistical classification,” 1986.

[13] “ADNI-D Protocol Alzheimer ’ s Disease Neuroimaging Initiative – The ADNI Depression Project ‘ Characterizing Cognitive Decline in Late Life Depression ’ Study Protocol ADNI-D Protocol ADNI Depression Project : Schedule of,” pp. 1–39, 2014.

[14] K. Ito, M. M. Hutmacher, and B. W. Corrigan, “Modeling of Functional Assessment Questionnaire (FAQ) as continuous bounded data from the ADNI database,” J. Pharmacokinet. Pharmacodyn., vol. 39, no. 6, pp. 601–618, 2012, doi: 10.1007/s10928-012-9271-3.

[15] B. C. Riedel, P. M. Thompson, R. D. Brinton RD. Age, APOE and sex: Triad of risk of Alzheimer's disease. J Steroid Biochem Mol Biol, vol. 160, pp. 134-47, 2016, doi: 10.1016/j.jsbmb.2016.03.012.Age.

[16] M. Kivipelto, F. Mangialasche, and T. Ngandu, “Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease,” Nat. Rev. Neurol., vol. 14, no. 11, pp. 653–666, 2018, doi: 10.1038/s41582-018-0070-3.

[17] N. Zhao et al., “Alzheimer’s Risk Factors Age, APOE Genotype, and Sex Drive Distinct Molecular Pathways,” Neuron, vol. 106, no. 5, pp. 727-742.e6, 2020, doi: 10.1016/j.neuron.2020.02.034.

[18] M. Ten Kate et al., “Secondary prevention of Alzheimer’s dementia: neuroimaging contributions,” Alzheimers. Res. Ther., vol. 10, no. 1, p. 112, 2018, doi: 10.1186/s13195-018-0438-z.

[19] A. F. Jorm et al., “Occupation type as a predictor of cognitive decline and dementia in old age,” Age Ageing, vol. 27, no. 4, pp. 477–483, 1998, doi: 10.1093/ageing/27.4.477.

[20] M. T. Ferretti et al., “Sex and gender differences in Alzheimer’s disease: current challenges and implications for clinical practice: Position paper of the Dementia and Cognitive Disorders Panel of the European Academy of Neurology,” Eur. J. Neurol., vol. 27, no. 6, pp. 928–943, 2020, doi: 10.1111/ene.14174.

[21] E. K. Naderali, S. H. Ratcliffe, and M. C. Dale, “Obesity and alzheimer?s disease: A link between body weight and cognitive function in old age,” Am. J. Alzheimers. Dis. Other Demen., vol. 24, no. 6, pp. 445–449, 2009, doi: 10.1177/1533317509348208.

[22] L. Mosconi et al., “Lifestyle and vascular risk effects on MRI-based biomarkers of Alzheimer’s disease: A cross-sectional study of middle-aged adults from the broader New York City area,” BMJ Open, vol. 8, no. 3, 2018, doi: 10.1136/bmjopen-2017-019362.

[23] S. Norton, F. E. Matthews, D. E. Barnes, K. Yaffe, and C. Brayne, “Potential for primary prevention of Alzheimer’s disease: An analysis of population-based data,” Lancet Neurol., vol. 13, no. 8, pp. 788–794, 2014, doi: 10.1016/S1474-4422(14)70136-X.

[24] B. Shanmugham and G. Alexopoulos, “Geriatric Depression,” Biol. Depress. From Nov. Insights to Ther. Strateg., no. September, pp. 317–339, 2008, doi: 10.1002/9783527619672.ch12.

[25] R. M. Tappen, M. Rosselli, and G. Engstrom, “Evaluation of the functional activities questionnaire (FAQ) in cognitive screening across four american ethnic groups,” Clin. Neuropsychol., vol. 24, no. 4, pp. 646–661, 2010, doi: 10.1080/13854040903482855.

[26] S. Sindi et al., “The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia,” Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit., vol. 1, no. 3, pp. 328–333, 2015, doi: 10.1016/j.dadm.2015.06.005.

[27] P. Johnson et al., “Genetic algorithm with logistic regression for prediction of progression to Alzheimer’s disease,” BMC Bioinformatics, vol. 15, no. Suppl 16, p. S11, 2014, doi: 10.1186/1471-2105-15-S16-S11.

[28] E. Moradi, I. Hallikainen, T. Hänninen, and J. Tohka, “Rey’s Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer’s disease,” NeuroImage Clin., vol. 13, pp. 415–427, 2017, doi: 10.1016/j.nicl.2016.12.011.