Project Based Learning

Performance Evaluation of Computing Systems

(17M11CS122)

Analysis of Experimental Designs on Alzheimer's Disease

Submitted By:

Pranav Gupta	19803021	B13
Garvita Nagpal	19803019	B13
Sushrut Khajuria	22303001	M.Tech

Submitted To

Dr. Kavita Pandey

Department of Computer Science,
Jaypee Institute of Information Technology University, Noida

18 April 2023

ABSTRACT

The purpose of this study was to study the experimental design and full factorial design that are crucial in scientific research because they enable researchers to systematically explore and analyze the effects of multiple factors on a particular outcome. Experiment design helps to ensure that the experiment is structured in a way that allows for valid conclusions to be drawn from the data collected. Full factorial design is a method that allows researchers to examine all possible combinations of the factors being tested, which can help to identify interactions between factors that may not be apparent otherwise. This type of design also ensures that all possible effects of each factor are considered, which can lead to more accurate and reliable results. Overall, using experimental design and full factorial design can help to minimize bias and errors in scientific research and increase the rigor and validity of the conclusions drawn.

INDEX

1. Introduction 3
2. DesignsUsed 3
3. Data Analysis and Feature Selection 6
4. Methodology 10
5. Conclusion 19
6. References 20

I. INTRODUCTION:

Alzheimer's disease is a progressive neurodegenerative disorder that affects millions of people worldwide. It is the most common cause of dementia and is characterized by a decline in cognitive function, including memory loss, language difficulties, and impaired judgment. The disease typically develops slowly, with symptoms becoming more severe over time. Although the exact cause of Alzheimer's disease is not yet fully understood, research suggests that a combination of genetic, environmental, and lifestyle factors may play a role. Despite advances, in research, there is currently no cure for Alzheimer's disease, and available treatments focus primarily on managing symptoms. The impact of Alzheimer's disease on individuals, families, and society as a whole is significant, highlighting the urgent need for continued research to better understand the disease and develop more effective treatments.

II. DESIGNS USED:

Experimental Design

Experimental design is a systematic approach to planning, conducting, and analyzing scientific experiments in order to draw valid conclusions and make meaningful inferences about a particular phenomenon or process. The main objective of experimental design is to control and manipulate the variables involved in an experiment in order to test hypotheses and theories.

A well-designed experiment typically involves several key components, including the identification of the research question, the formulation of a testable hypothesis, the selection of appropriate study participants or subjects, the definition of the variables and their operational definitions, the random assignment of subjects to groups, the selection of appropriate measures and instruments, and the establishment of procedures for data collection, analysis, and interpretation.

Experimental design can be used in a wide range of scientific fields, including biology, psychology, sociology, physics, chemistry, and engineering. It is a powerful tool for exploring causal relationships between variables and testing the effectiveness of interventions or treatments.

In order to conduct a successful experiment, it is important to have a solid understanding of experimental design principles, as well as the relevant statistical techniques and methods for data analysis. By carefully planning and executing experiments, researchers can obtain reliable and valid results that contribute to our understanding of the natural world.

Full Factorial Design Method

A full factorial design is a type of experimental design used in statistical analysis to study the effect of multiple independent variables on a dependent variable. In this design, all possible combinations of the independent variables are tested to identify which combination of variables has the most significant impact on the dependent variable.

A full factorial design typically involves two or more independent variables, also called factors, each with two or more levels. For example, if we want to investigate the effect of temperature and humidity on plant growth, temperature and humidity would be the two factors, and each factor would have two or more levels (low, medium, and high).

To perform a full factorial design, all possible combinations of the independent variables and their levels are tested. For example, if temperature has three levels and humidity has two levels, there would be six possible combinations (low temperature/low humidity, low temperature/medium humidity, low temperature/high humidity, medium temperature/low humidity, medium temperature/medium humidity, and medium temperature/high humidity).

The dependent variable is measured for each combination of the independent variables, and the results are analyzed using statistical techniques to determine the main effects of each independent variable and their interactions. The main effects indicate the individual impact of each independent variable on the dependent variable, while the interactions indicate the joint effect of two or more independent variables on the dependent variable.

The advantage of a full factorial design is that it allows for the identification of all main effects and interactions between independent variables, making it a powerful and efficient method for testing multiple hypotheses simultaneously. However, the number of experimental conditions can quickly become large, making it difficult and costly to perform in practice.

2^k Factorial Design Method

2^k factorial design is a statistical method used in experimental design to study the effects of two-level factors on a response variable. In this method, k factors are chosen, each of which has two levels. The two levels can be represented by a 0 or a 1, where 0 represents the absence of a factor and 1 represents the presence of a factor.

The2^k factorial design involves running a full factorial experiment, which means that all possible combinations of the levels of the factors are tested. For example, if $k = 3$, there would be 2^3 or 8 possible combinations of the factors. These combinations are called treatment combinations and are represented as factorials in the design.

The design allows researchers to identify the main effects of each factor, as well as the interactions between factors. The main effects are the differences in the response variable due to changes in a single factor, while the interactions are the effects that occur when two or more factors are combined.

One advantage of the 2^k factorial design is that it requires fewer runs or trials than other experimental designs, such as a full factorial design, which can be time-consuming and expensive. It also allows researchers to study multiple factors simultaneously, which can lead to a better understanding of the relationship between variables.

The results obtained from the 2^k factorial design can be analyzed using statistical methods, such as analysis of variance (ANOVA), to determine the significance of the main effects and interactions. This information can be used to make predictions about the response variable under different conditions and to optimize the factors for a desired outcome.

Overall, the 2^k factorial design is a useful method for studying the effects of multiple factors on a response variable and can provide valuable insights for researchers in a variety of fields.

Fractional Factorial Design Method

Fractional factorial design is a statistical method used in experimental design to study the effect of multiple variables on a response or outcome. It is an efficient way to reduce the number of experimental runs while still maintaining a certain level of accuracy in the results.

In a fractional factorial design, only a subset of the possible combinations of the variables is tested. This allows researchers to explore a larger number of variables while minimizing the number of experimental runs needed. Fractional factorial designs can be constructed using specific algorithms such as the Plackett-Burman design or the Taguchi method.

The key advantage of using fractional factorial designs is that they require fewer experiments, which can reduce costs and save time. Additionally, they are less prone to errors compared to full factorial designs, where all possible combinations of variables are tested. However, it is important to note that fractional factorial designs may not identify all interactions between variables, and the results may be less precise than those obtained from a full factorial design.

To use fractional factorial designs effectively, it is important to carefully select the variables to include in the design and choose an appropriate fraction of the full factorial design to test. This requires careful consideration of the objectives of the experiment and the resources available for conducting it.

2^{k-p} fractional factorial Design

2^{k-p} fractional factorial designs are a type of experimental design used in statistics and engineering to determine which factors affect a process or product. In this design, K is the number of factors being tested, and p is the number of runs that are not included in the design.

The design works by systematically varying the levels of the factors being tested, while holding all other factors constant. The results of the experiments are then analyzed using statistical methods to determine which factors have a significant effect on the process or product.

The term "fractional factorial" refers to the fact that not all possible combinations of factor levels are tested. Instead, a subset of the possible combinations is tested, which reduces the number of experiments required. This can save time and resources, particularly when there are many factors being tested.

For example, a 2^{3-1} fractional factorial design would involve testing all possible combinations of three factors at two levels (low and high), except for one combination, which is not tested. This would result in a total of seven experiments. The results of these experiments would be used to determine which factors have a significant effect on the process or product.

2^{k-p} fractional factorial designs can be very useful for screening a large number of factors to determine which ones have the most significant effects. However, they do have some limitations. They are not capable of detecting interactions between factors, and they may not be suitable for all types of experiments.

III. Data Analysis and Feature Selection:

In this initiative, data analysis is essential to gaining important insights and educating the public about the condition. The strategy includes gathering data from Kaggle and processing it in various ways. After that, we examine the data using statistical methods and tools, such as logistic regression. By analysing the data, we hope to spot patterns, correlations, and trends in the information and make inferences about the information obtained from Kaggle and released under the Database Contents License (DbCL) v1.0.

We can make sure that our results are statistically significant and serve as a strong basis for suggestions by employing rigorous data analysis procedures.

The data used in this particular analysis is as follows:

([Alzheimer features](https://www.kaggle.com/datasets/brsdincer/alzheimer-features)

Dataset Description:

1. **Group** --> Class: That whether this particular person is Demented or Non-demented

2. **Age** --> Age: Tells about the age of the person

3. **EDUC** --> Years of Education: How many years did he or she spend studying.

4. **GENDER**→ Whether this particular person is male or female.

5. **SES** --> Socioeconomic Status / 1-5: A way of describing people based on their education, income, and type of job.

6. **MMSE** --> Mini-Mental State Examination: The Mini-mental state examination is scored on a scale of 0-30 with scores > 25 interpreted as normal cognitive status.

- Severe cognitive impairment: 0-17
- Mild cognitive impairment: 18-23
- No cognitive impairment: 24-30

Interpretation of the mental status examination must take into account the patient's native language, education level, and culture as these factors can affect performance.

7. **CDR** --> Clinical Dementia Rating: The CDR is based on a scale of 0–3: no dementia (CDR = 0), questionable dementia (CDR = 0.5), MCI (CDR = 1), moderate cognitive impairment (CDR = 2), and severe cognitive impairment (CDR = 3).

8. **eTIV** --> Estimated total intracranial volume: Total intracranial volume (TIV/ICV) is an important covariate for volumetric analyses of the brain and brain regions, especially in the study of neurodegenerative diseases, where it can provide a proxy of maximum pre-morbid brain volume

9. nWBV --> Normalize Whole Brain Volume: nWBV stands for Normalize Whole Brain Volume. It is a measure used in neuroimaging research to account for individual differences in brain size when comparing brain volumes across different individuals or groups. The brain volume is the total volume of the brain, including both grey and white matter.

10. ASF --> Atlas Scaling Factor: ASF stands for Atlas Scaling Factor. It is a term commonly used in neuroimaging research to refer to a normalization factor that is applied to brain images to account for differences in brain size or scale among individuals. The Atlas Scaling Factor is calculated by comparing the size of the brain image to a reference brain template or atlas.

```
data.head(10)
```

	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
0	Nondemented	M	1.317594	-0.210170	-0.485925	-0.067810	-0.714052	2.745522	-0.907990	-2.209649
1	Nondemented	M	1.445715	-0.210170	-0.485925	0.709145	-0.714052	2.840114	-1.301670	-2.259770
5	Nondemented	F	1.445715	1.156475	0.404469	0.191175	-0.714052	-1.550071	-0.540555	1.807170
6	Nondemented	F	1.701959	1.156475	0.404469	-0.067810	-0.714052	-1.633534	-0.330592	1.936052
7	Nondemented	M	0.420741	-0.893492	1.294863	0.191175	-0.714052	1.087379	-0.488064	-1.092673
8	Nondemented	M	0.805106	-0.893492	1.294863	0.450160	0.594356	1.154150	-0.514310	-1.142793
9	Nondemented	M	1.061350	-0.893492	1.294863	0.709145	-0.714052	1.143021	-0.671782	-1.135633
13	Nondemented	F	2.086325	-0.210170	-0.485925	0.709145	-0.714052	-1.232909	-0.855499	1.348923
14	Nondemented	F	2.342568	-0.210170	-0.485925	0.450160	-0.714052	-1.316373	-0.724273	1.463485
15	Demented	M	-1.116721	-0.893492	-0.485925	-0.067810	0.594356	-0.203525	1.978999	0.095905

Figure 1: Dataset Snippet

The above dataset has 317 rows and 10 columns. (After Data Cleaning)

A pipeline for developing data science models must include feature engineering. The adage "More data leads to a better machine learning model" is valid in terms of instances but not features. Preparing pertinent characteristics for a powerful machine learning model takes up the majority of a data scientist's workday. Many duplicate characteristics in a raw dataset can affect how well a model performs. To train a strong machine learning model, the feature selection component of feature engineering entails eliminating irrelevant features and selecting the optimal collection of features. The problem of the "curse of dimensionality" is avoided through feature selection algorithms, which lower the dimensionality of the data.

Logistic Regression was used to establish a relation between the numerical features and the target categorical value i.e. Group which has two classes NonDemented and Demented.

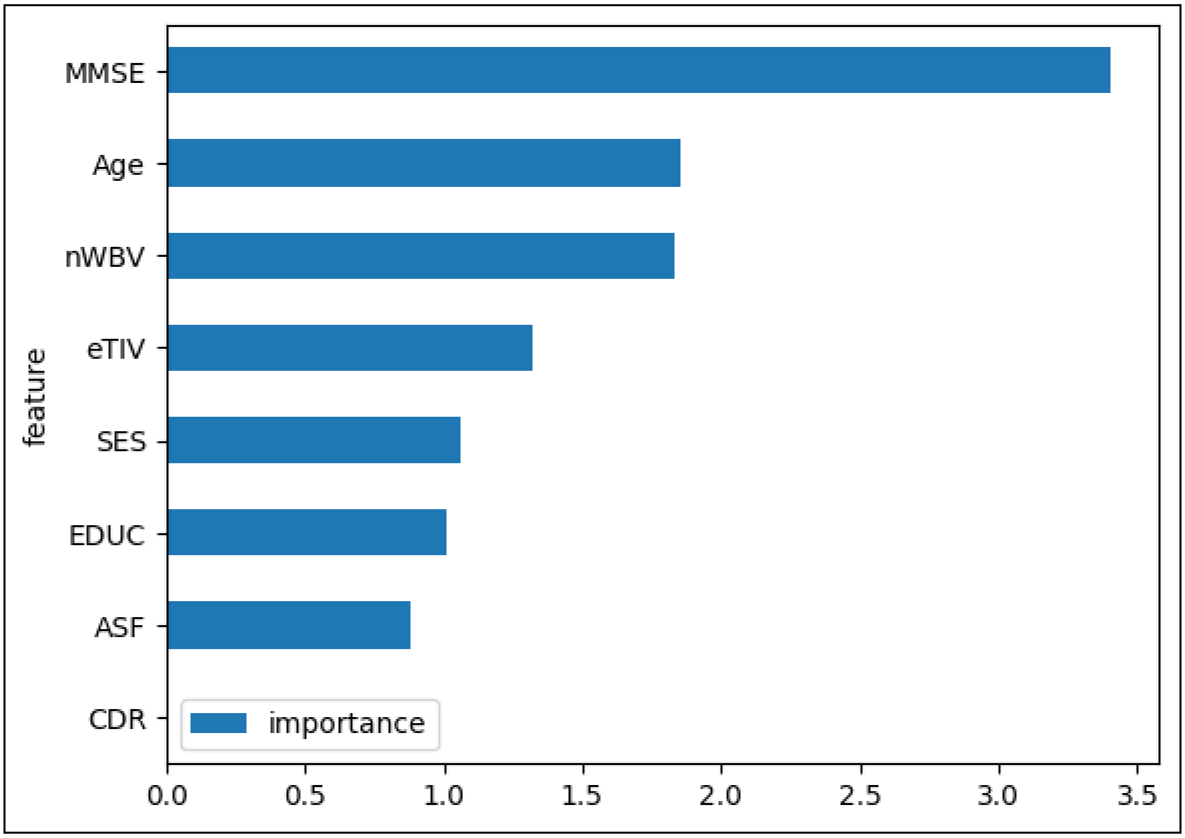

Figure 2: Importance of features in sorted order

Then the importance of each feature was calculated and a sorted list is as follows:

4 0.002894

7 0.257747

1 0.296294

2 0.311396

5 0.387364

6 0.537235 (nWBV)

0 0.544508(Age)

3 1.000000 (MMSE)

Name: importance, dtype: float64

The value of Importance sorted in ascending order is as mentioned above:

The features where importance was greater than 0.5 was selected for the full factorial analysis and one more feature was taken into consideration for fractional design.

The dataset had only one categorical feature and it was highly significant w.r.t to the target variable. We used Chi-square analysis to determine whether or not the gender designation of "M/F" was pertinent to Alzheimer's prediction, and the results indicate that it is fairly important. As a result, we have included "Gender" in our list of features.

```
[8] from scipy.stats import chi2_contingency
```

```
[10] X = data.values[:,1]
     y = data.values[:,0]
```

```
# X = X.astype('category')
# y = y.astype('category')

# Create an empty list to store the results of the chi-square test
chi2_results = []

# Loop through each feature column and calculate the chi-square statistic and p-value
contingency_table = pd.crosstab(X, y)
chi2, p_value, *_ = chi2_contingency(contingency_table)
chi2_results.append(('M/F', chi2, p_value))

# Convert the list of results into a DataFrame for easy manipulation
chi2_df = pd.DataFrame(chi2_results, columns=['Feature', 'Chi2', 'p-value'])

# Sort the DataFrame by p-value in ascending order
chi2_df.sort_values(by='p-value', ascending=True, inplace=True)

# Print the results
print(chi2_df)
```

```
  Feature       Chi2   p-value
0     M/F  22.747496  0.000002
```

Figure 3: Chi-Square Analysis for categorical feature

IV. METHODOLOGY

First, as we conducted analysis for the 2^k factorial design, the unnecessary features were eliminated, and the remaining features were divided into two levels each. Our dataset was transformed into the values 0 and 1 using the following encoding strategy: if the data value is less than the mean, it is regarded to be zero; otherwise, it is considered to be 1. The aggregate was then run to analyse the y value for each category, and the findings are as follows.

```
#CATEGORICAL FEATURES
#Encode Gender
df.loc[df["Group"] == "Nondemented", "Group"] = 1 #nonDemented
df.loc[df["Group"] == "Demented", "Group"] = 0 #demented

#Encode Gender
df.loc[df["M/F"] == "M", "M/F"] = 1 #Male
df.loc[df["M/F"] == "F", "M/F"] = 0 #Female

#NUMERICAL FEATURES
#encode Age
mean_age = df['Age'].mean()
print("Mean Age: ",round(mean_age, 2))
df.loc[df["Age"] < mean_age, "Age"] = 0 #ALTM age less than mean
df.loc[df["Age"] >= mean_age, "Age"] = 1 #AGTM

#encode MMSE
mean_MMSE = df['MMSE'].mean()
print("Mean MMSE: ",round(mean_MMSE, 2))
df.loc[df["MMSE"] < mean_MMSE, "MMSE"] = 0
df.loc[df["MMSE"] >= mean_MMSE, "MMSE"] = 1

#encode nWBV
mean_nWBV = df['nWBV'].mean()
print("Mean nWBV: ",round(mean_nWBV, 2))
df.loc[df["nWBV"] < mean_nWBV, "nWBV"] = 0
df.loc[df["nWBV"] >= mean_nWBV, "nWBV"] = 1
```

Figure 4: Encoding of features in two levels

Then we grouped the dataset and counted the value of 0 and 1 from the Group column for each class.

We used the formula of y from [1].

$$y = \frac{target\ 1\ with\ this\ condition}{target\ 0\ or\ 1\ with\ this\ condition}$$

Both the y values for the 4 feature dataset and the 5 feature dataset are calculated using this formula. Two subprocesses, the count of tuples in each category and the count of tuples marked 1 in that category, are used to complete this phase. The precise value of y is then determined by dividing both numbers.

Table 1: Y values for 4 feature dataset

M/F	Age	MMSE	nWBV	y
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0.8
0	0	1	1	0.852
0	1	0	0	0.111
0	1	0	1	0.333
0	1	1	0	0.964
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0.222
1	0	1	0	0.714
1	0	1	1	0.643
1	1	0	0	0.08
1	1	0	1	0.333
1	1	1	0	0.69
1	1	1	1	0.875

Table 2: Y values for 5 feature dataset

M/F	Age	MMSE	nWBV	eTIV	y	FRAC
0	0	0	0	0	0	
0	0	0	0	1	0	"00001"
0	0	0	1	0	0	"00010"
0	**0**	**0**	**1**	**1**	**0**	
0	0	1	0	0	0.75	"00100"
0	0	1	0	1	0.833	
0	0	1	1	0	0.812	

0	0	1	1	1	1	"00111"
0	1	0	0	0	0.133	"01000"
0	1	0	0	1	0	
0	1	0	1	0	0.333	
0	**1**	**0**	**1**	**1**	**0**	"01011"
0	1	1	0	0	0.957	
0	1	1	0	1	1	"01101"
0	1	1	1	0	1	"01110"
0	1	1	1	1	1	
1	0	0	0	0	0	"10000"
1	0	0	0	1	0	
1	0	0	1	0	0.4	
1	0	0	1	1	0	"10011"
1	0	1	0	0	0.75	
1	0	1	0	1	0.7	"10101"
1	0	1	1	0	0.6	"10110"
1	0	1	1	1	0.667	
1	1	0	0	0	0	
1	1	0	0	1	0.167	"11001"
1	**1**	**0**	**1**	**0**	**0**	"11010"
1	1	0	1	1	0.333	
1	1	1	0	0	0.25	"11100"
1	1	1	0	1	0.76	
1	1	1	1	0	0	
1	1	1	1	1	1	"11111"

Table 3: Full Factorial Design

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD	Y
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	0
1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	0
1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	80
1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	85.2
1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1	11.1
1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	33.3
1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	96.4
1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	100
1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	0
1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	22.2
1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	71.4
1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	64.3
1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	8
1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	33.3
1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	69
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	87.5
761.7	-50.3	115.5	545.9	89.9	574.9953	-88.5	27.9	-11.5	49.3	-49.5	14.9	8.1	-22.7	-1.3	46.3	
47.606	-3.144	7.219	34.119	5.619	35.937	-5.531	1.744	-0.719	3.081	-3.094	0.931	0.506	-1.419	-0.081	2.894	
	158.156	833.823	18625.699	505.171	20663.488	489.471	48.665	8.271	151.881	153.165	13.868	4.097	32.217	0.105	134.004	41822.081
	0.378	1.994	44.536	1.208	49.408	1.17	0.116	0.02	0.363	0.366	0.033	0.01	0.077	0	0.32	100

Table 4: Variation Explained by each feature

Feature	Variation	Percentage
SSA	158.156	0.378
SSB	833.823	1.994
SSC	18625.699	44.536
SSD	505.171	1.208
SS_AB	20663.488	49.408
SS_AC	489.471	1.17
SS_AD	48.665	0.116
SS_BC	8.271	0.02
SS_BD	151.881	0.363
SS_CD	153.165	0.366
SS_ABC	13.868	0.033
SS_ABD	4.097	0.01
SS_ACD	32.217	0.077
SS_BCD	0.105	0
SS_ABCD	134.004	0.32
Total Variation	41822.081	100

Table 5: Variation Explained by each feature in sorted order

Feature	Variation	Percentage
SS_AB	20663.488	49.408
SSC	18625.699	44.536
SSB	833.823	1.994
SSD	505.171	1.208
SS_AC	489.471	1.17
SSA	158.156	0.378

SS_CD	153.165	0.366
SS_BD	151.881	0.363
SS_ABCD	134.004	0.32
SS_AD	48.665	0.116
SS_ACD	32.217	0.077
SS_ABC	13.868	0.033
SS_BC	8.271	0.02
SS_ABD	4.097	0.01
SS_BCD	0.105	0
Total Variation	41822.081	100

Thus, Interactions AB and feature C are the most important from the above Analysis.

Where C: MMSE and A & B: Gender and Age

Table 6: Fractional Factorial Design

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	E	Y
1	**-1**	**-1**	**-1**	**-1**	1	1	1	1	1	1	-1	-1	-1	-1	**1**	**0**
1	**-1**	**-1**	**-1**	**1**	1	1	-1	1	-1	-1	-1	1	1	1	**-1**	**0**
1	**-1**	**-1**	**1**	**-1**	1	-1	1	-1	1	-1	1	-1	1	1	**-1**	**75**
1	**-1**	**-1**	**1**	**1**	1	-1	-1	-1	-1	1	1	1	-1	-1	**1**	**100**
1	**-1**	**1**	**-1**	**-1**	-1	1	1	-1	-1	1	1	1	-1	1	**-1**	**13.3**
1	**-1**	**1**	**-1**	**1**	-1	1	-1	-1	1	-1	1	-1	1	-1	**1**	**0**
1	**-1**	**1**	**1**	**-1**	-1	-1	1	1	-1	-1	-1	1	1	-1	**1**	**100**
1	**-1**	**1**	**1**	**1**	-1	-1	-1	1	1	1	-1	-1	-1	1	**-1**	**100**
1	**1**	**-1**	**-1**	**-1**	-1	-1	-1	1	1	1	1	1	1	-1	**-1**	**0**
1	**1**	**-1**	**-1**	**1**	-1	-1	1	1	-1	-1	1	-1	-1	1	**1**	**0**
1	**1**	**-1**	**1**	**-1**	-1	1	-1	-1	1	-1	-1	1	-1	1	**1**	**70**
1	**1**	**-1**	**1**	**1**	-1	1	1	-1	-1	1	-1	-1	1	-1	**-1**	**60**

1	**1**	**1**	**-1**	**-1**	1	-1	-1	-1	-1	1	-1	-1	1	1	**1**	**16.7**
1	**1**	**1**	**-1**	**1**	1	-1	1	-1	1	-1	-1	1	-1	-1	**-1**	**0**
1	**1**	**1**	**1**	**-1**	1	1	-1	1	-1	-1	1	-1	-1	-1	**-1**	**25**
1	**1**	**1**	**1**	**1**	1	1	1	1	1	1	1	1	1	1	**1**	**100**
660	-116.6	50	600	60	552.0578	-123.4	36.6	-10	30	120	-33.4	106.6	43.4	90	113.4	
41.25	-7.288	3.125	37.5	3.75	34.504	-7.713	2.288	-0.625	1.875	7.5	-2.088	6.663	2.713	5.625	7.088	

Table 7: Variations Explained by each feature

Feature	Variation	Percentage
SSA	849.839	1.809
SSB	156.25	0.333
SSC	22500	47.887
SSD	225	0.479
SS_AB	19048.416	40.541
SS_AC	951.846	2.026
SS_AD	83.759	0.178
SS_BC	6.25	0.013
SS_BD	56.25	0.12
SS_CD	900	1.915
SS_ABC	69.756	0.148
SS_ABD	710.329	1.512
SS_ACD	117.766	0.251
SS_BCD	506.25	1.077
SS_ABCD	803.836	1.711
Total Variation	46985.547	100

Table 8: Confoundings

ABCD=E
I=ABCDE
A=BCDE
B=ACDE
C=ABDE
D=ABCE
AB=CDE
AC=BDE
AD=BCE
BC=ADE
BD=ACE
CD=ABE
ABC=DE
ABD=CE
ACD=BE
BCD=AE
ABCD=E

Table 9: Variation Explained by each feature in sorted order

Feature	Variation	Percentage
SSC	22500	47.887
SS_AB	19048.416	40.541
SS_AC	951.846	2.026
SS_CD	900	1.915
SSA	849.839	1.809
SS_ABCD	803.836	1.711

SS_ABD	710.329	1.512
SS_BCD	506.25	1.077
SSD	225	0.479
SSB	156.25	0.333
SS_ACD	117.766	0.251
SS_AD	83.759	0.178
SS_ABC	69.756	0.148
SS_BD	56.25	0.12
SS_BC	6.25	0.013
SS_T	46985.547	100

Again, similar results are obtained from this analysis as well that MMSE and interaction of Age and gender contribute the most to the target i.e. Alzheimer's.

V. CONCLUSION

Two designs were successfully concluded on the above dataset. Both the designs showed a similar trend thus focusing on the correctness of the results derived from both the designs.

In Full factorial test Interation of factor A & B was the most significant, followed by feature C.

Whereas the same was observed for fractional factorial test. Here just a minor deviation from highest variation prevails but the overall result is similar i.e, C is the most important feature followed by interaction of age and gender.

This sums up that Mini mental state exam score is an important parameter for judging the sane along with the interation of age and gender.

VI. REFERENCES

[1] Aggarwal, S., Pandey, K. & Senior Member, IEEE. Determining the representative features of polycystic ovary syndrome via Design of Experiments. 81, 29207–29227 (2022). https://doi.org/10.1007/s11042-022-12913-0

[2] Raj Jain. The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling, John Wiley and Sons, Inc., New York, NY, Reprint Edition, © 2014

[3] Obaidat, Boudriga, "Fundamentals of Performance Evaluation of Computer and Telecommunication Systems ", 2010, Wiley