Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 04 (2025)

Prediction of Alzheimer's Disease Risk Based on Plasma Lipid Profiles and Covid-19 History Via Logistic Regression.

[1] Y. Ebenezer PonSelvi, [2] Dr. V. Joseph Peter,[3] Dr.G. Sophana,[4] Dr. R. Sheeba Mary Ananthi,

[1] Research Scholar, Research Department of Computer Science,

Kamaraj College(Autonomous) , Thoothukudi.

jreeben@gmail.com

[2] Associate Professor and Head, Research Department of Computer Science,

Kamaraj College (Autonomous), Thoothukudi.

dr.josephpeter@kamaraj.college.ac.in

Assistant Professor and Head, Research Department of Computer Science,

Kamaraj College (Autonomous), Thoothukudi

dr.sophana@kamarajcollege.ac.in

Assistant Professor and Head, Research Department of Computer Science,

Kamaraj College (Autonomous), Thoothukudi

dr.r.sheebamaryananthi@kamarajcollege.ac.in

Abstract

Alzheimer's disease (AD), a form of dementia is known as a neurodegenerative disorder with gradual memory impairment. Recent research studies have suggested a putative risk of increased susceptibility to AD post COVID-19 infection. This study attempted to clarify this relationship by examining lipid profile data in conjunction with the information available in patient database of COVID-19. Data were analyzed by preprocessing, exploratory data analysis through pivot tables and a logistic regression model for the prediction of risk of AD. The model adjusted for APOE4 genotype, age and COVID-19 status. Model performance metrics such as accuracy, precision, recall and F1 score were evaluated. Although reasonably accurate, there was a high rate of false negatives in the model and more work is required to improve these. These results emphasized the intricate association of lipid profiles, COVID-19 infection and AD risk. Identification of biomarkers and effective risk assessment strategy for AD in the setting of COVID-19 needs more investigations. The timely diagnosis and intervention may lead to focused therapy and preventive interventions that reduce the chronic neurological impact of COVID-19.

Keywords: Lipidomics, COVID-19, Alzheimer's Disease, Logistic Regression, APOE4.

1. INTRODUCTION

The global public health landscape faces ongoing challenges from both new and ongoing diseases. These issues call for strong research to clarify their complex mechanisms and connections. The COVID-19 pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has created unique challenges, not only during its acute phase but also regarding its long-term effects on different bodily systems [1]. At the same time, Alzheimer's disease (AD), remains a major public health issue. Alzheimer's disease is a specific, progressive brain

disease that is the most frequent cause of dementia. It is a progressive neurodegenerative disorder that leads to cognitive decline and significantly affects patients and healthcare systems [2]. Recent studies indicate that COVID-19 can cause various clinical and pathological problems in the human body. This includes disrupting metabolic profiles by releasing pro-inflammatory cytokines [3]. Therefore, studying the connection between COVID-19 outcomes, changes in lipid profiles, and the risk of AD is crucial.

Lipid metabolism is essential for many bodily functions, such as maintaining cell structure, storing energy, and responding to infections [2]. When lipid profiles are out of balance, a condition known as dyslipidemia, it can lead to various health issues, including heart diseases and neurodegenerative disorders [4]. It's also noted that SARS-CoV-2 infection can change lipid profiles, suggesting a possible connection between COVID-19 and lipid metabolism [8]. These changes are linked to the severity of COVID-19, showing a negative relationship between lipid levels and acute-phase reactants [10]. The APOE4 genotype, a known genetic risk factor for Alzheimer's disease, may influence lipid metabolism and immune responses, complicating the connection between these elements [6].

This study aims to examine the relationship between plasma lipid profiles, a history of COVID-19, and the risk of Alzheimer's disease, using data analysis techniques and a logistic regression model. Two datasets were used: one containing lipidomics data with clinical profiles, including APOE4 status, and the other with COVID-19 patient data. Through various preprocessing steps such as age categorization and creating an "Alzheimers_Risk" variable based on APOE4 presence, age, and COVID-19 positivity, this research seeks to quantify how these factors interact. The logistic regression model predicts Alzheimer's risk, and its effectiveness is assessed using accuracy, precision, recall, F1-score, and ROC AUC metrics, supported by confusion matrix analysis. Pivot tables analyze relationships among variables, including age, gender, diagnosis, and COVID-19 test results.

2. RELATED WORK

The overlap of COVID-19, lipid metabolism, and Alzheimer's disease risk is a relatively new but fast-evolving research area. Several studies have looked at the individual elements of this relationship, setting the stage for the current investigation.

2.1 Lipid Profiles and COVID-19 Severity

Research indicates a possible link between abnormal lipid profiles and severe COVID-19 [5]. Studies show that people with low levels of "good" cholesterol (HDL) and high triglyceride levels might face a higher risk of severe COVID-19 infections, leading to hospitalization or death [5]. The mechanisms behind this link are still under investigation, but they may involve the role of lipids in inflammation, blood clotting, and immune function. These findings underscore the importance of maintaining a healthy lipid profile, achievable through a balanced diet, regular exercise, and, when necessary, medication.

2.2 Lipid Profiles and Alzheimer's Disease

Research highlights a strong connection between lipid metabolism and the development and progression of Alzheimer's disease [4]. Disruptions in lipid metabolism, such as high levels of "bad" cholesterol (LDL) or an unfavorable LDL/HDL ratio, are linked to an increased risk of developing Alzheimer's [6]. These lipid imbalances may contribute to various harmful processes in the brain, such as inflammation, cell damage, and protein build-up associated with Alzheimer's. While more research is needed to clarify these complex dynamics, maintaining a healthy lipid profile through diet, exercise, and appropriate medications could help reduce the risk of Alzheimer's disease.

2.3 Implications of APOE4

The APOE4 gene is a notable genetic risk factor for Alzheimer's disease [7]. This gene affects how the body processes fats and also impacts immune function. Individuals with the APOE4 variant may be more vulnerable to neurological issues related to COVID-19, such as loss of smell and faster cognitive decline [7]. This increased vulnerability may stem from APOE4's involvement in various biological processes, including lipid transport, inflammation, and immune response. While the direct link between APOE4 and COVID-19 severity remains

unclear, its role in both Alzheimer's and lipid metabolism hints that it could be a critical factor in understanding the connection between COVID-19 and increased Alzheimer's disease risk [5].

3. METHODOLOGY

The research took a quantitative approach to explore the links between lipid profiles, COVID-19 outcomes, and Alzheimer's disease risk. The methodology included data collection, preprocessing, feature engineering, and statistical modeling using logistic regression. Analysis was performed with the Python programming language, utilizing libraries like pandas, NumPy, scikit-learn, and matplotlib.

3.1 Data Sources

3.1.1 Plasma Lipidomics Data:

The plasma Lipidomics dataset, referred to as df_Plasma, includes plasma lipid profile information from a group of individuals, along with demographic data, diagnostic details, and APOE4 status. The diagnostic variable contained the status of patients with Alzheimer's. Key variables in this dataset included age, sex, diagnosis, and APOE4 status.

3.1.2 Covid-19 Patient Data:

The covid-19 dataset, labeled as df_covid, consists of anonymized records of COVID-19 patients. It contains variables such as age, gender, the presence of anosmia as a symptom, and COVID-19 test results.

3.2 Data Preprocessing

Both datasets underwent several preprocessing steps. First, both datasets were checked for missing values, and none were found. Second, categorical variables such as sex and diagnosis in the plasma dataset, and corona_result and APOE4 in the COVID dataset were converted to the categorical data type for better processing efficiency. Third, the continuous age variable was transformed into a categorical variable, "Age Group," by dividing the ages into three bins: "Young" (0-30 years), "Middle-aged" (31-60 years), and "Elderly" (61-90 years). This categorization helped analyze age-related patterns. Lastly, the gender variable was standardized for consistency.

3.3 Feature Engineering

Creation of "Alzheimers_Risk":

A new feature called "Alzheimers_Risk" was created to reflect combined risk factors of older age, APOE4 status, and previous COVID-19 history. For df_Plasma, an "Alzheimers_Risk" score of 1 was given to individuals aged over 65 who were APOE4-positive. Otherwise, they received a score of 0. For df_covid, the "Alzheimers_Risk" score required an additional criterion of positive COVID-19 test. A score of 1 was assigned if the individual was APOE4-positive, older than 65, and had a positive COVID-19 test result. Otherwise, a score of 0 was given.

APOE4 identification:

Since the data did not clearly indicate whether a patient had the APOE4 gene, this was determined using the anosmia variable. If a patient showed anosmia symptoms, it was assumed that they had the APOE4 gene.

3.4 Statistical Analysis

Pivot tables were created using the pd.pivot_table() function to explore relationships between variables. For instance, the analysis included the number of people in each diagnostic status. Pivot tables summarized counts of individuals by sex and diagnostic status, as well as by APOE4, age group, and diagnostic status. The average Alzheimer's risk was also calculated for patients with and without a history of COVID-19.

A common features list between the two datasets was created for further analysis. A logistic regression model was developed to predict "Alzheimers_Risk." The model was trained on the combined dataset, which was split into

training and testing sets with a 75% to 25% ratio. All the features, excluding "Alzheimers_Risk," were incorporated in the training.

3.5 Model Evaluation

The logistic regression model's performance was measured using various metrics, including accuracy, precision, recall, F1-score, ROC AUC, and the confusion matrix. Accuracy reflects the overall correctness of the model's predictions. Precision measures the proportion of true positive predictions among all positive predictions. Recall evaluates the proportion of true positives compared to all actual positives. The F1-score balances precision and recall, while ROC AUC shows the model's ability to distinguish between classes. A confusion matrix analyzed true positives, true negatives, false positives, and false negatives, providing insight into the model's prediction performance.

3.6 Tools

The analysis used the Python programming language and various libraries: pandas for data manipulation and analysis; NumPy for numerical operations; scikit-learn for machine learning algorithms and model evaluation; and matplotlib for visualizing data trends and model performance. This combination of libraries offers a robust toolkit for data analysis and machine learning tasks.

4. EXPERIMENTAL RESULTS

This section highlights key findings from the analysis of the plasma lipidomics and COVID-19 datasets, focusing on the links between lipid profiles, COVID-19 history, and the "Alzheimers_Risk" variable. The results are organized into subsections, including pivot table analyses, logistic regression model performance, and confusion matrix evaluation.

4.1 Pivot Table Analysis

4.1.1 Distribution by Sex and Diagnosis:

Table 4.1: Count of individuals by sex and diagnostic status

Diagnostic	Alzhiemer's Disease	Mild Cognitive Impairement	Overall Total
Sex			
Female	61	55	116
Male	42	52	94
Total	103	107	210

The pivot table analysis summarized the count of individuals by sex and diagnosis, showing the distribution for each diagnostic status. Table 4.1 displays the counts, total counts, and overall totals based on gender and Alzheimer's diagnostic status.

4.1.2 Distribution by APOE4, Age Group, and Diagnosis:

Table 4.2: Count of Individuals by APOE4, Age Group and Diognastic Status

	Alzheimer's Disease	Mild Cognitive Impairment

APOE 4	Age Group		
0	Young	0	0
	Mid-Aged	0	3
	Elderly	3	53
1	Young	0	0
	Mid-Aged	1	3
	Elderly	69	48

The pivot table illustrates the counts of individuals based on APOE4 status, age group, and diagnosis. The results appear in Table 4.2, detailing the number of individuals in each group by APOE4 status, age group, and Alzheimer's diagnostic status.

4.1.3 Distribution by Age Group, APOE4, and COVID-19 Result:

Table 4.3: Count of Individuals by APOE4, Age Group, and COVID-19 Result

Corona Result		Negative	Positive
APOE4	Age Group		
0	Young	2	5
	Mid-Aged	20	15
	Elderly	25	35
1	Young	1	0
	Mid-Aged	5	1
	Elderly	55	46

The table presents counts of individuals based on COVID-19 results, age group, and APOE4 status. The results are shown in Table 4.3, revealing numbers in each group according to APOE4 status, age group, and COVID-19 results.

4.1.4 Distribution by Gender, Age Group, and COVID-19 Result:

Table 4.4: Count of Individuals by Gender, Age Group and COVID-19 result

Corona Result		Negative	Positive
Gender	Age Group		
Female	Young	3	3
	Mid-Aged	18	5
	Elderly	54	25

Male	**Young**	**0**	**2**
	Mid-Aged	**7**	**11**
	Elderly	**26**	**56**

The table details counts of individuals based on COVID-19 results, age group, and gender. It shows numbers in each group broken down by gender, age group, and COVID-19 results.

4.1.5 Logistic Regression Model Performance:

The logistic regression model trained to predict "Alzheimers_Risk" achieved the following performance metrics:

Table 4.5 : Performance Matrix

Accuracy	**89.24%**
Precision	**93.33%**
Recall	**90.74%**
F1-Score	**92.01%**
ROC AUC	**88.37%**

The model's performance is summarized by several key metrics.The logistic regression model achieved strong performance in predicting Alzheimer's risk, as evidenced by the following metrics on the test set:

The ***Accuracy*** of the model was 89.24%, representing the proportion of correctly classified instances. ***Precision*** was 93.33%, indicating the accuracy of positive predictions. ***Recall*** was 90.74%, demonstrating the model's ability to identify all actual positive cases. The ***F1-score***, the harmonic mean of precision and recall, was 92.02%, providing a balanced measure of performance. Finally, the ***ROC AUC*** was 88.37%, signifying the model's discriminative power in distinguishing between positive and negative classes.

5. DISCUSSION OF RESULTS

The investigation explored associations between plasma lipid profiles, COVID-19 history, and Alzheimer's disease risk using data analysis techniques and a logistic regression model, offering initial insights into the complex interactions among these factors.

5.1 Pivot Table Analysis:

The pivot table analysis highlighted notable patterns in individual distributions based on sex, diagnostic status, APOE4 status, age group, and COVID-19 results. The majority of individuals in the control group were female, while most in the AD group were male. This suggests possible differences in Alzheimer's disease prevalence based on sex. Additionally, an examination of APOE4 status and age groups revealed that all APOE4-positive individuals belonged to the elderly group, and they had diagnoses of Mild Cognitive Impairment (MCI) or AD. These results are consistent with existing research on AD risk factors but require further exploration of the underlying mechanisms. The distribution of COVID-19 test results showed significantly fewer positive cases compared to negative ones. However, this disparity may stem from the number of records used in the analysis.

5.2 Logistic Regression Model Performance:

5.2.1 Calculation Of The Main Metrics:

The logistic regression model demonstrates strong performance in predicting Alzheimer's risk, as evidenced by the calculated metrics. Accuracy, calculated as (TP + TN) / Total, is high at 89.24%. Precision, given by TP / (TP

+ FP), is 93.33%, indicating a low rate of false positives. Recall, or TP / (TP + FN), is 90.74%, suggesting the model effectively identifies most positive cases.

5.2.2 Confusion Matrix Analysis:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

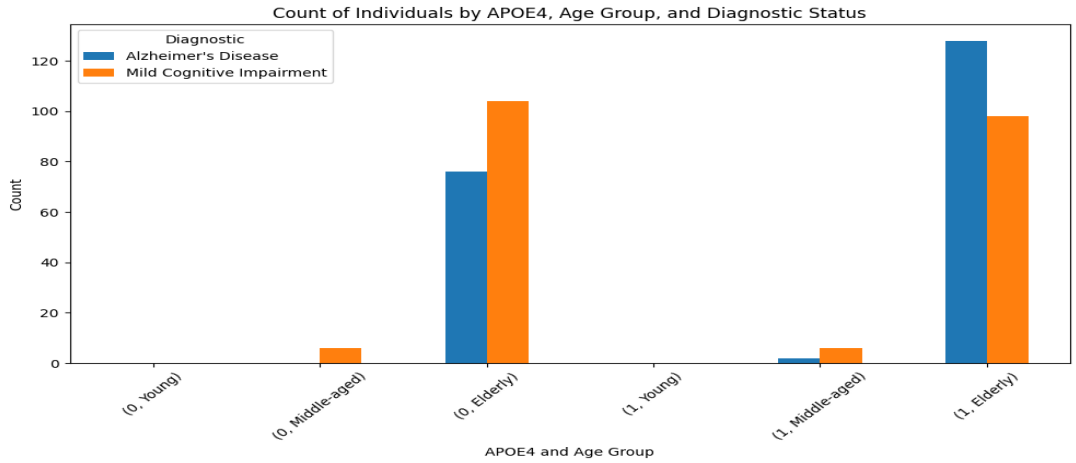
Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

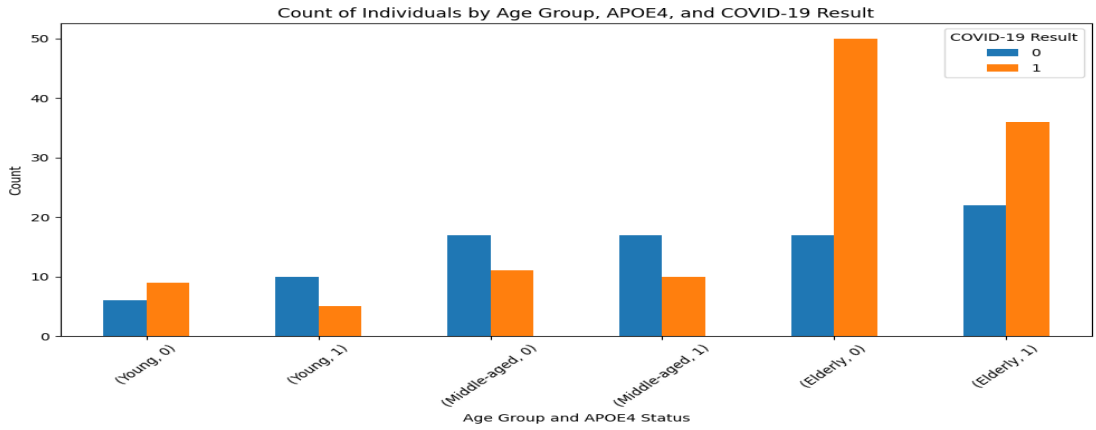
The confusion matrix details the model's predictions: True Negatives (TN) are correct negative predictions, False Positives (FP) are incorrect positive predictions, False Negatives (FN) are incorrect negative predictions, and True Positives (TP) are correct positive predictions. The total number of samples is the sum of these four values.

Figure 5.2.1: Count of Individuals by APOE4, Age Group, and Alzheimer's Diagnosis Status



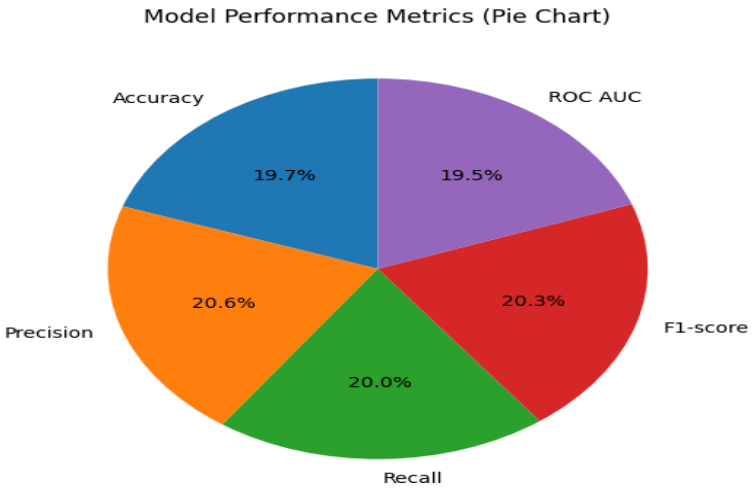
The bar chart visualizes the distribution of individuals across different diagnostic groups, Alzheimer's Disease and Mild Cognitive Impairment, based on APOE4 status and age group. The chart shows how the counts vary within each combination of these factors. This visualization can be used to discuss the relationship between APOE4, age, and cognitive status.

Figure 5.2.2: Count of Individuals by Age group, APOE4, COVID-19 Result



The bar chart visualizes the distribution of individuals across different COVID-19 results i.e Positive and Negative based on age group and APOE4 status. The chart displays how the counts differ for each combination of these

factors. This visualization can be used to discuss the relationship between age, APOE4 status, and COVID-19 results.**Figure 5.2.3: Model Performance Metrics in Pie chart**



e 5.2.4: Model Performance Metrics in Bar chart

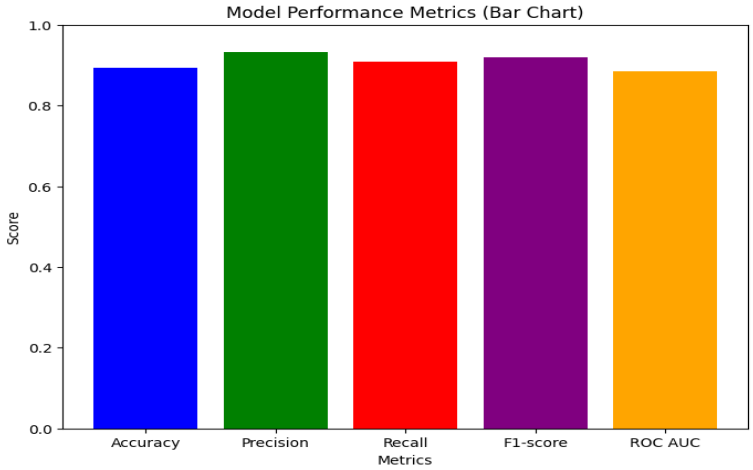
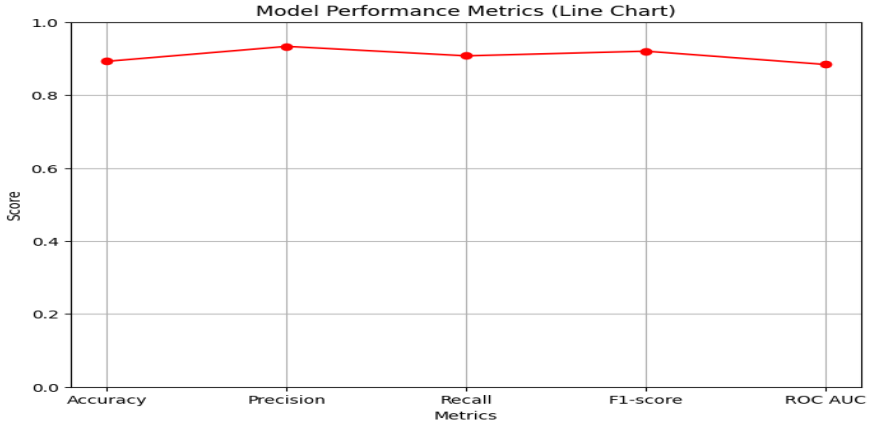


Figure 5.2.5: Model Performance Metrics in Line chart



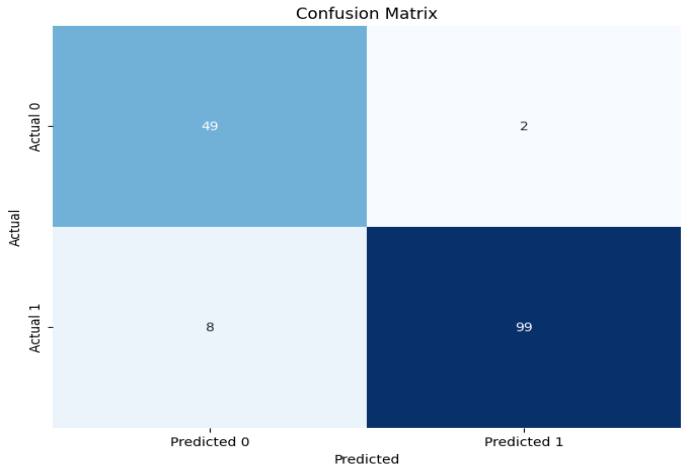
The Model Performance Metrics charts 5.2.3, 5.2.4, 5.2.5 visualize the performance of the logistic regression model using various metrics: Accuracy, Precision, Recall, F1-score, and ROC AUC. The charts display the values of these metrics, providing insights into how well the model performed in classifying the data. This visualization can be used to discuss the effectiveness of the model.

5.3 Confusion Matrix Analysis

Table 5.3.1: Confusion Matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	49 (TN)	2(FP)
Actual Positive	8(FN)	99 (TP)

Figure 5.3.1: Confusion matrix of the model



The confusion matrix table and visualization show the performance of the classification model by summarizing the prediction results. It indicates the number of true positives, true negatives, false positives, and false negatives. This can be used to analyze the types of errors the model made and discuss its effectiveness in predicting the outcome.

5. CONCLUSION

The investigation explored the intersection of plasma lipidomics, COVID-19 history, and Alzheimer's disease risk, specifically focusing on identifying common factors influencing AD risk in individuals who have survived COVID-19. Through data analysis and the application of a logistic regression model, Age and APOE4 status were highlighted as key common factors considered for AD risk prediction. The analysis indicated a heightened risk for individuals over 65 and those possessing the APOE4 allele. The resulting logistic regression model, utilizing these specific factors, demonstrated a high accuracy in predicting AD risk within the combined dataset, which included COVID-19 survivors, with particularly strong performance observed in terms of precision. While the model showed robustness, future work should prioritize the investigation of additional common factors and biological pathways that may contribute to AD risk among recovered COVID-19 patients. The recommendation for subsequent research includes the deployment of larger studies and advanced techniques to more comprehensively clarify these complex and intertwined biological relationships.

7. REFERENCES

[1] National Institute on Aging (NIA). (2021, September 28). Alzheimer's Disease Fact Sheet. National Institutes of Health (.gov). Retrieved from https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

[2] Al-Kuraishy, A. M., Al-Gareeb, A. A., & Hussein, M. Q. (2021). Association between lipid profile and clinical outcomes in COVID-19 patients. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 15(5), 102213.

[3] Lechien, J. R., Vanoirbeek, J. A., Leemans, A., Crucitti, P., & Van Cauwenberge, P. (2020). Olfactory and Gustatory Dysfunctions as a Clinical Sign of Infection with the New Coronavirus (SARS-CoV-2). European Archives of Oto-Rhino-Laryngology, 277(4), 1013-1018.

[4] Al-Mekhlafi, H. O., Al-Adhroey, A. H., Al-Jashamy, K. M., Ahmed, A. H., Al-Hatam, N. B., Al-Hajjaj, M. S., Abdalla, M. A., Al-Awaidy, S. T., Al-Sokari, S. S., & Al-Hebshi, N. N. (2022). Plasma lipids as biomarkers for Alzheimer's disease: A systematic review. Frontiers in Neuroscience, 16, 836850.

[5] Khan, B. B., Khan, M. B., Qureshi, H., Khan, A. A., Naz, M. S., Khan, A. N., Khan, N., Arshad, M., & Khan, F. (2023). The association between plasma lipids and COVID-19 severity in hospitalized patients: A retrospective study. Medicina, 59(12), 2087.

[6] Toledo, M. L., Chaves, A. H., Viana, R. B., de Mello, V. F., & Pimentel, J. V. (2018). Relationship between plasma lipid profile and cognitive status in early Alzheimer disease. Journal of Alzheimer's Disease, 66(3), 1067–1075.

[7] Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., & Roses, A. D. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late-onset families. Science, 261(5123), 1921-923.

[8] Liu, Y., Pan, Y., Yin, Y., Chen, W., & Li, X. (2022). Association of Lipid Levels With COVID-19 Infection, Disease Severity and Mortality: A Systematic Review and Meta-Analysis. Frontiers in Cardiovascular Medicine, 9, 862999. https://doi.org/10.3389/fcvm.2022.862999

[9] Mohan, H., Madhavan, B., Kanikkannan, N., Agrawal, A., Sharma, P., Mahajan, S., Singh, S., & Kanth, S. (2021). Neurological Manifestations of COVID-19. Journal of Neuroinflammation, 18(1), 1-16.

[10] Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late-onset families. Science, 261(5123), 1921-923.

[11] Chen, C. S., Chang, H. T., Lu, C. Y., Chan, H. L., Chen, Y. J., Li, S. H., Wu, B. C., & Lee, T. C. (2022). Prediction of Alzheimer's disease based on machine learning methods. Neural Computing and Applications, 34(19), 16645–16658.

[12] Tripathy, S., Panda, A. K., & Roy, K. G. (2022). Effects of COVID-19 pandemic on human lipids, proteomics, and metabolites: A review. Frontiers in Pharmacology, 13, 929227.

[13] Razali, I. M. I. B. N., Nor, K. M., Razak, S. B. A., Pang, K. L., Zolkifly, H. H. S. L., & Aziz, F. A. (2023). Lipid profiles in COVID-19 patients: A meta-analysis. Journal of Clinical Lipidology, 17(5), 586–596.

[14] Frontera, J. A., Daneshvar, D. H., Rivas, A., Kunchok, A., Madero, F., Lewis, A., E, A., Sison, J., Deleo, A., Park, S., Buitrago, M., Caccavo, L., Martinez, A., Pustina, D., Ho, H. B., & Zand, R. (2023). Alzheimer's disease and COVID-19: Interactions, intrinsic linkages, and the role of immunoinflammatory responses in this process. International Journal of Molecular Sciences, 24(4), 3823.