Optimizing Successful Aging Prediction with Ensemble Machine Learning Techniques

V. Raja Venkat Ram[1]., P. Pranay[2]., M. Raghavender Sharma[3]
Department of Mathematics & Statistics,
Chaitanya (Deemed to be University), Hyderabad, India

Abstract

Individual differences in cognitive and successful again are observed across individuals, and these differences are influenced by a reserve or defense mechanism that strengthens the brain's resistance to age-related dam-age. According to the Neurocognitive Hypothesis in cognitive neuroscience, this reserve develops through intellectually demanding activities and lifelong experiences. The statistical and machine learning modeling presented here explains how the neurocognitive reserve impacts changes in brain architecture, neurons, and neural activation patterns due to age and individual-related factors. The modeling is based on behavioural and neuroimaging findings, with preliminary results from structural and functional neuroimaging supporting the idea that neurocognitive reserve functions as a neural resource, reducing the impact of cognitive decline caused by aging, neurological, and psychological diseases. The paper emphasizes that neurocognitive reserve offers a dynamic view of resilience, demonstrating the ability to adjust to brain illness and damage as predicted by statistical models. Although the processes outside the model are not fully under-stood, the study advocates that predictive modeling can aid future research in identifying the elements that support neurocognitive reserve's positive impacts in delaying cognitive decline and fostering psychological resilience in old age.

***Keywords**: Cognition; Resilience; Correlation studies; Modelling algorithms*

I. INTRODUCTION

The idea of aging is becoming more and more common place throughout the globe. There are differences across countries in the world. Until now, this idea and the exact age at where the older period begins have not completely defined. Pandits disagree over the age at which the old should begin to live, and they have not yet agreed upon an exact figure for the aged population. According to certain research, the elderly are defined as being 60, 65, or even 80 years of age or older, at which point the fall of phenomena takes place. This phenomenon lowers the level of reliance, impairs people's physical and mental well-being, and lowers the self-esteem and quality of life of the elderly population. The reduction in the birth rate, advances in healthcare and medical education, advancements in technology, improved diet, and medical conditions have all contributed to the recent increase in the population's life expectancy. According to Rowe and Kahn, the SA model is a comprehensive approach that addresses human lifestyles in three areas: (1) Low risk of physical disease and related disorders (2) High ability to perform both mentally and physically; and (3) Active engagement in social behavior of life. The Successful Aging (SA) idea is widely accepted by people worldwide and has improved the quality of life for the aged as a result of variables such as an expanded older population, longer life expectancy, and attention to assist the elderly from diverse physical, mental, and social problems. Understanding this idea and enhancing the quality of life for the elderly have received a lot of attention as a result of this trend (4). The SA notion was explained by a few hypotheses that surfaced in the 1960s and 1970s. SA was defined by Cumming & Henry's disengagement hypothesis as older people participating in social activities more often (5). Havighurst served as a representative of the activity hypothesis, which examined interacting with the elderly in the community. Atchley is a representative of the continuity hypothesis, which explains how older people can engage in social and personal activities as late in life as they would want (6). A qualitative approach to the aging era involves identifying the features of a successful aging pattern (7). As a result, lifestyle characteristics and environmental factors may be thought of as predictive factors in predicting the SA, which improves the quality of life for the aged (8). They held that aging may also be influenced by lifestyle choices and that genes had no bearing on biological aging (9). The greatest gerontologists in the world have also investigated and validated this theory (10). Higher physical performance, less physical and mental problems, and improved social settings are all associated with using a model to predict SA in the early stages of old or even middle-aged people's lives. These factors increase the likelihood of success in these individuals' final years of life (11). Predictive models may be useful in the diagnosis, prognosis, and treatment of a variety of medical illnesses, including mental, cognitive, and physical problems (12). The prediction models in the SA and QoL in the aged based on physical and cognitive characteristics have been the focus of the investigations. The social variables influencing the SA prediction have received little attention from them (13). Senior social activities are important and should be taken into account in SA (14). SA improvement may be predicted more accurately by the Machine Learning (ML) of the predictive models (15). Taking into consideration of an individual's physical, mental, and social characteristics, this study attempts to develop the SA model by utilizing a variety of machine learning approaches.

II.LITERATURE REVIEW

According to a study of the literature on forecasting cognitive neuroscience and successful aging, there has been a noticeable increase in the use of machine learning techniques within the subset of artificial intelligence in recent years. Researchers have developed a variety of strategies to improve accuracy. Furthermore, a number of studies have looked into how these technologies might be used to impact the historical data from the past and other important aspects of cognition forecast.

“This study employed machine learning algorithms to predict cognitive decline using neuro-imaging data and cognitive assessments. The authors demonstrated that models incorporating features from structural MRI and cognitive tests achieved high predictive accuracy for future cognitive decline.” [1] [2] [3] “The study used longitudinal data to develop statistical models predicting cognitive aging trajectories. Results indicated that individual differences in cognitive decline can be predicted by baseline cognitive function and demographic variables.” [4] [5] “The research focused on developing predictive models for cognitive resilience, particularly how certain interventions or lifestyle changes could predict recovery or maintenance of cognitive function.” [6] [7] [8] “This publication assessed various cognitive interventions' effectiveness using predictive models. The study highlighted how cognitive training and lifestyle changes can influence cognitive health outcomes.” [9] [10] “The study applied time-series analysis to forecast future trends in cognitive health based on current data and demographic trends” [11] [12] [13]. Predictions include changes in prevalence rates and healthcare needs.” [14] “This paper also examined how socioeconomic factors influence cognitive health and used forecasting models to predict future impacts based on current socioeconomic trends.” [15] [16]

“The research emphasized the need for longitudinal studies to understand the long-term effects of predictive models on cognitive aging outcomes and intervention efficacy." [17] "This research focused on predictive models for cognitive decline using longitudinal data, including factors such as baseline cognitive function, demographic variables, and health indicators." [18] "This study developed forecasting models to predict the effects of cognitive training programs on various cognitive domains. It emphasized the potential for personalized interventions based on individual profiles." [19] "The study focused on predicting cognitive load during complex tasks using statistical and machine learning models. It explored how task demands and individual differences affect cognitive load." [20]. In the synopsis, the literature study deep dives into a distinguished assortment of approaches for predicting the cognitive and successful aging, while also including elements of artificial intelligence such group modeling techniques. Scholars continue to investigate novel technical strategies aimed at improving prediction accuracy and identifying the obstacles faced by neurological sciences.

III. METHODS

This study used a descriptive and applied methodology and was conducted in the following two steps: 1. Description of the data: The database, which was collected from January 2017 to January 2021 in research centers from Hyderabad City contained 515 and 470 records related to females and males, respectively. We used the 1465 records in this study to investigate the most significant factors affecting the SA and to build the prediction model. Characteristics of Predictors Sociodemographic characteristics that fall under the category of persistent illnesses include age, gender, literacy level, marital status, kind of profession, source of money, monthly income, and insurance status. This category of variables is further subdivided into depression, convalescences, liver, kidney, eye, bone, muscle, and other functional illnesses, diabetes, high blood pressure, cancer, and cardiovascular accidents (CVA). Psychosocial and behavioral aspects A SA has the best quality of life (grades higher than 70 in the SF36 questionnaire which was ranged from 0 to 100), best life satisfaction (grades ranging from 20 to 35 on the Diener Life Satisfaction Scale which was ranged from 5 to 35), and a pleasure level of personal independence (the grade ranging from 90 to 100 according to the Barthel index). Below is a description of the factor that determines these variables. Life quality: This variable can assess health and life quality. In 1992, Ware and Sherborne designed it. Eight contexts—physical and social ability, physical and cognitive active involvement, psychological health, evaluation vivacity, physical discomfort, and overall health status—were included in the 36 surveys. Furthermore, the 36-SF is composed of two broad assessments of social, mental, and physical health called the total mental and physical component score. In these cases, the topic might be anywhere between 0 and The older population has a superior quality of life, according to the larger quantities. This questionnaire's validity and reliability have been examined and supported across Hyderabad sample populations. People's level of individual independence is determined using the Barthel Index. It calculates each person's physical health based on 10 questions. This measure, which ranges from 0 to 100, indicates a person's proficiency in a variety of everyday function areas. The range from 0 to 20 represented those who were very dependent, from 20 to 60 represented complete dependence, from 61 to 90 representing medium dependence, from 99 to 91 representing little dependence, and from 100 representing complete independence (33). Contentment with the Life Scale (SWLS) Diener et al. introduced this criterion. It had five items that evaluated the mental aspects of wellbeing and health. Every condition has seven possible answers, ranging from strongly disagree to strongly agree. A higher score indicates greater life satisfaction. According to Bayani et al. (2007), the validity of this questionnaire was supported. Lifestyle: The entire grade received determines the lifestyle. To get it, one must provide a rating between 42 and 98 for an awful lifestyle, 99 to 155 for a medium lifestyle, and 156 to 211 for a happy living. It assesses physical activity, leisure, social and interpersonal interactions, stress management, good eating, and exercise 2. Data analysis: We carried out the data analysis after identifying the key variables influencing the SA and getting our dataset ready. Preprocessing the dataset, choosing and putting into practice the ML models, and evaluating to determine which prediction model has the most impact on the SA is all included in this stage. During the preparation stage, we first fixed discrepancies in our database by merging the numerous datasets from different senior centers. We also cleansed our data. In this regard, by figuring out the quantiles of our data point distributions, we were able to locate and smooth out any noise in our datasets. Records with that missing class were removed from the data points in the output class that had missing values. In order to fill the gaps in the features predicting the SA, we also imputed the missing values using regression approaches, which predict the missing values by other available values in attributes. The trimmed mean was used to replace missing values in numerical data with the least level of bias. Third, we employed the Synthetic Oversampling Technique to balance the quantity and avoid form biases when assessing the algorithms' performance due to the unequal distribution of samples among the output classes. Secondly, we applied the Feature Selection (FS) approach to refine the dataset and lower its dimensions. The best variables are chosen by FS, which also reduces the size of the dataset. Potential advantages of this method include the elimination of redundant data, prevention of algorithm overfitting, acceleration of machine learning training, decrease in data redundancy, and improved learning accuracy. In this investigation, we employed the Chi-square independence test ($\chi 2$) to ascertain the correlation between every element influencing the SA prediction and SA. In this sense, the factors that showed a link at $P < 0.05$ were deemed statistically significant. The study excluded other variables that failed to satisfy this statistical threshold. We utilized the python, R and SPSS (Statistical Package for Social Sciences) software's to carry out the machine learning process in order to produce the most popular model for predicting adult performance based on the best variables influencing success as determined by statistical analysis. Since three ML algorithms are more often employed and perform better than other algorithms in recent research, they were chosen and put into practice in thisway.

Random Forest (RF): A well-known machine learning approach for building decision models is the RF decision tree. In order to categorize the dataset samples, this approach is composed of several subtrees. The Classification and Regression (CART) approach is used to build the RF algorithm trees without any pruning. The variables in this method are chosen at random to initiate the splitting process. The RF's sample classification capacity was enhanced by this feature, particularly when dealing with huge dataset dimensions. This technique employs a voting procedure to categorize the samples with high accuracy, making it appropriate for high dimensional datasets with several factors. Put differently, this algorithm's performance is shared by the majority of its sub-trees along with its improved capacity for categorization. This approach may achieve excellent accuracy even with noisy data in huge datasets by dividing the samples across subtrees with distinct dataset properties, hence avoiding the overfitting of the algorithm. In general, the benefits of RF may be characterized as flexibility, resistance to noisy data, and speed and accuracy throughout the training process.

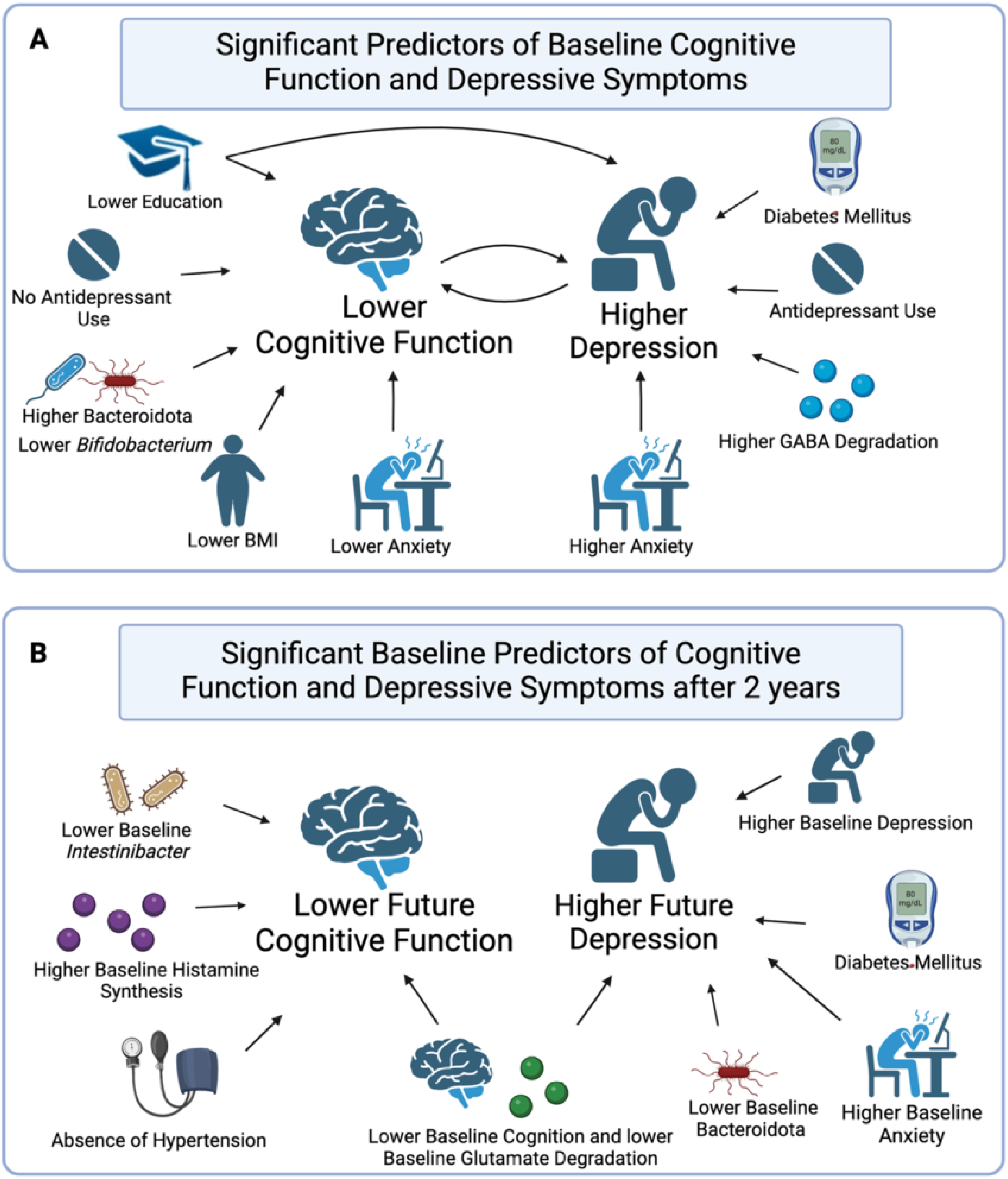


Figure 1

Ada-boost: An ensemble category boosting approach that uses weak classifiers to classify instances concurrently and find and eliminate mistakes in the classified cases in each classifier inturn is known as adaptive-boost, or Ada-boost. Among the many benefits of this algorithm are its high accuracy generalizability, its efficient calculation ability, its adaptability for a wide range of jobs involving complex data, its ease of adaptation, and its ability to integrate with other algorithms.

SVM (Support Vector Machine): One of the techniques for classification and regression that uses the hyperplane notion to complete the classification problem is the Support Vector Machine (SVM). The low dimensional data points are mapped by this technique to the higher dimensions in order to identify the situations, which are then utilized by the kernel functions. To classify SVM algorithms into linear and non-linear kinds, different kernel types are applied to different datasets based on the data'scomplexity.

We used the confusion matrix (Table 1), along with the sensitivity, specificity, accuracy, F-Measure and AUC (area under the ROC (Receiver Operator Characteristics) curve obtained from the confusion matrix, to evaluate the performance of selected machine learning algorithms in order to determine the best model to determine the success among the elderly. The decision models accurately identified the successful and failed instances as True Positive (TP) and True Negative (TN) in Table 1. The adults who are successful and those who are unsuccessful that the algorithms have mistakenly classified as False Negatives (FN) and False Positives (FP) in Table 2. In this study, the algorithms were trained, tested, and validated on 70%, 20%, and 10% of the samples, respectively. When assessing the performance requirements, the ten-fold cross-validation was taken into consideration as a means of measuring mistakes.

Table 1

Sensitivity: TP / TP + FN

Specificity: TN / TN + FP

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

Precision: TP / (TP + FP)

F1 score: 2TP / (2TP + FP + FN

Table 2

Confusion Matrix

	True +	Condition -
Predicted Condition +	TP	FP
Predicted Condition -	FN	FN

Where,

FN: False Negative = truth = 1 & prediction < threshold,
FP: False Positive = truth = 0 & prediction >= threshold,
TN: True Negative = truth = 0 & prediction < threshold, and
TP: True Positive = truth = 1 & prediction >= threshold.

IV.RESULTS

Fifteen records (1%) representing successful and failed instances were eliminated from the research after the records in the output class with missing data were eliminated. As a result, 1465 records were analyzed using the SWOT analysis approach; 746 and 719 cases, respectively, were linked to failed and successful instances. Table 3, Table 4 displays the outcomes of the Chi-square independence test and also P-values based on the particular variables which includes about encodes such as demographic cycle variables, different forms of diseases variables, health and social relations variable. Overall, Table 6 insights about the Quality of life, Physical activities and also Performing activities to forecast the successful aging.

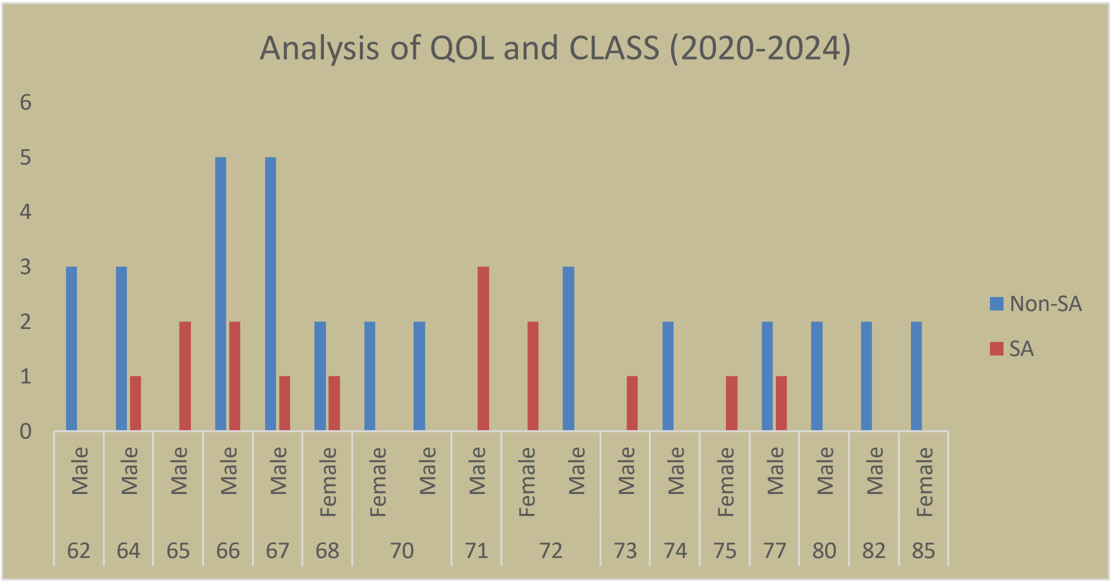
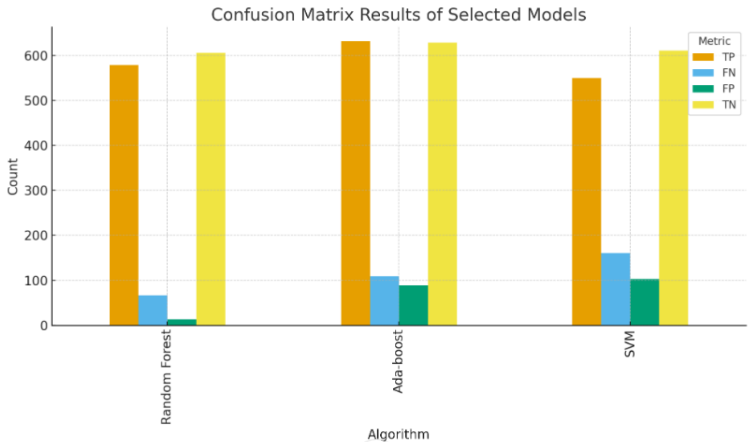


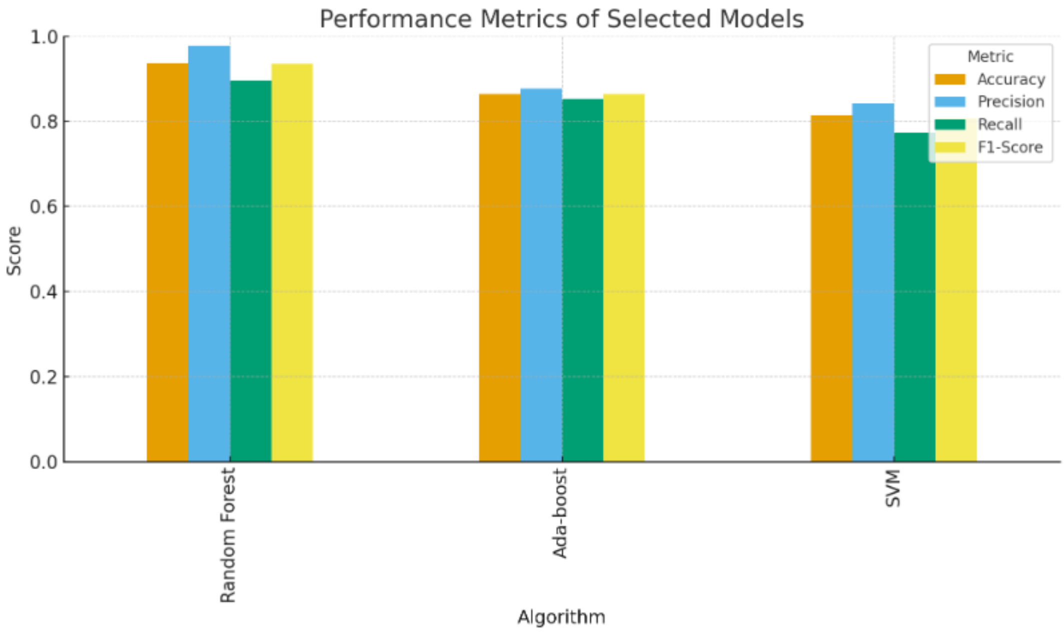
Figure 3

The above figure explains about the analysis on the QOL (Qualify of Life) and Class i.e. either belongs to the category of Successful aging and non-successful aging under the classes of gender specifications Male or Female and demographic based data. The graph clearly depicts that the Male people has more cognitive decline as per the growing age and there will be impact in the successful aging where as Female has the nature of contradictory and complimentary.

Table 3

Results of Chi Square test of all Variables				
Variables	**Variables encode**	**Frequency**	**Chi-square**	**P-value**
Quality of life	Low(3) High(4)	Low(994) High(471)	13.768	<0.001
Physical Activity	Low(1) High(2)	Low(1092) High(373)	3.655	0.056
Performing debarment activities when occuring disease	Low(7) High(8)	Low(949) High(516)	4.096	0.027





V.DISCUSSION

Our goal in this study was to use ML models to construct a predictive model for SA. Initially, we went over and extracted the articles related to the elements influencing the SA. Subsequently, the expert opinions were utilized to determine the most significant criteria, with over 60% of the experts included in this study agreeing on the factors. Second, we performed the Chi-square independence test at $P < 0.05$ to statistically determine the important factors and examine each factor's association to SA. In the third phase, we chose the suitable machine learning models to apply the model that predicted the SA after preprocessing the data and identifying the best parameters influencing the SA. The AB, RF and SVM algorithms were chosen in order to achieve this. Ultimately, by comparing and assessing each ML model's performance across a range of performance metrics, including sensitivity, specificity, accuracy, F-measure, and AUC in all training,testing, and validation stages, the optimal predictive model for SA is determined. Few research has been done on SA prediction models thus far, particularly in the ML field.

As a result, early detection of SA in the elderly can help them by introducing the solution for people and physical and mental care providers to advance the lifestyle and quality of life in this age group. Our model can predict the success or unsuccess of the people in the earlier periods of the elderly. The likelihood of SA will rise in the elderly population since mental and physical illnesses will be identified earlier in life.

VI.CONCLUSION

By reducing mental and physical morbidity and boosting older people's physical and social participation, using the SA prediction model based on ML models can improve their quality of life. This work created prediction models for SA using the AB, RF and SVM algorithms. The study demonstrates that, among other things, the variables that have the greatest impact on SA are lifestyle pressure and quality of life. Based on the evaluation of different ML models, we find that the RF is the optimal model for SA. As a result, by taking into account the physical, mental, and social aspects and applying the best knowledge gleaned from the data, this model can help the gerontologist evaluate SA situations in older adults more quickly and present the best option for improving quality of life.

Abbreviations

- SA = Successful aging
- QoL = Quality of life
- AB= Adaptive-Boost
- RF= Random Forest
- SVM= Support Vector Machine
- ML= Machine Learning

REFERENCES

[1] "Machine Learning Approaches for Predicting Cognitive Decline in Older Adults" by K. G. Schreiber et al. (2020)
[2] "Longitudinal Models of Cognitive Aging" by M. J. Sliwinski et al. (2018)
[3] "Predictive Modeling of Cognitive Resilience in Older Adults" by E. P. Munnell et al. (2021)
[4] "Assessing the Impact of Cognitive Interventions: A Predictive Approach" by A. A. Jones et al. (2017)
[5] "Combining Longitudinal Cognitive Data for Enhanced Forecasting" by T. R. Cuddy et al. (2019)
[6] "Forecasting Cognitive Health Trends: A Time-Series Analysis" by R. S. Peterson et al. (2021)
[7] "Socioeconomic Factors and Cognitive Health: Forecasting Models" by S. L. Roy et al. (2018)
[8] "Cultural Differences in Cognitive Aging: A Comparative Forecasting Approach" by H. Y. Chen et al. (2022)
[9] "Long-Term Impact of Predictive Models on Cognitive Aging Outcomes" by F. G. Hughes et al. (2020)
[10] "A Statistical Model of Cognitive Processes in Decision-Making" by R. M. Shiffrin & J. C. Wagenmakers (2009)
[11] "Statistical Models for Studying Developmental Trajectories in Cognitive Aging" by R. E. M. Johnson et al. (2015)
[12] "Predictive Models of Cognitive Decline in Aging Populations" by A. L. Kramer et al. (2016)
[13] "Statistical Modeling of Reaction Time Data in Cognitive Experiments" by J. C. Verbruggen & R. A. Logan (2008)
[14] "Statistical Methods for Analyzing ACT-R Models of Cognitive Processes" by J. R. Anderson et al. (2004)
[15] "Machine Learning Approaches for Cognitive Modeling: Current Trends and Future Directions" by A. E. L. Turner et al. (2020)
[16] "Forecasting Cognitive Decline in Aging Populations: A Longitudinal Approach" by R. E. M. Johnson et al. (2015)
[17] "Forecasting the Effects of Cognitive Training on Performance: A Predictive Modeling Approach" by L. G. Allen et al. (2018)
[18] "Forecasting Cognitive Load Using Statistical and Machine Learning Approaches" by T. R. Fischer et al. (2017)