

## Research Article

# Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test

A. Revathi <sup>1</sup>, R. Kaladevi <sup>2</sup>, Kadiyala Ramana <sup>3</sup>, Rutvij H. Jhaveri <sup>4</sup>,  
Madapuri Rudra Kumar <sup>3</sup> and M. Sankara Prasanna Kumar <sup>3</sup>

<sup>1</sup>Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai, India

<sup>2</sup>Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India

<sup>3</sup>Department of Computer Science and Engineering, Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India

<sup>4</sup>Department of Computer Science and Engineering, Pandit Deendayal Energy University, Gandhinagar, India

Correspondence should be addressed to Rutvij H. Jhaveri; [rutvij.jhaveri@sot.pdpu.ac.in](mailto:rutvij.jhaveri@sot.pdpu.ac.in)

Received 1 December 2021; Revised 1 January 2022; Accepted 5 January 2022; Published 20 January 2022

Academic Editor: Celestine Iwendi

Copyright © 2022 A. Revathi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Elderly people are the assets of the country and the government can ensure their peaceful and healthier life. Life expectancy of individuals has expanded with technological advancements and survey tells that the elderly population will become double in the year 2030. The noninfectious cognitive dysfunction is the most important risk factor among elderly people due to a decline in their physiological function. Alzheimer, Vascular Dementia, and Dementia are the key reasons for cognitive inabilities. These diseases require manual assistance, which is difficult to provide in this fast-growing world. Prevention and early detection are the wise solution for the above diseases. Diabetes and hypertension are considered as main risk factors allied with Alzheimer's disease. Our proposed work applies a two-stage classification technique to improve prediction accuracy. In the first stage, we train a Support vector machine and a Random Forest algorithm to analyze the influence of diabetes and high blood pressure on cognitive decline. In the second stage, the cognitive function of the person with the possibility of Dementia is assessed using the neuropsychological test called Cognitive Ability Test (CAT). Multinomial Logistic Regression algorithm is applied to CAT results to predict the possibility of cognitive decline in their postlife. We classified the risk factor using the operational definitions: "No Alzheimer's," "Uncertain Alzheimer's," and "Definite Alzheimer's". SVM of stage 1 classifier predicts with an accuracy of 0.86 and Random Forest with an accuracy of 0.71. Multinomial Logistic algorithm of stage 2 classifier accuracy is 0.89. The proposed work enables early prediction of a person at risk of Alzheimer's Disease using clinical data.

## 1. Introduction

Physical health and mental health carry equal importance in human life. Elderly people are normally affected with cardiovascular disease, cancer, diabetic, arthritis, depression, kidney disease, pulmonary disease, dementia and Alzheimer's disease. Dementia is a cognitive decline in mental ability, which severely affects routine life. A person suffering from dementia is always in need of someone to accomplish his everyday activity since the disease affects cognitive function in multiple domains. Alzheimer's Disease (AD) is one among the overall general neurodegenerative cortical

dementia. The incurable neurodegenerative disorder primarily affects the elderly population. It gradually progresses from mild cognitive impairment to Alzheimer's and other kinds of Dementia. The projections are specifically high in South Asian countries such as India and China. The rise in AD disease is proportionate to the elderly population and it is foreseen that 5% to 7% of elders are affected by dementia. By 2050, 1 in 5 persons of low- and middle-income countries will be above 60 ages which may escalate the disease population [1].

Dementia will be an inevitable result of demographic transition and it causes damage to the brain cells. The stages

of dementia span start with no cognitive decline to severe decline. The different types of dementia are Alzheimer's Disease (AD), Vascular Dementia (VaD), Frontotemporal Dementia, etc. This impairment affects the capacity of synapses to converse with one another which in turn affects person's thinking, emotion, and behavior. Various sorts of dementia align with a specific type of brain cell decay in brain regions. A significant level of specific proteins presents inside and outside of synapses, making it difficult for brain cells to remain healthy and to connect with others. The foremost section to be affected is "*Hippocampus*" region of the brain cell, which is the central point of learning and memory in the cerebrum. This is the reason why cognitive decline is perhaps the initial indication of Alzheimer's Disease. There is no effective handling or treatment available for the disease. The feasible option is to train the population with related risk factors and the defending factors.

People affected by diabetes are growing exponentially and it is expected that 640 million people will be affected by the year 2040 [2,3]. As indicated by the World Alzheimer Report 2014, people who had hypertension in their midlife (individuals age around 40–64 years old) were bound to create vascular dementia in later life [4]. Choked or decreased blood flow to the brain is the basic symptom of dementia. Many people with diabetes have brain changes that are a hallmark of Alzheimer's disease. Hypertension causes hurt on the heart and veins and it happens when the power of blood pushing against within our veins is excessively high. This causes the cells to work tougher, which makes them less effective. A recent exploration in a journal named Neurology Trusted Source shows that elderly people have more average BP that is likely to create tangles and plaques in brain. There exists some evidence for a relation of SBP with AD, specifically tangles [5,6].

Multifactor analysis predicts Alzheimer's disease more precisely by extracting heterogeneous information present in health records. It is feasible to predict AD using administrative, clinical information rather than images. Machine learning algorithms are the ideal alternative to apply to a large volume of health data [7]. The focus of our proposed work is twofold: (i) Predicting people with possibilities of Alzheimer in their late life by doing careful analysis on various risk factors associated with Alzheimer's. (ii) Conducting a neuropsychological test called Cognitive Ability Test (CAT) to assess the cognitive decline of a person [8]. The proposed work considers general health data available in "Data World" repository. We apply 2-stage classifier algorithms in the proposed work. In the first stage, support vector machine learning algorithm and Random Forest algorithm are used to find the associated risk factor of individuals. In the second stage, to enhance the prediction accuracy, cognitive ability test was conducted among the people identified by the stage 1 classifier. The cognitive ability of a person is estimated using CAT test, which contains simple yes or no type questions, values ranging from 0 to 30. The CAT test results are applied to Multinomial Logistic Regression to classify the severity of the disease. The score between 25 and 30 is classified as "No Dementia," between 13 and 24 as "Uncertain Dementia," and less than

13 as "Severe Dementia." The proposed work combines multiple factors associated with Alzheimer's to predict the possibility of disease more accurately.

The paper is ordered as follows: Related work on dementia and Alzheimer's disease is explored in chapter 2. Chapter 3 describes the proposed work that includes the relation between Alzheimer's with type 2 diabetes and hypertension dataset, which relates to our claim. The application of multinomial logistic expression on CAT test results to enhance the prediction process is also explored. Chapter 4 justifies the results and relevant discussions. Conclusion of the present work and extension is mentioned in chapter 5.

## 2. Literature Survey

Mild Cognitive Impairment (MCI) leads to Alzheimer's and various kinds of Dementia in later life. Exceptional intelligent inability of the Alzheimer's diseased patient weighs more burden on family members and public. It has a physical, psychological, social, and economic impact. Careful review was conducted in various aspects such as cause of disease, the different test applied, clinical diagnosis procedure, statistical techniques used, AI/Machine learning techniques used, and so on in order to find the research gap. The research findings are tabulated in Table 1.

The summary leads to an understanding of the correlation between diseases such as diabetes, hypertension, depression, and cognitive impairment. Few drugs are also identified as the main cause of Alzheimer's and related dementia. Various statistical techniques such as ANOVA, *t*-test, Kaplan–Meier estimates (survival estimation function), and QUADAS-2 (diagnosis test against ref value) are used to analyze the data. The main issue to be addressed by the use of a statistical tool is sampling error present in the dataset. This sampling error in the data set would lead to wrong conception. Application of suitable machine learning algorithm will provide optimal solution. The extensive review shows the importance of analyzing cognitive level of the patient and also the role of machine learning algorithm for prediction and classification. Our proposed research work considers diabetes and hypertension details of the patients and applies suitable machine learning algorithm and cognitive ability test to predict the risk of Alzheimer in person's late life.

## 3. Proposed Work

The proposed model aims at early prediction of cognitive decline of the people using cognitive data, clinical data, and physical data from his history. Dementia has a lengthy preclinical period during which there are no perceptible cognitive impairments, but neurogenerative changes are happening. Therefore, it is essential to identify individuals at high risk of dementia in an earlier stage to protect them from the possibility of disease in their late life [16]. Population based study among precise age group supports our understanding with fewer possible bias. Studies looking at the mid-age people with chronic diseases are particularly helpful since the chances of dementia development are higher for

TABLE 1: Summary of research findings and problem formulation.

Author(s)	Journal details, year of publication	Description and findings	Issues and identification of research avenue
Subha [9]	Journal of Clinical and Diagnostic Research, 2012	Dataset contains thirty males and females above 50 years, with and without diabetes. Authors applied unpaired <i>T</i> -test and one-way ANOVA using SPSS to find the correlation between age, sex, diabetes, HbA1C level, and cognitive status. The 3MS-MMSE test was also conducted	SPSS lacking efficient handling of regression data. Data analytics is limited. Appropriate machine learning algorithm would support efficient data analytics
Arevalo-Rodriguez et al. [10]	Cochrane Database of Systematic Reviews, 2015	1569 MCI patients are considered in the case study. The authors assessed Alzheimer's, Lewy body dementia, vascular dementia, and front to temporal dementia among the patients. They used QUADAS -2 tool and search methods in their research. Finally, an MMSE test was conducted to diagnose whether MCI advanced to dementia or Alzheimer's	Early diagnose of MCI would help to prevent the disease
Park et al. [11]	NPJ Digital Medicine, 2020	Korean national health insurance service data between the year 2002 and 2010 with persons age >65 was used in their research proposal. It includes 4,894 clinical features which includes laboratory values, medication codes, ICD-10 codes, history of illness of the person and his family, and sociodemographics. The authors developed RF algorithm, SVM, and logistic regression models for performance comparison. MMSE test was applied to find the severity of Alzheimer's disease	There exists +ve correlation among level of urine protein and alzheimer incident. Tolfenamic acid and zotepine positively correlated with incident AD. Nicametate citrate, was negatively associated with incident AD
Alencar et al. [12]	Diabetology and Metabolic Syndrome, 2010	Total of 346 outpatients with type 2 diabetes which includes 216 females of average age $58.6 \pm 12$ is considered for the case study. Authors conducted ANOVA test, ANCOVA test with important features such as diabetes, glycemic control, A1C, drugs used by patients, and smoking	It is perceived that the duration of disease has a direct association with a decline in cognition
Zhao et al. [13]	BMC Endocrine Disorders, 2020	The authors conducted a cohort study among type 2 diabetes patients with age >55 years. The data set is divided into three groups with reference to the level of HbA1C: HbA1c < 7.7%, HbA1c between 7.8% and 8%, and HbA1c greater than 8%. Univariate and multivariate regression analysis was done to find the correlation between the level of HbA1C and cognitive decline	It is noted that HbA1c greater than 8% is an important factor to determine the level of cognitive decline
Moore et al. [14]	PLoS One, 2019	The data set of TADPOLE grand challenge in association with ADNI is taken in their case study. ADAS-13 score and normalized ventricles volume were used to analyze the severity of the Alzheimer's disease. The random forest model is simulated in their study.	The outcome of the model was effective and comparable with other methods. However, image processing adds overhead.
Altaf et al. [15]	Biomedical Signal Processing and Control, 2018	Clinical data and MR imaging available in Alzheimer's disease neuroimaging initiative (ADNI) dataset is analyzed using a multiclass classification algorithm. MMSE test is included finally.	The images are classified into three different classes: AD, normal, and MCI. Overhead due to image processing needs to be addressed.
Exalto et al. [16]	Alzheimer's Dement, 2014	Retrospective cohort study was conducted among 9480 Kaiser Permanente members from 1964 to 1973 of age 40–55(CAIDE). The midlife vascular risk factors are analyzed using C statistic and Kaplan–Meier estimates to predict dementia	The disease prevention strategies need to pinpoint life course perspective on maintaining vascular health

TABLE 1: Continued.

Author(s)	Journal details, year of publication	Description and findings	Issues and identification of research avenue
Matioli et al. [17]	Dement Neuropsychol, 2017	Brazilian aging brain study group between 2004 and 2015 containing 1,037 subjects with diabetes mellitus is considered. Multivariate logistic regression algorithm is used to analyze diabetes. Neuropathological examination is done using immunohistochemistry to identify vascular dementia (VaD) and AD. The CDR value and IQCODE is used as reference metrics	The large clinic pathological study results have shown that diabetes was not directly associated with dementia
Cholerton et al. [18]	Diabetes Spectrum 2016	The level of cognitive decline in older adults with type 2 diabetes is analyzed using preclinical pathophysiological processes	Detected and focused pathological changes in early stages due to type 2 diabetes which cause dementia in later stage
Shaji et al. [19]	Indian J Psychiatry, 2010	Dementia aging, mental health issue, and other late life issues were carefully reviewed in Indian Journal of Psychiatry (IJP) from 1958 to 2009. The authors categorized their review as editorials, research reports, and other articles. They used search strategies to accomplish the task	The authors identified the various key risk factors which cause depression and dementia
Lee 2020 [20]	Journal of the American Geriatrics Society, 2020	A subsample of more than 3,000 respondents aged 60 and higher from LASI- the longitudinal aging study in India is considered. LASI is a nationwide health survey which is conducted among people aged 45 to review their social and economic well-being	Disability assessment for dementia is analyzed using the depth of late-life cognition
Cunningham et al. [21]	Ulster med J, 2015	19,765 people from northern Ireland were analyzed for VD, AD, dementia with Lewy bodies and frontotemporal lobar dementia	Advanced cognitive decline, pathophysiological processes, and biomarkers are explored in their study. Application of cognitive decline test would give more insights
Thunell et al. [22]	The Journal of the Alzheimer's association, 2021	The authors included 29 drug classes span of 11 therapeutic areas and 404 human studies. They applied drug therapy to find drug classes related to increasing or declining ADRD risk (Alzheimer's disease and related dementias)	Among 13 drug classes, 50% or more reported consistent effects on the risk of ADRD
Hane et al. [23]	Journal of Alzheimer's Disease, 2017	The authors reviewed more than 300 prominent AD research proposals. They divided the work into three parts detailed as Part 1 represents pathogenesis on a molecular and macroscale, part 2 represents genetics and epidemiology, and part 3 includes diagnosis and cure. They applied MMSE and MoCA tests	The medical imaging techniques consider pathogenesis and genetics to identify the risk associated with Alzheimer's disease

them. Midlife hypertension increases the risk of lacunar infarcts and stroke, which in turn increases the risk VaD.

The existing system requires imaging data or fluid collection, which imposes a delay in early detection. Huge measure of Electronic Health Records available in structured and unstructured manner supports timely diagnosis and decisions. Collection of administrative, electronic medical data requires less amount of time. Viable use of information and attaining precise outcome is the major challenge in different fields, particularly in medical field. Utilization of Machine learning is found in almost all fields like image processing, language automation, computer vision, e-business, etc. The advent of predictive models of machine learning can be applied to these valuable digitized health records for the early risk prediction of VAD and AD.

Chronic diseases like diabetics, blood pressure, heart problems, and kidney infection are increasing worldwide. It was witnessed that diabetics and blood pressure have strong relation with cognitive decline in elderly people [24,25]. The helpless diabetic control and bad adherence to physician instructions are the primary reason for the elevation of AD or dementia in their late life [26, 27]. The early detection aids to prevent AD with the help of proper diabetic control, drugs, cognitive training, and so forth. Reference [28] research finding states that there is a direct connection between glucose dysregulation and neurodegeneration. Diabetes is viewed as a key risk factor for cognitive impairment and few investigations prove that cognitive dysfunction influences both older and younger persons with diabetes [29, 30]. Type 2 diabetes patients ought to be

routinely assessed for their intellectual capacity since a span of infection could be related to a decrease in cognizance.

Physicians do not screen the psychological capacity of the patient until he gets compliant from the patient or from the patient's family. After the beginning indications, patients consult doctor. During this period, dementia is moved to a moderately advanced stage. Timely diagnosis would help the patients to overcome disease progression. The Cognitive Ability Test (CAT) is a brief neuropsychological screening test that provides an outline of cognitive function. The CAT test helps the physician to assess the cognitive function of the patient in the early stage itself [13]. The proposed work applies a support vector machine learning algorithm to identify people with a high risk of cognitive impairment in their late life and they are exposed to CAT screening tests [31, 32]. The test results are analyzed with the help of Multinomial logistic regression to classify them as "Severe dementia," "Uncertain Dementia," and "No Dementia."

### 3.1. Data and Methodology

**3.1.1. Data Set.** The primary focus of the proposed work is to provide health care service to the elderly population residing in resource poor areas. People with ages between 40 and 65 years are considered as mid-age people in our case study. The proposed work mainly considers hypertension, diabetes as the most common risk factor for cognitive decline. The appropriate data set available in "Data World" repository is taken and filtered with the required features. Plausible cross-sectional examination provides the best technique for analyzing a causal connection between diabetics, blood pressure (BP), and the occurrence of dementia. To enhance the analysis, we consider two age classifications, namely midlife <65 years and late life >65 years.

The emphasis on midlife is especially relevant for dementia counteraction for two reasons. (i) Midlife is sufficiently early to make an association between risk factors and Alzheimer's before the initiation of neurodegeneration. (ii) A few examinations presented the connection between raised BP in midlife (age 40–64 years) and the beginning of dementia and AD in their late life. This study considers general health data available in "Data world," the world's largest collaborative data community. The database consists of 19 features and 2361 patients records whose snippet is depicted in Table 2.

**3.1.2. Feature Selection.** The data set contains 19 features describing age, cholesterol, glucose, etc. Since Hemoglobin A1c (HbA1c) is the important measure of long-term control of glucose in our body, it was mainly considered in the early identification of AD [30]. Along with HbA1c, the patient's systolic and diastolic BP was examined. The list of features and their descriptions are given in Table 1. HbA1c value greater than 6.5 is considered diabetes positive. The given data set is separated into two distinctive sets with respect to the age to assess the midlife attributes and their association with AD. Exploratory Data Analytics is performed to summarize the main characteristics of data and to find important features

with the help of visual aids. Multivariate analytics is performed to understand different features and their interaction. Table 3 summarizes the features description.

The correlation factor associated with each pair of features helps to extract relevant attributes for study. The highly influential factors such as age, gender, HbA1c, glucose, systolic pressure, diastolic pressure, and cholesterol are considered in our case study.

**3.2. Process Flow.** The process flow of the proposed model to predict the level of cognitive decline is shown in Figure 1. The diabetes and pressure data set collected from "Data World" is preprocessed to remove redundant information and missing values. The highly influencing features are extracted with the help of correlation values. We apply 2-stage classification model to determine the cognitive decline more accurately. In the first stage, the selected set of features is applied to the classifier algorithm to identify the associated risk among the population. Support Vector Machine and Random Forest algorithm are used for risk classification. In the second stage, we apply CAT among the people identified in stage 1. The Multinomial Logistic Regression algorithm examines the CAT results and medical care is provided for the people predicted as "Severe Alzheimer". CAT.

#### 3.2.1. Algorithm

Step 1: Input: Patients with Blood Pressure, Diabetes dataset.

Filter 40–65 age group data.

Handle inconsistent and missing data

Output: Preprocessed data

Step 2: Identify correlation between features of the dataset using multivariate analysis

Step 3: Do Initial classification for Alzheimer Disease using stage1 classifier

Step 4: If AD possible perform CAT test on those patients

Step 5: Perform second level classification to find AD Dementia, Uncertain Dementia, No-Dementia using stage 2 classifier

**3.2.2. Flowchart.** The process flow of the proposed model to predict the level of cognitive decline is represented as a flowchart in Figure 2.

**3.2.3. Support Vector Machine.** SVM is the commonly used supervised classifier, which classifies data in N-dimensional space using a hyperplane. It has been applied in enormous healthcare applications in predicting diseases from structural data [33]. Figure 3 shows the classification graph of SVM.

The line function  $y = ax + b$  helps to easily differentiate linearly separable data. The SVM uses the line equation transformed into a hyperplane which is applied in the prediction process. The model tries to find out optimal bias and variance for both train and test data set. The

TABLE 2: Snippet of Diabetics registry.

Id	Chols	Stab-glu	Hdl	...	Glyhb	Age	Height	Weight	Bp-s	Bp-d
1000	203	82	56	...	4.31	46	62	119	120	82
1001	165	97	24	...	4.44	29	64	218	140	85
1002	228	92	37	...	4.64	58	61	243	185	92
1003	78	93	12	...	4.63	67	67	121	170	95
1005	249	90	28	...	7.72	64	68	179	158	85
1008	248	94	69	...	4.81	34	71	186	125	80
1011	195	92	41	...	4.84	30	69	180	161	112
1015	227	75	44	...	3.94	37	59	170	121	81
1016	177	87	49	...	4.84	45	69	166	128	86
1022	263	89	40	...	5.78	55	63	198	132	82
1024	242	82	54	...	4.77	60	65	164	130	90
1029	215	128	34	...	4.97	38	58	191	145	83
1030	238	75	36	...	4.47	27	60	167	120	80
1031	183	79	46	...	4.59	40	59	173	132	85
1035	191	76	30	...	4.67	36	69	179	122	82
1036	213	83	47	...	3.41	33	65	154	120	96
1037	255	78	38	...	4.33	50	65	187	126	83
1041	230	112	64	...	4.53	20	67	159	152	92
...	...	...	...	...	...	...	...	...	...	...

TABLE 3: List of features and their descriptions.

Feature name	Description and values	% Missing	Mean	Standard Deviation
ID	Patient ID	0%	NA	NA
Chols	Overall cholesterol	<1%	207.275	44.715
Stab-glu	Stabilized glucose	0%	107.338	53.798
HDL	High-density lipoprotein	<1%	50.267	17.301
Glyhb	Glycosylated hemoglobin	0%	5.59	2.243
Age	Patients age (Years)	0%	46.774	16.436
Ht	Patients height (inches)	<1%	65.979	3.927
Wt	Patients weight (pounds)	0%	177.349	40.392
Bp-s	Patients systolic BP	= 1.28%	137.148	22.997
Bp-d	Patients diastolic BP	= 1.28%	83.286	13.582

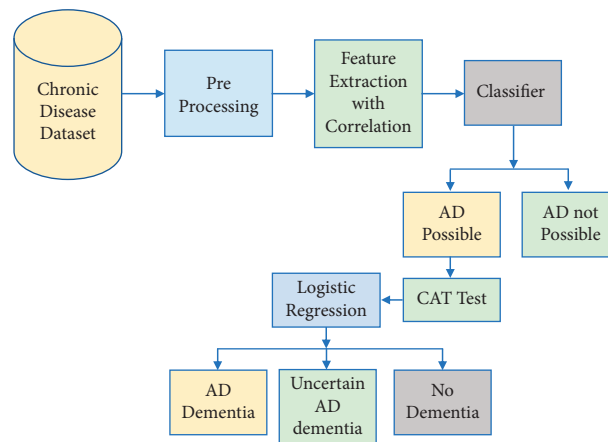


FIGURE 1: Process flow diagram.

comprehensive review has proven that support vector machine provides good performance for big data and healthcare applications.

**3.2.4. Random Forest.** The Random Forest algorithm trains  $n$  different decision trees with different data subset and tuning parameters [34]. It combines the output of all  $n$  trees

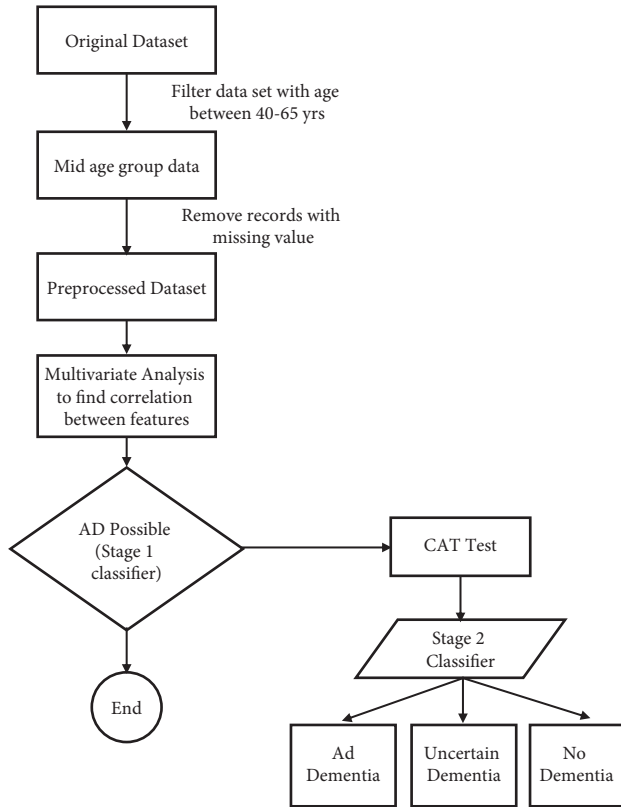


FIGURE 2: Flow chart.

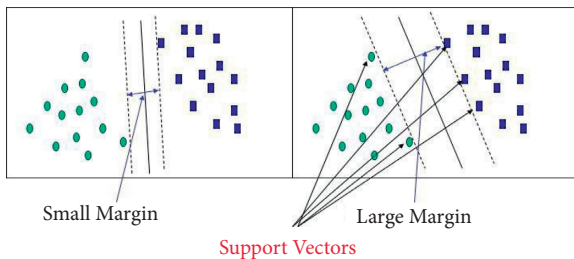


FIGURE 3: SVM classification graph.

with the help of a voting mechanism. Hence, it is also called Ensemble learning. The working principle of Random Forest algorithm is depicted in Figure 4.

**3.3. Mental Ability Test.** The cognitive decay of a person ranges from mild to severe. The primary causes include medications, disorder among blood vessels, despair, and dementia. Dementia represents a severe loss of mental functioning and the common type is Alzheimer. Cognition of a person includes a blend of processes in the brain involved in all facets of his life. It includes his memory capacity, thinking skill, language, and talent to learn new things. A cognitive ability test is performed to examine the cognitive impairment of a person. With the help of a detailed review conducted to screen the cognitive function, we framed multiple questions to check the decline in mental

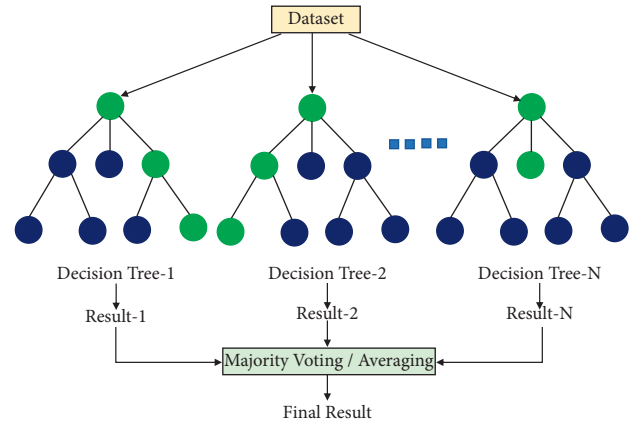


FIGURE 4: Working of random forest algorithm.

function, and the test questionnaire is given in Table 4 [35]. The CAT test scores from 25 to 30 are considered as normal [36]. Items address orientation, memory, attention, recall, naming objects, responding to verbal and written commands, writing a sentence, and copying a figure are the tasks considered in CAT to evaluate the cognitive status of persons. The informant accompanied the patients, and the questions are administered to the informants without unduly alarming the patient.

The maximum CAT score is 30 points. A score of 25 to 30 suggests no cognitive decline, 13 to 24 recommends moderate decline, and less than 12 indicates severe cognitive decline. In every year, the CAT score of Alzheimer’s diseased person declines about two to four points on average. The snippet of CAT dataset is shown in Table 5.

**3.3.1. Multinomial Logistic Regression on CAT.** The multinomial logistic regression model is applied to predict the severity of illness with respect to the correlation existing among the dependent variables as “Severe dementia,” “Uncertain Dementia,” and “No Dementia.” The multinomial logistic regression is applicable for the class of probe, which has more than two outcomes. Our proposed model owns three different outcomes. For N different outcomes, there are n-1 models developed as a set of independent binary regression. One outcome is referred to as Pivot class, and others are regressed against this reference class.

The probabilities for the N categories are estimated based on dependent variables.

$$\begin{aligned}
 pr(Y_i = k | X_i) &= \beta_1, \beta_2, \dots, \beta_n \\
 &= \frac{\exp(\beta_0 k + X_i \beta' k)}{\sum_{j=1}^n \exp(\beta_0 k + X_i \beta' k)}, \tag{1}
 \end{aligned}$$

where Y is the dependent variable and X is the set of explanatory variables,  $\beta_k$  is the regression coefficient for the k<sup>th</sup> category of Y.

Based on the estimated probability the output is categorized by the algorithm with reference to the threshold.

TABLE 4: Cognitive ability test-questionnaire.

S. No.	Question description	Yes/no	Weighted score
1	Does the patient place items properly?	Y	1
2	Does the patient able identify the present date, day, month, year?	Y	2
3	Does the patient comfort level change when they are in new places?	Y	1
4	Does the patient able to manage their medication schedule?	Y	2
5	Does the patient able to manage time while doing tasks?	Y	1
6	Does the patient confuse about certain things?	Y	2
7	Does the patient able to understand context?	Y	1
8	Does the patient confuse to identify known persons?	Y	2
9	Does the patient experience difficulty to recognize people familiar to them?	Y	1
10	Does the patient behavior is different from their earlier stages?	Y	1
11	Does the patient have imaginations?	Y	2
12	Does the patient forget to do regular tasks?	Y	2
13	Does the patient have problem in counting numbers or figures?	Y	2
14	Does the patient able to manage position or direction?	Y	1
15	Does the patient has shown less priory or interest towards hobby or passion?	Y	1
16	Does the patient understand situations or explanations?	N	1
17	Does the patient forget recent activities?	N	1
18	Does the patient have any cognitive issues previously?	Y	2
19.	Does the patient not able to recall main or important occasions?	Y	2
20	Does the patient not able to recollect some important days in his life.	Y	2

TABLE 5: Snippet of CAT dataset.

Subject	Gender	CAT	ageAtEntry	CDR	Memory	dx1
OAS30124	Female	16	79.14579	1	1	'AD dementia'
OAS30124	Female	21	79.14579	0.5	0.5	'AD dementia'
OAS31129	Female	20	68.07666	1	1	'AD dementia'
OAS31129	Female	28	68.07666	1	1	'AD dementia'
OAS31129	Female	29	68.07666	0.5	0.5	'AD dementia'
OAS31129	Female	29	68.07666	0.5	1	'AD dementia'
OAS30865	Female	24	73.697464	2	2	'AD dementia'
OAS30865	Female	28	73.697464	0.5	1	'AD dementia'
OAS31003	Male	30	63.92608	0.5	0.5	'AD dementia'
OAS31003	Male	30	63.92608	1	0.5	'AD dementia'
OAS31003	Male	30	63.92608	0.5	0.5	'AD dementia'
OAS31003	Male	30	63.92608	0.5	0.5	'AD dementia'
OAS30025	Female	23	64.81314	0.5	0.5	'AD dementia'
OAS31001	Female	25	62.984257	0.5	0.5	'Uncertain dementia'
OAS30331	Male	27	67.482544	0.5	0.5	'Uncertain dementia'
OAS30331	Male	17	67.482544	1	1	'AD dementia'
OAS30331	Male	27	67.482544	0.5	0.5	'Uncertain dementia'
OAS30596	Female	29	72.29295	0.5	0.5	'Uncertain dementia'
...	...	...	...	...	...	...

## 4. Experiments and Results

*4.1. Stage 1 Classifier.* In stage 1 we train SVM and Random Forest algorithm to diagnose chronic disease and to identify the associated risk. The data set contains 2361 records of mid-age people. The glyhb values in the range between 4% and 5.6% are considered as normal values, and between 5.7% and 6.4% informs more chance of being affected with diabetes. Values above 6.5% mean they have diabetes. Patient's systolic and diastolic pressure are the other important factors to be considered in the development of Alzheimer's. Systolic pressure less than 120 mm Hg and diastolic pressure less than 80 mm Hg is

considered as normal value and the range 120–139 of systolic and 80–89 of diastolic is the prehypertension values. Persons having >140 mm Hg of systolic and >90 mm Hg of diastolic pressure are considered as having hypertension. The chosen data set contains 526 records of persons with no diabetics and pressure, 1187 records of persons having diabetics or pressure, and 648 records of patients having both diabetics and pressure. Since the presence of either pressure or diabetes increases the chance of dementia, the total 1835 patients having either diabetics or pressure or both exposed to CAT test to assess their cognitive power. The details of records are visually represented in Figure 5.



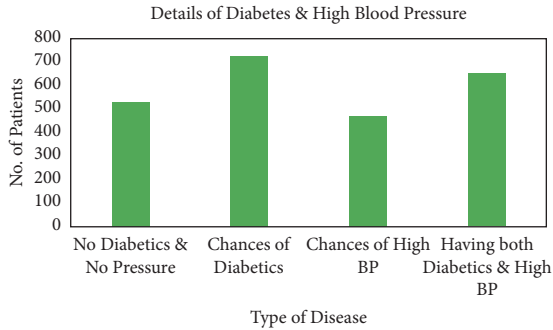


FIGURE 5: Patient count on Diabetes and Pressure.

The performance of the proposed classifier model is analyzed using confusion matrix. The total accuracy, sensitivity and specificity are calculated using the formula given.

(i) Total Accuracy (TC)

$$TC = \frac{TP + TN}{TP + FN + TN + FP} \quad (2)$$

(ii) Sensitivity (SN)

Sensitivity (or) True Positive Rate and Specificity (or) True Negative rate are calculated to assess the performance of the proposed model. True positive rate reveals the details of how effectively the model identifies actual positives present in the data set. Specificity measure is given as follows:

$$SN = \frac{TP}{TP + FN} \quad (3)$$

(iii) Specificity

Specificity (or) True Negative rate measure is given as follows:

$$SP = \frac{TN}{TN + FP} \quad (4)$$

The terms TN, TP, FP, and FN denote True negative (person with no chance of Alzheimer is identified as 'No Alzheimer'), True positive (person subjected to Alzheimer is predicted as 'Alzheimer'), False Positive (healthy person is detected as 'Alzheimer'), and False Negative (person suffering from Alzheimer is identified as healthy), respectively.

**4.2. Performance Analysis of Stage 1 Classifier.** The proposed SVM classifier outperforms with 0.90 AUC value and for Random Forest AUC is 0.74. The probabilistic classifier shows the tradeoff between sensitivity and specificity. Table 2 shows the performance comparison of SVM and Random Forest algorithm used in our case study. NDP represents patients with No Diabetics and No Pressure, CRD represents patients having either diabetes or pressure, HDP represents patients Having both Diabetics and Pressure. Performance comparison of each measure is given in Table 6. The same is visually represented in Figures 6–8.

In Random Forest algorithm, it is important to consider the subsampling of data points in the tree construction process. More subsampling or no subsampling results in inconsistent effects. It is possible to enhance the accuracy of Random Forest algorithm by varying the parameters. Due to the unavailability of sample data set, it is not probable to fine-tune the parameters for RF in our case study.

**4.3. Stage 2 Classifier.** The CAT test result dataset contains a minimum age of 47 years and a maximum of 96 years. Clinical Dementia Rating shortly termed as CDR is a numeric scale used to quantify the severity of dementia indications and its score ranges from zero (none) to 3(severe). Summarization of CAT data set is provided in Table 7 and the same is visually represented in Figure 9.

The confusion matrix of multinomial logistic regression is considered for analysis. True Positive Rate and True Negative rate for the three different classes of output is given as follows:  $TPR_{No\ Alzheimer} = 98\%$ ,  $TPR_{Uncertain\_Alzheimer} = 85\%$  and  $TPR_{Alzheimer} = 81\%$ . True Negative rates of the three different classes are given as follows.  $TNR_{No\ Alzheimer} = 86\%$ ,  $TNR_{Uncertain\_Alzheimer} = 73\%$ , and  $TNR_{Alzheimer} = 75\%$ . The results of the proposed model show that the model can predict with improved accuracy provided ample amount of dataset for training.

**4.4. Analysis with Bench Mark Models.** The study of neurodegenerative diseases caused by the ageing of brain systems necessitates brain banking. The Brazilian Aging Brain Study Group's Brain Bank collects a large number of elderly brains and their related disorders. It encourages researchers to look at a variety of aspects of ageing brain processes and related neurodegenerative illnesses.

Table 8 represents the performance analysis of our proposed model with existing benchmark models. The table explores the detail of the data set used by different authors and the employed machine learning algorithms. Reference [13] considers people above 65 years from 2002 to 2010. The main features considered in their case study include Implantable Cardioverter Defibrillator -10 codes, laboratory results, medication codes, sociodemographics, illness of a person, and his family [37]. They have trained and tested dataset with random forest, logistic regression, and SVM and to predict Alzheimer's incident in 1, 2, 3, and 4 subsequent years. For comparison average of 4 years is taken. The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge is organized in association with Alzheimer's Disease Neuroimaging Initiative (ADNI) to find people at risk of Alzheimer's. The historical measurements of the people were considered to predict future implications. TADPOLE challenge facilitates early identification of Alzheimer disease with the help of appropriate algorithm [14] used data from the TADPOLE grand challenge and claimed their result with benchmark SVM, which produces 62% AUC and classification accuracy of 52%. Reference [15] collected records of Brain Bank of the Brazilian Aging Brain Study Group between 2004 and 2015. Among 1,037 subjects, diabetes was present, with 279 participants (27%). They

TABLE 6: Performance comparison of SVM and Random Forest.

Algorithm	Accuracy			Sensitivity			Specificity		
	NDP	CRD	HDP	NDP	CRD	HDP	NDP	CRD	HDP
SVM	0.89	0.86	0.84	0.78	0.82	0.88	0.93	0.78	0.89
RF	0.81	0.78	0.79	0.62	0.83	0.71	0.82	0.75	0.78

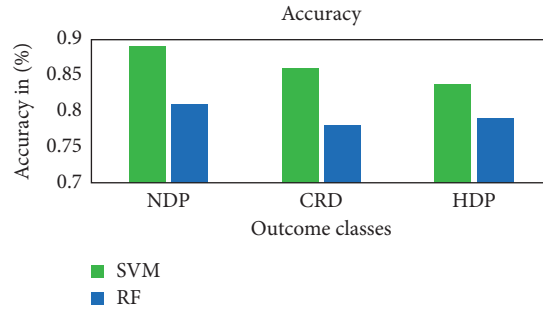


FIGURE 6: Accuracy.

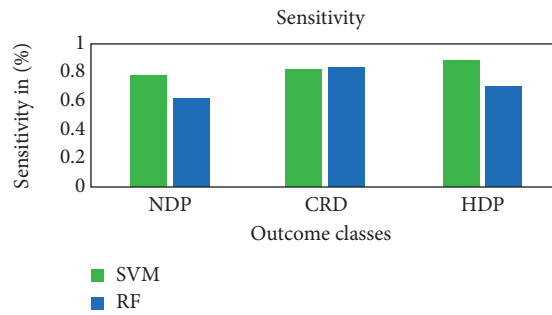


FIGURE 7: Sensitivity.

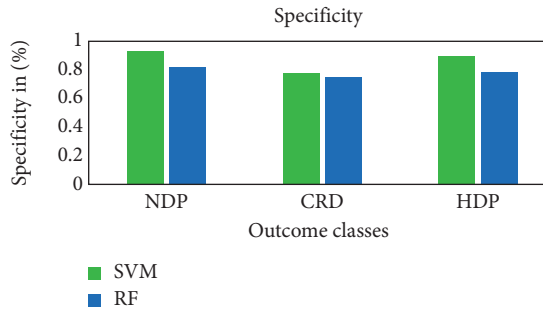


FIGURE 8: Specificity.

TABLE 7: Summarization of CAT dataset.

CAT score level	Patient count	Cognitive status
25–30	1017	No dementia
13–24	603	Uncertain dementia
0–12	215	Severe dementia

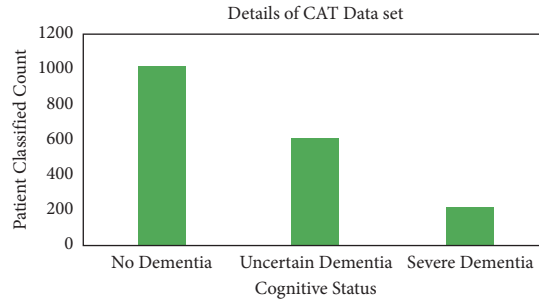


FIGURE 9: Summarization of CAT dataset.

TABLE 8: Analysis with Bench mark models.

Method	Dataset	ML algorithm	Accuracy (%)	AUC (%)
Park et al. [11]	Korean National Health Insurance Service database between 2002 and 2010	Linear regression	66.8	69.5
		Support vector machine	67.1	70.58
		Random forest	70.12	76.1
Moore et al. [14]	TADPOLE grand challenge dataset	Random forest	73	82
Matioli et al. [17]	Diabetes data from brazilian aging brain study group between 2004 and 2015	Multivariate logistic regression	95	-
		Support vector machine	0.86	90
Proposed method	Diabetics registry data in kaggle repository and data set generated through CAT test	Random forest	0.71	0.74
		Multinomial logistic regression	0.89	0.92

proved there is no association between diabetes and dementia (OR = 1.22; 95%CI = 0.81–1.82;  $p = 0.34$ ) based on the multivariate analysis.

## 5. Conclusion

Automated healthcare techniques support physicians in making correct decisions on patient care in resource poor rural areas. The timely identification of risk factors with the help of AI based model's safeguards the person from late life Alzheimer's. The availability of an appropriate dataset with relevant attribute is a cumbersome process in the development of a more accurate model. The proposed method supports the statistically significant diagnosis of persons at risk for Alzheimer's disease simply based on administrative health records. It allows earlier and accurate screening for further clinical testing. Our proposed work analyzes the influence of hypertension and diabetes on Alzheimer's disease. Support Vector Machine algorithm is more suitable when the dataset is not continually distributed. The performance of SVM is relatively good due its convex optimization nature. Survey conducted on the population with chronic disease for cognitive assessment provides the degree of cognitive decline in the community. The CAT test results are analyzed with the help of multinomial logistic regression to exactly identify the possibility of Alzheimer's in patient's late life. To achieve optimum accuracy of the model, a large sample size is essential. In the future, the proposed work may be extended with more classifiers by accumulating a huge

volume of samples and an increased number of surveys on CAT tests. Time series survey among the population for CAT test will further improve the precision of prediction.

## Data Availability

The datasets used and/or analyzed during the current study are available in the following repository: <https://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Diabetes.html>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Authors' Contributions

A. Revathi was responsible for conceptualization, data curation, formal analysis, methodology, software, and writing—original draft; R.Kala Devi was responsible for supervision, writing—review and editing, project administration, and visualization; Kadiyala Ramana was responsible for software, validation, writing—original draft, methodology, and supervision; Rutvij H.Jhaveri was responsible for supervision, writing—review and editing, and visualization; Madapuri Rudra Kumar was responsible for data curation, investigation, resources, and software; M.Sankara Prasanna Kumar was responsible for visualization, investigation, formal analysis, and software.

## References

- [1] S. Bhattacharya, P. K. R. Maddikunta, Q.-V. Pham et al., "Deep learning and medical image processing for coronavirus (covid-19) pandemic: a survey," *Sustainable Cities and Society*, vol. 65, Article ID 102589, 2021.
- [2] G. J. Biessels and F. Despa, "Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications," *Nature Reviews Endocrinology*, vol. 14, no. 10, pp. 591–604, 2018.
- [3] C. Qiu, E. V. Strauss, J. Fastbom, B. Winblad, and L. Fratiglioni, "Low blood pressure and risk of dementia in the kungsholmen project: a 6-year follow-up study," *Archives of Neurology*, vol. 60, no. 2, pp. 223–228, 2003 Feb.
- [4] B. Jani and C. Rajkumar, "Ageing and vascular ageing," *Postgraduate Medical Journal*, vol. 82, no. 968, pp. 357–362, 2006.
- [5] J. R. Marden, E. R. Mayeda, E. J. Tchetgen Tchetgen, I. Kawachi, and M. M. Glymour, "High hemoglobin A1c and diabetes predict memory decline in the health and retirement study," *Alzheimer Disease and Associated Disorders*, vol. 31, no. 1, pp. 48–54, 2017.
- [6] P. Ramesh, K. Dunn, W. Eberle, and D. Chaung, "Cognitive health prediction on the elderly using sensor data in smart homes," in *Proceedings of the 31st International FLAIRS Conference*, pp. 317–322, Melbourne, Florida, USA, May 2018.
- [7] K. R. Lakshmana, F. Khan, D. Sadia, S. B. Shahab, M. Amir, and I. Ebuka, "Recurrent neural network and reinforcement learning model for covid-19 prediction," *Frontiers in Public Health*, vol. 9, 2021.
- [8] D. Ngabo, D. Wang, W. Dong, E. Ibeke, and C. Iwendi, "Tackling pandemics in smart cities using machine learning architecture," *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 8444–8461, 2021.
- [9] N. K. Subha, "Assessment of the cognitive status in diabetes mellitus," *Journal of Clinical and Diagnostic Research*, vol. 6, no. 10, pp. 1658–1662, 2012.
- [10] I. Arevalo-Rodriguez, N. Smailagic, M. Roqué i Figuls et al., "Mini-mental state examination (MMSE) for the detection of alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)," *Cochrane Database of Systematic Reviews*, vol. 3, Article ID CD010783, 2015.
- [11] J. H. Park and H. Cho, J. H. Kim, M. M. Wall, Y. Stern et al., "Machine learning prediction of incidence of alzheimer's disease using large-scale administrative health data," *NPJ Digital Medicine*, vol. 3, p. 46, 2020.
- [12] R. C. Alencar, R. A. Cobas, and M. B. Gomes, "Assessment of cognitive status in patients with type 2 diabetes through the mini-mental status examination: a cross-sectional study," *Diabetology & Metabolic Syndrome*, vol. 2, no. 1, p. 10, 2010.
- [13] L. Zhao, C. Han, Z. Zheng, S. L. Xiu, and P. Chan, "Risk of mini-mental state examination (MMSE) decline in the elderly with type 2 diabetes: a Chinese community-based cohort study," *BMC Endocrine Disorders*, vol. 20, p. 129, 2020.
- [14] P. J. Moore, T. J. Lyons, and J. Gallacher, "Random forest prediction of Alzheimer's disease using pairwise selection from time series data," *PLoS One*, vol. 14, no. 2, pp. 1–14, 2019.
- [15] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, 2018.
- [16] L. G. Exalto, C. P. Quesenberry, D. Barnes, M. Kivipelto, G. J. Biessels, and R. A. Whitmer, "Midlife risk score for the prediction of dementia four decades later," *Alzheimers Dement*, vol. 10, no. 5, pp. 562–570, 2014.
- [17] M. N. P. D. S. Matioli, C. K. Suemoto, R. D. Rodriguez et al., "Association between diabetes and causes of dementia: evidence from a clinicopathological study," *Dement Neuro-psychol*, vol. 11, no. 4, pp. 406–412, 2017.
- [18] B. Cholerton, L. D. Baker, T. J. Montine, and S. Craft, "Type 2 diabetes, cognition, and dementia in older adults: toward a precision health approach," *Diabetes Spectrum*, vol. 29, no. 4, pp. 210–219, 2016.
- [19] K. S. Shaji, V. P. Jithu, and K. S. Jyothi, "Indian research on aging and dementia," *Indian Journal of Psychiatry*, vol. 52, no. Suppl1, pp. S148–S152, 2010 Jan.
- [20] J. Lee, Y. K. Pranali, J. Banerjee et al., "Design and methodology of the longitudinal aging study in india-diagnostic assessment of dementia (LASI-DAD)," *Journal of the American Geriatrics Society*, vol. 68, 2020.
- [21] E. L. Cunningham, B. McGuinness, B. Herron, and A. P. Passmore, "Dementia," *Ulster Medical Journal*, vol. 84, no. 2, pp. 79–87, 2015.
- [22] J. Thunell, Y. Chen, G. Joyce et al., "drug therapies for chronic conditions and risk of alzheimer's disease and related dementias: a scoping review," *Alzheimer's & Dementia The Journal of the Alzheimer's association*, vol. 17, no. 1, pp. 41–48, 2021.
- [23] F. T. Hane, M. Robinson, Y. L. Brenda, B. Owen, Z. Leonenko, and M. S. Albert, "Recent progress in alzheimer's disease research, Part 3: diagnosis and treatment," *J Alzheimers Dis*, vol. 57, no. 3, pp. 645–665, 2017.
- [24] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, "3D convolutional neural network and stacked bidirectional recurrent neural network for alzheimer's disease diagnosis," in *Predictive Intelligence in Medicine*, I. Rekić, G. Unal, E. Adeli, and S. Park, Eds., Springer Nature Switzerland AG, Berlin, Germany, pp. 138–146, 2018.
- [25] A. Ramirez, S. Wolfsgruber, C. Lange et al., "Elevated HbA1c is associated with increased risk of incident dementia in primary care patients," *J Alzheimers Dis*, vol. 44, no. 4, pp. 1203–1212, 2015.
- [26] T. R. Gadekallu, N. Khare, S. Bhattacharya et al., "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, 2020.
- [27] C. Tzourio, "Hypertension, cognitive decline, and dementia: an epidemiological perspective," *Dialogues in Clinical Neuroscience*, vol. 9, no. 1, pp. 61–70, 2007.
- [28] S. Ahmed Khan and M. A. Jabbar, "Improved classification techniques to predict the co-disease in diabetic mellitus patients using discretization and apriori algorithm," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN, vol. 8, no. 11, pp. 2278–3075, 2019.
- [29] S. E. Nilsson, S. Read, S. Berg, B. Johansson, A. Melander, and U. Lindblad, "Low systolic blood pressure is associated with impaired cognitive function in the oldest old: longitudinal observations in a population-based sample 80 years and older," *Aging-Clinical & Experimental Research*, vol. 19, no. 1, pp. 41–47, 2007.
- [30] K. R. Kruthika and H. D. Rajeswari, "Maheshappa, multistage classifier-based approach for alzheimer's disease prediction and retrieval," *Informatics in Medicine Unlocked*, vol. 14, pp. 34–42, 2019.
- [31] X. Li, J. Dai, S. Zhao, W. Liu, and H. Li, "Comparison of the value of mini-cog and mmse screening in the rapid identification of chinese outpatients with mild cognitive

- impairment,” *Medicine (Baltimore)*, vol. 97, no. 22, Article ID e10966, 2018.
- [32] P. T. Trzepacz, H. Hochstetler, S. Wang, B. Walke, and A. Saykin, “Relationship between the montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults,” *BMC Geriatrics*, vol. 15, no. 107, 2015.
- [33] G. Battineni, N. Chintalapudi, and F. Amenta, “Machine learning in medicine: performance calculation of dementia prediction by support vector machines (SVM),” *Informatics in Medicine Unlocked*, vol. 16, Article ID 100200, 2019.
- [34] I. Celestine, B. Ali Kashif, P. Atharva, and R. Sujatha, “Chatterjee jyotir moy, pasupuleti swetha, mishra rishita, pillai sofia, jo ohyun, covid-19 patient health prediction using boosted random forest algorithm,” *Frontiers in Public Health*, vol. 8, 2020.
- [35] Y. Maki, T. Yamaguchi, and H. Yamaguchi, “Symptoms of early dementia-11 questionnaire (SED-11Q): a brief informant-operated screening for dementia,” *Dement Geriatr Cogn Disord Extra*, vol. 3, pp. 131–142, 2013.
- [36] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “Mini-mental state”. a practical method for grading the cognitive state of patients for the clinician,” *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.
- [37] J. C. De la Torre, “Alzheimer disease as a vascular disorder,” *Stroke*, vol. 33, no. 4, pp. 1152–1162, 2002.