A Novel Approach to Predict the Longevity in Alzheimer's Patients Based on Rate of Cognitive Deterioration using Fuzzy Logic Based Feature Extraction Algorithm

Mutyala Sridevi [1†] and Arun Kumar B.R.[2††],
Sridevim@bmsit.in Arunkumarbr@bmsit.in
BMS Institute of Technology and Management, Bangalore, India

Summary
Alzheimer's is a chronic progressive disease which exhibits varied symptoms and behavioural traits from person to person. The deterioration in cognitive abilities is more noticeable through their Activities and Instrumental Activities of Daily Living rather than biological markers. This information discussed in social media communities was collected and features were extracted by using the proposed fuzzy logic based algorithm to address the uncertainties and imprecision in the data reported. The data thus obtained is used to train machine learning models in order to predict the longevity of the patients. Models built on features extracted using the proposed algorithm performs better than models trained on full set of features. Important findings are discussed and Support Vector Regressor with RBF kernel is identified as the best performing model in predicting the longevity of Alzheimer's patients. The results would prove to be of high value for healthcare practitioners and palliative care providers to design interventions that can alleviate the trauma faced by patients and caregivers due to chronic diseases.
Key words:
Alzheimer's, cognitive ability, fuzzy logic, feature extraction, machine learning.

1. Introduction

Alzheimer's is a chronic and progressive disease that brings gradual cognitive impairment to the patient which is assessed using various biological and behavioural biomarkers. The symptoms and rate of deterioration in cognitive abilities vary from person to person and has high impact on longevity of the patient. In order to deal with this kind of scenario, caregivers join social media communities built around the disease to get emotional and social support. Social media interactions are providing a huge source of secondary data, especially in the healthcare domain. Studies or research based on retrospective longitudinal studies can select disease communities on social media as one of the sources of data which can be analysed to detect chronic disease trends and patient behaviour.

On the other hand, machine learning is one of the many tools available with data science to build learning models trained on past data and apply that experience to predict/classify the unseen data with good amount of generalizability. These two concepts of social media interactions and machine learning can be combined to analyse participant interactions/data obtained from the interactions on social media to derive valuable insights into patient behaviour or disease trends. The details about deterioration in cognitive levels can't be reported or recorded precisely as the phenomena occurs gradually over time. This associates uncertainty to the data while developing the learning models. The current work focuses on designing an algorithm to extract features from raw data when the data collected for analysis has uncertainty/imprecision associated with it. Fuzzy logic based approaches comes to rescue in scenarios where data uncertainties exist and provides an efficient way to work with imprecise data. The proposed fuzzy logic based algorithm designed for feature extraction, when used as a prior step to fitting the machine learning models with the data, exhibits good performance when compared to the entire set of features. Linear and non-linear regression techniques with tuned hyper parameters were used to fit the data with all features and with reduced features. All the models have shown significant improvement in performance when combined with the proposed algorithm to extract the features by reducing the dimensionality except Lasso regression which explains the variance in the data to be same in both the cases. Support Vector Regressor (SVR) with Radial Basis Function (RBF) kernel performs the best among other fuzzy feature extracted models in predicting the longevity/final stage of Alzheimer's patients. These predictions would help the healthcare professionals, terminal care providers and caregivers to design and execute interventions that can prolong the deterioration in cognitive capabilities to maintain better life style that can reduce trauma caused by the disease to the patients and caregivers. The results would also help in providing personalized terminal care

Manuscript received August 5, 2021
Manuscript revised August 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.8.11

facilities in the palliative care centres. This section is followed by literature survey, problem context, data collection and participant demographics, feature extraction, experimental results and discussion followed by conclusion.

2. Literature Survey

Very less number of models have been developed to analyse the disease progression or rate of deterioration in cognitive capabilities of Alzheimer's patients. The research performed by [1] described quantitatively, the Alzheimer's disease's natural progression. Also, the dropout patterns of patients from the clinical trials were characterized with the help of parametric survival models. The work in [2] presents a disease progression model of Alzheimer's disease and the effect of education level on the entire spectrum of disease. It explains the decline in cognitive abilities right from subjective cognitive impairment (SCI) to the final stage of Alzheimer's Disease Dementia (ADD). The authors in [3] modelled Alzheimer's disease markers exhibiting mixed effects which is used to extract the global as well as individual marker trajectories for training population and then unseen subject specific models were invoked using a layered "marker signature". They focused on bringing out markers that are patient specific models that are shown to give better predictions of transition time to Alzheimer's disease than entire population centric models.

The article [4] discusses about a strata of statistical models related to Alzheimer's disease progression and applies these models to the longitudinal cognitive scores. The models were non-linear disease progression models that address the mixed-effects, model the disease stages, and the patient-centric changes of cognitive ability as the indirect variables. Their results provided new perspectives of biomarkers for staging the patients. The authors in [5] developed a biomarker generation by using partial least squares regression method and also classification method that works on classifying the patients into discrete subgroups to predict cognitive trajectories of individuals. The research performed in [6] analysed Alzheimer's patients for changes in cognitive abilities based on deterioration of attention, visuo-spatial, and executive functions and observed that these factors show faster deterioration rates than other cognitive functions.

The study performed in [7] was based on Alzheimer's disease messages delivered through the social media channel like YouTube. It identified a gap between viewer's interests and available information and suggested that the videos should focus to meet user requirements and provide further information on related resources. The work in [8] highlights the value of social media as data source for deeper insights into the effect of a dementia diagnosis on caregivers and relatives. It used Reddit posts related to dementia diagnosis to understand the communities in which these people engage and the content of the posts were assessed with automated topic gathering technique.

The present research focuses on assessing the rate of cognitive ability degeneration with respect to Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) that are quite evident and are reported by caregivers on social media communities to obtain emotional, social and moral support, and its impact on the longevity of patients with Alzheimer's disease. An algorithm based on fuzzy logic has been designed for feature extraction in order to address uncertainty and imprecision in the reported data. Further, regression learning models are trained on the data to predict longevity of the patients and in almost all the models, reduced features using the fuzzy logic based feature extraction performed better than models trained on all the features.

3. Problem Context

Alzheimer's is a progressive disease which goes through various stages viz. Stage 1 through Stage 7, latter being the advanced stage. The details of conditions related to cognitive abilities at various stages are presented in Table 1.

Table 1: Conditions related to cognitive abilities encountered at different stages of Alzheimer's disease and the probable duration they last for

Stages	***Condition***	***Duration***
Stage 1	No Impairment, Normal Outward Behaviour	2-4 years
Stage 2	Very Mild Decline	
Stage 3	Mild Decline	
Stage 4	Moderate Decline	2-10 years
Stage 5	Moderately Severe Decline	
Stage 6	Severe Decline	1-3 years
Stage 7	Very Severe Decline	

The mild or early stage, generally, would continue for 2 to 4 years, moderate or middle stage for 2 to 10 years and severe or late stage would continue for 1 to 3 years [9] based on life style, care received and other related factors. The deterioration can be assessed with the help of biomarkers related to Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) through which the status of cognitive abilities is more obvious than compared to other biological parameters. Moreover, these are the highly discussed behavioural aspects on social media by the caregivers as the deterioration in these two kinds of activities increases the patient's dependence on the caregiver which actually changes the priorities and lifestyle of the caregiver as well. This provides rich source of behaviour related data of Alzheimer's patients in disease communities for healthcare research.

4. Data Collection and Participant Demographics

The present research is based on a retrospective longitudinal study using machine learning techniques to predict the final stage of Alzheimer's patients based on cognitive ability deterioration modelling. The data points are collected from Facebook support groups/communities built around Alzheimer's in which the participants, mostly caregivers, join the group when they feel the need for emotional and social support while giving care to their loved ones. The caregivers, generally join the group during stage 4-5 of the patient as they have to increase the level of care that have to be given during these stages. Most of the caregivers resort to social media platforms for getting help related to changing and diverse symptoms, life-style, caregiving tips etc. The observation and data collection period focuses on the support group participants who give care to patients in stage 6 to 7 encountering severe decline in cognitive skills. The care givers share information like when the patient was diagnosed (during stage 1 to 3), gender of the patient, for what activities/instrumental activities of daily living they are requiring assistance, the duration between the first diagnosis and the deterioration in various activities, the level of depression they are undergoing, and the year of final stage, in the form of posts, comments, and replies. This data of the patients over a timeframe of 7 years was collected using the information available on timeline by using appropriate search text to retrieve relevant information.

The data about the patients whose last stage is reported are considered for the experiment as the objective of the research is to predict the age at final stage based on the rate of cognitive deterioration. This would help in understanding the relation between the rate of cognitive deterioration and the final stage, and interventions to prolong the disease progression/cognitive deterioration can be designed to maintain a better life-style for the victims. [10] The data with respect to deterioration of cognitive abilities in ADL and IADL of the patients as reported by caregivers in terms of approximate months after diagnosis is collected along with age at which they were diagnosed with the disease, gender, and the year of final stage. The depression score is recorded based on Geriatric Depression Scale (GDS) on a scale of 1 to 15 with 15 indicating extreme depression [11].

Among the data points collected, 19 are female and 14 are male patients with an average age at which they were diagnosed being 67.52 years. Table 2 and Table 3 presents the details about the description of activities under ADL and IADL [12].

Table 2: Activities of Daily Living

Activity	*Description*
ADL1	Toilet
ADL2	Feeding
ADL3	Dressing
ADL4	Grooming
ADL5	Physical Ambulation
ADL6	Bathing

Table 3: Instrumental Activities of Daily Living

Activity	*Description*
IADL1	Ability to use telephone
IADL2	Shopping
IADL3	Preparation of food
IADL4	Housekeeping
IADL5	Laundry
IADL6	Independence in travel/transportation
IADL7	Responsibility in taking medicines
IADL8	Capability to handle finances

Very less number of data points are available on the public communities with the required information. The data points are anonymised before performing the analysis.

5. Feature Extraction

Fuzzy sets are potential tools for handling uncertainty and imprecision. While monitoring the progress of Alzheimer's disease in a patient, the recording of the time when a certain cognitive ability in ADL and IADL started deteriorating, may not be accurate as the deterioration happens eventually. Fuzzy modelling of the data will help to address this uncertainty efficiently. Moreover, having too many parameters results in problems during analysis and prediction tasks. The current work designed an algorithm for feature extraction using fuzzy logic to address the above issues of uncertainty and high-dimensionality.

Fuzzy sets are generally described by words using which we can represent data in a linguistic manner. Based on the details presented in Table 1, fuzzification of data with the level of cognitive ability deterioration (in months) is performed with linguistic variables. The universe of discourse defined over 132 months (11 years) is partitioned into fuzzy subsets, using trapezoidal membership function which is linear as well as trapezoidal to account for the vagueness in the duration reported. The trapezoidal membership function for a fuzzy number V can be represented by vector $[v_1, v_2, v_3, v_4]$ which can have the membership function given in Equation (1).

$$\mu_V(a) = \begin{cases} 0, & a < v_1, \\ \frac{a - v_1}{v_2 - v_1}, & v_1 \le a \le v_2, \\ 1, & v_2 \le a \le v_3, \\ \frac{a - v_4}{v_3 - v_4}, & v_3 \le a \le v_4, \\ 0, & a > v_4 \end{cases} \quad (1)$$

The details of fuzzification process is presented in Table 4.

Table 4: Fuzzification for the deterioration in cognitive abilities

Linguistic Term	*Severity Description*	*Vector for Trapmf*	*Defuzzified Centroid*
'VH'	Very High deterioration	[0,0,24,36]	15.2
'H'	High deterioration	[24,36,48,60]	42
'M'	Moderate deterioration	[48,60,72,84]	66
'L'	Low deterioration	[72,84,96,108]	90
'VL'	Very Low deterioration	[96,108,132,132]	116.8

The fuzzy representation of the linguistic terms is represented by Figure 1.

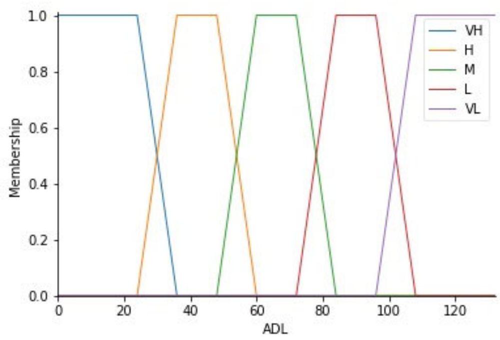

Fig. 1 Fuzzy membership representation for deterioration rate in ADL/IADL

The following Algorithm 1 is designed to extract the features con_ADL (consolidated ADL) and con_IADL (consolidated IADL) from the raw features ADL1 to ADL6 and IADL1 to IADL8, respectively.

Algorithm 1: ADL_Feature_Extraction
Global variable: (key, centroid) pairs of LinguisticTerm and associated Centroid value

```
function conADL_calc(arr,col):
Input: Each row values under ADL/IADL along
with the number of columns
Output: Consolidated ADL/IADL
for each value in arr:
   for each pair in (key,centroid):
      deviation<-value – pair(centroid)
      (key,deviationList)<- (key,deviation)
   end for
   low_deviation<-key,min(deviationList))
   dev_ADL <- (low_deviation(key),value)
end for
for each (key,value) in low_deviation:
   (key,avg) <-  (key,avg(value))
end for
if avg in all (key,avg) are similar:
   key_val <- all keys from (key,avg)
   for i,j in (key,centroid):
      for k in key_val:
         if k = = i:
            list <- collect all j
      end for
   end for
   con_ADL <- avg(list)
else
   less_deviation <- (key,min(avg))
   key_val <- less_deviation(key)
   for i,j in (dev_ADL):
      for k in key_val:
         if k= =i:
            list <- collect all j
      end for
```

```
        end for
        con_ADL <- avg(list)
    end if
    return con_ADL
```

The algorithm presents the steps followed to extract the features for ADL and IADL from the raw data. The deviation of each ADL/IADL value from the centroid of the trapezoidal membership function is calculated, the linguistic terms associated with average minimum deviation are recorded, minimum average linguistic term is selected and average of the ADL/IADL values associated with that linguistic term is calculated to present the consolidated ADL/IADL. These extracted features using fuzzy logic address the issue of uncertainty in data and reduce the dimensionality of the data for better analysis.

6. Experimental Results and Discussion

Once the features are extracted, preliminary investigation on the data characteristics is conducted. A strong positive correlation is observed between the age at which a person was diagnosed with Alzheimer's and the longevity/final stage which is presented in Figure 2.

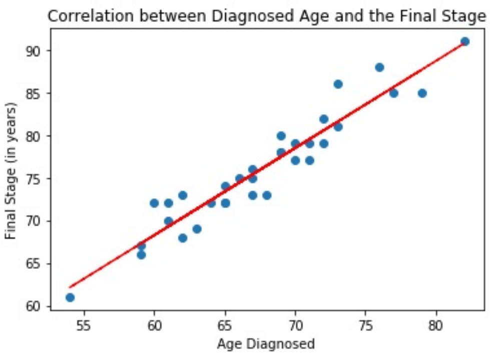

Fig. 2 The correlation between Age at which a person was diagnosed with Alzheimer's and Final Stage

The distribution of gender with respect to the age at which a person was diagnosed indicates that the female patients are acquiring the disease at an earlier age when compared to the male patients. This is depicted by the Figure 3.

Fig. 3 The distribution of Gender with respect to Diagnosed Age

The gender-wise deterioration in cognitive abilities, measured in months, with respect to ADL and IADL presented by the Figure 4 shows that the degeneration in ADL is encountered later than in IADL in both the genders whereas the degradation is faster in female victims rather than in male victims in both the cases.

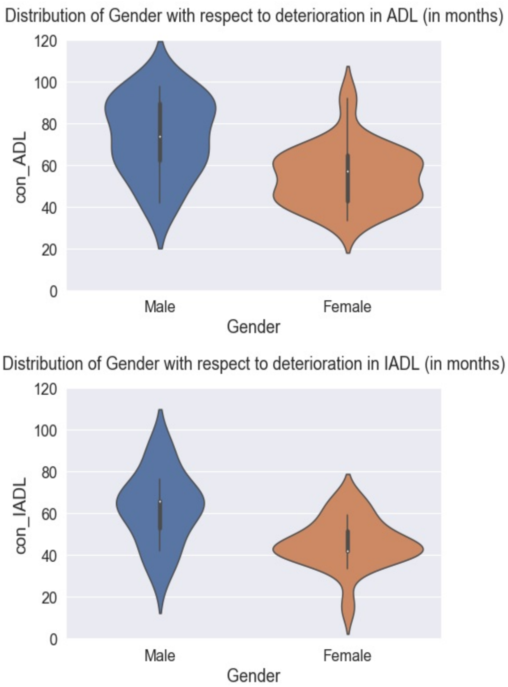

Fig. 4 Distribution of Gender with respect to ADL and IADL

Various regression techniques are used to train the models on full set of features and the reduced features as well, to predict the longevity/final stage of the patient. The correlation between the reduced features is calculated to see if there is any correlation among the predicting features which would actually reduce the reliability of the model. The correlation between the features is depicted through the Figure 5. From the Figure 5, it can be understood that 'con_ADL' and 'con_IADL' are strongly correlated that would have the same effect on the predicted feature 'FinalStage'. Removing each of these features in turn leads to select consolidated ADL for further model building as it gives better performance metrics when compared to considering consolidated IADL.

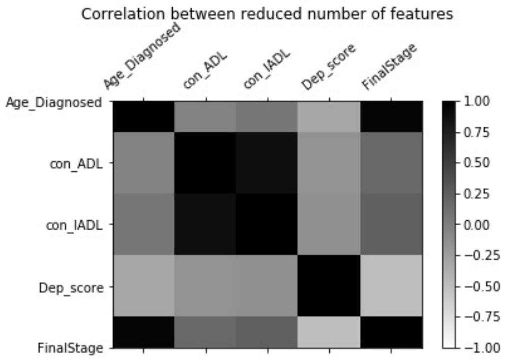

Fig. 5 Correlation graph for the reduced number of features

The regression techniques namely Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression, and Support Vector Regression (SVR) using RBF and Linear kernels are used to fit the data at hand by fine-tuning the hyper parameters. The regularization parameter α is set to 0.01 in the model with full features and to 0.001 in the model with reduced features for both Ridge and Lasso regression techniques. The parameters 'c' and $'\gamma'$ are tuned to 100 and 0.1 respectively, for RBF kernel of SVR, and to 100 and 0.001 respectively, for Linear kernel of SVR in both models with full features and reduced features. Leave One Out cross-validation technique is used as the number of data points is less. The performance metrics, Root Mean Square Error (RMSE) and the coefficient of determination, R-Squared value (R^2), for both basic model and the model with reduced features using the proposed algorithm are obtained and presented in the Table 5 and Figure 6, respectively.

Table 5: Mean RMSE with cross-validation and RMSE with held-out data obtained with all features and with features extracted using proposed algorithm

	Model with all features		***Model with features extracted using proposed algorithm***	
	Mean RMSE of cross-validation	***RMSE with held-out data***	***Mean RMSE of cross-validation***	***RMSE with held-out data***
Linear Regression	1.29	1.73	1.19	0.88
Ridge Regression	1.19	0.95	1.19	0.89
Lasso Regression	1.00	0.82	1.19	0.89
Polynomial Regression	2.55	1.49	1.88	0.99
Support Vector Regression (Kernel = 'rbf')	1.72	0.93	1.40	0.76
Support Vector Regression (Kernel = 'linear')	0.99	1.05	1.22	0.89

The following Figure 6 shows the R-Squared value obtained from the learning models fit on all features and fit on features extracted using the proposed algorithm. The RMSE of the models with features extracted using the proposed algorithm is lesser when compared to model with all features except in the case of Lasso Regression. The R-Squared values for held-out test set, with features extracted using the proposed algorithm, in linear, ridge, polynomial, and support vector regressor with 'rbf' and 'linear' kernels are showing better performance than with all features used.

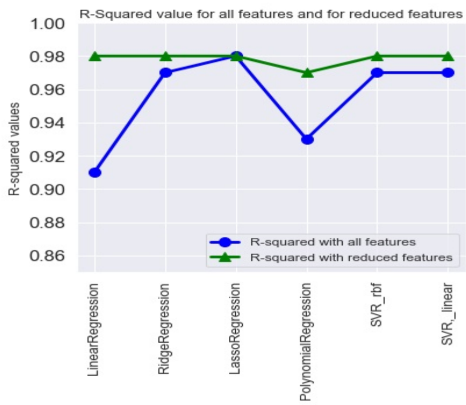

Fig. 6 Coefficient of determination – the R-Squared value for all features and for reduced features

In Lasso regression, the performance in both the scenarios doesn't vary as it performs equally well even with more number of features as with few variables that have medium or large effect, which is its inherent characteristic, though the RMSE increases a little with reduced number of features. The least RMSE is obtained with SVR using RBF kernel which shows its efficiency in predicting the longevity, given the rate of deterioration in cognitive abilities with other related features.

7. Conclusion

The present research focuses on modelling the cognitive deterioration in Alzheimer's patients with respect to their Activities and Instrumental Activities of Daily Living and assess their rate of influence in predicting the longevity of the patients. The data collected from social media communities was used to train various learning models with full set of features and reduced features extracted using the proposed algorithm based on fuzzy logic. In almost all the cases, models fit on reduced features has given better performance when compared to models fit on full set of features. It is observed that the onset of Alzheimer's disease happened earlier in age in females than in males and the deterioration of cognitive abilities related to IADL is much faster when compared to ADL irrespective of the gender. The Support Vector Regressor model with RBF kernel gives better performance with minimal Root Mean Square Error (RMSE) and best R – Squared value in predicting the longevity/final stage of Alzheimer's patients. Future work is directed towards stabilizing the learning models by using ensemble methods as the results may vary based on size of the sample and the data points selected into a sample as well.

Acknowledgement

We would like to acknowledge the support by Dr. Meenakshi Banerjee, Ph.D. (National Institute of Mental Health and Neuro Sciences – NIMHANS), Assistant Professor, Jindal School of Psychology and Counselling, O.P. Jindal Global University, for providing the knowledge on geriatric mental health, assessment indexes and assistance in the collection of relevant material, to progress with the research. Her specialized knowledge in Cognitive Behaviour Therapy and work experience in special clinical units such as Adult Psychiatry and Neuropsychology led our research in the right direction.

References

[1] William-Faltaos, D., Chen, Y., Wang, Y., Gobburu, J., Zhu, H.: *Quantification of disease progression and dropout for Alzheimer's disease.* International Journal of Clinical Pharmacology and Therapeutics 51(2), 20-31 (2013)

[2] Kim, K.W., Woo, S.Y., Kim, S. et al.: *Disease progression modelling of Alzheimer's disease according to education level.* Scientific Reports 10**,** Article No. 16808 (2020)

[3] Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D.: *Instantiated mixed effects modeling of Alzheimer's disease markers.* Neuroimage 15(142), 113-125 (2016)

[4] Raket, LL.: *Statistical Disease Progression Modeling in Alzheimer Disease*. Frontiers in Big Data 3(24), (2020)

[5] Joseph, G., Susan, M.L., William, J.J., Peter, T., Zoe, K.: *Modelling Prognostic Trajectories of Cognitive Decline due to Alzheimer's Disease.* NeuroImage: Clinical 26, (2020)

[6] Zhao, Q., Zhou, B., Ding, D., Teramukai, S., Guo, Q., Fukushima, M., et al.: *Cognitive Decline in Patients with Alzheimer's Disease and Its Related Factors in a Memory Clinic Setting, Shanghai, China*. PLoS ONE 9(4), (2014)

[7] Tang, W., Olscamp, K., Choi, SK., Friedman, D.B.: *Alzheimer's Disease in Social Media: Content Analysis of YouTube Videos.* Interactive Journal of Medical Research 6(2), (2017)

[8] Gkotsis, G., Mueller, C., Dobson, R.J.B., Hubbard, T.J.P., Dutta, R.: *Mining Social Media Data to Study the Consequences of Dementia Diagnosis on Caregivers and Relatives.* Dementia and Geriatric Cognitive Disorders 49, 295-302 (2020)

[9] Mary, E.E.: *The Progression of Alzheimer's Disease: What Are the Stages?* Healthline, (2018) https://www.healthline.com/health/stages-progression-alzheimers (accessed Jul. 24, 2021)

[10] Douglas, W.S.: *Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease - Practical Neurology.* Practical Neurology, (2019) https://practicalneurology.com/articles/2019-june/preclinical-prodromal-and-dementia-stages-ofalzheimers-disease (accessed Jul. 24, 2021)

[11] Douglas, W.S.: *Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease - Practical Neurology.* Practical Neurology, (2019) https://practicalneurology.com/articles/2019-june/preclinical-prodromal-and-dementia-stages-ofalzheimers-disease (accessed Jul. 24, 2021)

[12] R.T. Brink, T.L., Yesavage, J.A., Lum, O., Heersema, P., Adey, M.B., *Screening tests for geriatric depression.* Clinical Gerontologist 1, 37–44 (1982)

[13] Lawton, M.P., Brody, E.M., *Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living.* Gerontologist 9(3), 179–186 (1969)

Mutyala Sridevi is Assistant Professor and Research Scholar, VTU Research Centre in the Department of MCA at BMS Institute of Technology and Management, Bangalore, India. She is pursuing Ph.D. degree from VTU, Belagavi. She is Consultant for Larsen & Toubro, India in the area of Data Science. Her research interests include Social Media Data Analytics, Machine Learning, and Data Science. She published/presented 15 research papers in national/ international peer reviewed journals/ conferences among which some papers are scopus indexed.

Dr. Arun Kumar B.R. is Professor, Department of CSE and Research Supervisor, VTU Research Centre in the Department of MCA at BMS Institute of Technology and Management, Bangalore, India. He has published 70+ research papers in National/International Journals. He has presented and published nearly 20+ papers in the National/International Conferences/proceedings including IEEE international conferences. He has authored one Book and 6 Book chapters on contemporary topics. Presently he is guiding 4 Ph.D. scholars under VTU, delivered nearly 50 expert talks, working as MCA-BOS Chairman, VTU. He is a research paper reviewer/editorial board member for various indexed journals. His research interest includes Data Science, Computer Networks and Block Chain Technology.