Informatics in Medicine Unlocked 6 (2017) 28–35

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Devising an interpretable calibrated scale to quantitatively assess the dementia stage of subjects with alzheimer's disease: A machine learning approach

C.R. Aditya[a,c,*], M.B. Sanjay Pande[b,c]

[a] Department of CSE, VVIET, Mysuru, Karnataka, India
[b] Department of CSE, SITAR, Channapatna, Karnataka, India
[c] Visvesvaraya Technological University, Belgaum, Karnataka, India

ARTICLE INFO

Keywords:
Alzheimer's disease
Dementia
Multifactor dimensionality reduction
Knowledge discovery
Similarity measure
Affiliation analysis

ABSTRACT

Background: Machine learning and data mining techniques have been successfully applied on MRI images for detecting Alzheimer's disease (AD). But only a few studies have explored the possibility of AD detection from non-image data. These studies have applied traditional data visualization and classification algorithms. There is a need for new sophisticated learning algorithms, for detecting and quantifying the severity of AD by exploring the complex interactions between the features in AD subjects.

Method: In this work, a supervised learning model to effectively capture the complex feature interactions, in the sample space of AD data, is presented for knowledge discovery. The discovered knowledge is further used to quantify the similarity of a test subject to the demented class.

Results: Evaluation of the proposed model, on OASIS database of Alzheimer's subjects, validates the well established risk factors and identifiers for AD: Age, Socio-Economic Status, MMSE Score and Whole Brain Volume. The Test subjects are affiliated to either non-demented (ND) or AD class, with non-overlapping and measurable similarity indices: Female ND (CDR=0) [0.48–2.90], Female AD (CDR=0.5) [90.16–774.51], Female AD (CDR=1) [1633.90–7182.23], Female AD (CDR=2) [55258.51–66382.44], Male ND (CDR=0) [0.69–3.66], Male AD (CDR=0.5) [99.18–647.51] and Male AD (CDR=1) [3880.16–6519.40].

Conclusion: The outcome of the work clearly demonstrates that, supervised learning model can be used effectively to quantify the severity of AD on a standard measurable scale. This scale of distance can be used as a supplement for clinical dementia rating.

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia disease involving degeneration of the brain which is irreversible and gradually ends up with the complete brain failure. According to the statistics of Alzheimer's Association, AD accounts for 60–80% of the dementia cases [1]. In 2006, there were 26.6 million sufferers world-wide, and is expected to double by 2030 and triple by 2050 as projected by world health organisation. AD is predicted to affect 1 in 85 people globally by 2050, and at least 43% of prevalent cases need a high level of care [2,3]. Aging and other factors increase the possibility of neuron degeneration and can lead to AD. As the world is evolving into an aging society, the burdens and impacts caused by AD on families and the society will be increasingly pronounced. Studies have shown that AD is influenced by several factors such as age, education and socio-economic status. In normal aging, whole-brain volume decline begins in early adult hood and accelerates in advanced aging [4–8]. Preferential volume loss of gray matter [9] and regionally specific thinning of the cortex are also noted [10]. Level of education, sex, socioeconomic status, and cardiovascular health have been identified as contributing factors in volume decline in advanced aging, suggesting that subclinical health conditions contribute to age related changes in brain structure [11–15]. Individuals with clinically diagnosed AD show substantially reduced over all brain volumes relative to age matched peers as well as regional volume loss that has been well documented in the hippocampal formation, among other regions [16–20].

Many models have been created to analyse and detect AD from MRI images [21–23]. Techniques such as neural networks, support vector machines, decision tree classifiers have been successfully applied on MRI images to find the region of interest responsible for causing AD

* Corresponding author at: Department of CSE, VVIET, Mysuru, Karnataka, India.
E-mail address: aditya.ise@gmail.com (C.R. Aditya).

http://dx.doi.org/10.1016/j.imu.2016.12.004
Received 9 September 2016; Received in revised form 29 December 2016; Accepted 30 December 2016
Available online 03 January 2017

[24–26]. But only few studies have explored the possibility of AD detection from non-image data using machine learning approaches [27,28]. The challenges faced by these machine learning approaches are, capturing complex interactions among features in the data set and the ability to handle high-dimensional data(if required). Even though the existing approaches (both image based and non-image based methods) classify the test subjects to either normal or abnormal class, they fail to quantify the degree of abnormality on a standard scale. In medical data analysis such as detecting AD, it is important that a standard measurable scale is formulated to compare the severity of the disease. It is required to affiliate whether the sample is normal or abnormal on a measurable scale. Computational models provide a better means of modelling complex systems (such as the nervous system, neuro-disorders, etc.). To arrive at a good model for analysis, we need to design the knowledge base (KB) and ensure accuracy in the classification of data samples. Creation of a reference KB implies consolidating a huge database of control/training set into knowledge parameters, which provide meaningful and useful comprehension for later analysis. In this study we explore the possibility of analysing AD subjects from non-image data using supervised learning approach, and quantify their abnormality using measurable similarity indices. The improvements needed over existing models, which are addressed by the proposed model, are as follows:

1. Inter-feature relationship has to explored and captured in a presentable knowledge base.
2. Affiliation has to be carried out and represented through a suitable similarity index.
3. High Dimensional data (If any) must be handled.

The main theme of this research paper is to devise a computational model for critical analysis of an input (a test subject). Two phases are involved in the suggested computational model: (i) Learning Phase (ii) Recognition Phase. Learning phase involves building up a reference KB using known demented and non-demented (ND) subjects. In the recognition phase a test subject is contrasted with the reference KB to decide the label (ND/AD), and further the degree of belongingness (affinity) of the subject with respect to AD class.

This research communication is organized in the following way: Section 2 provides a detailed picture of computational model and the dataset used for evaluating the proposed model. Section 2.1 describes the data-set used for validating the model. *Multi-factor Affiliation Analysis* [29], a powerful computational model for KB creation and affiliation analysis of target data is explained in Section 2.2. Section 3 includes experimentation results and Section 4 includes useful discussions. Section 5 provides the conclusion.

2. Materials and methods

2.1. Data-set

The proposed model is evaluated on data obtained from the Open Access Series of Imaging Studies [30]. The dataset consists of a collection of 354 observations for 142 subjects aged 60–96. For some patients the observations are recorded more than once. The dataset contains both men and women subjects who are all right-handed. The data also includes the *education level* (EDUC) and *socio-economic status* (SES) of the subjects. Moreover, some other medical statistics exist in the dataset, including *intracranial volumes* (eTIV) and *brain volumes* (nWBV) of the subjects. The two groups of subjects are demented and non-demented in which the patient has the AD or not, respectively. Patients that develop the AD during the tests are grouped as *converted*.

The features in the dataset are explained in Table 1. *Clinical Dementia Rating* (CDR) can only take values 0, 0.5, 1 and 2. CDR being equal to 0 corresponds to non-demented subject. CDR being equal to 0.5, 1 and 2 corresponds to very mild dementia, mild dementia and moderate dementia respectively. This medical test carries significance to entitle a subject as Alzheimer patient. The *minimental state examination* (MMSE) is a questionnaire test that has 30 questions covering cover arithmetic, memory, and orientation to examine the cognitive situations of individuals.

2.2. Multifactor affiliation analysis (MAA)

As pointed out in introduction section, the improvements needed over existing models are the ability to effectively capture inter-feature relationships in data and to handle high dimensional data(if any). Also the knowledge base has to be further used in quantifying the similarity of a test subject with a target class. Non-parametric methods such as multifactor dimensionality reduction (MDR) and combinatorial partitioning method [31–33] were developed to handle high dimensional data and to uncover complex relationships between the data features. MAA, an MDR based model, effectively captures the inter-feature relationships in the form of a knowledge-base containing affiliation weights, which can be further used for quantifying the test subjects on a measurable scale. MAA uses the Case-Control ratio comparison concept of MDR [34] to reduce the $m*n$ feature space into $m*1$ dimensional space. As the feature dimension is reduced, a knowledge base, which captures the complex inter-feature relationships, is generated in the form of *multi-factor affiliation table* (MAT). The affiliation of a test subject with the AD class is carried out by comparing the feature values of the test subject with the knowledge in MAT.

2.2.1. MAT generation

Initially the feature values are converted to a value 'n' or 'a', depending on whether the values are close to the mean of ND or AD subjects respectively. In the changed representation space of data, the possible combinations of values for any two selected features are: $n-n$, $n-a$, $a-n$ and $a-a$. The occurrence count of each of these combinations in ND and AD subjects is compared, and selected features are merged to form a new feature. The value of new merged feature for each combination of values in old features and multifactor affiliation weight, calculated using equation (1), is updated in the MAT. This process is repeated till $m*n$ dimensional dataset gets reduced to $m*1$ dimensional dataset. MAT thus created will contain one row for each combiation of feature values and one column for every new feature created. So a MAT generated for an $m*n$ dataset will have 4 rows and $n-1$ columns. The interpretation of MAT is summarized in Table 2.

Multifactor affiliation weight (MAW), updated for each combination of values in selected features, quantifies the degree of occurrence of a value combination in AD group. MAW is designed such that it varies between zero to one, where one is the highest quantifying factor for a value combination. Let $ND_{occurance}$ represent the occurrence count of a value combination in ND subjects and $AD_{occurance}$ represent the same in AD subjects. Then MAW exhibits the following properties:

Property 1: If $AD_{occurance} = 0$ and $ND_{occurance} = m$, then $MAW=0$. A value zero for MAW indicates that the occurance of a value combination is unique to ND subjects, and AD subjects do not exhibit that pattern of combination.

Property 2: If $AD_{occurance} = m$ and $ND_{occurance} = 0$, then $MAW=1$. If the degree of occurance of a value combination is unique and at its maximum value for AD subjects, then MAW must be equal to one.

Property 3: If $AD_{occurance} >= ND_{occurance}$ then $0.5 < MAW < 1$, else $0 < MAW < 0.5$. As the occurance count of a value combination increases in AD subjects then the value of MAW increases towards 1, else it decreases towards 0.

$$MAW = 0.5 + \left[\frac{AD_{occurance}}{2*m} - 0.5*\frac{ND_{occurance}}{2*m}\right] \quad (1)$$

Table 1
Features and their value ranges in OASIS AD database.

Feature	Explanation	Value range
SubjectID	Unique Identifier for individual subjects	1–142 (number of subjects)
MRI ID	Unique identifier for each test. One subject may have more than one MRI ID	1–354
Group	Class label	Non-demented or demented or converted
Visit	Number of times a subject has visited for test	1–5
MRDelay	Delay of visit by a subject since last visit (Number of days)	183–1707 days between visits
CDR	Clinical dementia rating	0 or 0.5 or 1 or 2
Gender		Male or female
Age	Age of a subject at the time of test observation	
EDUC	Education level of test subject	Higher value denotes higher education level
SES	Socio Economic Status assessed by hollingshead index of social position	1–5
MMSE	Mini Mental State Examination Value assessed through questionnaire	0–30
eTIV	Estimated total intracranial volume	1488 ± 176.13
nWBV	Normalized whole-brain volume, expressed as a percent of all voxels ("constant" for any value of estimated total intracranial volume)	0.730 ± 0.037
ASF	Atlas Scale Factor; volume scaling factor for brain size ("constant" for any value of estimated total intracranial volume)	1.195 ± 0.138

Table 2
Interpretation of the values in multifactor-affiliation table(MAT). MAT data Dj_i indicates data from row-j and column-i.

		Affiliation of test/training sample		
FIT data	Feature interaction value	Inferred from previously observed features' values	Inferred from present feature value	Net/overall affiliation
D1i	0	Non-demented	Non-demented	Non-demented
D2i	0	Non-demented	Demented	Non-demented
D3i	0	Demented	Non-demented	Non-demented
D4i	0	Demented	Demented	Non-demented
D1i	1	Non-demented	Non-demented	Demented
D2i	1	Non-demented	Demented	Demented
D3i	1	Demented	Non-demented	Demented
D4i	1	Demented	Demented	Demented

The training data set to be used in knowledge-base (MAT) creation will consist of 2*m subjects with 'm' equal ND and AD subjects. So the factors $AD_{occurance}/(2*m)$ and $ND_{occurance}/(2*m)$ can never be greater than 0.5. Further the MAW should quantify the occurance of a value combination with respect to AD subjects as compared with ND subjects, and must be equal 1 if such an observed combination is strongly affiliated to the AD class. Since only 50% of the training subjects are AD subjects, a value 0.5 is added as a term to the equation (1) and 0.5 weightage is given to the factor $ND_{occurance}/(2*m)$.

The steps for creating the multifactor affiliation table by performing dimensionality reduction are as follows.

Input: n-dimensional feature values [f1, f2,... fi,... fn,] for 'm' AD and 'm' ND samples.

Output: Multifactor Affiliation Table.

Step 1: Calculate μ_{ni} and μ_{ai} for each feature f_i in the sample space, where μ_{ni} and μ_{ai} denote the *ith* feature value mean of ND and AD subjects respectively.

Step 2: Convert all feature values of training subjects to either 'n': if they are close to the value of μ_{ni}, or to 'a': if the values are closer to μ_{ai}.

Step 3: For any two selected features in the changed representation space of data, calculate the occurance count of each possible value combination in ND($ND_{occurance}$) and AD ($AD_{occurance}$)subjects.

Step 4: Using the occurance counts calculated in Step-2, merge the two features into one. If $AD_{occurance}$ is greater than $ND_{occurance}$ then the value of new merged feature (multifactor-affiliation value) will be 'a', else 'n'.

Step 5: Calculate multifactor-affiliation weight (MAW) for all combinations of feature values using equation (1).

Step 6: Create a new column in MAT and update it with "multifactor-affiliation value, multifactor-affiliation weight" pair for all the four possible old features' value combinations.

Repeat steps 3–6 till the n-dimensional feature space is reduced to a single dimensional space.

2.2.2. Affiliation analysis

The cognition stage involves representing the feature wise knowledge parameters of the training group using mean and standard deviation, which forms the reference base. The recognition stage involves testing whether a presented test subject belongs to the healthy (ND) group or not. The most important objective, in either case is to estimate the amount/degree of affiliation of the sample to the reference base (in case it is healthy) or to estimate the amount/degree of being away from the reference base. In this work we have employed a Euclidean distance based similarity measure to estimate the degree of a test subject being away from the reference base (ND group) on a measurable scale.

The Eucledian distance [35] between two points (p,q) and (r,s) is given by

$$Eucledian_{distance} = \sum \sqrt{(p-r)^2 + (q-s)^2} \quad (2)$$

We use the idea behind eucledian distance to find the feature wise distance of a subject with respect to the mean of the feature in consideration. Let μ_i denote the mean value for the *ith* feature in a training data set containing equal number of ND and AD samples. Given the mean value μ_i along with the feature value f_i of a test sample, the *Individual Feature Distance* is calculated using $((\mu_i - f_i)^2)$, which denotes the farness/closeness of a subject from the reference base in *ith* dimension. The net/Overall distance ($NetFeature_{distance}$) for a test subject is the sum of all the individual feature distances calculated using equation (3). It is the final affiliation distance after processing all the feature values ($i = n - 1$ in equation (3), where 'n' is the total number of features in dataset).

$$FD_{i+1} = \sqrt{(FD_i)^2 + (\mu_{i+1} - f_{i+1})^2} \quad (3)$$

Association/Dissociation of a subject with respect to degree of affiliation is done through calculated *multifactor-affiliation weights*. The class assigned to the subject based on the previously observed $i - 1$ features, and the class assigned to the subject based on newly observed feature 'i' will form the row index to the MAT for affiliation analysis of a test subject. The column index will be one less than the number of features assessed for affiliation analysis of the test subject. An affilia-

tion value '*n*' or '*a*' obtained from the MAT, for the given row and column indices, will affiliate the test subject to ND/AD class based on observed features. The weight, obtained along with the affiliation value, quantifies the degree of similarity of a test subject with reference to the AD group.

The affiliation algorithm used in our model estimates the degree of test subject being away from the ND group on a measurable scale. When a test subject is presented for affiliation analysis, the feature values are initially compared with the respective feature means of training subjects, calculated during KB creation stage, and are converted to a value '*n*' or '*a*' based on the closeness of the value to ND mean or AD mean respectively. This conversion step will assign the subject to either ND or AD class across each feature. But the overall affiliation, after observing all features, is calculated using equation (4).

$$AD_{AffiliationDistance} = \frac{Affiliation_{weight}}{n-1} * NetFeature_{distance} \quad (4)$$

The $Affiliation_{weight}$ gives overall weightage for the degree of affiliation of test subject to the AD class. The features of the test subjects are merged iteratively into one new feature by comparing the values with the knowledge base (MAT), and the multifactor-affiliation weights from the MAT are iteratively added as the feature values are merged to get $Affiliation_{weight}$.

The overall steps involved in affiliation analysis are as follows.

Input: Multifactor Affiliation Table and n-dimensional feature values [f1, f2,... fi,... fn,] for a test sample.

Output: $AD_{AffiliationDistance}$ for the test subject.

Step 1: Calculate the $NetFeature_{distance}$ using equation (3) by iteratively summing up the individual feature distances.

Step 2: Convert all the test subject feature values ($f_1, f_2 \ldots f_i$) to either '*n*' or '*a*', based on the closeness of feature value to the mean of ND (μ_{ni}) or AD (μ_{ai}) subjects calculated in MAT generation phase. Initialize $Affiliation_{weight}$ to 0.

Step 3: Select two features f_i and f_{i+1}, and access the MAT based on the selected two feature values (*row index*) and number of features merged (*column index*).

Step 4: Merge the two selected features into a new feature. New merged value will be the *multifactor-affiliation value* present in the accessed cell of MAT.

Step 5: Add the *multifactor-affiliation weight* in the MAT entry to $Affiliation_{weight}$. Repeat Steps 2–5 till the n-dimensional feature space of test subject is reduced to a single dimensional space.

Step 6: Calculate the $AD_{AffiliationDistance}$ by using equation (4)

3. Results

Experimentation was carried out on the Alzheimer's data obtained from OASIS. Separate experimentations were conducted for male and female subjects. From the data set selected, the ND and AD subjects with missing values were discarded and only the subjects with complete data entries were input to the experimentation. The subjects in the dataset are characterized using the CDR scale [36,37]. Group attribute which describes the subjects as ND or AD, based on CDR value, is considered as target label for all experiments. The subjects for training and testing phase were input without considering the *subject-ID*, *MRI-ID* and *visit* attributes of the subjects. Two different samples of a same subject were considered to be different samples for training the system. In total the subjects were assessed with respect to the five attributes - Age, Education, SES, MMSE and nWBV. Since the subjects in the data set are only right handed, the *Hand* attribute was not considered for experimentation, as it would not affect either the knowledge-base creation phase or affiliation analysis of a test subject. Tables 3 and 4 illustrate the feature-wise parameters derived from the chosen subjects for experimentation. μ_{ni}, σ_{ni}, max_{ni} and min_{ni} represent the mean, standard-deviation, maximum and minimum values, of i^{th} feature, in ND subjects. μ_{ai}, σ_{ai}, max_{ai} and min_{ai} represent the same in AD subjects.

Table 3
Feature-wise knowledge parameters for Female subjects.

Feature	Female ND samples				Female AD samples			
	μ_{ni}	σ_{ni}	max_{ni}	min_{ni}	μ_{ai}	σ_{ai}	max_{ai}	min_{ai}
Age	78.32	8.79	97	60	77.12	6.63	98	67
Educ	15.22	2.90	23	12	13.0	3.37	18	8
SES	2.52	1.06	4	1	3.06	1.22	5	1
MMSE	29.32	0.83	30	27	24.08	6.51	30	15
nWBV	0.744	0.03	0.82	0.684	0.71	0.04	0.77	0.65

Table 4
Feature-wise knowledge parameters for Male subjects.

Feature	Male ND samples				Male AD samples			
	μ_{ni}	σ_{ni}	max_{ni}	min_{ni}	μ_{ai}	σ_{ai}	max_{ai}	min_{ai}
Age	7.96	7.28	92	60	76.66	7.92	92	62
Educ	15.21	2.74	20	1	14.23	3.41	20	6
SES	2.41	1.15	4	1	2.71	1.19	4	1
MMSE	2.91	1.02	30	26	25.71	4.99	30	15
nWBV	0.73	0.03	0.83	0.66	0.71	0.03	0.80	0.64

The following insights are shown by the knowledge parameters (Mean and standard deviation) obtained:

- The SES values of AD subjects are distributed across a higher value range. Compared to ND subjects, AD subjects exhibit lower nWBV values.
- The Educ level in male ND and AD subjects is same. But the Educ level of female AD subjects is relatively lesser than AD subjects.
- The distribution of Age and MMSE values are same for ND and AD subjects (both male and female). The demented and non-demented subjects cannot be effectively distinguished from these two features.

The first experimentation was carried out for female subjects. 50 ND and 50 CE subjects were chosen as training data for generating MAT. For the affiliation analysis, 50 subjects of both ND and AD were chosen randomly. Table 5 shows the MAT generated for the chosen female training set. For the male subjects a separate experimentation was carried out with 60 ND and 60 CE subjects for generating MAT. Table 6 shows the MAT generated for the chosen male training set. For affiliation analysis randomly chosen 120 subjects, comprising of both ND and AD subjects, were considered. The individual feature distances calculated during the process of affiliation analysis is tabulated in Tables 7 and 8. The results are presented in "*minimum-maximum*" format, the *minimum* corresponding to the smallest affiliation distance computed over all the test subjects and the *maximum* corresponding to the largest affiliation distance. It is evident from the results that ND subjects cannot be discriminated from AD subjects through individual feature distances.

The multi-factor affiliation tables, generated for female and male subjects, provide us with the following insights:

- The mean of age for ND subjects is greater than AD subjects. But it is obvious that, if the people of lower age suffer from dementia then older people also suffer from the dementia with equal or higher

Table 5
Multifactor Affiliation table for 50 ND and 50 AD female subjects.

n, 0.44	n, 0.43	n, 0.21	n, 0.27
a, 0.50	n, 0.44	a, 0.65	n, 0.39
n, 0.43	n, 0.49	n, 0.45	a, 0.60
a, 0.63	a, 0.64	a, 0.69	a, 0.74

Table 6
Multifactor Affiliation table for 60 ND and 60 AD male subjects.

n, 0.45	n, 0.44	n, 0.31	n, 0.34
a, 0.54	a, 0.51	a, 0.60	n, 0.4
a, 0.50	n, 0.48	n, 0.42	a, 0.58
a, 0.50	a, 0.55	a, 0.65	a, 0.67

Table 7
Feature-wise distance computations for female subjects.

Feature	Female ND Samples [Min, Max]	Female AD Samples [Min, Max]
Age	0.84, 49.15	0.84, 51.78
Educ	5.2, 51.86	1.46, 51.86
SES	12, 38	12, 62
MMSE	2.13, 15.46	2.13, 95.46
nWBV	0.10, 45.77	0.10, 54.22

Table 8
Feature-wise distance computations for male subjects.

Feature	Male ND Samples [Min, Max]	Male AD Samples [Min, Max]
Age	0.10, 53.02	0.10, 53.02
Educ	5.59, 34.16	1.54, 65.82
SES	13.88, 52.77	13.88, 52.77
MMSE	0.66, 20.66	0.66, 94
nWBV	0.74, 55.80	0.30, 55.80

probability. Therefore we consider that the possibility of dementia increases with age as true. We can also state that AD is not a disease caused due to aging, but it is a type of dementia that can occur to people at any age.

- The older people with less education are more likely to be demented. In case of females the highest probability recorded in MAT is 0.63, and is 0.54 in case of male subjects. But it should be noted that the probability of dementia in male subjects is not satisfactorily high, as compared to female subjects.
- As social-economic status increases for less educated male subjects, the probability of dementia does not increase significantly. The educated male subjects, who are less likely to be demented, will have a probability of 0.51 with the increase in their SES. The same is not true for female subjects. Only, older female subjects with less education and increased SES are affected by alzheimer's disease (0.64 probability). The SES has less effect on more educated females 0.44–0.49 probability).
- Reduced MMSE value in both male and female subjects increase the possibility of dementia.
- Further observation of nWBV reveals that the subjects (both male and female) with lesser nWBV value are more likely to be demented.

Table 9 summarizes the $AD_{affiliation}$ distances obtained for male and female subjects. ND and AD subjects (both male and female) have overlapping sets of affiliation distances, and are not clearly separable on a measurable scale. The $AD_{affiliation}$ distances have to be refined further for proper quantification of test subjects. CDR has become a *De-Facto* standard for categorizing subjects as either ND or AD, and the subjects used in this experimentation are also characterized using the CDR scale. We propose to use the obtained $AD_{affiliation}$ distances in conjunction with CDR for clearly distinguishing ND and AD subjects on a measurable distance scale. The $AD_{affiliation}$ distances obtained through our model are refined with a refining weightage factor '*W*' as:

Refined $AD_{AffiliationDistance}$ = W times $AD_{AffiliationDistance}$.

With W = 0.1, 10, 100 and 1000 for subjects having CDR = 0, 0.5, 1 and 2, respectively.

The refined $AD_{affiliation}$ distances obtained for male and female subjects are summarized in Table 10.

The range of affiliation distances, after refining with the factor '*W*', obtained for female ND and AD subjects are distinguishable into two non-overlapping sets separated by clearly distinguishable range. Male ND(CDR=0) and AD(CDR=0.5) subjects have overlapping sets of $AD_{affiliation}$ distances and are not clearly separated. When observed closely, only two ND male subjects (shown in Table 11) were yielding affiliation distances of far higher values than others from the same ND set. The CDR value for both is 0.5, denoting mildly demented subjects, but the group label for them is non-demented in the data set used. Hence after considering the two subjects as outliers, the experimentation was re-iterated without those two subjects. The knowledge parameters created did not significantly vary from the previously obtained values. The refined range of affiliation distances obtained for AD subjects was clearly separable from ND subjects. The final, refined $AD_{affiliation}$ distances for male and female subjects is shown in Table 12. The affiliation distances of all test subjects are shown in Fig. 1.

Table 9
Range of $AD_{affiliation}$ distances for test subjects.

Subjects	AD affiliation distance			
	ND [Min, Max]	AD (Very Mild) [Min, Max]	AD (Mild) [Min, Max]	AD (Moderate) [Min, Max]
Female	4.89, 29.08	09.01, 77.45	16.33, 71.82	55.25, 66.38
Male	7.64, 40.10	10.65, 63.86	38.96, 68.11	–

Table 10
Range of $AD_{affiliation}$ distances for test subjects after scaling with respect to CDR.

Subjects	AD affiliation distance			
	ND [Min, Max]	AD (Very Mild) [Min, Max]	AD (Mild) [Min, Max]	AD (Moderate) [Min, Max]
Female	0.48, 2.90	90.16, 774.51	1633.90, 7182.23	55258.51, 66382.44
Male	0.76, 4.01	106.53, 638.66	3896.42, 6811.79	–

4. Discussion

From the insights obtained from KB generated, following inferences can be drawn:

Inference 1: AD is not a disease caused due to aging, but it is a type of dementia that can occur to people at any age. The possibility of dementia increases with aging.

Inference 2: Female subjects with less education are more likely to be demented. Compared to female subjects, male subjects tend to be less affected from their education level.

Inference 3: The SES has less effect on more educated females. Males are generally affected by their SES irrespective of their education level.

Inference 4: The distribution of MMSE parameter values for ND and AD subjects are same. But the MAT infers that a subject is likely to be more demented if MMSE score is less, provided, other features (age, education, SES) indicate the possibility of dementia.

Inference 5: Subjects (both male and female) with lesser nWBV value are more likely to be demented.

Comparison of etiologic fractions for low education level with other well established risk factors for AD has suggested that low education level may be the most significant risk factor for AD [38]. Lower

Table 11
Outlier subjects having higher $AD_{affiliation}$ distances compared with others from the set of ND male subjects.

Subject ID	MRI ID	Group	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
OAS2_0005	OAS2_0005_MR2	Non-Demented	83	12	4	29	0.5	1701	0.711	1.032
OAS2_0017	OAS2_0005_MR3	Non-Demented	81	12	3	27	0.5	1814	0.759	0.968

Table 12
Refined range of $AD_{affiliation}$ distances for female and male test subjects.

	AD affiliation distance			
Subjects	ND [Min, Max]	AD (Very Mild) [Min, Max]	AD (Mild) [Min, Max]	AD (Moderate) [Min, Max]
Female	0.48, 2.90	90.16, 774.51	1633.90, 7182.23	55258.51, 66382.44
Male	0.69, 3.66	99.18, 647.51	3880.16, 6519.40	–

educational attainment in women is associated with an increased risk of dementia-related death independently of common risk behaviours and comorbidities [39]. Low SES during early life is associated with reduced body size and delayed function [40,41]. Low SES is likely to be related to impairment of brain development as well. Deficient brain development, could increase the risk of dementia late in life, the result of reduced brain reserve capacity [42,43]. The knowledge base created by our model also shows that education level and SES can be the cause for dementia in subjects, and the extent to which features influence AD may vary across male and female subjects.

Previous studies have recognised that mental status tests scores of an individual is strongly associated with the individual's educational attainment [44–46]. Some have even suggested that individuals with lower education level may perform more poorly because of impaired test-taking ability [47–49], with more highly educated persons better able to compensate for early deficits [50]. MMSE is discriminated well between CDR stages 0.5, 1, 2, and 3 but performs poorly in the separation between CDR stages zero and 0.5. MMSE can be used as surrogate measure for the CDR for the staging of dementia in AD [51]. The distribution pattern recorded and zones created in our experimentation clearly indicate that MMSE cannot be used as an important parameter for assessing AD. nWBV, is known to decline across the adult life span with acceleration in advanced aging [4,5,52,7,8].

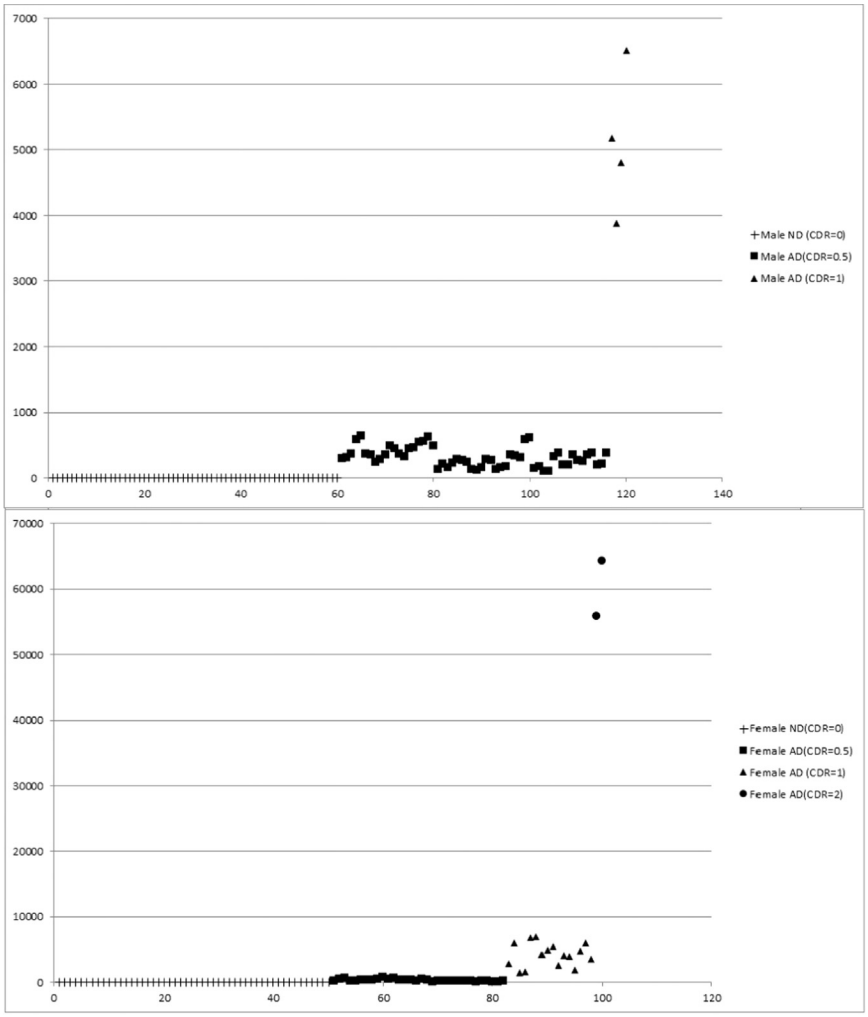

Fig. 1. Affiliation distances obtained for ND and AD subjects using *Multifactor Affiliation Analysis*. Y–axis represents affiliation distance and X–axis represents subject ID of test subjects used for experimentation.

Table 13
Comparision of proposed and existing AD assessing models.

Computational models	Comparision parameters		
	Dimensionality reduction of data	Logical rules or feature interaction details extracted	Quantifies severity of AD in test subjects
Classification Tree	No	Yes	Partially
Generic-Distribution Zones	No	No	Yes (<85% accuracy)
Multifactor Afiliation Analysis	Yes	Yes	Yes (>95% accuracy)

Unsurprisingly, the insights from the knowledge base are in agreement with those described in previous studies.

In *Multifactor Affiliation Analysis*, the affiliation of a test subject is assessed across each dimension by comparing a feature value with the mean of training subjects. The *multifactor-affiliation weights* calculated, by considering the inter-feature relationships, denote the extent to which the features collectively influence the class of a subject. The results obtained show that, it is possible to quantify the severity of AD on a standard measurable scale. But the quantified distances have to be scaled with proper weights, with respect to CDR of subjects, to separate ND and AD subjects on a measurable scale of affiliation distances.

The efficacy of the proposed *MAA*-model was compared with *classification tree analysis* [27] and *generic-distribution zones analysis* [28]. The parameters used for comparing the models were high dimensional data handling capability, exploring complex interactions between data features and quantifying the severity of AD. Table 13 summarizes the efficacy of each model, and it is evident that the efficacy of our model is high compared with other two models.

5. Conclusion

Classifying a test subject, under diagnosis for Alzheimer's disease (AD), to either non-demented(ND) or demented class is a complex issue. In this paper, an exploratory data analysis approach is proposed to carry out affiliation analysis of AD subjects. The explicit need was to cognize a reference knowledge base(KB) which could be used in the recognition stage, for an accurate discrimination of the test subject as being ND or AD. The proposed computational model explores the training set of subjects, and captures the feature interaction details and each feature's impact in causing AD in the form of a KB. The present computational model is able to match the test subject with the KB, thus understanding the closeness/farness of the subject with respect to the AD class. Overall affiliation of test subjects is done by combining the individual weights, derived from feature wise interaction details, in the KB. The efficacy of the proposed model, as compared with those existing in literature, is high in terms of high dimensional data handling capability, exploring complex interactions between data features and quantifying the severity of AD.

The value of the distance computed, for affiliation of test subjects to either ND or AD class, can be read on an interpretable scale to recognize the severity of the progression of the disorder. This scale of distance can be used as a supplement for clinical dementia rating. The affiliation distances, obtained from proposed model, for quantitatively assessing the test subjects is: Female ND (CDR=0) [0.48–2.90], Female AD (CDR=0.5) [90.16–774.51], Female AD (CDR=1) [1633.90–7182.23], Female AD (CDR=2) [55258.51–66382.44], Male ND (CDR=0)[0.69–3.66], Male AD (CDR=0.5) [99.18–647.51] and Male AD (CDR=1) [3880.16–6519.40].

We have proposed a possible approach for devising a distance scale for quantifying AD from non-image data, using machine learning approach. The proposed model works on numeric input values. Researchers can further use MRI image-based methods as a pre-processing step for the proposed model. Useful information from images can be extracted in the form of numeric values and used as input to our proposed model for better results. In future, experts from across the globe can apply the proposed model on standardized AD data, and devise a calibrated scale of AD affiliation distances.

Acknowledgement

The authors would like to thank Dr. B S Mahanand (Dept. of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India) for his expertise and able guidance in analysing the experimental works.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.imu.2016.12.004.

References

[1] Alz.org, Alzheimer's and dementia., Online URL ⟨http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp⟩. [accessed 2.02.16]
[2] Brookmeyer R, Johnson E, Ziegler-Grahamm K, Arrighi HM. 01-02-01: forecasting the global prevalence and burden of alzheimer's disease. Alzheimer'S Dement 2007;3(3):S168.
[3] Alzheimer.ca, World health organisation repot, Online URL ⟨http://www.alzheimer.ca/en/Get-involved/Raise-your-voice/WHO-report-dementia-2012⟩. [accessed 24.01.13]
[4] DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, Beiser A, D'Agostino R, Wolf PA. Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. Neurobiol Aging 2005;26(4):491–510.
[5] Fotenos AF, Snyder A, Girton L, Morris J, Buckner R. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and ad. Neurology 2005;64(6):1032–9.
[6] Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal mapping of cortical thickness and brain growth in normal children. J Neurosci 2004;24(38):8223–31.
[7] Durston S. A review of the biological bases of adhd: what have we learned from imaging studies?. Ment Retard Dev Disabil Res Rev 2003;9(3):184–95.
[8] Blatter DD, Bigler ED, Gale SD, Johnson SC, Anderson CV, Burnett BM, Parker N, Kurth S, Horn SD. Quantitative volumetric analysis of brain mr: normative database spanning 5 decades of life. Am J Neuroradiol 1995;16(2):241–51.
[9] Raz N, Williamson A, Gunning-Dixon F, Head D, Acker JD. Neuroanatomical and cognitive correlates of adult age differences in acquisition of a perceptual-motor skill. Microsc Res Tech 2000;51(1):85–93.
[10] Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, Morris JC, Dale AM, Fischl B. Thinning of the cerebral cortex in aging. Cereb Cortex 2004;14(7):721–30.
[11] Fotenos AF, Mintun MA, Snyder AZ, Morris JC, Buckner RL. Brain volume decline in aging: evidence for a relation between socioeconomic status, preclinical alzheimer disease, and reserve. Arch Neurol 2008;65(1):113–20.
[12] Raz N, Rodrigue KM, Acker JD. Hypertension and the brain: vulnerability of the prefrontal regions and executive functions. Behav Neurosci 2003;117(6):1169.
[13] Koga H, Yuzuriha T, Yao H, Endo K, Hiejima S, Takashima Y, Sadanaga F, Matsumoto T, Uchino A, Ogomori K, et al. Quantitative mri findings and cognitive impairment among community dwelling elderly subjects. J Neurol, Neurosurg Psychiatry 2002;72(6):737–41.
[14] Coffey CE, Saxton J, Ratcliff G, Bryan R, Lucke J. Relation of education to brain size in normal aging implications for the reserve hypothesis. Neurology 1999;53(1), [189-189].
[15] Coffey CE, Lucke JF, Saxton JA, Ratcliff G, Unitas LJ, Billig B, Bryan RN. Sex differences in brain aging: a quantitative magnetic resonance imaging study. Arch Neurol 1998;55(2):169–79.
[16] Jack CR, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, Smith GE, Ivnik RJ, Kokmen E. Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. Neurology 1997;49(3):786–94.
[17] Raz N, Gunning FM, Head D, Dupuis JH, McQuain J, Briggs SD, Loken WJ, Thornton AE, Acker JD. Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. Cereb Cortex 1997;7(3):268–82.

[18] Killiany RJ, Moss MB, Albert MS, Sandor T, Tieman J, Jolesz F. Temporal lobe regions on magnetic resonance imaging identify patients with early alzheimer's disease. Arch Neurol 1993;50(9):949–54.
[19] Jack CR, Petersen RC, O'Brien PC, Tangalos EG. Mr-based hippocampal volumetry in the diagnosis of alzheimer's disease. Neurology 1992;42(1), [183-183].
[20] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein H, Vermersch P, Kuiper M, Steinling M, Wolters EC, Valk J. Atrophy of medial temporal lobes on mri in "probable" alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 1992;55(10):967–72.
[21] Mahanand B, Suresh S, Sundararajan N, Kumar MA. Alzheimer's disease detection using a self-adaptive resource allocation network classifier. In: Proceedings of the International Joint Conference on of the Neural Networks (IJCNN), IEEE; 2011, pp. 1930–1934.
[22] Zhang Y, Wang S, Dong Z. Classification of alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. Prog Electromagn Res 2014;144:171–84.
[23] Babu GS, Suresh S, Mahanand B. Alzheimer's disease detection using a projection based learning meta-cognitive rbf network. In: Proceedings of the International Joint Conference onNeural Networks (IJCNN), IEEE; 2012. p. 1–8.
[24] Mahanand B, Suresh S, Sundararajan N, Kumar MA. Identification of brain regions responsible for alzheimer's disease using a self-adaptive resource allocation network. Neural Netw 2012;32:313–22.
[25] Kubota T, Ushijima Y, Nishimura T. A region-of-interest (roi) template for three-dimensional stereotactic surface projection (3d-ssp) images: Initial application to analysis of alzheimer disease and mild cognitive impairment. In: International Congress Series, vol. 1290. Elsevier. 2006. p. 128–34.
[26] Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hänninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala E-L, et al. Hippocampus and entorhinal cortex in mild cognitive impairment and early ad. Neurobiol Aging 2004;25(3):303–10.
[27] Ertek G, Tokdil B, Günaydın İ. Risk factors and identifiers for alzheimer's disease: A data mining analysis.In: Advances in Data Mining. Applications and Theoretical Aspects, Springer, 2014, pp. 1–11.
[28] Aditya CR, Pande MBS. An algorithmic approach for alzheimer's disease detection from non-image data.
[29] Aditya CR, Pande MBS. Multifactor affiliation analysis: a multifactor dimensionality reduction based learning model for knowledge discovery and similarity measure in 2-way data classification. Int J Comput Appl 2015;130(5):1–5.
[30] Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. J Cogn Neurosci 2010;22(12):2677–84.
[31] Ritchie Marylyn D, Motsinger Alison A. Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene - gene interactions in human genetics and pharmacogenomics studies. Hum Genom 2006;2(5):318–28.
[32] Alison A. Motsinger, Marylyn D. Ritchie, Multi factor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies, Pharmogenomics.
[33] Ritchie MD, Motsinger AA. Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies.
[34] Wikipedia, Multifactor dimensionality reduction — Wikipedia, the free encyclopedia, 2014, Online URL ⟨https://en.wikipedia.org/w/index.php?Title=Multifactor_dimensionality_reduction&oldid=594067410⟩. [accessed 15.09.15]
[35] Wikipedia, Euclidean distance — Wikipedia, the free encyclopedia, 2015. Online URL ⟨https://en.wikipedia.org/w/index.php?title=Euclidean_distance&oldid=680026113⟩. [accessed 15.0915]
[36] Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage alzheimer disease. Arch Neurol 2001;58(3):397–405.
[37] Morris, JC. The clinical dementia rating (cdr): current version and scoring rules. Neurology.
[38] Mortimer JA, Graves AB. Education and other socioeconomic determinants of dementia and alzheimer's disease. Neurology 1993;43(suppl 4):S39–S44.
[39] Russ TC, Stamatakis E, Hamer M, Starr JM, Kivimäki M, Batty GD. Socioeconomic status as a risk factor for dementia death: individual participant meta-analysis of 86 508 men and women from the UK. Br J Psychiatry 2013;203(1):10–7.
[40] Frisch RE. Population food intake and fertility. Science 1978;199(4324):22–30.
[41] Peck A, Vågerö DH. Adult body height and childhood socioeconomic group in the swedish population. J Epidemiol Community Health 1987;41(4):333–7.
[42] Katzman R, Terry R, DeTeresa R, Brown T, Davies P, Fuld P, Renbing X, Peck A. Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. Ann Neurol 1988;23(2):138–44.
[43] Mortimer JA. What are the risk factors for dementia, dementia and normal aging. Cambridge: Cambridge University Press; 1994. p. 208–29.
[44] Kahn RL, Pollack M, Goldfarb AI. Factors related to individual differences in mental status of institutionalized aged. Psychopathology of aging. New York: Grune & Stratton; 1961. p. 104–13.
[45] Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975;23(10):433–41.
[46] Gurland BJ. The borderlands of dementia: the influence of sociocultural characteristics on rates of dementia occurring in the senium. Aging 1981;15:61–84.
[47] Anthony JC, LeResche L, Niaz U, Von Korff MR, Folstein MF. Limits of the mini-mental state'as a screening test for dementia and delirium among hospital patients. Psychol Med 1982;12(02):397–408.
[48] Escobar JI, Burnam A, Karno M, Forsythe A, Landsverk J, Golding JM. Use of the mini-mental state examination (mmse) in a community population of mixed ethnicity: cultural and linguistic artifacts. J Nerv Ment Dis 1986;174(10):607–14.
[49] Kittner SJ, White LR, Farmer ME, Wolz M, Kaplan E, Moes E, Brody JA, Feinleib M. Methodological issues in screening for dementia: the problem of education adjustment. J Chronic Dis 1986;39(3):163–70.
[50] Mortimer J. Do psychosocial risk factor contribute to alzheimer's disease?. Etiology of dementia of Alzheimer's type.
[51] Perneczky R, Wagenpfeil S, Komossa K, Grimmer T, Diehl J, Kurz A. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. Am J Geriatr Psychiatry 2006;14(2):139–44.
[52] Sowell ER, Thompson PM, Toga AW. Mapping changes in the human cortex throughout the span of life. Neuroscientist 2004;10(4):372–92.