

Alzheimer Disease Prediction through Unsupervised Classification

Mr Parikshith Nayaka S K^{#1}, Dr Dayanand Lal N^{#2}, Mr. Kiran Ramaswamy^{*3}, Ms. Nida Kousar^{#4},
Dr. Nijaguna G S^{@5}, Mr. Zameer Adhoni^{#6}

^{#1,2,4}Assistant Professor, Department of Computer Science and Engineering, GITAM University,
Bengaluru

^{#6}Research Scholar, Department of Computer Science and Engineering,
GITAM University, Bengaluru

^{*3}Assistant Professor, Department of Electrical and Computer Engineering at Ambo University
ETHIOPIA

^{@5}Associate Professor, Department of Information Science and Engineering, SEACET, Bengaluru
npariksh@gitam.edu, dnarayan@gitam.edu, kiran.ramaswamy@ambou.edu.et, knida@gitam.edu
gsniju@gmail.com, zameeradhoni@yahoo.co.in

Abstract— Alzheimer is considered to be more challenging to the researchers; it is complicated to predict the disease at the beginning stage and analyse it. But due to the availability of colossal brain image samples and neural dimensions, making a diagnosis is much more comfortable and practical. Researches show that machine learning technologies are more beneficial and useful in analysing this disease. The unsupervised classification, which is motivated by unsupervised learning, will help to learn raw data intelligently. This paper proposes unsupervised classification using k-means clustering, which gives a good result. The proposed work is analysed based on performance measures like Accuracy (ACC), Sensitivity (SCN), Specificity (SPE).

Keywords— *Alzheimer disease, R Programming Language, Unsupervised learning, Sparse filtering, Machine Learning, Accuracy (ACC), Whitening, Mild Cognitive Impairment (MCI)*

I. INTRODUCTION

The whole of mankind is living in the 21st century, which consists of higher and advanced technologies exclusively in the analysing and scanning of the affected region using MRI (magnetic resonance imaging) [2]. By classification of the pertinent data, we can prevent the future impact on an individual (affected one) [3]. Current methods are not so useful in predicting disease. But the disease can be treated positively through undergoing many types of research and development of diagnosis technics.

Some prediction can be performed by considering the [4] dimension areas of given samples. Machine learning [5] can be implemented for classification by comparing the sample images of the human brain along with primary instruction of expertise. Techniques like CNN (convolution neural networks) [6-8] are playing a satisfactory role for researchers in predicting the biological condition of an individual and diagnose according to the symptoms. Image dimensions [9], grey image [10-11] are more beneficial compared to the other methods. Focusing on sparse filtering can lead to easy classification sample.

II. RELATED WORKS

Saraf et al. proposed CNN (convolution neural network) [12] for predicting Alzheimer disease. Data is gathering from the biological signs and aspects 3-dimensional convolutional neural network is developed convolution from 3-dimension autoencoder, the model will be trained to be determined. The changes were occurring in the brain.

Hosseini et al. proposed a model which is trained in such a way that, it scans the difference by comparing with typical images[13] of the brain with the currently affected region of the brain. The outcome result can be analysed based on the colouring of the affected regions. The dark shaded region represents the affected regions.

Brosch et al. proposed a manifold based learning model[14]. Here the model is trained in such a way that it has the ability to predict the outcome result by comparing the similarity present from stored data. If a number of similarities are less, it indicates that the region of the brain, if the number of similarities is more, it shows the percent of chance that the region gets affected and finally the result is represented in the form of graphical chart representation.

Jing et al. proposed longitudinal studies[15] of social network and dementia are considered to be a supportive cause

in approaching protective and effective in social networks on dementia, but these are not directly/indirectly controlled traits that have looked at the health impacts of social network on dementia because if we are considering the social network as a key, taking the behaviour of an individual as a sample, which is considered to be more time consuming (later disease will progress rapidly, leads to uncontrolled condition), but finally with these experimental observations they were able to observe many symptoms in affected individuals.

Andrew proposed social contact as a mechanism [16] through which the chances of affecting AD would be reduced. More social contact leads to the organisation and re-building of memory which leads to the reactivation of neuron cells.

III. PROPOSED METHOD

In this work, K means clustering is used for predicting the Alzheimer disease. Some of the parameters like the number of features, digression of sample weight, dispersion compensation, rate of learning through the experiencing the same task many times, rate of change in sample stability are used. If an inappropriate adjustment in these parameters will undoubtedly affect the accuracy of the model. Nigam describes the unsupervised training [17] in sparse filtering, here the number of features plays an essential role in deciding the filtering following derivatives about filtering the data.

Fig.8 represents that the particular samples area of the brain is taken and implemented into the training model as input data. The model tries to classify the different patterns based on its experience. Classification is based on the unsupervised learning model. Trained images are filtered using the concept of sparse filtering [18]. In this work, K means clustering is used for predicting the Alzheimer disease.

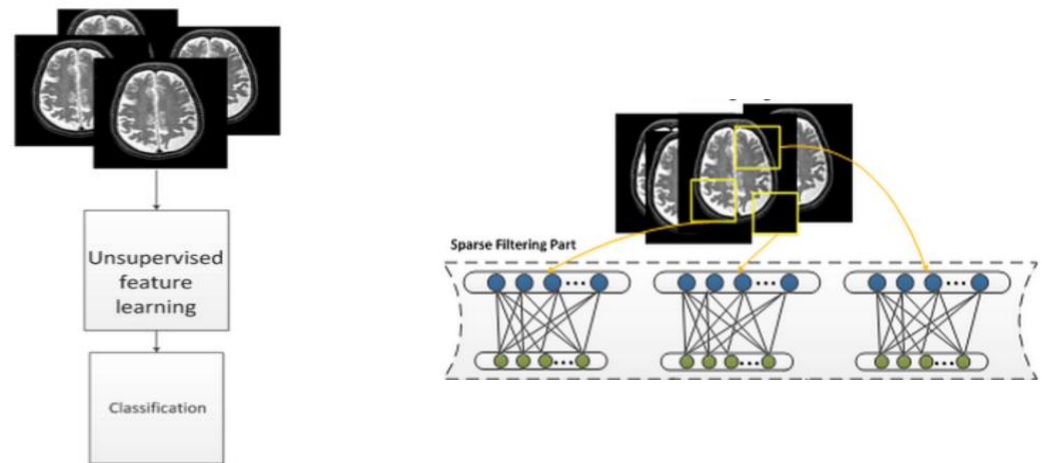


Figure1: Unsupervised concept plays an essential role in the classification of data; it takes the raw image and classifies them into clustered data format. Sparse filtrations part is the vital method in sorting the sample data, here the classification occurred based on the values of an individual pixel.

A. Mathematical Representation of Sparse Filtration of Data

Let X be the sample data,

$$\{X^i\}^{M_i=1}$$

Light filtering calculates the linear features of each sample:

$$F_i = W_1 * X^i$$

L_p norm of t is formulated as $\|t\|_p$ (therefore)

$$\|t\|_p = (|t_1|^p + \dots + |t_n|^p)^{1/p} \quad t = t_1 + t_2 + \dots + t_n$$

F_L matrix's feature matrix:

$$F_L = F / \|F\|_2$$

$$F^{-i} = F^{-i} / \|F^{-i}\|_2$$

The weight in the (equation 1) is solved with optimization of L_p norm of cost function restriction for the sample data, which can be expressed as:

$$\text{Min} \sum_{i=1}^M \|F^i\|$$

Concluding the (equation 1), which can be expanded as:

$$F_L = g(W_1 * X^i)$$

$$\left[\begin{array}{l} X^i \in R^{N \times 1}; \text{ is the sample data.} \\ M \text{ is the number of samples.} \end{array} \right]$$

$$\left[\begin{array}{l} F_i \text{ corresponds to } L^{\text{th}} \text{ feature } i^{\text{th}} \text{ sample.} \\ \text{--(equation 1)} \end{array} \right]$$

$$[\quad]$$

$$\text{--(equation 2)}$$

$$\text{--(equation 3)}$$

$$\text{--(equation 4)}$$

$$\text{--(equation 5)}$$

B. What is the K-Mean Clustering Algorithm

K -mean clustering algorithm was developed by MacQueen [1], clustering is done with the help of Euclidean distance, k-mean algorithm and unsupervised learning. Using Euclidean data, all sample are re-clustered according to the centre point until iterative operation terminates.

It is a simple and effective unsupervised algorithm used to solve the problems related to grouping the data samples. The given data can be calculated through the number of centrioles. C (centre point) should be constated because alter in centre point can give a different result, which affects the algorithm. Based on the distance between the data points (samples) and C, the clustering will be conducted, and the process is continued(looped) until the number of data point becomes zero. This algorithm is more effective in reducing factual errors (square errors), which can be implemented by:

$$J(V)=\sum_i \sum_j (|| x_i-x_j ||)^2 \quad (|| x_i-x_j ||)=\text{Euclid distance}$$

1) K-Mean Clustering Algorithm in Alzheimer Disease:

Alzheimer disease, one of the most common neurogenetic disease seen in old age people. There is a need to detect the disease before it starts affecting an individual. The detection is possible by collecting the data of previous samples in document form, which consist of a critical pattern of disease in subjected biodata. We apply k -mean clustering algorithm to data sample, collected from ADNI dataset to classify the sample into, effected and unaffected patter(cells). The pattern can be clustered(grouped) Into AD (effected) and MRI (normal) patterns. This key pattern plays a vital role in early detection of neurogenetic disease. So, it can be diagnosed and cured in an early stage.

2) A parameter used While clustering:

- i. Clustering of k sample with data key (x) can be shown by lowering the (J) parameters. J value can be mentioned minimum if the clustering data lay near to the centre later data key x is classified into clusters(groups).

$$[J=\min (\sum_k \sum x \in C_k W_x \text{dist} (x, O_k))] \quad \text{-----equation 1}$$

Wx: Obtain minimum J value.

dist(x,ok): Distance function.

X: Pixel data.

O_k:Centre.

C_k : (C_{k,1},C_{k,2},...,C_{k,n})

- ii. Distance function can be calculated with the help of Euclidean distance for K-mean clustering.

$$[\text{dist} (x, ok)=(x_{ij,1}-O_{k,1})^2+(x_{ij,2}-O_{k,2})^2+\dots+(x_{ij,n}-O_{k,n})^n] \quad \text{-----equation 2}$$

- iii. Loop operation is repeated, as shown in the flow diagram. The flowchart represents the iterative distance control operation in order to lower the J parameter.

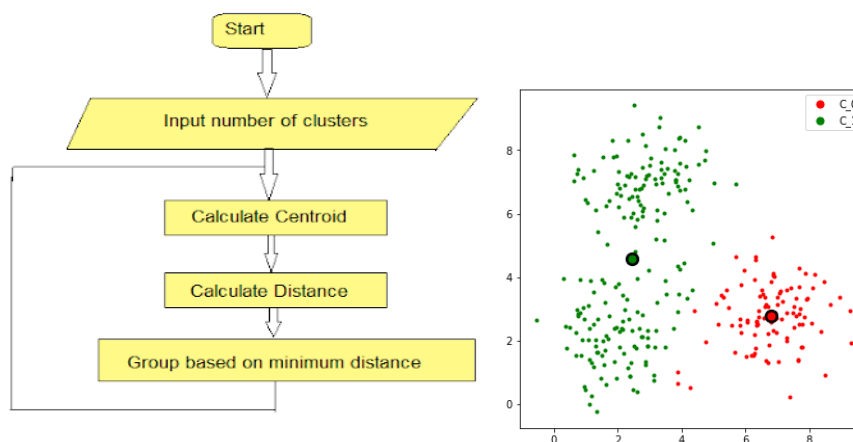


Figure 2: Representation of k-mean clustering flowchart, centrioles and clusters.

Above flowchart describe the clustering of data using the k-mean clustering method. A higher number of data forms a higher number of centrioles, based on the distance between the data keys and centrioles, clusters(group) are formed. The process consists of loop operation, which runs until data keys get empty later as a default operator get terminated.

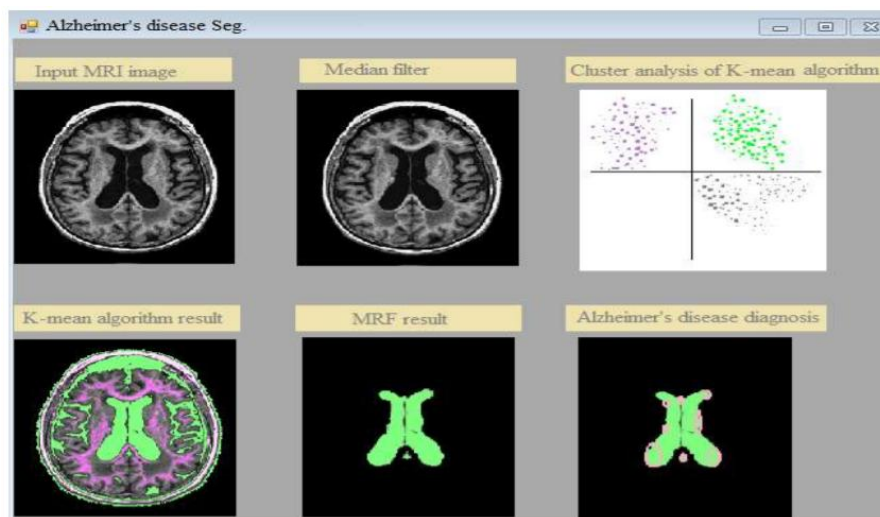


Figure 3: Representation of grey image comparison using k-mean classification.

In the process of clustering, collection of data sample plays a key role. Following diagram represents the data classification based on the comparison. MRI image is given as an input, which is in grey image form (black-white image), the k-man algorithm is applied to the image and results are compared based on cluster analysis prediction.

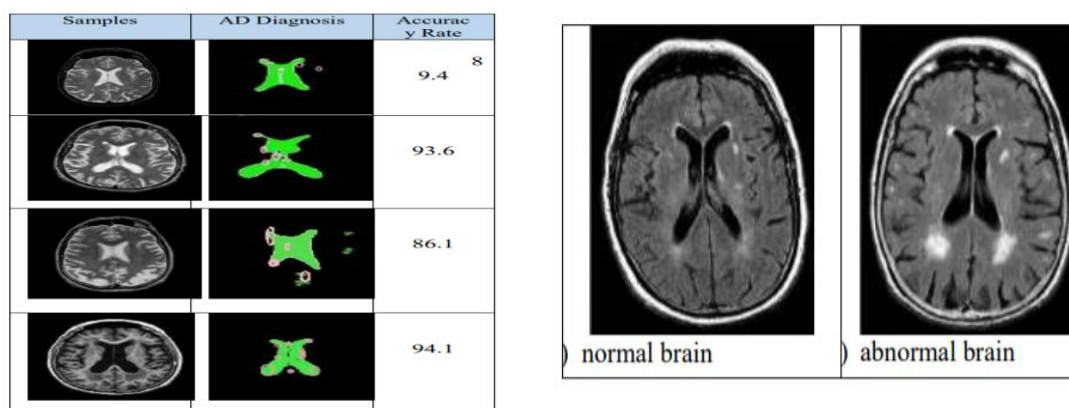


Figure 4: Accuracy rate obtained as a result of cluster analysis prediction and image analysis. Following prediction represents the difference between the typical sample and algorithm applied sample. As a result of the cluster analysis, the accuracy rate is calculated, and the classification of the sample is performed.

C. Data Sets

In real life, the number of AD patient [19] and [20] is about 50%, and a number of MCI patient was more than 50% and, it was recorded to 40% of the patient has a possibility of suffering from AD, 50% not exposed to the disease their change was low. At the first stage of learning light filtering [21] was so effective in extracting the different features of name and crud brain images and the image was trained to a model-based pixel through local image and feature.

Name	AD N= 51	MCI Converter (N= 43)	MCI Non-converter (N= 56)	HC (N= 52)
Female/male	18/33	15/39	17/39	18/34
Age (mean $\pm$ SD)	75.2 $\pm$ 7.4 [59–88]	75.7 $\pm$ 6.9 [58–88]	75.0 $\pm$ 7.1 [55–89]	75.3 $\pm$ 5.2 [62–85]
Education (mean $\pm$ SD)	14.7 $\pm$ 3.6 [4–20]	15.4 $\pm$ 2.7 [10–20]	14.9 $\pm$ 3.3 [8–20]	15.8 $\pm$ 3.2 [8–20]
MMSE (mean $\pm$ SD)	23.8 $\pm$ 2.0 [20–26]	26.9 $\pm$ 2.7 [20–30]	27.0 $\pm$ 3.2 [18–30]	29 $\pm$ 1.2 [25–30]
CDR (mean $\pm$ SD)	0.7 $\pm$ 0.3 [0.5–1]	0.5 $\pm$ 0.0 [0.5–0.5]	0.5 $\pm$ 0.0 [0.5–0.5]	0.0 $\pm$ 0.0 [0–0]

Table1: clinical information of the image

Criteria	Normal-person	MCI-person	AD-person
MMSE score	24–30	24–30	20–26
CDR score	0	0	0.5–1.0
Depressed	No	–	–
MCI	No	Yes	Yes
Dementia	No	Yes	Yes
Memory complaint	–	Yes	–

Table2: clinical criteria for patient

According to the prediction of clinical information of the image between AD in male and female, Table 1 represents the resultant outcomes. The result is based on the images. In Table 2, The following prediction table is the resultant output is based on clinical criteria for the patient between an average person and AD affected person which are classified based on the criteria like MMC scores, depression level, MCI, Dementia and medical complaint

IV. RESULTS AND DISCUSSION

Scatter filtering plays a vital role in these sections, and using the filtering technique, the effect of the proposal method was investigated for extracting the features. Four categories like (AD V/S HC), (MCI V/S HS), (A D V/A MCI), (MCI-C V/S MCI-NC). The obtained data sample set is randomly classified into 9-10 subdivisions, later 9/10 is used for training the model, and the remaining one is used for testing. In order to get an average ratio, the process is repeated many times, and accuracy was seen between 85-92%, it will be apparent that the data set can classify the set of AD. But even higher accuracy can be obtained by giving more detailed image dimensions.

Models	Hidden units	Accuracy (ACC)
AD vs. HC		
MRI	500-50-10	0.876 ± 0.017
PET	1000-50-30	0.879 ± 0.023
CSF	50-3	0.845 ± 0.014
CONTACT	500-100-20	0.902 ± 0.019
MCI vs. HC		
MRI	100-100-20	4 ± 0.018
PET	300-50-10	0.694 ± 0.014
CSF	10-3	0.704 ± 0.019
CONTACT	100-50-20	0.752 ± 0.017
AD vs. MCI		
MRI	1000-50-30	0.662 ± 0.021
PET	100-50-10	0.678 ± 0.020
CSF	10-1	0.687 ± 0.011
CONTACT	100-100-20	0.712 ± 0.019
MCI-C vs. MCI-NC		
MRI	100-100-10	0.561 ± 0.030
PET	100-100-10	0.614 ± 0.028
CSF	30-2	0.601 ± 0.021
CONTACT	500-50-20	0.625 ± 0.025

Table 3: classification accuracy derived from SP classification and the relevant structure as the number of hidden units

A single stage of learning method, whitening was used virtually in scatter filtering training process. To compare the performance F-score representation is considered to be more effective in measuring the classification, here it declares numeric 1 for the best result and numeric 0 for the worst result outcomes. The outcomes are depending on the accuracy (ACC), sensitivity (SN) and specificity (SPE). Considering the health conditions proposed concept derived from F-score [22] is applied, and it was successful, both average character and whitening are necessary for proposed methods.

ACCURACY: $(TP + TN) / (TP + TN + FP + FN)$.
 SENSITIVITY: $(TP) / (TP + FN)$.
 F-SCORE : $(2TP) / (2TP + FP + FN)$.

TP = TRUE-POSITIVE
 TN=TRUE-NEGATIVE
 FP=FLASE-POSITIVE
 FN=FALSE-NEGATIVE

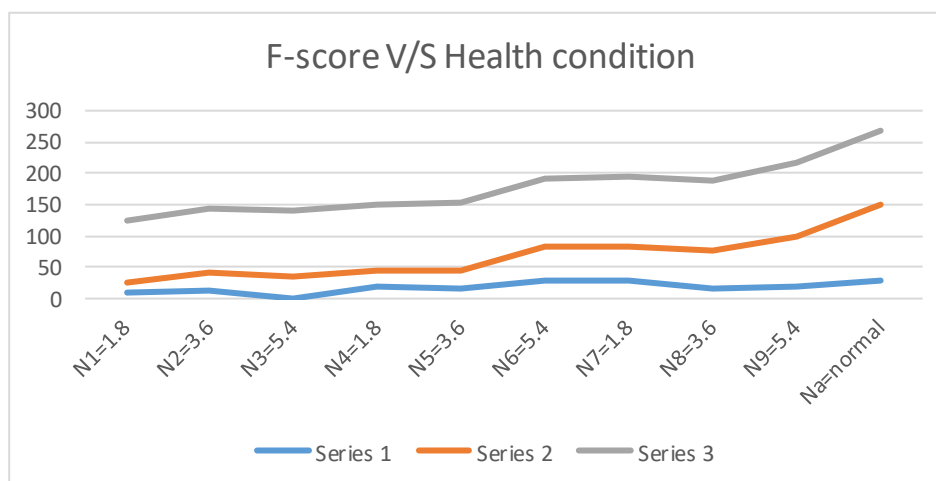


Figure 5: F-score of the Alzheimer diseases data set using the non-whitening method, a method with aggregate features and the proposed method

Graphical representation of the classification of data, which is displayed in F-score format. Y-axis consists of F-score, and X-axis consists of health condition. Y-axis values are selective; they are less than 1 number, the x-axis represents the pixel values of individual neurons networks, based on that the classification of data takes place. Higher the value, higher will be the selective changes and vice versa. Blue plotted line is not whitening, red plotted line is proposed and green plotted line are aggregate features. Based on the neural network value, the health conditions are approached.

VI. CONCLUSION AND FUTURE WORK

Alzheimer disease prediction is inevitable for starting the treatment earlier for such disease-prone patients. The k-mean clustering unsupervised algorithm helps in improving the accuracy rate of the Alzheimer disease prediction. The accuracy can be further improved by incorporating deep learning techniques and by altering the structure of the network.

REFERENCES

1. Alyssa A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities.
2. Chen M, Mao S, Liu Y. Big data: a survey. Mob Netw Appl.
3. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care.
4. Silly S, Zhang Y. Medical big data: neurological diseases diagnosis through medical data analysis.
5. Poldrack RA, Gorgolewski KJ. Making big data open: data sharing in neuroimaging. Nat Neurosci.
6. Glenner GG. Alzheimer's disease. In: Biomedical advances in ageing. Springer.
7. Baum LW, Chow HLA, Cheng KK. Nanoparticle contrast agent for early diagnosis of Alzheimer's disease by magnetic resonance imaging (MRI).
8. Sabri O, et al. Florbetaben PET imaging to detect amyloid-beta plaques in Alzheimer's disease: phase 3 study. Alzheimer's Dement.
9. Li R et al. Deep learning-based imaging data completion for improved brain disease diagnosis. In: International conference on medical image computing and computer-assisted intervention
10. Socher R. Recursive deep learning for natural language processing and computer vision.
11. Yu D, Deng L. Automatic speech recognition: an in-depth learning approach. Berlin: Springer.
12. Sarraf S, Tofighi G. Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks. arXiv preprint arXiv:1607.06583. 2016.
13. Hosseini-Asl E, Gimel' farb G, El-Baz A. Alzheimer's disease diagnostics by an intensely supervised adaptable 3D convolutional network. arXiv preprint arXiv:1607.00556. 2016.
14. Brosch T, Tam R, A. s. D. N. Initiative. Manifold learning of brain MRIs by deep learning. In: International conference on medical image computing and computer-assisted intervention
15. Jing Wu ORCID Icon, Caroline Hasselgren, Anna Zettergren, Henrik Zetterberg, Kaj Blennow, Ingmar Skoog & Björn Halleröd, Received 15 Feb 2018, Accepted 30 Sep 2018, Published online: 27 Dec 2018
16. Andrew Sommerlad, Séverine Sabia, Archana Singh-Manoux, Glyn Lewis, Gill. Published: August 2, 2019
17. Suk H-I, Lee S-W, Shen D, A. S. D. N. Initiative. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal. 2017;37:101–13
18. Ngiam J, Chen Z, Bhaskar SA, Koh PW, Ng AY. Sparse filtering. In: Advances in neural information processing systems. 2011. pp. 1125–33

19. Ngiam J, Chen Z, Bhaskar SA, Koh PW, Ng AY. Sparse filtering. In: Advances in neural information processing systems. 2011. pp. 1125–33
20. Held E, Cape J, Tintle N. Comparing machine learning and logistic regression methods for predicting hypertension using a combination of gene expression and next-generation sequencing data. In: BMC proceedings. BioMed Central, vol. 10, no. 7. 2016. p. 34.
21. Parikshith2020, A modern themed system for patients security of data exposure in semi-convinced servers in the cloud. International Journal of Emerging Trends in Engineering Research, 2020, 8, 4123-4127. <https://doi.org/10.30534/ijeter/2020/15882020>
22. Mr. Manoj K, Ms. Akhila R, Mr. Ganesh M, Mr. Parikshith Nayaka S K. (2020). Social Media Sentimental Analysis using Machine Learning. International Journal of Advanced Science and Technology, 29(05), 2654 - 2662. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/11364>
23. Risacher S, et al. Alzheimer's disease neuroimaging initiative (ADNI). Neurobiol Aging. 2010;31:1401–18.
24. Hu C, Ju R, Shen Y, Zhou P, Li Q. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In: Communications (ICC), 2016 IEEE international conference on, IEEE. 2016. pp. 1–6.
25. https://en.wikipedia.org/wiki/K-means_clustering