

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

# Regression for Identifying Genetic Biomarkers Using Brain Image Segmentation

Sujatha E<sup>1</sup>, Gayathri M<sup>2</sup>, Hansika V<sup>3</sup>, Renuka R<sup>4</sup>,

Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,  
Tamil Nadu, India<sup>1</sup>

U.G Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,  
Tamil Nadu, India<sup>2,3,4</sup>

**ABSTRACT:** Alzheimer's disease (AD) is the most common form of dementia among older people. Dementia is a brain disorder that seriously affects a person's ability to carry out daily activities. A novel sparse regression method for identifying genetic biomarkers using Brain-Wide and Genome-Wide association is used. Specifically, relationships in genetic features and in brain imaging features, are identified by Support Vector Machine classification technique. Regression applies to account the correlations among the genetic variables, by which it can be possible to identify the representative Single Nucleotide Polymorphisms that are highly associated with the brain imaging features. It optimizes the objective function by jointly tackling low-rank regression and variable selection in a framework. In order to resolve homogeneity in MR images N3 algorithm is adapted. To segment the structural MR images into three different tissues, grey, white matter, and cerebrospinal fluid using features from Accelerated Segment Test Algorithm. Preprocessing approaches such as Gaussian filter is used to enhance the quality of the image. The experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset showed that the proposed method could select the important SNPs to more accurately estimate the brain imaging features than the state-of-the-art methods. The advantage of the proposed method is that it will remove the noise and to find out the specific region which is affected by the Alzheimer's disease.

**KEYWORDS:** Regression, Brain Image Segmentation, Alzheimer's Disease Neuroimaging Initiative (ADNI)

### I. INTRODUCTION

Alzheimer's disease grows slowly. It first involves the parts of the brain that control thought, memory and language. People with this disease may have trouble remembering things that happened recently or names of people they know. Over 5.5 million Americans of all ages have Alzheimer's disease in 2016. This includes an estimated 5.3 million people age 65 and older and approximately 200,000 individuals under age 65 who have younger-onset Alzheimer's. One in ten people age 65 and older (10 percent) has Alzheimer's disease. In the proposed System low-rank constraint and sparse graph representation regularization in SNPs along with a structured sparsity constraint in a linear regression framework helped to effectively utilize the inherent information in genetic data and brain imaging data, and thus finding informative associations. In the Existing method, pairwise univariate analysis measures the correlation between individual phenotype and an isolated genotype without considering the potential correlations on the phenotypes and the genotypes. Regularized ridge regression conducts the imaging genetic association study via ordinary least square estimation, which considers the correlations among the variables and but ignore the correlations among the corresponding responses. In Proposed method considers the interlinked structures among Single Nucleotide Polymorphisms (SNPs) to output interpretable results. It is limited in the sense that they didn't jointly consider the relational information in a unified framework. The main challenge in current imaging-genetic association study comes

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

from the large number of variables from both brain imaging data and genetic data, thus requiring appropriate statistical techniques such as regression, variable selection and sparsity constraint.

## II. RELATED WORKS

In 2017 X Zhu proposed in a novel relational regularization feature selection method for joint regression and classification in AD diagnosis. It focuses on joint regression and classification for Alzheimer's disease diagnosis and propose a new feature selection method by embedding the relational information inherent in the observations into a sparse multi-task learning framework. It confirmed the key role of the APOE and TOMM40 genes but also highlighted some novel potential associations with AD.

In 2016 Y Zhu proposed in early diagnosis of Alzheimer's disease by joint feature selection and classification. The diagnosis of Alzheimer's disease (AD) from neuro imaging data at the pre-clinical stage has been intensively investigated because of the immense social and economic cost. In the past decade, computational approaches on longitudinal image sequences have been actively investigated with special attention.

In 2014 K Zia proposed common genetic variants and it is associated with 1p 13.2 with risk of autism. Autism is a highly heritable neurodevelopmental disorder, and known genetic variants, mostly rare, account only for a small proportion of cases. Here we report a genome-wide association study on autism using two Chinese cohorts as gene discovery pairment (MCI), which is an intermediate stage between normal control (NC) and AD..

In 2010 DH Ballard proposed in comparisons of multi-marker association methods to detect association between a candidate region and disease. The joint use of information from multiple markers may be more effective to reveal the region identified by disease.

In 2011 J Bratlin proposed in association of the Alzheimer's gene SORL1 with hippocampal volume in young, healthy adults. Alzheimer's disease is among the most common neurodegenerative disorders. The *SORL1* (sortilin receptor 1) gene is associated with the disease, but different variants seem to contribute used a gene-wide approach to test whether *SORL1* is associated with volume of the hippocampus, one of the first structures to be affected by Alzheimer's disease in young, healthy individuals, in an attempt to map potential pathways from gene to disease.

IN 2009 N Filippini proposed in anatomically-distinct genetic associations of APOE  $\epsilon$ 4 allele load. APOE  $\epsilon$ 4 is the best-established genetic risk factor for sporadic Alzheimer's disease (AD). Temporal regions appeared to show a dominant effect of APOE  $\epsilon$ 4 allele load instead of the additive effect previously strongly associated with age of onset. Regional variations with allele load may be related to different mechanisms for effects of APOE  $\epsilon$ 4 load on susceptibility and disease progression.

## III. PROPOSED SYSTEM

The main challenge in current imaging-genetic association study comes from the large number of variables from both brain imaging data and genetic data, thus requiring appropriate statistical techniques such as regression, variable selection and sparsity constraint. In this proposed a novel low-rank variable selection method in a regularization-based linear regression framework by taking correlations inherent in phenotypes and genotypes into account and also avoiding the adverse effect of noise and redundancy. Genotype data is used to regress the phenotype data with a least square regression to consider the correlations between the variables and the responses. Homogeneity using the N3 algorithm to select the region growing for affected area. A robust skull-stripping method was applied to extract only a brain on all structural MR images. The manual edition and intensity in homogeneity correction were followed for better quality. After repeating N3 algorithm three times to remove the cerebellum based on registration and intensity in homogeneity correction it is used that FAST algorithm to segment the structural MR images into three different tissues, gray matter, white matter, and cerebrospinal fluid. The proposed method validating the imaging-genetic analysis between the SNPs and the neuro imaging phenotypes. Finally the images are classified into normal or abnormal condition for the input image. The proposed method enhances the accuracy in detecting extract region of tumor present in brain and removes the skull region for brain image with high PSNR and MSE.

# International Journal of Innovative Research in Science, Engineering and Technology

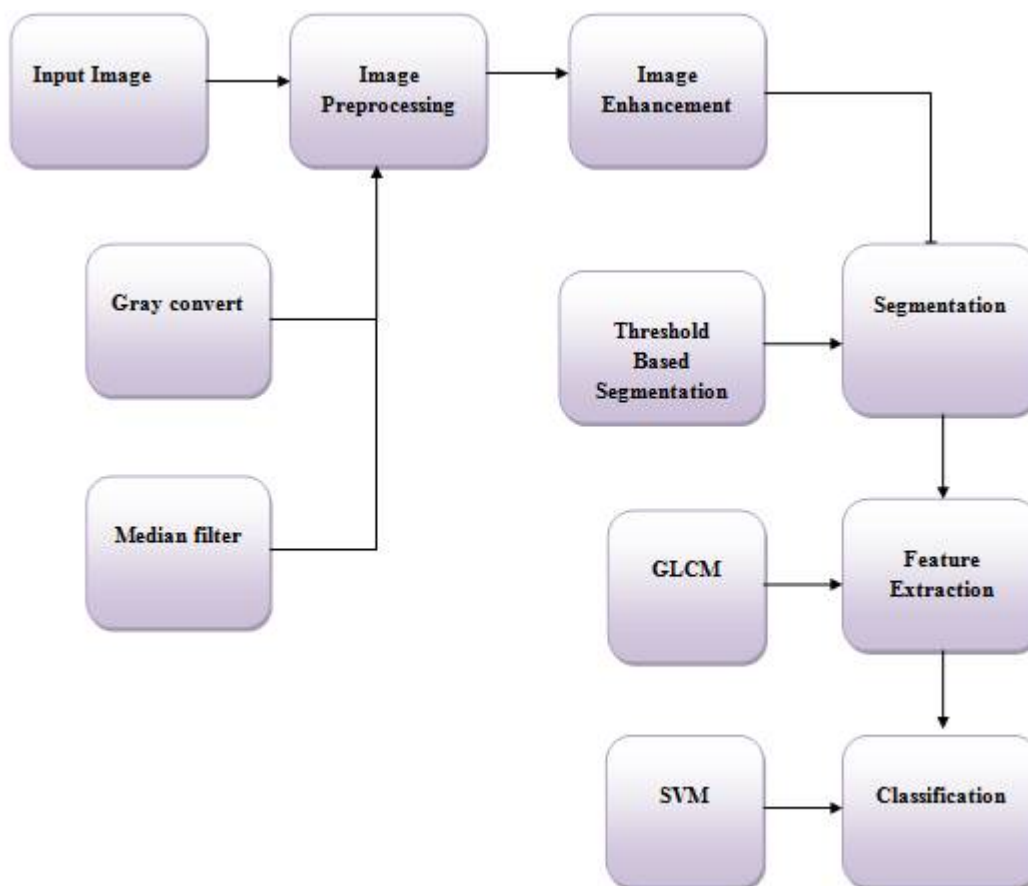
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

## IV. SYSTEM DESIGN

### BLOCK DIAGRAM:



### INPUT IMAGE:

Input Image is a one which is given as input. Here MRI image is given as input.

### IMAGE ENHANCEMENT

Image Enhancement is mainly used to process the input image in a way that is more specific than original image. It sharpens image features such as edges, boundaries to make a graphic display more suitable for analysis.

### MEDIAN FILTER

It is used to remove noise from image. It provides a better way to improve the results of preprocessing.

### SUPPORT VECTOR MACHINE

It is used to analyze data used for classification & regression analysis

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018

## SUBSYSTEMS

### IMAGE PREPROCESSING

Digital image processing is always an interesting field as it gives improved pictorial information for human interpretation and processing of image data for storage, transmission, and representation for machine perception. Image Processing is a technique to enhance raw images received from cameras/sensors placed on satellites, space probes and aircrafts or pictures taken in normal day-to-day life for various applications. This field of image processing significantly improved in recent times and extended to various fields of science and technology.

In preprocessing the steps are as follows

#### 1. Image Resampling

In MRI image Resampling is performed. It is the process of reducing or increasing the number of pixels in an image. It can change the image file size as well as the image resolution.

#### 2. Filtering

Filtering is a technique for modifying or enhancing an image. Here we can filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement.

### IMAGE SEGMENTATION:

There is a linear relationship between SNPs and ROIs, but the data are often found to have complex nonlinear relationship. In this case, even though a number of literature have reported that the sparsity-inducing regularization may implicitly result in nonlinear relationship, an explicit assumption of nonlinear relationship (e.g., mapping the original data into kernel space by kernel functions). Second, only selected 16 ROIs related to AD to conduct SNP selection in this work. It should have other ROIs which are also in relation to SNPs in imaging-genetic analysis. Segmentation is one of the key problems in image processing. Image segmentation is the process that subdivides an image into its constituent parts or objects. The level to which this subdivision is carried out depends on the problem being solved, i.e., the segmentation should stop when the objects of interest in an application have been isolated e.g., in autonomous air-to-ground target acquisition, suppose our interest lies in identifying vehicles on a road, the first step is to segment the road from the image and then to segment the contents of the road down to potential vehicles. Image thresholding techniques are used for image segmentation.

### FEATURE EXTRACTION

The QC criteria for the SNP data include: 1) call rate check per subject and per SNP marker; 2) gender check; 3) sibling pair identification; 4) the Hardy-Weinberg equilibrium test; 5) marker removal by the minor allele frequency; and 6) population stratification. In this paper, we used the MaCH software to impute the missing SNPs satisfied the QC step.

The feature extraction techniques are developed to extract features in synthetic aperture radar images. This technique extracts high-level features needed in order to perform classification of targets. Features are those items which uniquely describe a target, such as size, shape, composition, location etc. Segmentation techniques are used to isolate the desired object from the scene so that measurements can be made on it subsequently. Quantitative measurements of object features allow classification and description of the image.

### CLASSIFICATION

In Supervised classification, these areas are commonly referred as training sites because the spectral characteristics of these known areas are used to 'train' the classification algorithm for eventual land cover mapping of remainder of the image. Multivariate statistical parameters are calculated for each training site. Every pixel both within and outside these training sites is then evaluated and assigned to a class of which it has the highest likelihood of being a member. In this classification function finally detects Alzheimer's disease present or not. If a disease is present it will display the area in which the disease is present. If the disease is not present it will display the result as no tumor found.

# International Journal of Innovative Research in Science, Engineering and Technology

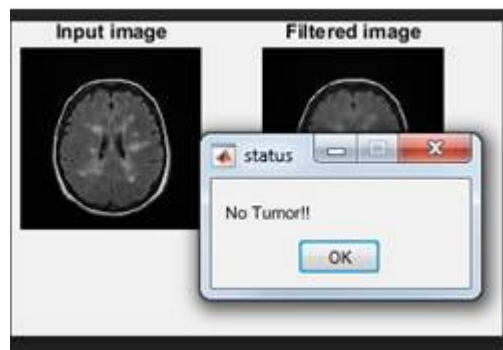
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

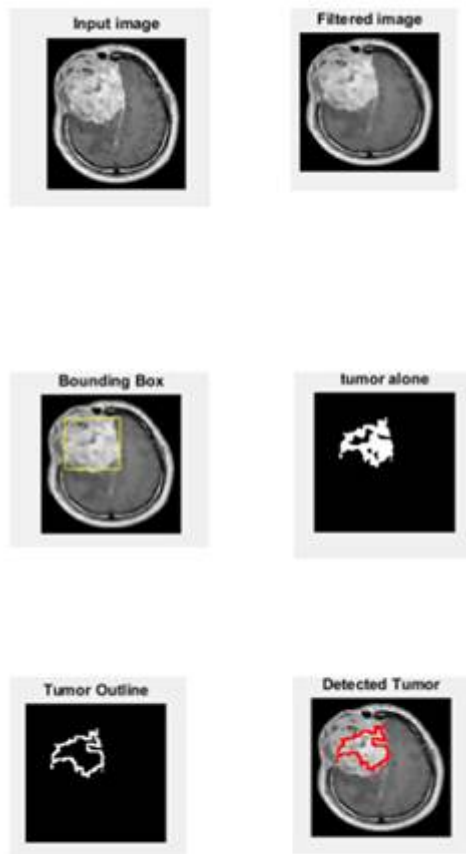
Vol. 7, Special Issue 2, March 2018

## V. RESULTS

### SCREENSHOTS NORMAL IMAGE



### ABNORMAL IMAGE



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Special Issue 2, March 2018



## VI. CONCLUSION

Although the human and societal cost of AD is staggering, there is hope that earlier and better diagnosis, increased knowledge of its natural history with support of the patient and family throughout the disease stages, effective symptomatic drugs, and potentially effective disease modification strategies will have a dramatic impact on the number of persons affected in the future, and the quality of life of persons currently affected. The fast pace of research and development in AD is unique in neurological history, and should lead to a better future for aging populations. The proposed method has been demonstrated to outperform all the competing methods in our experiments, the performance of our proposed framework can be further improved for SNP selection. We have to detect the tumor region by using morphological operation. Compare to existing method proposed method to give good performance. To detect the Alzheimer's disease this is present in the brain using the low-rank constraint. The experimental results on the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset showed that the proposed method could select the important SNPs to more accurately.

## REFERENCES

- [1] N. Filippini, et al., "Anatomically-distinct genetic associations of APOE  $\epsilon$ 4 allele load with regional cortical atrophy in Alzheimer's disease," *NeuroImage*, vol. 44, no. 3, pp. 724–728, 2009.
- [2] X. Zhu, et al., "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Med. Image Anal.*, vol. 38, pp. 205–214, 2017.
- [3] M. Vounou, et al., "Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease," *NeuroImage*, vol. 60, no. 1, pp. 700–716, 2012.
- [4] S. L. Rosenthal, et al., "Beta-amyloid toxicity modifier genes and the risk of Alzheimer's disease," *Amer. J. Neurodegenerative Disease*, vol. 1, no. 2, pp. 191–198, 2012.
- [5] Y. Zhu, X. Zhu, M. Kim, D. Shen, and G. Wu, "Early diagnosis of Alzheimer's disease by joint feature selection and classification on temporally structured support vector machine," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2016, pp. 264–272.
- [6] K. Xia, et al., "Common genetic variants on 1p13.2 associate with risk of autism," *Mol. Psychiatry*, vol. 19, no. 11, pp. 1212–1219, 2014.
- [7] D. H. Ballard, J. Cho, and H. Zhao, "Comparisons of multi-marker association methods to detect association between a candidate region and disease," *Genetic Epidemiology*, vol. 34, no. 3, pp. 201–212, 2010.
- [8] J. Bralten, et al., "Association of the Alzheimer's gene SORL1 with hippocampal volume in young, healthy adults," *Amer. J. Psychiatry*, vol. 168, no. 10, pp. 1083–1089, 2011.
- [9] X. Zhu, X. Li, S. Zhang, Z. Xu, L. Yu, and C. Wang, "Graph PCA hashing for similarity search," *IEEE Trans. Multimedia*, doi: 10.1109/TMM.2017.2703636.
- [10] D. P. Hibar, et al., "Voxelwise gene-wide association study (vGeneWAS): Multivariate gene-based association testing in 731 elderly subjects," *NeuroImage*, vol. 56, no. 4, pp. 1875–1891, 2011.