

# NMF-SVM Based CAD Tool for the Diagnosis of Alzheimer's Disease

Ms. Tejal A. Fuse<sup>1</sup>, Mr. Nikita D. Jayasignpure<sup>2</sup> and Prof. Pragati D. Pawar<sup>3</sup>

<sup>1</sup>Ms. Tejal A. Fuse Electronics & Telecomm. Department  
J.D.I.E.T. Yavataml

<sup>2</sup>Ms Nikita D. Jayasignpure Electronics & Telecomm. Department  
J.D.I.E.T. Yavataml

<sup>3</sup>Prof. Pragati D. Pawar M.E. (Electronics)  
J.D.I.E.T. Yavataml

## ABSTRACT

*This paper presents a novel computer-aided diagnosis (CAD) technique for the early diagnosis of the Alzheimer's disease (AD) based on non-negative matrix factorization (NMF) and support vector machines (SVM) with bounds of confidence. For the study and classification of functional brain images the CAD tool is designed. For this purpose, two brain image databases are selected first one : a single photon emission computed tomography (SPECT) database and second is positron emission tomography (PET) images, both of them containing data for both Alzheimer's disease (AD) patients and healthy controls. The Fisher discriminant ratio (FDR) and nonnegative matrix factorization (NMF) are used for feature selection and extraction of the most relevant features. The NMF-transformed sets of data are classified by means of a SVM-based classifier with bounds of confidence for decision. The proposed NMF-SVM method gives classification accuracy up to 91% with high sensitivity and specificity rates.*

## 1.INTRODUCTION

ALZHEIMER's disease (AD) is the most common cause of dementia in aged people and affects more than 30 million individuals worldwide. Dementia refers to a progressive deterioration of thinking abilities severe enough to interfere with social, occupational and intellectual functions[2]. The particular evolution of AD patients and their increasing dependence on the close affective environment provokes an important social repercussion. In the next 50 years the prevalence of AD is expected to triple. Functional imaging modalities are often used with the aim of achieving early diagnosis, although this early diagnosis remains as a demanding task. Emission computed tomography images have been widely employed in biomedical research and clinical medicine during the last decade. These emission-based functional images reproduce a map of physiological functions providing information about physiological phenomena and their location in the body. In this work, two different modalities are used for brain image acquisition: Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT). Both techniques are non-invasive, which produce a three-dimensional image of functional processes in the body, such as blood perfusion or glucose metabolism, by means of emitting radio nuclides (tracers)[3]. In both techniques, PET and SPECT, all these detected emissions are processed and a three-dimensional image of the region under study is obtained. Computer-aided diagnosis (CAD), are procedures in medicine that assist doctors in the interpretation of medical images. It helps to scan digital images and is a relatively young interdisciplinary technology combining elements of artificial intelligence and digital image processing with radiological image processing. CAD systems help physicians by either identifying patterns that might have been overlooked or by providing a road map of suspicious areas. There are two different approaches for designing CAD systems for the Alzheimer's disease diagnosis- The first approach is the statistical parametric mapping (SPM) tool, developed specifically to study by comparing groups of images. The second approach is based on the analysis of the images, the analysis of regions of interest (ROI), feature extraction and posterior classification in different classes by means of some discriminative functions. To overcome the problems of above two approaches a supervised learning-based CAD system applied to functional imaging consists of several important stages:

- 1) Functional image acquisition and normalization
- 2) Feature selection and reduction
- 3) Classification, with a train and test strategy.

Along with the adequate description of its forming techniques for feature selection, extraction and for posterior classification, design and proper validation of a CAD tool for AD diagnosis, is the main issue in this work.

## 2.FUNCTIONAL IMAGE ACQUISITION AND NORMALIZATION

Each voxel of a brain functional 3-D image contains information of the corresponding brain point. However, not all the voxels have the same level of relevance in terms of discrimination between groups of subjects. In this case, two groups of subjects are defined: Alzheimer’s disease patients, label as AD, and subjects not affected by this disease, label as NOR.

### 2.1 Intensity Normalization

Previous to any kind of feature selection, the data sets have to be normalized in intensity in order to be able to compare images according to their voxel normalized intensity levels. Regarding the intensity normalization, the normalization to the maximum intensity level may introduce problems in some images that can have peak intensity values due to noise. Thus, these images are badly normalized as the normalization is based on wrong noisy voxels. In this work, in order to avoid these possible normalization errors, the intensity normalization is based on the mean value of a group of voxels with the highest intensity values. Thus, the mean value of the 0.1% voxels with the highest intensity levels is selected for the intensity normalization.

## 3.FEATURE SELECTION AND REDUCTION

An initial feature selection based on discrimination capability is typically selected, obtaining a vector of discriminant voxels for each participant. In addition, the selected discriminant voxel vectors can be projected onto a different subspace. This subspace is chosen so that only a few variables represent the most discriminant features of each patient images in each database. These steps are described below.

### 3.1 Fisher Discriminant Ratio for Feature Selection

The Fisher discriminant ratio (FDR) criterion is characterized by its separation. For the two-class case, it may be defined as follows:

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 - \sigma_2^2} \quad (1)$$

Where  $\mu_i$  and  $\sigma_i^2$  denote the class mean value and variance for each input variable, respectively. For a given variable, the ratio value grows as the difference of the mean values of each two classes increases or the cumulative scattering in each class decreases. In the case of the functional images, the voxels that satisfy a particular FDR threshold level are selected as the most discriminative variables [4].

### 3.2 Nonnegative Matrix Factorization for Feature Reduction

Nonnegative matrix factorization (NMF) is a technique for finding linear representations of nonnegative data, being a useful decomposition tool for multivariate data. This technique is especially suitable for nonnegative data sets such as functional images and for the PET and SPECT brain images. Given a nonnegative data matrix A, NMF finds an approximate factorization  $V \approx WH$  into nonnegative matrices W and H. Nonnegative matrix factorization is a linear, nonnegative approximate data representation where the original database  $V = [V_1, V_2, \dots, V_M]$  (N by M elements) which consists of measurements (profiles) of nonnegative scalar variables, is approximated by a nonnegative matrix product, as given in

$$V \approx WH \quad (2)$$

where the matrix  $W = [W_1, W_2, \dots, W_K]$  has dimension  $N \times K$ , and the matrix  $H = [H_1, H_2, \dots, H_M]$  has dimension  $K \times M$ . Thus, each element of matrix is decomposed, as shown in,

$$V_{nm} = \sum_{k=1}^k W_{nk} H_{km} \quad (3)$$

An appropriate decision on the value of K is critical in practice, but the choice of K is very often problem dependent. In most cases, however, K is usually chosen such that  $K \ll \min(M, N)$  in which case WH can be seen as compressed form of data in [5]. This property yields a reduced-variable matrix H that represents A in terms of the NMF basis W. After NMF factorization, the data contained in (K by M elements) can be considered a transformed database with lower rank (K), than the original database V. The relative error (%) of the factorization can be computed by means of the comparison of matrix V and the approximation WH. The minimum number of vectors (K) in the NMF basis is selected so that a predefined level of relative error is not exceeded.

#### 3.2.1 Factorization Rule

Given the data matrix, the optimal choice of matrices W and H are defined to be those nonnegative matrices that minimize the reconstruction error between V and WH. A variety of error functions have been proposed some of the most useful are given below, in (4) and (5)

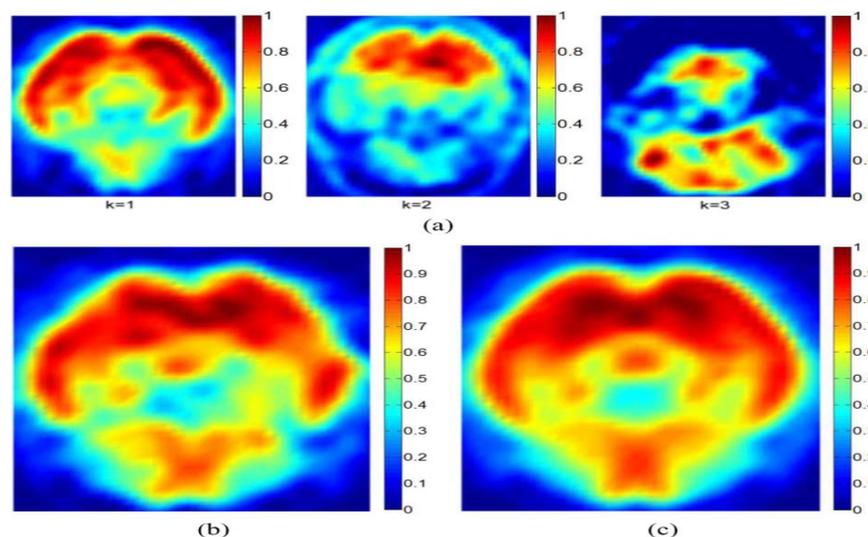
$$E_{rr1} = \frac{1}{NM} \|V - WH\|^2$$

$$E_{rr1} = \frac{1}{NM} \sum_{nm} (V_{nm} - (WH)_{nm})^2 \quad (4)$$

$$E_{rr2} = D(V \parallel WH)$$

$$E_{rr2} = \sum_{nm} (V_{nm} \log \frac{V_{nm}}{(WH)_{nm}} - V_{nm} + (WH)_{nm}) \quad (5)$$

where (4) is known as the Frobenius norm (reduction of the Euclidean distance), and (5) as the Kullback–Leibler divergence among others. The NMF process is subject to minimize the Err. The alternating least squares algorithms (ALS) is choose due to their fast convergence and lower iteration requirements for NMF in this work. In this work only the linear case is considered and NMF is selected, due to the simplicity of the proposed factorization and the preservation of a linear relation between the original space of features and the new one. Fig.1 provides the first three vectors of the NMF basis  $W$ , in the form of 2-D images, derived from one of the data sets used in this work, along with one particular transaxial slice and the one obtained from the NMF projection. The transaxial slices provided in Fig.1 are oriented from posterior (top) to anterior.



**Figure.1** NMF projection for the sametrans-axial slice of all the SPECT database subjects.(a) NMF eigenvectors for  $k=1, 2, 3$ . (b) One example subject slice. (c) Its NMF reconstruction

## 4.SVM BASED CLASSIFIER WITH BOUNDS OF CONFIDENCE

### 4.1SVM Background

Support vector machine (SVM) is a widely used technique for pattern recognition and classification in a variety of applications for its ability for detecting patterns in experimental databases. SVM has become an essential machine-learning method for the detection and classification of particular patterns in medical images[6].SVM techniques consist of two separate steps: first of all a given set of binary label training data is used for training; then new unlabeled data can be classified according to the learned behaviour. SVM separates a given set of binary label training data by means of a hyperplane that is maximally distant from the two possible classes (in our particular case, NOR an AD classes). The objective is to build a function with the training data  $F$ , as expressed in (6), able to properly classify new unclassified data

$$F: \mathbb{R}^N \rightarrow \{-1, 1\} \quad (6)$$

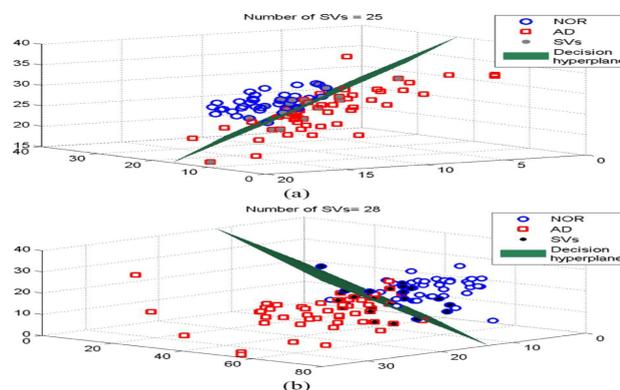
The training data are formed by  $p$  different profiles each one containing  $N$  variables, together with their proper label (NOR or AD)Thus, the training database can be expressed as

$$S_p = [(x_1, x_2, \dots, x_N), y]_p = x \cdot y ]_p \quad (7)$$

where  $x_N$  are the variables of the profile  $p$  and  $y$  the corresponding label.Linear discriminant functions define decision hyperplanes in the  $N$ -dimensional feature space

$$g(x) = w^T x + w_0 \quad (8)$$

where  $w$  is the weight vector that is orthogonal to the decision hyperplane and  $w_0$  is the threshold. The hyperplane is not unique and the selection process focuses on maximizing the generalization performance of the classifier. The vectors that define the separation hyperplane are called support vectors (SVs)[3].



**Figure.2.** Example of SVP\_N and SVN\_P distribution in terms of distance to hyperplane

**4.2 Bounds of Confidence**

SVM is based in the definition of a decision hyperplane in the N-dimensional space. The SVs are a subset of the training dataset and are chosen so that the decision hyperplane that is defined by means of them is the best one in terms of separability between two classes. The particular distance-like decisions in SVM make this machine learning approach especially suitable for the selection of a security region around the decision hyperplane in which decisions may not be adopted. This security region is defined by means of the selection of bounds of confidence for the decision. When the conventional SVM learning approach is performed and the resulting SVM classifier is trained with a set of training observations, the separation hyperplane of these data is defined. Ideally all the training data should be properly classified when they are re-entered in the SVM for test and classification. The hyperplane defined by all the SVs may not be able to properly define the class of one particular support vector. That means although the SVs help to define the best hyperplane in terms of separability some of them may not be in the right subspace defined by the hyperplane. In most of the cases, the two classes are not completely separable and there is some kind of overlapping between them in the space of interest. It is useful to compute the classification of all the SVs and to derive the error probability for the proper classification of the SVs of class +1 (SVPs) and SVs of class -1 (SVNs). If we denote SVP\_N as a support vector of class +1 wrongly labelled as -1 and SVN\_P on the contrary, the error probabilities for SVPs (PerrPos) and SVNs (PerrNeg) can be defined as follows:

$$P_{errPos} = \frac{N_{SVP\_N}}{N_{SVPs}} , \quad P_{errNeg} = \frac{N_{SVN\_P}}{N_{SVNs}}$$

where  $N_{SVN\_P}$ ,  $N_{SVP\_N}$ ,  $N_{SVPs}$ , and  $N_{SVNs}$  denote the number of vectors of each group. This error probability can be reduced if SVs which are nearer to the hyperplane are not considered. The consequence of not considering these SVs in the SVM with bounds of confidence is that all the observations located between the discarded SV which is furthest to the hyperplane and the hyperplane itself are not classified because they are in the security zone. Fig.2 reveals clearly this fact.

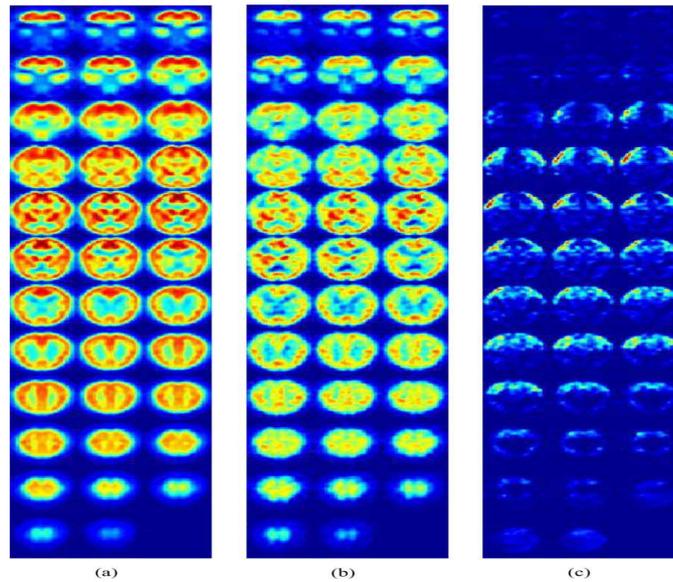
**5. IMPLEMENTATION**

**5.1 Functional Brain Image Data Sets**

In order to validate the performance and outcomes of the designed NMF-SVM based CAD tool for Alzheimer’s disease detection, two different databases are used. The first one involves SPECT brain images, whereas the second one consists of PET brain images. These two databases contain spatially normalized functional brain images of different subjects. This normalization step ensures that a given voxel in one patient refers to the same brain position than the same voxel in another patient. Then, the intensities of the functional images are normalized to the maximum intensity[7].

**5.1.1 SPECT Database**

SPECT is a nuclear medicine tomographic imaging technique using gamma rays. However, it is able to provide true 3-D information. This information is typically presented as cross-sectional slices through the patient, but can be freely reformatted or manipulated as required. Brain perfusion images were reconstructed from projection data by filtered backprojection (FBP) in combination with a Butterworth noise filter[4]. The SPECT images were labelled using two different labels: NOR for subjects without any symptom, and AD for Alzheimer’s patients. Fig.3 shows a set of trans-axial brain slices of the SPECT database, for one AD patient and one NOR subject.



**Figure.3** SPECT transaxial brain slices, oriented from posterior (top) to anterior (bottom) in each slice, and from ventral (top left) to dorsal (bottom right) in the complete set. (a) Man mask for NOR subjects. (b) Example AD patient. (c) Between-subjects variability between AD and NOR groups.

**5.2PET Database**

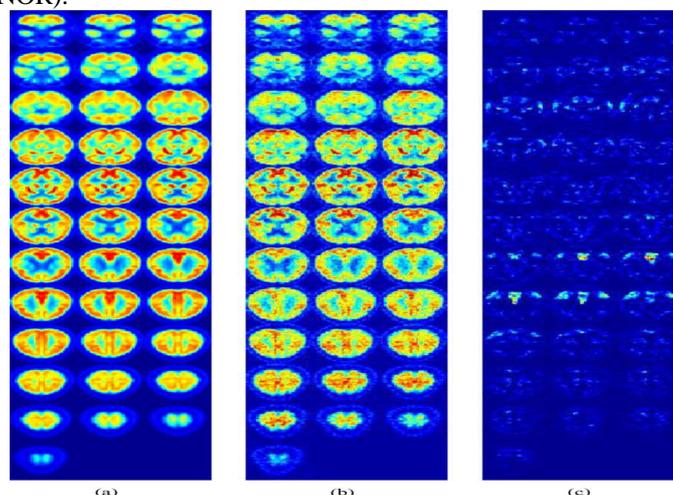
PET is a functional imaging technique that produces a three-dimensional image of functional processes in the body. This detects pairs of gamma rays emitted indirectly by a positron-emitting radionuclide (tracer), which is introduced into the body on a biologically active molecule. The computer analysis is used to construct three-dimensional images of tracer concentration within the body. If the biologically active molecule chosen for PET is fludeoxyglucose (FDG), analogue of glucose, the concentrations of tracer imaged will indicate tissue metabolic activity by virtue of the regional glucose. Fig. 4 show a set of transaxial plane brain slices of the PET database, for one AD patient and one NOR subject glucose.

**5.3 CAD Tool Evaluation**

In order to evaluate the developed NMF-SVM CAD tool in all its variations, the success rate (Acc), sensitivity (Sens), and specificity (Spec) are obtained, these SENS and SPECT defined as:

$$SENS = \frac{TP}{TP + FN}, \quad SPECT = \frac{TN}{TN + FP}$$

where TP is the number of positives (AD patients correctly classified); TN is the number of true negatives (NOR patients correctly classified); FP is the number of false positives (NOR classified as AD); FN is the number of false negatives (AD classified as NOR).



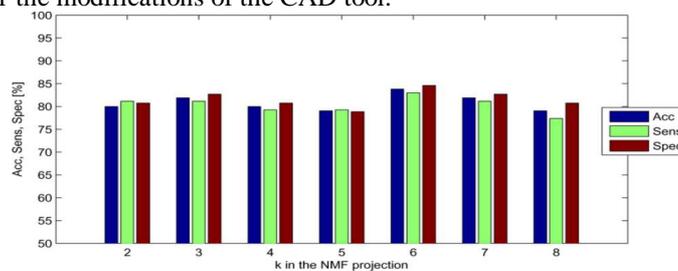
**Figure.3.** PET transaxial brain slices, oriented from posterior (top) to anterior (bottom) in each slice, and from ventral (top left) to dorsal (bottom right) in the complete set. (a) Mean mask for NOR subjects. (b) Example AD patient. (c) Between-subjects variability between AD and NOR groups.

## 6. EXPERIMENTAL RESULTS AND DISCUSSION

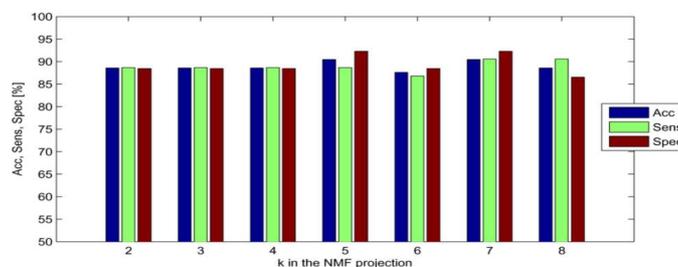
First of all a NMF-SVM based CAD tool is developed, with linear SVM as classifier. The proposed method is later enhanced with the addition of bounds of confidence for the classification decision. Finally a new SVM-based classifier is evaluated. In the PET database, only NOR and AD groups are considered for the validation of the CAD system, in order to avoid errors and the uncertainty of their actual state as AD or NOR.

### 6.1 Basic NMF-SVM CAD Tool

Figs.4 and 5 provide the experimental results of the basic NMF-SVM system, for a variety of K values in the NMF projection. As it is seen, levels in the range of 80% to 90% are achieved, for both databases. These results are considered as a reference for the modifications of the CAD tool.



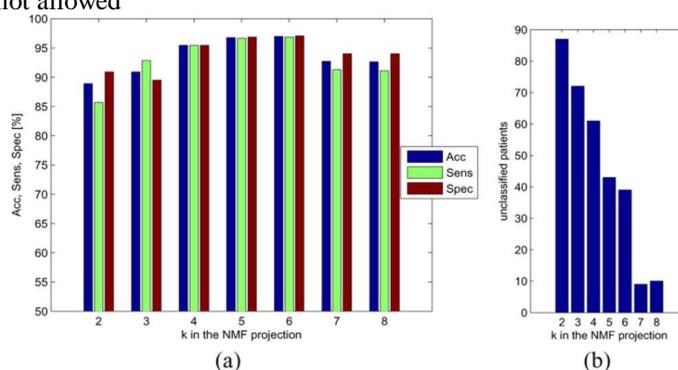
**Figure.4.** Performance of the basic NMF-SVM CAD system with the PET database, for different K values in NMF.



**Figure.5.** Performance of the basic NMF-SVM CAD system with the SPECT database, for different K values in NMF.

### 6.2 Basic NMF-SVM Tool With Bounds of Confidence

The addition of bounds of confidence to the SVM classifier permits the improvement of the results. Even though in this case some subjects are not classified, this approach is similar to real-life cases: sometimes some patients are difficult to be classified with reliability even by experts. Fig.6. show the diagnosis results of the basic NMF-SVM tool with bounds of confidence, along with the number of unlabeled patients due to the existence of a security region where classification decisions are not allowed



**Figure.6.** (a) Performance of the basic NMF-SVM CAD system with bounds of confidence in the PET database, for different K values in NMF, (b) number of subjects without classification in the NMF-SVM CAD system with bounds of confidence.

## 7. CONCLUSION

This paper presents a NMF-SVM based technique for computed aided diagnosis of Alzheimer’s disease. The proposed technique is based on the combination of nonnegative matrix factorization (NMF) for feature selection and reduction and SVM with bounds of confidence for classification. The feature reduction step provides a reduced set of variables representing the original data. This feature reduction is especially suitable for machine learning techniques such as SVM. Three different approaches for the classifier are provided, two of them including bounds of confidence and taking

advantage of the definition of a “security region” in the SVMhyperplane, where no decision is assumed. The validation results of the proposed NMF-SVM method yields up to 91% classification accuracy with high sensitivity and specificity values (upper than 85%) for both data sets.

## REFERANCES

- [1] P. Padilla, M. López, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer’s Disease, *ieeetransaction on medical imaging*, vol. 31, no. 2, february 2012
- [2] Aiswarya.V.S, Jemimah Simon, Diagnosis of Alzheimer’s disease in Brain Images using Pulse Coupled Neural Network, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-2, Issue-6, May 2013.
- [3] P. Padilla a, J.M. Górriza, J. Ramírez, E.W. Langb, R. Chavesa, F. Segoviaa, M. López, D. Salas-González, I. Álvarez, Analysis of SPECT brain images for the diagnosis of Alzheimer’s disease based on NMF for feature extraction, Department of Signal Theory, Networking and Communications, University of Granada, uentenueva s/n, Granada, Spain b CIML Group, Biophysics, University of Regensburg, Germany.
- [4] I. Álvarez, J. M. Górriz, J. Ramírez, D. Salas, M. López, C. G. Puntonet, and F. Segovia, “Alzheimer’s diagnosis using eigenbrains and support vector machines,” *IET Electron. Lett.*, vol. 45, no. 1, pp. 165–167, Feb.2009.
- [5] D. D. Lee and S. Seung, “Algorithms for non-negative matrix factorization,”*Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 556–562, 2001
- [6] A. Takemura, A. Shimizu, and K. Hamamoto, “Discrimination ofbreast tumors in ultrasonic images using an ensemble classifier basedon the adaBoost algorithm with feature selection,” *IEEE Trans. Med.Imag.*, vol. 29, no. 3, pp. 598–609,Mar.2010.
- [7] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-González, M. López, F.Segovia, R. Chaves, M. Gómez-Rio, and C. Puntonet, “18F-FDG PETimaging analysis for computer aided Alzheimer’s diagnosis,” *Inf. Sci.*,vol. 181, no. 4, pp. 903–916,2011.

## AUTHOR



**Ms. Tejal A. Fuse** persuing degree in Electronics & Telecommunication Engineering from J.D.I.E.T. Yavatmal



**Ms. Nikita D. Jayasignpure** persuing degree in Electronics & Telecommunication Engineering from J.D.I.E.T. Yavatmal



**Prof. Pragati D. Pawar** received the B.E. degrees in Electronics & Telecommunication Engineering from K.I.T.S, Ramtek and M.E. in Electronics from S.G.G.S. Nanded