



Toward Alzheimer's disease classification through machine learning

M. Rohini¹ · D. Surendran²

© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Alzheimer's disease (AD) and cognitive impairment due to aging are the recently prevailing diseases among aged inhabitants due to an increase in the aging population. Several demographic characters, structural and functional neuroimaging investigations, cardio-vascular studies, neuropsychiatric symptoms, cognitive performances, and biomarkers in cerebrospinal fluids are the various predictors for AD. These input features can be considered for the prediction of symptoms whether they belong to AD or normal cognitive impairment due to aging. In the proposed study, the hypothesis is derived for supervised learning methods such as multivariate linear regression, logistic regression, and SVM. Feature scaling and normalization are performed with features as initial steps for applying the parameters to derive the hypothesis. Performance metrics are analyzed with the implementation results. The present work is applied to 1000 baseline assessment data from Alzheimer's Disease Neuroimaging Initiative (ADNI) studies that give conversion prediction. The comparison of results in the literature suggests that the efficiency of the proposed study is highly advantageous in differentiating AD pathology from cognitive impairment due to aging.

Keywords Multivariate linear regression · Logistic regression · Support vector machine (SVM) · Feature scaling · Normalization · ADNI

1 Introduction

Alzheimer's disease causes different types of dementia. Symptoms prevail 20 years before the onset of disease. Only after many years of structural changes in brain, individuals notice memory loss, language problems, and basic brain skills (cognitive skills). Symptoms are due to neuron damage in various parts of the brain that are responsible for memory, cognitive function, and thinking and learning. It interferes with elderly people in performing everyday activities. At this point called AD dementia progression, neurons in the part of the brain damage the activities like individual identity, planning events, and sports. Furthermore, the basic motions like walking and swallowing are found difficult, and finally bed-bound and death occurred (Kang et al. 2013).

A normal adult's brain has 100 billion neurons, each of which is long and widely branched. The extension of branching helps the neurons to form connections with other neurons. The connection between neurons is called synapses. Information flow occurs through chemical substances released by one neuron and detected by the receiver. Within the connectivity, there are 100 trillion synapses present. They allow overall communication of information such as memory, thoughts, sense, skills, and emotions (Brooks and Loewenstein 2010).

Accumulation of abnormal proteins called beta-amyloid outside the neurons (beta plaques) and the accumulation of abnormal form of the protein tau (called tau tangles) inside the neuron are the main causes of changes occurring in the brain. Beta-amyloid protein causes cell death brain damage. The increase in these two proteins is due to a genetic mutation. It accumulates 22 years before AD, glucose metabolism decreases before 18 years of AD, and brain atrophy decreases 13 years before the onset of disease. Initially, the brain compensates without symptoms; later, nerve cell damage becomes a more significant cause for all behavioral and all abnormalities (Brooks and Loewenstein 2010).

Communicated by V. Loia.

✉ M. Rohini
rohini@skcet.ac.in

¹ Sri Krishna College of Engineering and Technology, Coimbatore, India

² KPR Institute of Engineering and Technology, Coimbatore, India

Prediction of disease symptoms is of a recent research area and can be solved using machine learning algorithms. The present research is toward utilizing features from the ADNI database whose symptoms are cognitively impaired. Using machine learning classifiers, symptoms of AD patients are identified from the cognitively normal aged community.

2 Materials and methods

The dataset utilized in this methodology is obtained from the ADNI database (Alzheimer's Disease Neuroimaging Initiative). It provides a stand methodology for biomarker identification and supplies high data availability to researchers. It provides structural and functional changes of the brain in the shape of the hippocampus regions of the brain and neurons. It also provides data about the deposition of Amyloid Beta protein and tau so that it can be utilized to check if it is abnormal or not. It also details the data of different types of brain atrophy. It provides white matter and glucose metabolism readings of brain regions (Kang et al. 2013). The data generated by ADNI are featured on LONI and are provided to several growing investigators. The cores of ADNI include ADNI-1, ADNI-Go, and ADNI-2. All the cores provide biomarkers through validation and statistical analysis. ADNI also provides pathological and genetic features of disease toward disease progression. These features lead to ADNI as the large utilized database for data sharing by worldwide researchers. It provides features in three stages: the MRI data (Magnetic Resonance Imaging), PET (Positron Emission Tomography), and genetic arrangements about disease progression. The proposed study features include demographic clinical data and APOE genotype of each subject. The training data include subjects involved in ADNI, the multicentered study, with the aim of prediction of Alzheimer's disease, mild cognitive impairment, early and late AD and elderly but cognitively normal individuals (Peterson et al. 2006). Disease progression is followed up through various AD interventions, imaging studies, and biomarkers. These disease types have been enlisted from three cores of ADNI-1, ADNI-GO, and ADNI. The study consists of a swarm of data from imaging, genetic, assessments, medical history, and subject's characteristics. These data are available from LONI. These inputs are considered as feature vectors that are extracted to train our supervised learning models (Cuingnet et al. 2011).

Given in Fig. 1 are the images acquired from ADNI image repository, scanned images of cognitively normal, AD affected, and mild cognitively impaired individuals. Acquisition Planes are AXIAL, SAGITTAL, CORONAL. Acquisition Type = 3D; Coil = HE; Field Strength = 1.5

T; Flip Angle = 8.0 degree; Manufacturer = SIEMENS; Matrix X = 192.0 pixels; Matrix Y = 192.0 pixels; Matrix Z = 160.0; Mfg Model = Symphony; Pixel Spacing X = 1.25 mm; Pixel Spacing Y = 1.25 mm; Pulse Sequence = IR/GR; Slice Thickness = 1.20000004 76837158 mm; TE = 3.609999895095825 ms; TI = 1000.0 ms; TR = 3000.0 ms; Weighting = T1 (Cuingnet et al. 2011);

3 Methods

3.1 Multivariate linear regression

MRI imaging data are used for feature extraction obtained from the ADNI image database. All three modalities of imaging data are collected. Image collection features are large, and many feature images are not appropriate for the classification of disease. Cooper et al. (2015) proposed the multitask feature selection for feature vector extraction and dimensionality reduction. Feature selection is commonly used for dimensionality reduction, as well as for the removal of irrelevant features. Feature selection for multiple regression/classification variables is very helpful to suppress noises in the individual variables (Fig. 2).

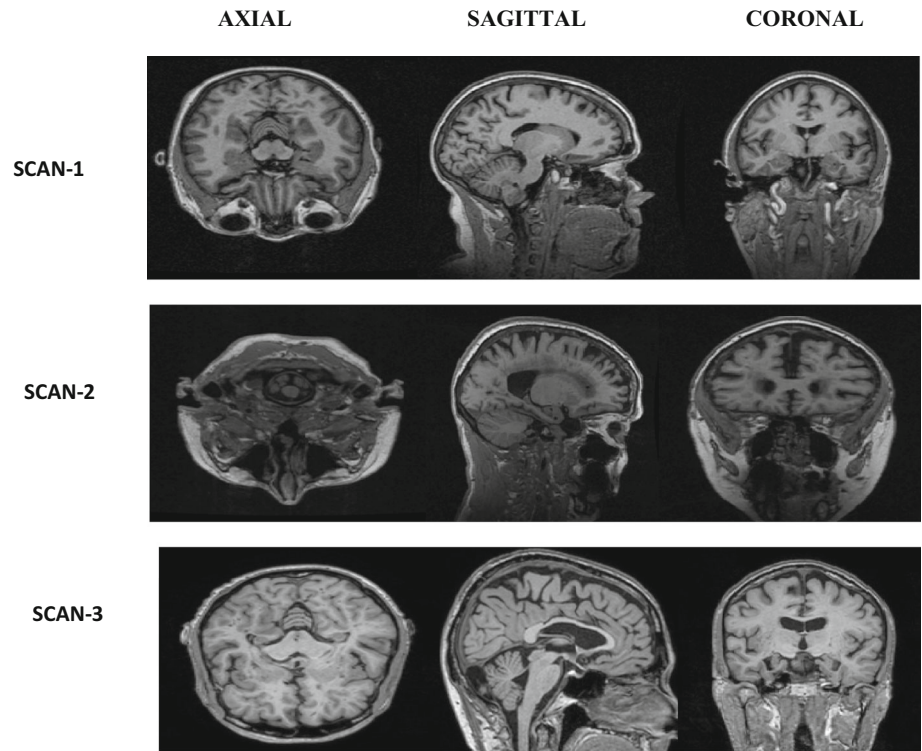
Multiple features of elderly inhabitants acquired from the ADNI database are taken for training the model using multivariate linear regression. The dataset consists of multiple features enabling the classification and is powerful with the features participant id(P_id), MMSE(MMSE score), APOE4(APOE4 genotype), and mod(modality). Y is the prediction group deciding whether the features belong to AD class or the symptoms are due to normal aging.

Feature vectors are taken as X_a, X_b, X_c, X_d and the class label that is predicted is Y_g . Here, n is the number of input feature vectors derived, i.e., $n = 4$ and m is the number of the training data sample, i.e., $m = 1000$. X_i denotes the feature in i_{th} training example. Here X_1 is a four-dimensional vector that consists of PTID, MMSE score, APOE4, and mod of first training example, i.e., $X \in R_N$ dimensional feature vectors. The hypothesis that models a function for predicting the output class by mapping the feature vectors is given below,

$$H(\theta_{MR}) = \theta_0 + \theta_1 X_a + \theta_2 X_b + \theta_3 X_c + \theta_4 X_d = \theta^T X \quad (1)$$

where $X = \begin{pmatrix} X_0 \\ X_a \\ X_b \\ X_c \\ X_d \end{pmatrix} \in R_5$, $\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} \in R_5$. This hypothesis predicts the individuals and their conversion from normal cognitive impairment. The hypothesis is derived from multivariate linear regression for the elderly affected

Fig. 1 Acquisition planes are axial, sagittal, coronal of cognitively normal aged individual, AD individual, and mild cognitively impaired individual (MCI)



individual. The parameters of the hypothesis are initialized such that the cost of the model predicting the pathology is termed as

$$\text{Cost}(\theta_{MR}) = 1/2000 \sum (H(\theta_{MR})(X_i) - Y_i)^2 \quad (2)$$

Parameters $\theta_0, \theta_1, \theta_2.. \theta_j$ are updated simultaneously such that $j = 0, 1..n$. Cost of the model evaluates the error for single data from imaging studies and genetic assessments that are the predictors and the loss functions for all the training examples are calculated by updating all the parameters simultaneously $\theta_{bias}, \theta_{ptid}, \theta_{mmse}, \theta_{apoe4}, \theta_{mod}$ for evaluating the cost of the proposed model. The following defines gradient descent for our model that compares each features using the hypothesis of multiple variables

$$\theta_{MR} = \theta_{MR} - \alpha(\partial/\partial\theta G(\text{cost}(\theta_{MR}))) \quad (3)$$

$$\theta_{bias} = \theta_0 - \alpha \left[1/1000 \sum_{i=1}^{1000} (H(X_i) - y_i) \right] X_{0i} \quad (4)$$

$$\theta_{ptid} = \theta_0 - \alpha \left[1/1000 \sum_{i=1}^{1000} (H(X_i) - y_i) \right] X_{ai} \quad (5)$$

$$\theta_{mmse} = \theta_0 - \alpha \left[1/1000 \sum_{i=1}^{1000} (H(X_i) - y_i) X_{bi} \right] \quad (6)$$

$$\theta_{apoe4} = \theta_0 - \alpha \left[1/1000 \sum_{i=1}^{1000} (H(X_i) - y_i) X_{ci} \right] \quad (7)$$

$$\theta_{mod} = \theta_0 - \alpha \left[1/1000 \sum_{i=1}^{1000} (H(X_i) - y_i) X_{di} \right] \quad (8)$$

3.2 Feature scaling and mean normalization

It is performed on the feature vectors where the features APOE and MMSE are scaled for approximately the range of values between $-1 \leq x \leq 1$. The gradient descent run on the loss function of the model takes more iterations in reaching the global minimum. It is performed to check whether all the features are the same scale of values that enables the gradient descent applied on the cost function of the model to reach a global minimum in fewer iterations and the loss function is hence reduced (Sperling and Johnson 2013).

In addition to feature scaling in which each feature is divided by its maximum value, mean normalization is performed with features. It is performed by replacing each feature X_{ai}, X_{bi} with difference between feature and mean to make each parameter with zero mean value, i.e., $-0.5 \leq X_a \leq 0.5, -0.5 \leq X_b \leq 0.5, -0.5 \leq X_c \leq 0.5, -0.5 \leq X_d \leq 0.5$. S_i is the range of values in which each feature vector ranges between minimum and maximum.

$$X_{ai} = X_{ai} - \text{mean}(X_{ai})/S_i \quad (9)$$

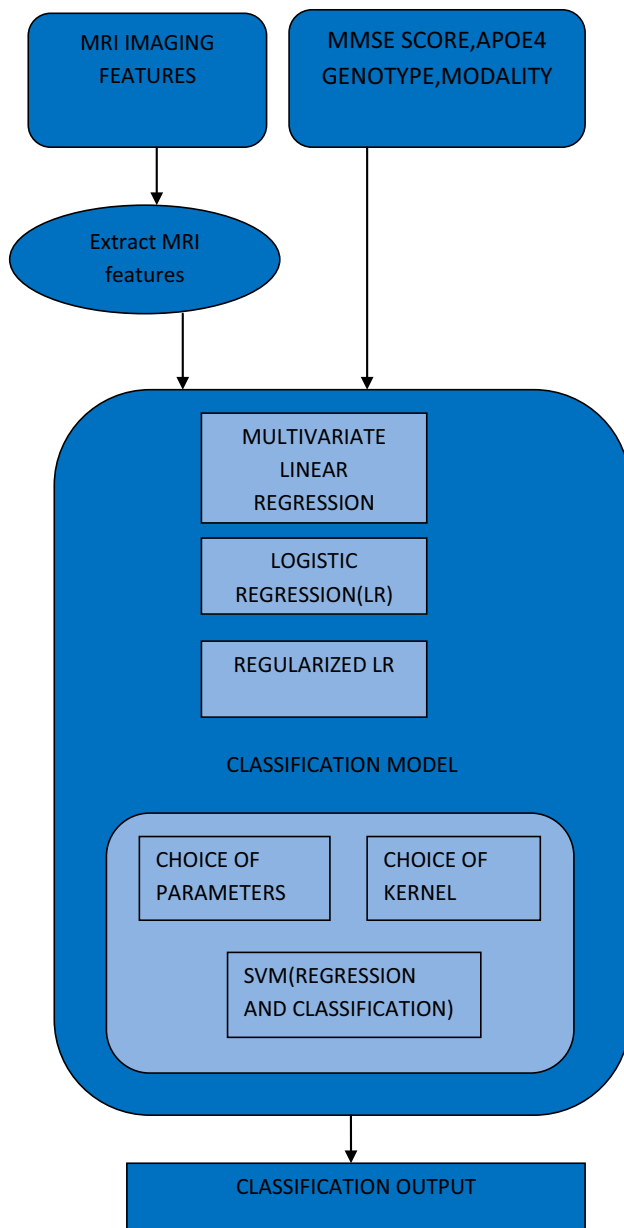


Fig. 2 The architecture of the machine learning based AD prediction

Feature scaling and mean normalization are optimization techniques applied in the proposed study to apply gradient descent on cost function to reach a global minimum with fewer iterations and hence the model learns the training data at a faster rate and minimum loss. With the implementation, it is noted that the cost function converges if gradient descent decreases less than 0.0001 for each iteration. The convergence depends on the learning rate passed to gradient descent function (Figs. 3, 4).

3.3 Logistic regression (LR)

Logistic regression (LR) predicts disease symptoms and classifies its pathology classes from several features. A linear feature predictor is applied to fit the LR model and each subject's disease class is predicted. Unlike the classification model that predicts a hypothesis that can be $0 > h_c > 1$. In the present study, the logistic regression model is applied for classifying the feature where the hypothesis predicts classes that must be $0 \leq H_{LR} \leq 1$, where H_{LR} is the hypothesis of the LR model. Despite the multivariate linear model implemented, the LR classification model ignores discrete-valued outcome classes. It is constructed with intuition $0 \leq H_{LR} \leq 1$. This can be achieved by applying linear multivariate regression and its hypothesis to fit into LR function as follows.

$$H_{LR}(\theta) = G_{LR}(\theta^T X) = 1 / (1 + e^{-\theta^T X}) \quad (10)$$

$$\text{Cost}(\theta_{LR}) = \begin{cases} -\log(H_{LR}(\theta)), & Y \in 1/AD \\ -\log(1 - H_{LR}(\theta)), & Y \in 0/NC \end{cases} \quad (11)$$

where $X = \begin{pmatrix} X_0 \\ X_a \\ X_b \\ X_c \\ X_d \end{pmatrix} \in R_5$, $Y \in \{Y_{AD}, Y_{NC}\}$, resulting in class

belonging to Alzheimer's disease pathology and cognitively normal aged individuals. The cost function of the LR model results in 0 if the hypothesis derived is 1, and the model incurs infinite cost if the hypothesis derived is 0. The prediction class Y resulted in AD by choosing $P(Y = 1|X; G_{LR}(\theta^T X))$ but $Y = AD$ is the penalty of the learning algorithm with an infinite cost. LR function applies the input features from the dataset to be mapped for $0 \leq H_{LR} \leq 1$ to convert discrete-valued feature vectors into binary classification function. However, the present work proposes a novel optimization algorithm that makes the LR model to learn the training examples faster and to scale with the number of features (Van Rossum et al. 2010).

3.4 Regularized LR

The input feature vector of our problem ranges from X_a, X_b, X_c, X_d with M training examples. When the hypothesis and loss function modules are implemented, the learned model fits well for training examples but failed to fit for testing and validation set. Hence, the hypothesis can be regularized such that the parameters are converted as a quadratic function, and the model fits for more number of datasets. The quadratic hypothesis model is defined by

Fig. 3 Accuracy in multivariate linear regression

```

Program paused. Press enter to continue.
Cost at initial theta (zeros): 0.693147
Expected cost (approx): 0.693
Gradient at initial theta (zeros):
-0.100000
-12.009217
-11.262842
Expected gradients (approx):
-0.1000
-12.0092
-11.2628

Cost at test theta: 0.218330
Expected cost (approx): 0.218
Gradient at test theta:
0.042903
2.566234
2.646797
Expected gradients (approx):
0.043
2.566
2.647

Program paused. Press enter to continue.
Cost at theta found by fminunc: 0.203498
Expected cost (approx): 0.203
theta:
-25.161272
0.206233
0.201470
Expected theta (approx):
-25.161
0.206
0.201

Program paused. Press enter to continue.
For given features with AD symptoms and normal aging, we predict probability of 0
.776289
Expected value: 0.775 +/- 0.002

Train Accuracy: 89.000000
Expected accuracy (approx): 89.0

octave:9> _
    
```

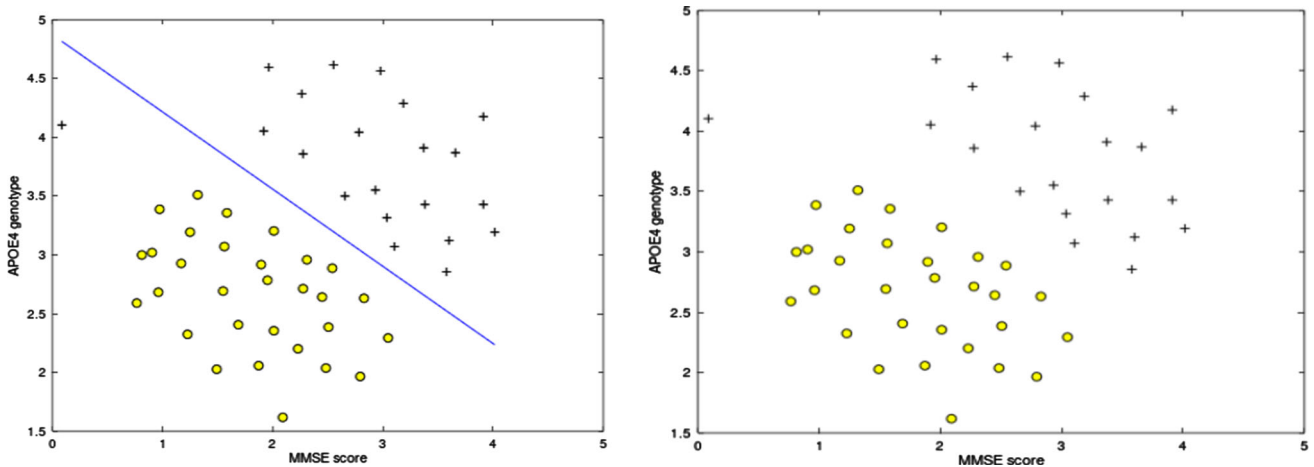


Fig. 4 Training the data in multivariate linear regression model

$$H_{rLR}(\theta) = G_{LR}(\theta_0 + \theta_1 X_a + \theta_2 X_a^2 + \theta_3 X_a^2 X_b + \theta_4 X_a^2 X_b^2 + \dots) \tag{12}$$

Hereby, in the proposed regularized LR model, the hypothesis is made more quadratic by including MMSE score and APOE genotype features to the parameters θ_3 and θ_4 . To decrease the cost of the model, the parameters θ_3 and θ_4 are reduced by values ≈ 0 . Hence, the regularized cost function of the model is defined by (Fig. 5)

$$\begin{aligned}
 \text{Cost}(\theta_{rLR}) = & - \left[\frac{1}{1000} \sum_{i=1}^{1000} (Y_i \log H\theta(X_i) + (1 - y_i) \right. \\
 & \left. \times \log(1 - H\theta(X_i)) + \lambda/2000 \left(\sum_{j=1}^4 \theta_j \right) \right] \tag{13}
 \end{aligned}$$

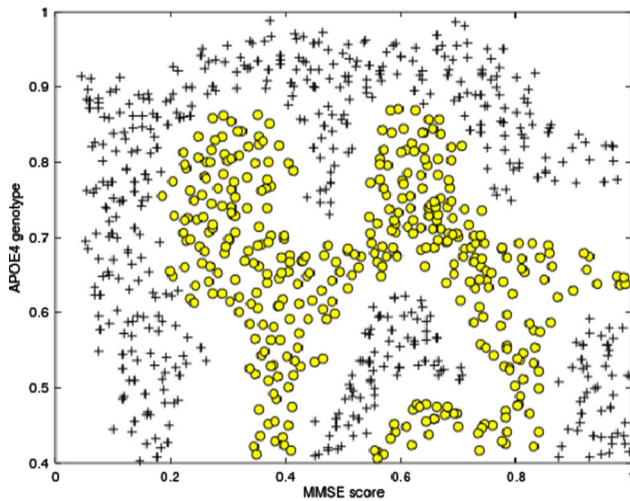
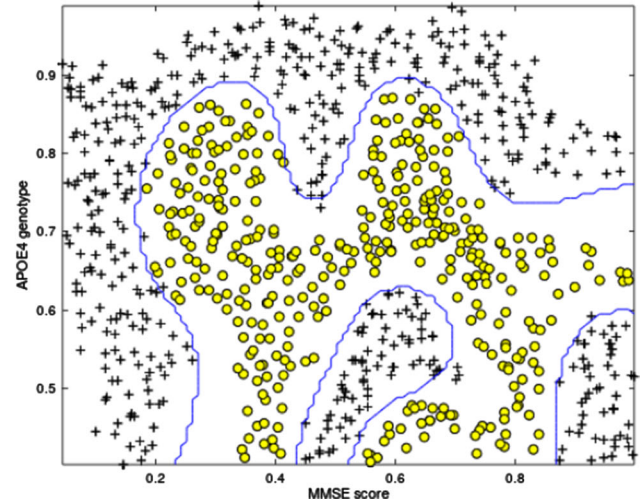


Fig. 5 Training the data logistic regression model

3.5 Support vector machine

The SVM learning algorithm is chosen to classify the feature set since it shows better performance compared to the previous logistic regression model. Here, we train the SVM hypothesis corresponding to the regression variable. Classification is performed to classification groups. This module utilizes Liblinear, LibSVM packages in solving SVM parameter θ . By implementing SVM, parameter CAD and the type of kernel are specified. In the previous modules, linear classification is performed with the hypothesis that is simple linear kernel implementation in SVM. The standard linear kernel with a larger number of features and very few training examples has led to overfitting with the linear kernel functions, i.e., when $X \in R_{n+1} > m$. Hence, Gaussian and Gaussian Radial Basis function modules are implemented for our ADNI dataset. ADNI database provides a large collection of elderly cognitive impaired datasets. Since the disease predictors involve only fewer features Gaussian kernel, SVM is a feasible option for classification (Kruthika et al. 2019), where $X \in R_{n+1}$ is less compared to the number of training examples. It leads to a more complex non-linear decision boundary; for classification, we define the Gaussian kernel function as follows



$$\text{Min SVM} \sum_{i=1}^{1000} Y_i \text{cost}(\text{Transpose}(\theta)X) \quad (14)$$

$$+ (1 - Y_i) \text{cost}(\text{Transpose}(\theta)X) + 1/2 \sum \theta_2$$

$$F(i)G_{\text{SVM}} = \exp(-\|X_i - X_j\|^2/2\sigma^2) \quad (15)$$

Given X evaluates features, $F(i) \in R_{m+1}$. This function should be iterated for M number of training examples that are the rows of the dataset, where X_i, X_j are featured in each row mapped to $F(i)$ that returns classification results of each row. This is implemented as a Gaussian kernel, and the parameters depend on feature scaling to be performed (Figs. 6, 7).

4 Discussions

Our proposed system predicts and classifies the ADNI features that are numerically valued as cognitively normal or under AD pathology. A Boolean variable is initialized for the class label by implementing the range of continuous features and classifying it. We prepared 1000 features for testing and 200 features for evaluation data, and classification is performed. We implemented multivariate linear

Fig. 6 Training the data in SVM

```

C:\Octave\OCTAVE~1.1\bin\octave-gui.exe
Loading and Visualizing Data ...
Program paused. Press enter to continue.

Training Linear SVM ...

Training .....
. Done!

Program paused. Press enter to continue.

Evaluating the Gaussian Kernel ...
Gaussian Kernel between Xa-MMSE score = [19; 21; 30], Xb-APOE4genotype = [1; 2;
1], sigma = 2.000000 :
0.324652
(for sigma = 2, this value should be about 0.324652)

```

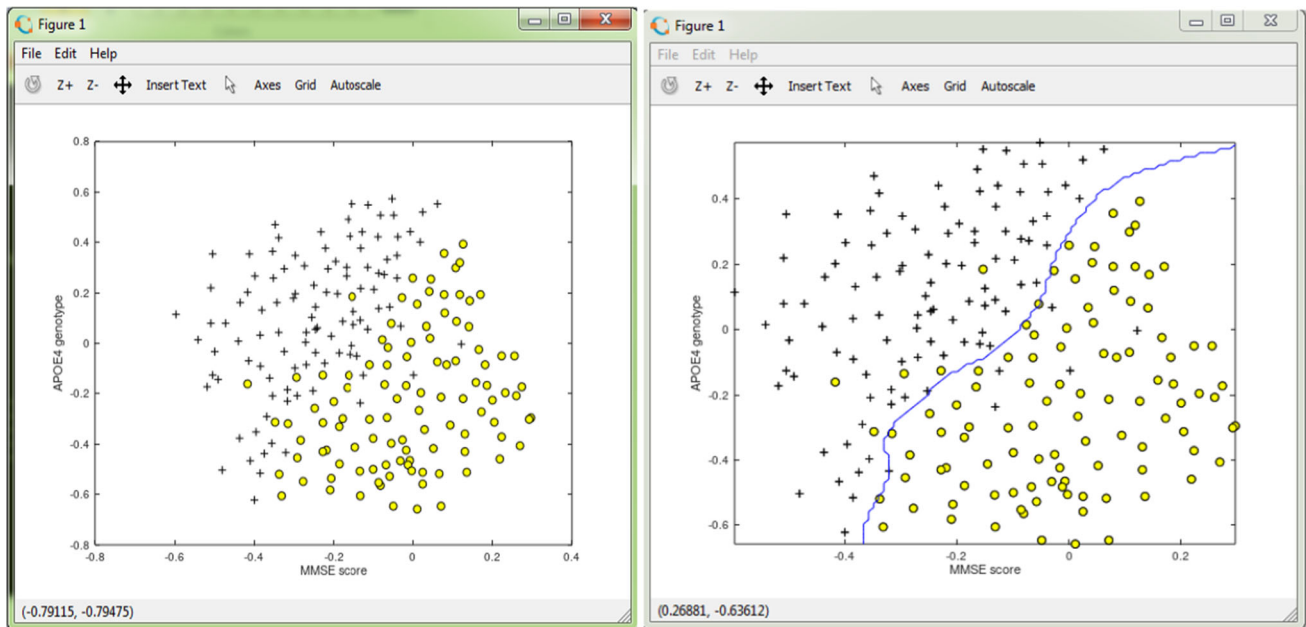


Fig. 7 Training the data using Gaussian kernel

regression for the training model that initially predicts continuous values. The accuracy and performance are improved by using logistic regression. Regression Loss is calculated that differentiates negative class that predicts cognitively normal controls as Alzheimer's individuals with a probability of 0.999. Logistic loss is calculated by evaluating features and labels as follows:

$$\text{Logistic loss} = \sum_{X_i, Y} -Y \log(Y_{AD}) - (1 - Y) \log(1 - Y_{AD}) \quad (16)$$

where Y_{AD} is the predicted group of feature labels from input feature X that falls under two class groups. Metrics such as model accuracy, ROC, and AUC are calculated using `LinearClassifier.evaluate()` method and plotted using the `matplotlib` library. Root mean square function and logistic regression loss functions are invoked in the evaluation module.

5 Experimental results

The proposed classification models are built using Jupyter notebook with Anaconda Python distribution, an open-source software. GNU Octave is used to evaluate the hypothesis of each classifier since it is an easy-to-use environment for mathematical evaluations. In early identification of symptoms of cognitive disability in the elderly community, we implement a machine learning classifier algorithm that takes features of cognitive abnormal individuals. Initially, we implemented using logistic

regression, a supervised learning method. Next, we implemented an SVM and chosen kernel based on the number of features. The results are cross-validated with testing data. AUC is derived such that it shows successful result in early prediction of AD. In the present study, the entire dataset is used for training, testing, and classification. The number of features chosen is dependent on choosing the regularization parameter in regression. Hence, few age-related features are used for training the model. This involves above 70% of ROI were chosen as features. For feature scaling, we used min-max normalization by selecting a feature that has a non-zero value to increase the regularization parameter. This improved the detection accuracy with available features. Thus, in this stage, regularized logistic regression outperformed than other existing methods. The results show an MMSE score corresponding to cognitively impaired individuals in high dimensional features playing major roles. For further prediction of brain cell atrophy, more specific features and scaling parameters are needed.

From the dataset of elderly individuals, their mild cognitive decline symptoms are similar to pathological symptoms. Hence, feature set is analyzed and cognitively normal fields are removed. After this preprocessing, the features are taken for training the classifier. Hence the neuron cell degeneration that is actual AD pathology and normal aging cognitive decline are differentiated. Present work also includes analyzing MRI imaging to identify age-related and cognitively impaired features. MRI imaging features are best utilized for such differentiation. In SVM, these features are also accountable for the kernel selection

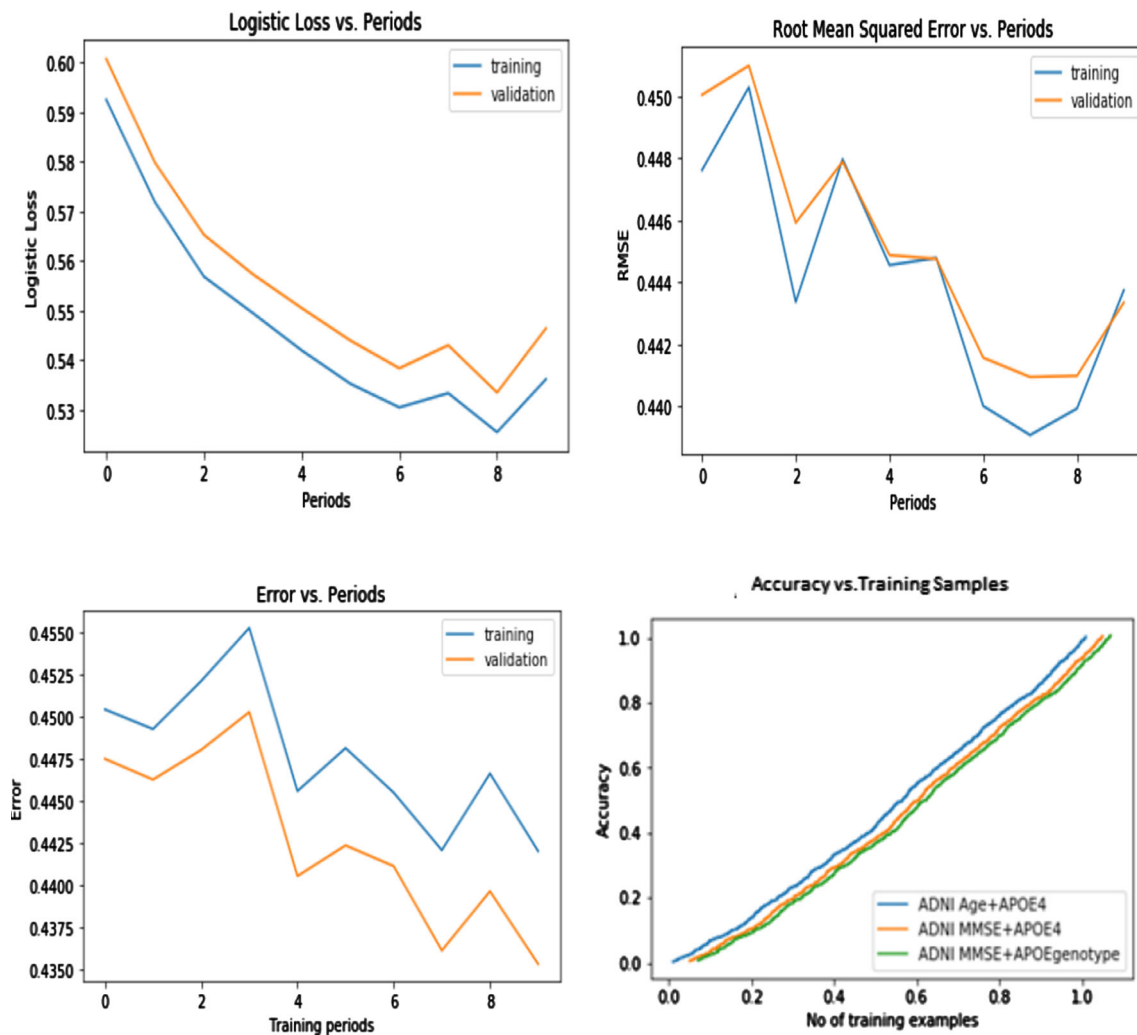


Fig. 8 Plots of training error, logistic regression loss, accuracy, and root mean squared error

algorithm. Hence, the performance using the selected kernel, differentiates different class labels more accurately with all classifiers.

The regression classifiers used to predict the pathology are checked for logistic regression loss that plots the misclassifications where the cognitive impaired elders without any brain cell atrophy are misclassified as AD patients. Accuracy, ROC, and the area under the ROC curve (AUC) are plotted for performance evaluation. Compared to previous studies utilizing MMSE, APO genotype using ADNI in our present system outperforms with Accuracy of 89% and AUC of 78% (Fig. 8).

6 Conclusion and future work

The prediction of pathology as AD or normal cognitive decline in older adults remains the challenging problem in cognitive medicine that is addressed using binary

classification. Our study proved to utilize simple and cost-effective machine learning models for good accuracy in disease prediction. It employs feature vectors with higher weights for classification and outperformed best results after preprocessing steps such as feature scaling and normalization. Based on the number of features and training dataset, choice of the kernel is decided and implemented in the SVM classifier. The clinical data are analyzed along with imaging modalities, and MRI has shown to be feasible in classification. It is noticeable from the work that the regression model is implemented with an advanced optimization hypothesis for gradient descent. It can be further improved by choosing the conjugate gradient, LFGS, and BFGS optimization techniques. These proved to reduce cost function and have shown better classification results in previous studies. This also employs implementing the hypothesis without a manual selection of learning parameters. The work also can be extended further with multi-class classification by predicting the feature under AD,

Dementia, Mild cognitive impairment, and Pre-mild cognitive impairment. One-Vs-All classification can be employed in regression and SVM classifiers to show the best multiclass classification results.

Acknowledgements Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

Compliance with ethical standards

Conflict of interest All the authors in the paper have no conflict of interest.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Brooks LG, Loewenstein DA (2010) Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. *Alzheimer's Res Ther* 2:28

- Cooper C, Sommerlad A, Lyketsos CG, Livingston G (2015) Modifiable predictors of dementia in mild cognitive impairment: a systematic review and meta-analysis. *Am J Psychiatry* 172:323–334
- Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert M-O, Chupin M, Benali H, Colliot O, Initiative ADN et al (2011) Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56(2):766–781
- Kang JH, Korecka M, Toledo JB, Trojanowski JQ, Shaw LM (2013) Clinical utility and analytical challenges in measurement of cerebrospinal fluid amyloid-beta(1-42) and tau proteins as Alzheimer disease biomarkers. *Clin Chem* 59:903–916
- Kruthika KR, Rajeswari, Maheshappa HD (2019) Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. *Inform Med Unlocked* 14:34–42
- Petersen RC, Parisi JE, Dickson DW, Johnson KA, Knopman DS, Boeve BF, Jicha GA, Ivnik RJ, Smith GE, Tangalos EG, Braak H, Kokmen E (2006) Neuropathologic features of amnesic mild cognitive impairment
- Sperling R, Johnson K (2013) Biomarkers of Alzheimer's disease: current and future applications to diagnostic criteria. *Continuum (Minneapolis Minn)* 19:325–338
- Van Rossum IA, Vos S, Handels R, Visser PJ (2010) Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design. *J Alzheimers Dis* 20:881–891

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.