

# Imputation And Classification Of Missing Data Using Least Square Support Vector Machines – A New Approach In Dementia Diagnosis

T.R.Sivapriya

Dept. of Computer Science Lady  
Doak College  
Madurai , India

A.R.Nadira Banu Kamal

Dept. of MCA  
TBAK College  
Kilakarai, India

V.Thavavel

Dept. of MCA  
Karunya University  
Coimbatore, India

**Abstract**— This paper presents a comparison of different data imputation approaches used in filling missing data and proposes a combined approach to estimate accurately missing attribute values in a patient database. The present study suggests a more robust technique that is likely to supply a value closer to the one that is missing for effective classification and diagnosis. Initially data is clustered and z-score method is used to select possible values of an instance with missing attribute values. Then multiple imputation method using LSSVM (Least Squares Support Vector Machine) is applied to select the most appropriate values for the missing attributes. Five imputed datasets have been used to demonstrate the performance of the proposed method. Experimental results show that our method outperforms conventional methods of multiple imputation and mean substitution. Moreover, the proposed method CZLSSVM (Clustered Z-score Least Square Support Vector Machine) has been evaluated in two classification problems for incomplete data. The efficacy of the imputation methods have been evaluated using LSSVM classifier. Experimental results indicate that accuracy of the classification is increases with CZLSSVM in the case of missing attribute value estimation. It is found that CZLSSVM outperforms other data imputation approaches like decision tree, rough sets and artificial neural networks, K-NN (K-Nearest Neighbour) and SVM. Further it is observed that CZLSSVM yields 95 per cent accuracy and prediction capability than other methods included and tested in the study.

**Keywords**- *Least Square Support Vector Machine; z-score; Classification; KNN; Support Vector Machine.*

## I. INTRODUCTION

Knowledge mining in databases especially medical databases of patient details consists of several steps like understanding the disease domain, forming the correct data set and cleaning the data, extracting of disease regularities hidden in the data thus formulating knowledge in the form of patterns or models, evaluation of the correctness and usefulness of results. Availability of large collections of medical data provides a valuable resource from which potentially new and useful knowledge can be discovered through data mining. Data Mining is increasingly popular as it holds to gain insight into the relationships and patterns hidden in the data. Patient records collected for diagnosis and prognosis typically encompass values of clinical and laboratory parameters and results of particular investigations specific to the disease

domain. Such data are not usually complete and inadequate due to inappropriate selection of parameters for the given task.

Development of Data Mining tools for medical diagnosis and prediction is an utmost of the hour. Patient database often has measurements of a set of parameters at different times, requesting temporal component to be taken into account in data analysis. In this study, patients have been under a longitudinal and cross-sectional monitoring to record data through various modalities like neuropsychological testing and Magnetic Resonance Imaging.

Researchers usually address missing data by including in analysis only complete cases i.e. those individuals who have no missing data in any of the variables required for that analysis. However, results of such analyses could be biased. Furthermore, cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, which in turn causes a substantial loss of precision and power leading to wrong diagnosis and treatment.

The risk of biased inclusion due to missing data depends on the reasons why data are missing. Reasons for missing data are commonly classified as: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). If it is plausible that data are missing at random, and not completely at random, analyses based on complete cases could be biased and such biases could be overcome using multiple imputation methods that allow individuals with incomplete data to be included in the analyses. Unfortunately, often it is not possible to distinguish between missing values at random and missing not at random in observed data. Therefore, biases caused by data set that are missing not at random can be addressed only by sensitivity analyses to examine the effect of different assumptions on missing data mechanism.

## II. RELATED WORK

Several Statistical and data mining methods have been used to analyse diagnosis of dementia. There are two traditional missing value imputation techniques. They are parametric and non-parametric imputation strategies. Parametric method is applied when relationship between conditional attributes is known. Non- parametric method is applied when the relationship between the conditional

attributes is unknown. Parametric methods like Nearest Neighbour [4][10][25] have been used for the prediction of missing attribute(s). Non-parametric technique such as empirical likelihood [32], clustering [26], Semi-parametric techniques [21][33] have also been applied for missing data imputation. Techniques like mixture model clustering [9], machine learning [12] have been used for imputing missing data. Multiple imputations [22] provide another way of finding missing values of attribute(s). In case of regression models, parametric regression imputation performs better if a dataset could be adequately and accurately modeled parametrically, or if users could correctly specify parametric forms for the dataset. Non-parametric imputation algorithm is found to be very effective when the user is unaware of the distribution of the dataset. Neural network method is regarded as one of non-parametric techniques used to compensate for missing values in sample surveys [24]. A non-parametric algorithm is useful only when form of relationship between conditional attributes and target attribute is not known a priori.

For imputation in medical databases, Jose et.al [11] have concluded that the methods based on machine learning techniques have been found to be suited for imputation of missing values and led to a significant enhancement of prognosis accuracy compared to imputation methods based on statistical estimation. In another approach, STATA v.10 [1] is used to impute missing data in patient database, lowest scores[8] of MMSE were used to fill missing values in diagnosis of dementia.

Several algorithms have been proposed as a solution for diagnosis of dementia. Kloppel et al. developed a supervised method using a support vector machine (SVM) in a high dimensional space [14], Trosset et al. proposed another semi-supervised learning method, which used multidimensional scaling (MDS) [30], Ceyhan et al. analyze the shape and size of hippocampus, where prominent neuropathological markers are shown to be present in AD [3]. In our previous study [27][28], we have investigated classification of dementia patients using SVM and an automatic supervised classification approach based on image texture analysis with Gabor wavelets as input to SVM, LS-SVM for distinguishing demented and non-demented patients.

This paper also evaluates approaches used to fill missing values and proposes a new and better approach to handle missing value situation and thereby enabling to feed correct input to the LSSVM classifier to get better prediction, diagnosis and treatment of the given data. The present study also examines the multiple biomarkers that contribute to dementia rather than concentrating on a single volume factor as described in the above studies through LS-SVM-PSO.

### III. MISSING DATA HANDLING MECHANISMS

Several methods have been applied in data mining to handle missing values in database. Data with missing values could be ignored, or a global constant could be used to fill missing values (unknown, not applicable, infinity), such as attribute mean, attribute mean of the same class, or an algorithm could be applied to find missing values[34]. Missing data imputation technique means a strategy to fill missing values of a data set in order to apply standard methods

which require completed data set for analysis. These techniques retain data in incomplete cases, as well as impute values of correlated variables.

Missing data imputation techniques are classified as ignorable missing data imputation methods, which include single imputation methods and multiple imputation methods, and non-ignorable missing data imputation methods which include likelihood based methods and the non-likelihood based methods. A single imputation method could fill one value for each missing value and it is more commonly used at present than multiple imputations which replace each missing value with several plausible values and better reflects sampling variability about actual value.

### IV. DATA SETS

OASIS provides brain imaging data that are freely available and used for distribution and data analysis [17]. This data set consists of a cross-sectional collection of 416 subjects covering adults in the age group 18 to 96 with early-stage Alzheimer's Disease (AD). For each subject, 3 or 4 individual T1-weighted MRI scans taken during a single imaging session are available.

The basic data source for the present studies is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI), a clinic-based, multicenter, which provides longitudinal study with blood, CSF, PET, and MRI scans repeatedly measured in 229 participants with normal cognition (NC), 397 with mild cognitive impairment (MCI), and 193 with mild AD during 2005-2007.

### V. IMPUTATION STRATEGIES

#### A. K-Nearest Neighbors (KNN) Imputation

If a training example contains one or more missing values, the *distance* between the example with missing values and all other examples is measured. Distance metric is a modified version of the *Manhattan distance* – distance between two examples is sum of the distances between the corresponding attribute values in each example. For discrete attributes, this distance is 0 if the values are the same, and 1 otherwise. In order to combine distances for discrete and continuous attributes, we perform a similar distance measurement for continuous attributes is performed– if the absolute difference between the two values is less than half of standard deviation, the distance is treated as 0; otherwise, 1.

The K complete examples closest to the example with missing values are used to choose a value. For a discrete attribute, the most frequently occurring value is used. For a continuous attribute, the average of the values from the K neighbors is used. In this study K value is determined by MMSE (Mini Mental State Examination) attribute distribution and set as 4 and 5 for demented and Non-Demented sets respectively.

#### B. Decision Tree

Decision tree is a classifier expressed as a recursive partition of the instance space. Decision trees are self-explanatory. They can handle both nominal and numeric input attributes and can handle datasets that may have errors and

missing values. C4.5 is an evolution of ID3[20]. It uses gain ratio as splitting criteria. Splitting ceases when the number of instances to be split is below a certain threshold. Error-based pruning is performed after the growing phase. C4.5 can handle numeric attributes. C4.5's distribution-based imputation (DBI)[19], is used in this study. MMSE score is the splitting criterion based on which patient details are classified. Further CDR( Clinical Dementia Rating)[17] is an essential attribute in dementia diagnosis.

### C. Back propagation algorithm

In our study a multilayered back-propagation neural network has been used (10 inputs from each of the 150 adolescents of the longitudinal and cross-sectional data set, comprising input patterns, and two binary outputs). The network was exposed to data, and parameters (weights and biases) have been adjusted to minimize error, using back-propagation training algorithm. The Input layer has 7 neurons, where each neuron represents reduced patient group. The number of neurons in the hidden layer is calculated based on the following equation :

$$N3 = ((2/3)*(N1))+N2$$

N1 represents number of nodes in the input layer; N2 represents number of nodes in the output layer; N3 represents number of nodes in the hidden layer.

### D. Support Vector Machines

SVM is a classification technique originated from statistical learning theory [5][31]. Depending on the chosen kernel, SVM selects a set of data examples (support vectors) that define the decision boundary between classes. SVM is known for excellent classification performance, though it is arguable whether support vectors could be effectively used in communication of medical knowledge to domain experts.

Standard formulation of support vector machines (SVMs) fails if data has missing values for any of the attributes. The present study examines methods by which data sets containing missing values can be processed using an SVM. This is typically accomplished by one of the two means namely ignoring missing data (either by discarding examples with a missing attribute value or discarding an attribute that has missing values), or using a process generally referred to as imputation through, by which a value is generated for the attribute. These techniques are typically carried out on data set prior to its being supplied to learning algorithm.

First SVM is trained to use all training examples that have no missing values[16]. Then ignoring original classification value from the data set, value of the attribute imputed is utilised as target value. It is to be noted that any other attribute that has missing value is ignored while generating this new training data set.

### E. LS-SVM

Least squares support vector machine (LS-SVM) [29] is a least squares version of support vector machine (SVM). In this technique estimated value of the missing value is obtained by solving a convex quadratic programming (QP) for classical SVMs. Least Squares SVMs (LS-SVMs) classifiers, in Suykens and Vandewalle. LS-SVM is a class of kernel based

learning methods. Primary goals of the LS-SVM models are regression and classification.

If the attribute has continuous values, LSSVM in regression mode is applied to study the data. If the attribute is discrete with only two values, standard LSSVM in classification mode is used. For discrete attribute with more than two values, special handling is required with the standard LSSVM technique of one-against- all. After an LSSVM is trained on each data set[18], then that model is used to classify or perform regression on examples of that attribute with missing values. If more than one LSSVM model generates a positive classification, selection is made on the basis of accuracy and sensitivity of the classifier.

### F. CZLSSVM imputation

In this study of automatic classification of dementia, filling missing values[12] is done through a combined approach to overcome overfitting of data. Several methods have reported in literature along with their own advantages and disadvantages. Our proposed method is a trial to give the best fit mechanism for filling in missing values in a patient database especially when data is collected over a period of time of several years along with several visits( a pool of cross section and time series data).

Data is clustered in two groups namely AD[15] and CN(Cognitively Normal). Z-score of the attribute MMSE is computed for each cluster in AD and CN. K-means clustering is an efficient algorithm applied in processing very large databases[6]. In a k-means cluster [2][7] constructed using similarity measure of MMSE, a missing value could be imputed based on (a) mean value of the corresponding attribute in other items contained in this cluster, or (b) similarity to nearest instance with a non-missing value (c) z-score of values in the cluster.

Steps :

1. Cluster the data sets based on MMSE in AD and CN groups using k-means algorithm
2. Find the mean and standard deviation for each cluster
3. Compute z-score for each cluster in each group

$$v' = \frac{v - \mu_A}{\sigma_A}$$

Where v' is the estimate of the missing value to be computed, v is observed value,  $\mu$  is mean and  $\sigma$  is the standard deviation of the cluster respectively.

4. Generate datasets with multiple imputation
5. Train LS-SVM with imputed values and check for classification accuracy
6. Evaluate the imputation strategy based on accuracy and sensitivity yielded by the classifier

Multiple imputation is done by LSSVM which is trained with various z-score values computed for each value of MMSE belonging to the demented group. Similarly, same procedure is repeated to find multiple values for missing attribute in non-demented group.

## VI. LSSVM – PSO CLASSIFIER

In standard SVMs and its reformulations, LS-SVM, regularization parameter and kernel parameters are called hyper-parameters, which play a crucial role to the performance of the SVMs. There exist different techniques for tuning the hyper-parameters related to regularization constant and parameter of kernel function. PSO (Particle Swarm Optimisation) is an evolutionary computation technique based on swarm intelligence [13]. It has many advantages over other heuristic techniques. This technique has an edge over distributed and parallel computing capabilities, escapes local optima and enables quick convergence.

LSSVM-PSO is trained and tested with multiple imputation values [23] in 5 different data sets A, B, C, D, E constructed from the existing data from ADNI and OASIS database. Four different models of the classifier are designed by varying the number of particles in PSO search that improves the quick convergence and classification. Models I, II, III and IV are evaluated based on their Sensitivity, Specificity and Accuracy to find the best-fit for the diagnosis of dementia.

### A. Optimization of LSSVM Parameters

In the case of LS-SVM with radial kernel function, optimized parameters are:  $\gamma$ , which is the weight at which testing errors are treated in relation to separation margin and parameter  $\sigma$ , which corresponds to width of the kernel function. It is unknown in advance what combination of these two parameters will achieve the best result of classification. In order to find the best values several techniques like Grid-Search, K-fold Cross-Validation, Particle Swarm Optimization have been in use. PSO provides better optimization than Grid-Search and K-fold method.

## VII. RESULTS

All classification results could have an error rate and on occasion will either fail to identify dementia or misclassify a normal patient as demented. It is common to describe this error rate by the terms true positive and false positive and true negative and false negative as follows:

*True Positive (TP)*: the classification result is positive in the presence of the clinical abnormality.

*True Negative (TN)*: the classification result is negative in the absence of the clinical abnormality.

*False Positive (FP)*: the classification result is positive in the absence of the clinical abnormality.

*False Negative (FN)*: the classification result is negative in the presence of the clinical abnormality.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) * 100\%$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) * 100\%$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100 \%$$

TP, TN, FP, FN, Sensitivity, Specificity and Accuracy are used to measure the performance of the classifiers. Experiments were carried out in MATLAB.

Imputation methods based on CZLSSVM-PSO method outperformed other imputation methods in the prediction of Dementia. Sensitivity and sensitivity analysis revealed a significant difference in percentage, error rate evaluation showed that the rate of error detected for CZLSSVM is significantly lower than KNN, BPN, C4.5 and SVM methods. Table 1 indicates the average error rate of imputation methods. Table 2 and 3 illustrate the accuracy of LSSVM-PSO classifier yielded by various imputation strategies in OASIS and ADNI databases respectively.

Table 4 and 5 depict that the overall performance of LSSVM-PSO classifier is high with the input of data imputed by the proposed CZLSSVM method compared to other methods. Out of the 4 models tested for classification as illustrated in Figure 1., Model 3 of LSSVM-PSO classifier is found to be very effective when combined with CZLSSVM method.

### Validation

A neural network model with 10 X 7 X 1 structure has been used in the present study to perform classification by setting aside 20% of the patterns (or observations) as validation (or testing) data. In this cross-validation approach, training is done repeatedly exposing the network to the remaining 80% of the patterns (training data) for several epochs, where an epoch is one complete cycle through the network for all cases. Data has been normalized before training. A network trained in this manner is considered generalizable, in the sense that it can be used to make estimate.

TABLE I. COMPARISON OF ERROR RATE OF IMPUTATION METHODS

| Imputation Methods | Average Error rate interval (OASIS) | Average Error rate interval (ADNI) |
|--------------------|-------------------------------------|------------------------------------|
| k-nn               | 2.7±0.22                            | 2.2±0.22                           |
| BPN                | 1.5±0.03                            | 1.2±0.13                           |
| C5.0               | 2.5±0.15                            | 1.5±0.25                           |
| SVM                | 0.5±0.04                            | 0.9±0.03                           |
| CZLSSVM            | 0.03±0.01                           | 0.23±0.11                          |

SOURCE : COMPUTED USING OASIS AND ADNI DATA  
NOTE : POOL OF CROSS-SECTION AND TIME SERIES DATA IS USED

TABLE II. CLASSIFICATION ACCURACY OF LSSVM-PSO FOR MULTIPLE IMPUTATION IN 5 DATASETS A, B, C, D, E SELECTED FROM OASIS DATABASE

| IMPUTATION METHODS | DATA SETS |         |         |         |         |
|--------------------|-----------|---------|---------|---------|---------|
|                    | A         | B       | C       | D       | E       |
| K-NN               | 75±0.4    | 78±0.7  | 80±0.2  | 79±0.33 | 76±0.08 |
| BPN                | 89±0.3    | 85±0.5  | 87±0.32 | 86±0.56 | 88±0.04 |
| C5.0               | 80±0.2    | 79±0.04 | 81±0.22 | 85±0.43 | 83±0.03 |
| SVM                | 89±0.7    | 90±0.25 | 91±0.4  | 89±0.28 | 85±0.06 |
| CZLSSVM            | 90±0.6    | 97±0.34 | 96±0.23 | 98±0.48 | 96±0.33 |

TABLE III. CLASSIFICATION ACCURACY OF LSSVM-PSO FOR MULTIPLE IMPUTATION IN 5 DATASETS A, B, C, D, E SELECTED FROM ADNI DATABASE

| IMPUTATION METHODS | DATA SETS |         |         |         |         |
|--------------------|-----------|---------|---------|---------|---------|
|                    | A         | B       | C       | D       | E       |
| K-NN               | 77±0.03   | 80±0.54 | 73±0.06 | 74±0.01 | 76±0.03 |
| BPN                | 85±0.21   | 87±0.43 | 83±0.04 | 88±0.03 | 89±0.01 |
| C4.5               | 80±0.11   | 81±0.29 | 83±0.4  | 84±0.03 | 84±0.05 |
| SVM                | 90±0.24   | 91±0.46 | 89±0.02 | 88±0.12 | 86±0.04 |
| CZLSSVM            | 95±0.2    | 97±0.02 | 98±0.31 | 95±0.04 | 98±0.14 |

Table IV. Comparison of Efficiency of LSSVM-PSO classifier for time series data set with multiple imputation strategies.

| MEASURE       | k-means | SVM | CZLSSVM | BPN | C4.5 |
|---------------|---------|-----|---------|-----|------|
| Sensitivity % | 89      | 92  | 95      | 89  | 88   |
| Accuracy %    | 90      | 90  | 96      | 90  | 89   |
| Specificity % | 89      | 9   | 97      | 90  | 85   |

TABLE V. COMPARISON OF EFFICIENCY OF LSSVM-PSO CLASSIFIER FOR CROSS-SECTION DATA SET WITH MULTIPLE IMPUTATION STRATEGIES

| MEASURE OF PERFORMANCE | IMPUTATION METHODS |     |         |     |      |
|------------------------|--------------------|-----|---------|-----|------|
|                        | k-means            | SVM | CZLSSVM | BPN | C4.5 |
| Sensitivity %          | 78                 | 94  | 94      | 89  | 85   |
| Accuracy %             | 79                 | 94  | 96      | 88  | 86   |
| Specificity %          | 80                 | 97  | 99      | 90  | 77   |

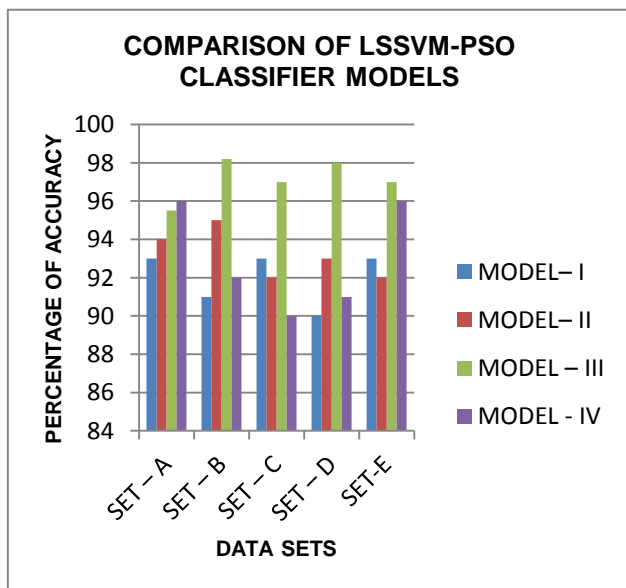


Fig. I. PERFORMANCE EVALUATION OF LSSVM-PSO CLASSIFIER MODELS

## VIII. CONCLUSION

Methods based on multiple imputation coupled with z-score and support vector machine classifier is found to be the most suited technique for imputation of missing values and led to a significant enhancement of prognosis accuracy compared to imputation methods based on k-NN, BPN, C4.5 and SVM procedures. Classification accuracy of LSSVM-PSO is very high when the missing values imputed by CZLSSVM are given as input as opposed to other methods in the diagnosis of dementia.

## ACKNOWLEDGEMENT

I would like to acknowledge the guidance, and expert opinion given by Dr. Sabesan, former Prof. and Head, Govt. Rajaji Hospital, Madurai and Dr.S.Vijayalakshmi M.Sc. Ph.D., former Professor and Head, School of Economics, Madurai Kamaraj University.

## REFERENCES

- [1] Anstey et al(2010). Estimates of probable dementia prevalence from population-based surveys compared with dementia prevalence estimates based on meta-analyses, BMC Neurology, 10:62.
- [2] Bankat M. Patil et al (2010) Missing value Imputation based on K-Means with Weighted Distance, Part I, CCIS 94, 600-609, Springer-Verlag Berlin Heidelberg (2010)
- [3] Ceyhan, E., Ceritoglu, C. et al. (2008) Analysis of metric distances and volumes of hippocampi indicates different morphometric changes over time in dementia of alzheimer type and nondemented subjects. *Technical Report*, Department of Mathematics, Koc University, Istanbul, Turkey..
- [4] Chi-Chun, H. and Hahn-Ming, L (2004) A Grey-Based Nearest Neighbor Approach for Missing Attribute Value Prediction. *Journal of Artificial Intelligence* 20,239-252.
- [5] Cristianini, N., Shawe-Taylor, J. (2000) An introduction to Support Vector Machines, Cambridge University Press, Cambridge.
- [6] Dan Li, Jitender Deogun, William Spaulding and Bill Shuart (2004) Towards Missing Data Imputation: A Study of Fuzzy K-means Clustering Method, ROUGH SETS AND CURRENT TRENDS IN COMPUTING , Lecture Notes in Computer Science, 3066/2004, , 573-579, DOI: 10.1007/978-3-540-25929-9\_70.
- [7] Fujikawa, Y., Ho, T.B. (2002) Cluster-based algorithms for dealing with missing values. In: Knowledge Discovery and Data Mining Conference, pp. 549--554. Springer, Berlin.
- [8] Harriet M M Smeding, Inge de Koning (2000), Frontotemporal dementia and neuropsychology:the value of missing values, *Journal of Neurol Neurosurg Psychiatry* 68, 726–730.
- [9] Hunt, L., Jorgensen, M. (2003) Mixture model clustering for mixed data with missing information. *Comput. Statist.Data Anal.* 41, 193–210.
- [10] ItoWasito, Boris Mirkin (2006) Nearest neighbours in least-squares data imputation algorithms with different missing patterns, *Computational Statistics & Data Analysis* 50, 926 – 949, doi:10.1016/j.csda.2004.11.009
- [11] Jose M. Jerez, Ignacio Molina, Pedro J. García-Laencina, Emilio Alba, Nuria Ribelles, Miguel Martín, Leonardo Franco (2010) Missing data imputation using statistical and machine learning methods in a real breast cancer problem , *Artificial Intelligence in Medicine* 50.
- [12] Kamakashi, L., Harp, S.A., Samad, T., Goldman, R.P. (1996) Imputation of missing data using machine learning techniques. In: Simoudis, E., Han, J., Fayyad, U. (Eds.), *Second International Conference on Knowledge, Discovery and Data Mining*. Oregon, 140–145.
- [13] Keneddy J. and Eberhart R. C. (1995) Particle swarm optimization, in *Proc. IEEE Int. Conf. Neural Networks*,, 1942–1948.
- [14] Kloppel, S., et al. (2008) Accuracy of dementia diagnosis -a direct comparison between radiologists and a computerized method, *Brain*, 131, 2969 -2974.
- [15] Little RJA.. (1995) Modeling the drop-out mechanism in longitudinal studies. *J. Am. Statist. Assoc.* 90: 1112–21

- [16]Mallinson, H. & Gammerman, A. (2003) Imputation Using Support Vector Machines <http://www.cs.york.ac.uk/euredit/>
- [17]Marcus, DS., Wang, TH., Parker J.M., Csernansky JG., Morris, JC., Buckner, RL. (2007) Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*, **19** 1498-1507.
- [18] Maytal Saar, Foster Provost (2007) Handling Missing Values when Applying Classification Models *Journal of Machine Learning Research* **8** 1625-1657
- [19] Qinbao Song ,Martin Shepperd, Xiangru Chen, Jun Liu (2008) Can k-NN imputation improve the performance of C4.5 with small software project data sets? A comparative evaluation *The Journal of Systems and Software* **81** 2361–2370.
- [20] Quinlan, J. R.: C4.5(1993) Programs for machine learning, Morgan Kaufmann, San Mateo, CA.
- [21][21] Robins, J. M., Rotnizky, A. & Zhao, L. P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**,106–121.
- [22] Rubin D.B (1987) Multiple imputation for nonresponse in surveys.. John Wiley and Sons.
- [23]Saar-Tsechansky M. and Provost F. (2007) Handling missing values when applying classification models. *Journal of Machine Learning Research*, **8**, 1625–1657.
- [24] Setiawan N.A, Venkatachalam P.A., Hani A.F.M. (2008) Missing Attribute Value Prediction Based on Artificial Neural Network and Rough Set Theory, ISBN: 978-0-7695-3118-2, International Conference on BioMedical Engineering and Informatics, DOI Bookmark: <http://doi.ieeecomputersociety.org/10.1109/BMEI.2008.322>.
- [25] Shichao Zhang (2011) Shell-neighbor method and its application in missing data imputation *Journal of Artificial Intelligence* **35**, 123-133.
- [26] Shichao Zhang , Jilian Zhang, et al (2008) Missing value imputation based on data clustering, *Transactions on Computer Science* 128-138.
- [27] Sivapriya T.R., Saravanan,V. and Ranjit Jeba Thangaiah, P. (2011) Texture Analysis of Brain MRI and Classification with BPN for the Diagnosis of Dementia, *Communications in Computer and Information Science*, Volume 204, Trends in Computer Science, Engineering and Information Technology, Part 1, 553-563.
- [28] Sivapriya, T.R., Saravanan, V. (2011) Dementia Diagnosis relying on Texture based features and SVM classification, *ICGST-AIML Journal*, **11**, 9-19.
- [29]Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B. , Vandewalle, J. ( 2002) Least Squares Support Vector Machines, World Scientific Publishing Company,.
- [30]Trosset,M., C. Priebe, Y. Park, and M. Miller. (2007) Semisupervised learning from dissimilarity data. *Technical Report* Department of Statistics, Indiana University, Bloomington, IN4705.
- [31]Vapnik, V. N. (1995) The Nature of Statistical Learning Theory. Springer-Verlag, New York,
- [32] Wang, Q. and Rao, J. N. K. (2002) Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.*, **30** 896-924.
- [33] Yongsong Qin, Shichao Zhang, Xiaofeng Zhu, Jilian Zhang and Chengqi Zhang (2007) Semi-parametric optimization for missing data imputation, *Journal of Artificial Intelligence* **27**, 79-88.
- [34] Zhang, C.Q., et al. (2007) An Imputation Method for Missing Values. *PAKDD, LNAI*, **4426**, 1080-1087.