Netw Model Anal Health Inform Bioinforma (2014) 3:58
DOI 10.1007/s13721-014-0058-x

ORIGINAL ARTICLE

Using MLP and SVM for predicting survival rate of oral cancer patients

Neha Sharma · Hari Om

Received: 25 January 2014 / Revised: 5 April 2014 / Accepted: 16 April 2014 / Published online: 6 May 2014

Abstract In this paper, we have attempted to build multilayer perceptron (MLP) and support vector machine (SVM) models for predicting survivability of the oral cancer patients who visit the ENT OPD. MLP and SVM have been applied in the past by few researchers for prediction of oral cancer using the genetic database. However, the database used for current research has the attributes like clinical symptoms, history of addiction, diagnosis, investigations, treatments and follow-up details which is gathered from presentations and review graphs related to oral malignancy from ENT and head and neck department. The MLP and SVM models are compared on the basis of various estimation criteria to identify the most effective model. Experimental result shows that accuracy of classification of SVM model is 73.56 %, whereas MLP model is 70.05 %; specificity of SVM model is 73.53 %, whereas MLP model is 65.36 %; and sensitivity of MLP model is 77.00 %, whereas SVM model is 73.56 %. SVM displays better results in terms of true negative, false negative, geometric mean of sensitivity and specificity, positive predictive value, geometric mean of positive predictive value and negative predictive value, precision, F-measure, area under receiver operating characteristics curve and lift and gain chart. Hence, it may be concluded that SVM is a most favourable model for predicting survival rate of oral cancer patients.

N. Sharma (✉)
Padmashree Dr. D.Y. Patil Institute of Master of Computer Applications, University of Pune, Akurdi, Pune 411044, Maharashtra, India
e-mail: nvsharma@rediffmail.com

H. Om
Computer Science and Engineering Department, Indian School of Mines, Dhanbad 826004, Jharkhand, India
e-mail: hari.om.cse@ismdhanbad.ac.in

Keywords Oral cancer · Data mining · Multilayer perceptron · Support vector machine · Classification · Early detection

1 Introduction

Data mining is the process of digging out significant, prospective and useful information or knowledge, from the huge data sets available in various forms at multiple locations. This technique presents the extracted information in a simplified manner with coherent structure that can be utilized to tackle classification, diagnosis or forecasting problems (Fayyad et al. 1996; Data Mining Curriculum 2006; Christopher 2010; Hastie et al. 2009). Traditionally, these issues were tackled by direct hands-on data analysis using standard statistical methods, but the continuously expanding volume of data has inspired the investigation and study of automatic data analysis using more complex and sophisticated tools which can operate directly on data. Therefore, the need gave origin to the concept called data mining, which identifies trends within data that go beyond simple analysis. By and large, the objective of the data mining process is to pull out information from a data set and represent it into a justifiable structure for future utilization (Data Mining Curriculum 2006). The data mining process also involves database and data management features, preprocessing of data, building a model and finding an inference, interestingness measurements, understanding the complexity, post-transforming of found structures, visualization and online updating (Data Mining Curriculum 2006). Association rules, decision trees, Gaussian mixture models, regression algorithms, neural networks (NN), support vector machines (SVM), Bayesian networks, etc., are few popular data mining technique used in many

domains to solve association, classification, segmentation, diagnosis and prediction problems (Han et al. 2011). Focus of the current research is prediction of the survival of oral cancer patients using data mining and therefore addressing the classification problem. This paper is organized as follows: the next section reviews various related literature; Sect. 3 highlights the material method; Sect. 4 covers the information about oral cancer; Sect. 5 gives the brief about multilayer perceptron (MLP) and SVM models; Sect. 6 presents the experimental results; whereas, Sect. 7 compares the model developed; Sect. 8 covers the conclusions and future work and finally acknowledgement is mentioned in the last section.

2 Literature survey

Several data mining techniques and algorithms were proposed for addressing classification problems after the Perceptron algorithm was proposed by Rosenblatt in 1957 (Rosenblatt 1957). Two most popular algorithms have been chosen for this research work. The first one is the multilayer perceptron (MLP) algorithm proposed by LeCun in 1987 and Rumelhart et al. in 1986, which became useful with the introduction of the back-propagation training (LeCun 1987; Rumelhart et al. 1986). The second one, is the SVM algorithm proposed by Vapnik in 1995 (Vapnik 1995) which supplied the classifier idea of large margin to improve the generalization performance for binary classification tasks. However, both MLP and SVM techniques have displayed excellent results in high dimensional classification problems (Lippmann 1987; Cristianini and Shawe-Taylor 2000). MLP networks and SVM are extensively used in various domains because they are very easy to train, very quick to use in classification decision process and usually accomplish good performances (Plamondon and Shirari 2000). The review of various literature indicates that data mining techniques like MLP and SVM has been used in the past by few researchers for prediction of oral cancer, but the data used was from genetic database (HariKumar et al. 2012; Chuang et al. 2011); whereas, the current research work has created a database by collecting data from variety of records available in the hospital.

In this section, the work done by various authors for early detection and prevention of oral cancer have been discussed. HariKumar et al. (2012), compares the classification accuracy of the TNM (tumour, lymph nodes, metastasis) staging system with that of Chi-square test and NN for breast cancer and oral cancer patients. Chi-square classification closely followed that of clinical investigation in comparison of TNM classification. Artificial neural networks (ANN) (MLP and RBF) are significantly more accurate than the TNM staging system when both use the TNM prognostic factors alone. Chuang et al. (2011) aimed at DNA repair genes. Single nucleotide polymorphisms (SNPs) dataset with 238 samples of oral cancer and control patients were chosen for disease prediction. All prediction experiments were conducted using the SVM. The result showed that the performances of the holdout cross-validation are superior to tenfold cross-validation, and the best classification accuracy is 64.2 %. Exarchos et al. (2012) used Dynamic Bayesian Networks to monitor the oral cancer evolvement and progression, as the network is capable to capture the temporal dimension of the disease and procure new informative biomarkers which correlate with the progression of the disease and identify early potential relapses. Gadewal and Zingde (2011) enabled fast retrieval of updated information by compiling the oral cancer gene database to include 374 genes by adding 132 gene entries.

Kent (1996) has used Genetic Algorithm to propose a diagnosis method for oral cancer. Kaladhar et al. (2011) used classification algorithms like CART, Random Forest, LMT, and Naïve Bayesian to predict oral cancer survivability. Milovic and Milovic (2012) present the applicability of data mining in health care and explain how these patterns can be used by physicians to determine diagnoses, prognoses and treatments for patients in healthcare organizations. Nahar et al. (2009) used three association rule mining algorithms, i.e. Apriori, Predictive apriori and Tertius algorithms, to extract the significant prevention factors for particular types of cancer. Shah and Kusiak (2007), facilitated proper treatment selection and drug development by analysing the gene expression data that leads to cancer identification and classification. Swami et al. (2011) collected the data from survey and formed the multidimensional association rule and its model of smoking habits, which is further used to take some preventive measures to reduce the smoking habit. Ha and Joo (2010) builds hybrid method combining association rule and classification trees which aims at helping physician to make fast and accurate classification of chest pain disease. RuthRamya et al. (2012) obtained the valuable rules and information in the case of chest pain by applying association rules into classification to improve the accuracy.

3 Material methods

The database for the current research work is created by collecting data through a retrospective chart review from ENT and head and neck department related to oral cancer, from the records of the cancer registries of tertiary care hospitals of western part of Indian subcontinent, from OPD data sheet and from archives of departments of histopathology, surgery and radiology were gathered. Clinical

details, personal history and habits were collected manually from the records to complete the datasheet of the patients. The data collection was done in non-randomized or non-probabilistic method, as all the data in the registries for the period of 5 years was considered. The dataset is based on the records of all the patients who have been reported with a lesion and treated at the centre from Jan 2004 and June 2009. The complete process of data preparation, data integration and data cleaning (i.e. removing missing values, noisy data and inconsistent data) to create the database of oral cancer patients has been presented in Sharma and Om (2012). The database has 33 variables and total of 1,024 records of patients that are used for the analysis. The data is used to develop data mining models and compare their performance for predicting the survival of oral cancer patients. The tool which is used for developing the predictive models is DTREG (http://www.dtreg.com) (pronounced D-T-Reg). It is a predictive modelling software that builds classification and regression decision trees, NN, SVM, GMDH polynomial networks, gene expression programs, K-means clustering, discriminant analysis and logistic regression models that can describe data relationships. The DTREG can be used to predict values for future observations and also has full support for time series analysis. It accepts a dataset in the form of table containing number of rows, whose columns represent attributes/variables. One of the variables is the "target variable" whose value is to be modelled and predicted as a function of the "predictor variables". The DTREG analyses the data and generates a model showing how best it predicts the values of target variable based on the values of predictor variables.

4 Oral cancer

Oral cancer, with its wide variable rate of incidences, has one of the highest incidences in the Indian subcontinent where it ranks among the top three types of cancer in the country (Elango et al. 2006). Age-adjusted rates of oral carcinoma in India is high, that is, 20 per 100,000 population and accounts for over 30 % of all cancers in the country (Sankaranarayanan et al. 2005). It is of major public health importance in India as it has been estimated that 83,000 new oral cancer cases occur here each year (Manoharan et al. 2010; Agrawal et al. 2012). However, early detection of oral cancer when lesion is less that 3 cm in size and shows no evidence of deep invasion or metastasis, dramatically increases the survival rates. Unfortunately, over the past decade there have been very little improvements in early diagnosis of oral cancer; more than 60 % of all oral cancers are extensive, late-stage malignancies at the time of diagnosis. Hence, treatment is more aggressive in trying to improve cure rates. But still it results in low treatment outcomes and considerable high cost to the patients who typically cannot afford this type of treatment (Khandekar et al. 2006). To achieve success in treatment, it is essential to determine the hidden patterns and trends in the oral disease data which can be collected from oral healthcare industry and subsequently "analysed or mined" to help healthcare practitioners for effective decision-making.

Oral cancer is a subtype of head and neck cancer, is any cancerous tissue growth located in the oral cavity (Werning 2007). There are several types of oral cancers, but around 90 % are squamous cell carcinomas (http://www.oralcancerfoundation.org/facts/index.htm), originating in the tissues that line the mouth and lips. No single factor causes an oral cancer. Oral malignancy is most likely caused by a combination of extrinsic and intrinsic factors acting in concert over a period of time. Historically, this has been attributed to widespread usage of smokeless tobacco, an established carcinogen. Well-supported by the evidence, great emphasis is being laid by the government and non-government organizations towards effective tobacco control. Over last several decades, billions of rupees have been spent to educate the public, implement laws effectively, rehabilitate tobacco growers, build cessation facility, create health infrastructure, etc., to reduce the smokeless tobacco usage. However, oral cancer that is an uncommon disease in the West, continues to be the major cause of cancer death in Indian men. About 2,000 deaths a day in India is tobacco related (http://www.oralcancerawareness.org).

The symptoms for an oral cancer at an earlier stage (Scully et al. 2008) are white or red patches inside the mouth or on lips, any sore which does heel for more than 14 days, bleeding in the mouth, pain when swallowing, and a lump in the neck. Treatments for oral cancers include surgery, radiation therapy and chemotherapy (http://www.yourtotalhealth.ivillage.com). But even that is not always successful as 70 % of the cases after treatment leads to relapse and the result is death and 5-year survival rate remains $\sim$40–50 % (Jemal et al. 2002; Woolgar et al. 1995). This study attempts to develop two data mining models and compare their efficiency for predicting survivability of oral cancer patients.

5 Data mining models

Data mining is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Multilayer perceptron and SVM are the models that are built and subsequently compared on the basis of experimental results and performances.

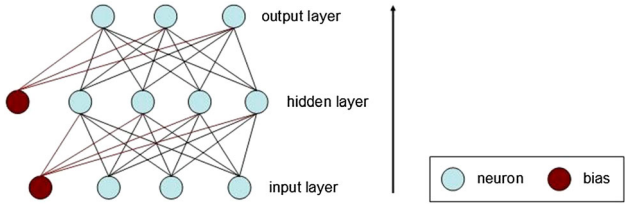

Fig. 1 Multilayer perceptron neural network

5.1 Multilayer perceptron model

The ANN is one of the most commonly used models based on human cognitive structure. Some different types of the ANN (multilayer perception, Radial Basis Function Neural Network and Kohonen's self-organizing map) are proposed to solve non-linear problem by learning. When used without qualification, the terms NN and ANN usually refer to a MLP Network. The diagram shown in Fig. 1 illustrates a perceptron network with three layers. This network has an input layer (at the bottom) with three neurons and a bias, one hidden layer (in the middle) with four neurons and a bias and an output layer (at the top) with three neurons. There is one neuron in the input layer for each predictor variable.

The network diagram shown above is a fully connected, three-layer, feed-forward, perceptron neural network. Fully connected network means that the output from each input and hidden neuron is distributed to all of the neurons in the following layer. Feed forward means that the values only move from input to hidden to output layers; no values are fed back to earlier layers. When there is more than one hidden layer, the output from one hidden layer is fed into the next hidden layer and separate weights are applied to the sum going into each layer.

5.2 Support vector machine model

An SVM performs classification by constructing an N-dimensional hyper-plane that optimally separates the data into two categories. SVM models are closely related to NN. In fact, an SVM model using a sigmoid kernel function is equivalent to a two-layer, feed-forward neural network. Using a kernel function, SVMs are an alternative training method for polynomial, radial basis function and multilayer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training. In the parlance of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used to define the hyper-plane is called a feature. The task of choosing the most suitable

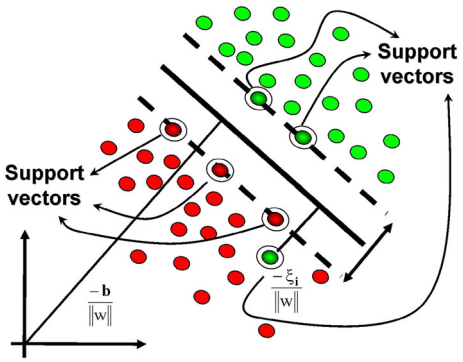

Fig. 2 Support vector machine

representation is known as feature selection. A set of features that describes one case (i.e. a row of predictor values) is called a vector. So the goal of SVM modelling is to find the optimal hyper-plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors near the hyper-plane are the support vectors and the same is shown in Fig. 2.

6 Experimental results

The database created is stored in comma separated values (csv) file format so that it can be used to build classification model using DTREG tool. All data rows were used for data subsetting. Number of data rows are 1,025 and total weight for all rows is 1,025. There are no rows with missing target or weight values. However, there is only one row with missing predictor value.

6.1 Building a model

There are two models that have been developed using classification technique and there are certain commonalities in both of them. Target variable for both the models is the attribute 'survival'. Classification technique is used for analysis, category weights are distributed over entire data file, misclassification costs are equal (unitary) and variable weights are also equal.

6.1.1 Multilayer perceptron neural network

Number of layers is three (input, hidden and output). Hidden layer and output layer activation function used in this model is logistic. Cross-validation method with ten-folds is used for validation, whereas network size evaluation is performed using fourfold cross-validation. The architecture of multilayer perceptron network is presented

Table 1 Architecture of multilayer perceptron network

Layer	Neurons	Activation	Min. weight	Max. weight
Input	48	Passthru	–	–
Hidden 1	3	Logistic	−1.277e+000	1.468e+000
Output	2	Logistic	−9.397e−001	1.157e+000

Table 2 Category weight of multilayer perceptron network

Category	Probability
Survival = D	0.4029268
Survival = A	0.5970732

Table 3 Training statistics of multilayer perceptron network

Process	Time	Evaluations	Error
Conjugate gradient	00:00:00.2	142,065	1.1888e−001

in Table 1. The category weights (prior probabilities) and training statistics of the network are given in Tables 2 and 3, respectively.

6.1.2 *Support vector machine*

Type of SVM model built is C-SVC. Radial Basis Function (RBF) is the SVM kernel function. Cross-validation method with tenfolds is used for validation. Search criterion used by the model is minimize total error. Total 144 points are evaluated during search. Minimum error found by search is 0.302439. SVM grid and pattern searches found optimal values for parameters: Epsilon = 0.001, $C = 40.7886087$ and Gamma = 0.01107173. Number of support vectors used by the model = 677.

6.2 Misclassification table

If the target variable is categorical then a misclassification summary table presents the number of rows with a particular category that were misclassified by the tree. There are two sections to the table—one section presents the actual data, whereas the other gives the information of misclassified data. Misclassification statistics for MLP and SVM for the training and validation data are shown in Tables 4, 5, 6 and 7. The tables also present the overall accuracy of the models.

6.3 Confusion matrix

Confusion matrix provides detailed information about how data rows are classified by the model. The matrix has a row and column for each category of the target variable. The categories shown in the first column are the actual categories of the target variable. The categories shown across the top of the table are the predicted categories. The numbers in the cells are the weights of the data rows with the actual category of the row and the predicted category of the column. The numbers in the diagonal cells are the weights for the correctly classified cases where the actual category matches the predicted category. The off-diagonal cells have misclassified row weights. Confusion matrix for training and validation data for MLP model and SVM is shown in Tables 8 and 9.

Table 4 Misclassification for MLP (training data)

Category	Actual		Misclassified			
	Count	Weight	Count	Weight	%	Cost
A	612	612	212	212	34.641	0.346
D	413	413	95	95	23.002	0.230
Total	1,025	1,025	307	307	29.951	0.300
Overall accuracy = 70.05 %						

Table 5 Misclassification for MLP (validation data)

Category	Actual		Misclassified			
	Count	Weight	Count	Weight	%	Cost
A	612	612	236	236	38.562	0.386
D	413	413	74	74	17.918	0.179
Total	1,025	1,025	310	310	30.244	0.302
Overall accuracy = 69.76 %						

Table 6 Misclassification for SVM (training data)

Category	Actual		Misclassified			
	Count	Weight	Count	Weight	%	Cost
A	612	612	162	162	26.471	0.265
D	413	413	109	109	26.392	0.264
Total	1,025	1,025	271	271	26.439	0.264
Overall accuracy = 73.56 %						

Table 7 Misclassification for SVM (validation data)

Category	Actual		Misclassified			
	Count	Weight	Count	Weight	%	Cost
A	612	612	162	162	26.471	0.265
D	413	413	109	109	26.392	0.264
Total	1,025	1,025	271	271	26.439	0.264
Overall accuracy = 69.27 %						

6.4 Sensitivity and specificity

The sensitivity and specificity report is generated only for classification problems (categorical target variable). One

Table 8 Confusion matrix (MLP model)

Actual category	Predicted category			
	Testing data		Validation data	
	A	D	A	D
A	400	212	376	236
D	95	318	74	339

Table 9 Confusion matrix (SVM model)

Actual category	Predicted category			
	Testing data		Validation data	
	A	D	A	D
A	450	162	431	181
D	109	304	134	279

category of the target variable is called the positive category, and the other is called the negative category. True positive (TP) means patients who are predicted as malignant among malignant patients. True negative (TN) means patients who are predicted as non-malignant among non-malignant patients. False positive (FP) means patients who are predicted as non-malignant among malignant patients. False negative (FN) means patients who are predicted as malignant among non-malignant patients. The sensitivity and specificity were calculated by TP, TN, FP and FN. Sensitivity means probability that the algorithms can correctly predict malignancy. Specificity means probability that the algorithms can correctly predict non-malignant. Sensitivity = TP/(TP + FN) and specificity = TN/(FP + TN). Survival = D is considered as a positive and survival = A is considered as negative for both the models developed. The positive/negative ratio, TP, TN, FP, FN, sensitivity, specificity, geometric mean of sensitivity and specificity, positive predictive value (PPV), negative predictive value (NPV), geometric mean of PPV and NPV, precision, recall, F-measure and area under receiver operating characteristics (ROC) curve for training and validation data for both the models is shown in Tables 10 and 11. Area under ROC curve for both the models is shown in Figs. 3 and 4.

6.5 Probability calibration

The probability calibration report shows how the predicted probability of a target category is distributed and provides a means for gauging the accuracy of predicted probabilities. The probability calibration report is generated only when a classification analysis is performed and there are two target categories [for e.g. survival = dead (D) or alive (A)]. If the model is accurate, the predicted probability of an event occurring should match the actual proportion of times that the event occurs.

Table 10 Sensitivity and specificity of MLP model

	Training data	Validation data
Positive/negative ratio	0.6748	0.6748
True positive (TP)	318 (31.02 %)	339 (33.07 %)
True negative (TN)	400 (39.02 %)	376 (36.68 %)
False positive (FP)	212 (20.68 %)	236 (23.02 %)
False negative (FN)	95 (9.27 %)	74 (7.22 %)
Sensitivity	77.00 %	82.08 %
Specificity	65.36 %	61.44 %
Geometric mean of sensitivity and specificity	70.94 %	71.01 %
Positive predictive value (PPV)	60.00 %	58.96 %
Negative predictive value (NPV)	80.81 %	83.56 %
Geometric mean of PPV and NPV	69.63 %	70.19 %
Precision	60.00 %	58.96 %
Recall	77.00 %	82.08 %
F-measure	0.6744	0.6862
Area under ROC curve	0.769	0.739

Table 11 Sensitivity and specificity of SVM model

	Training data	Validation data
Positive/negative ratio	0.6748	0.6748
True positive (TP)	304 (29.66 %)	279 (27.22 %)
True negative (TN)	450 (43.90 %)	431(42.05 %)
False positive (FP)	162 (15.80 %)	181 (17.66 %)
False negative (FN)	109 (10.63 %)	134 (13.07 %)
Sensitivity	73.61 %	67.55 %
Specificity	73.53 %	70.42 %
Geometric mean of sensitivity and specificity	73.57 %	68.97 %
Positive predictive value (PPV)	65.24 %	60.65 %
Negative predictive value (NPV)	80.50 %	76.28 %
Geometric mean of PPV and NPV	72.47 %	68.02 %
Precision	65.24 %	60.65 %
Recall	73.61 %	67.55 %
F-measure	0.6917	0.6392
Area under ROC curve	0.815	0.720

6.5.1 Multilayer perceptron

Probability calibration for survival = D and survival = A is same and is presented as follows:

Average weighted probability error for training data = 0.040631.

Average weighted squared probability error for training data = 0.050201.

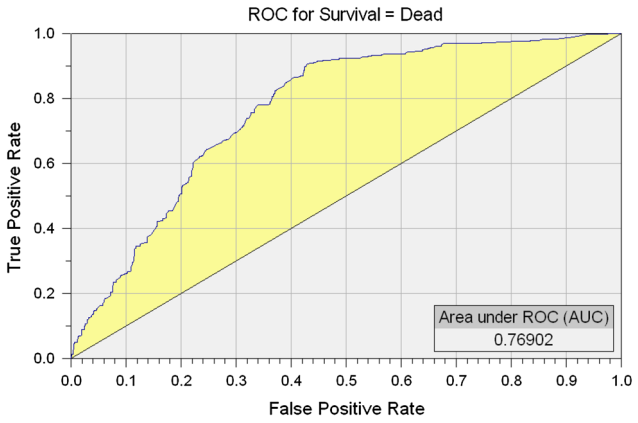

Fig. 3 ROC for multilayer perceptron model

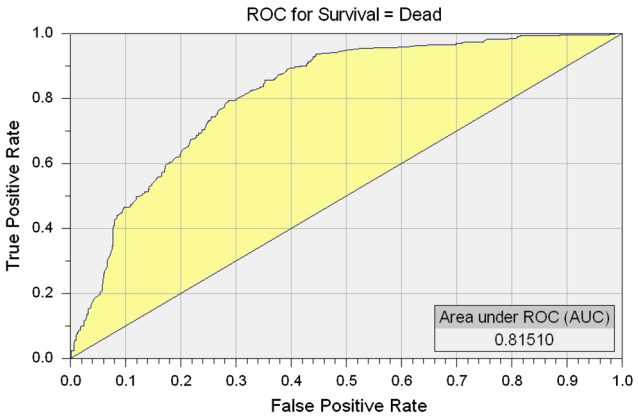

Fig. 4 ROC for support vector machine model

Average weighted probability error for validation data = 0.058048.

Average weighted squared probability error for validation data = 0.062428.

6.5.2 *Support vector machine*

Probability calibration for survival = D and survival = A is same and is presented as follows:

Average weighted probability error for training data = 0.073024.

Average weighted squared probability error for training data = 0.076266.

Average weighted probability error for validation data = 0.050982.

Average weighted squared probability error for validation data = 0.068804.

6.6 Probability threshold

The probability threshold report provides information about how different probability thresholds would affect target category assignments. The threshold report provides a convenient way to see the tradeoff between impurity and loss as the probability threshold is varied. The probability threshold report is generated only when a classification analysis is performed and there are two target categories. Probability threshold analysis is carried out for survival = D for both the models.

6.6.1 *Multilayer perceptron*

Area under ROC curve (AUC) for training data = 0.751660.

Threshold to minimize misclassification for training data = 0.459764.

Threshold to minimize weighted misclassification for training data = 0.459764.

Threshold to balance misclassifications for training data = 0.516274.

Area under ROC curve (AUC) for test data = 0.733567.

Threshold to minimize misclassification for test data = 0.493657.

Threshold to minimize weighted misclassification for test data = 0.493657.

Threshold to balance misclassifications for test data = 0.544517.

6.6.2 *Support vector machine*

Area under ROC curve (AUC) for training data = 0.829547.

Threshold to minimize misclassification for training data = 0.623342.

Threshold to minimize weighted misclassification for training data = 0.623342.

Threshold to balance misclassifications for training data = 0.490359.

Area under ROC curve (AUC) for test data = 0.744871.

Threshold to minimize misclassification for test data = 0.293013.

Threshold to minimize weighted misclassification for test data = 0.293013.

Threshold to balance misclassifications for test data = 0.473142.

6.7 Lift and gain

The lift and gain table is a useful tool for measuring the value of a predictive model. Lift and gain values are especially useful when a model is being used to target (prioritize) marketing efforts. The basic idea of lift and gain is to sort the predicted target values in decreasing order of purity on some target category and then compare the proportion of cases with the category in each bin with the overall proportion. In the case of a model with a

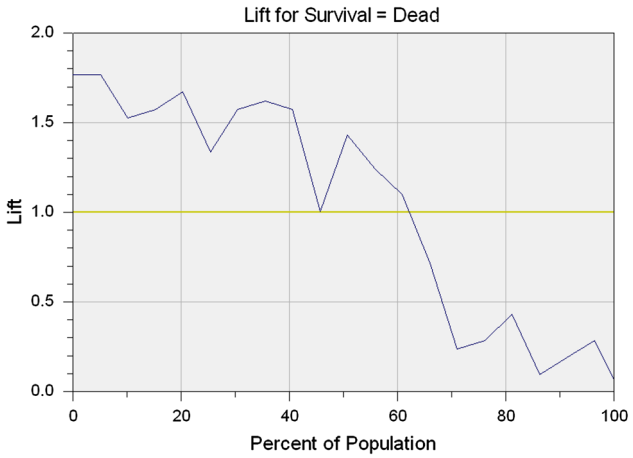

Fig. 5 Lift for multilayer perceptron model

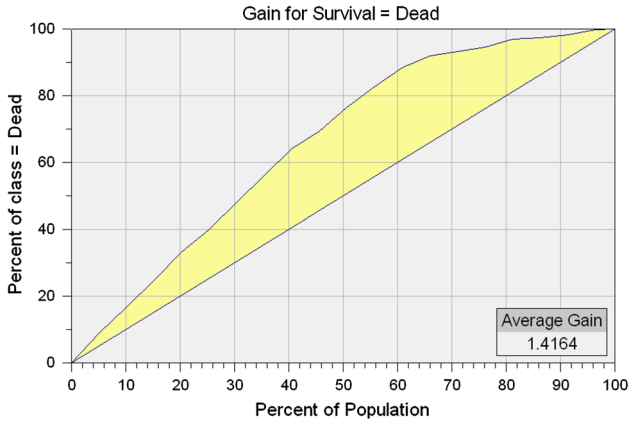

Fig. 6 Gain for multilayer perceptron model

continuous target variable, the predicted target values are sorted in decreasing target value order and then compared with the mean target value. The lift and gain values show how much improvement the model provides in picking out the best 10, 20 %, etc., of the cases.

6.7.1 Multilayer perceptron

For training data, average gain for survival = A is 1.278 and for survival = D is 1.362. Percent of cases with survival = A is 59.71 % and with survival = D is 40.29 %. For validation data, average gain for survival = A is 1.270 and for survival = D is 1.308. Percent of cases with survival = A is 59.71 % and with survival = D is 40.29 %. Lift and gain for survival = D is shown in Figs. 5 and 6.

6.7.2 Support vector machine

For training data, average gain for survival = A is 1.350 and for survival = D is 1.528. Percent of cases with survival = A is 59.71 % and with survival = D is 40.29 %.

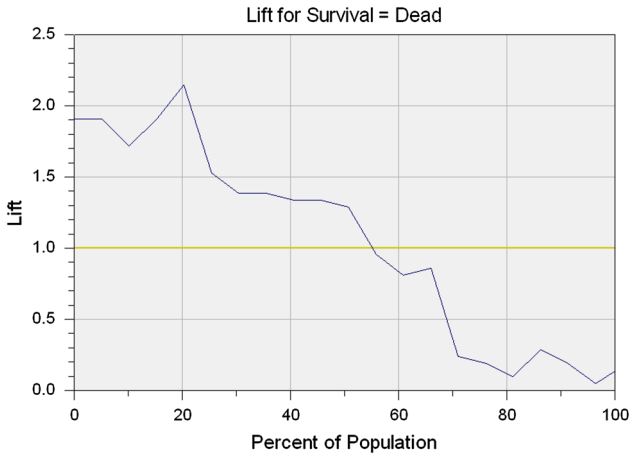

Fig. 7 Lift for support vector machine model

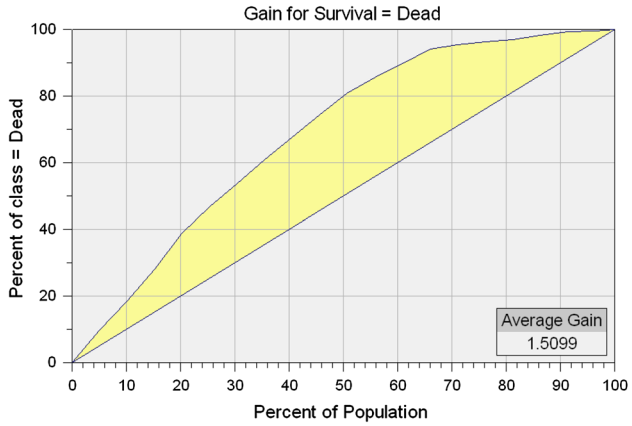

Fig. 8 Gain for support vector machine model

For validation data, average gain for survival = A is 1.272 and for survival = D is 1.360. Percent of cases with survival = A is 59.71 % and with survival = D is 40.29 %. %. Lift and gain for survival = D is shown in Figs. 7 and 8.

6.8 Analysis run time

Multilayer perceptron model took 00:06.76 and SVM model took 02:01.66 time to run the analysis.

7 Comparing the models

The attempt has been made to compare both the models on the basis of various criteria to assess the performance. MLP uses logistic as activation function at hidden and output layer where as radial basis function (RBF) is the SVM kernel function. SVM displays better results in terms of accuracy, true negative, false negative, specificity, geometric mean of sensitivity and specificity, positive

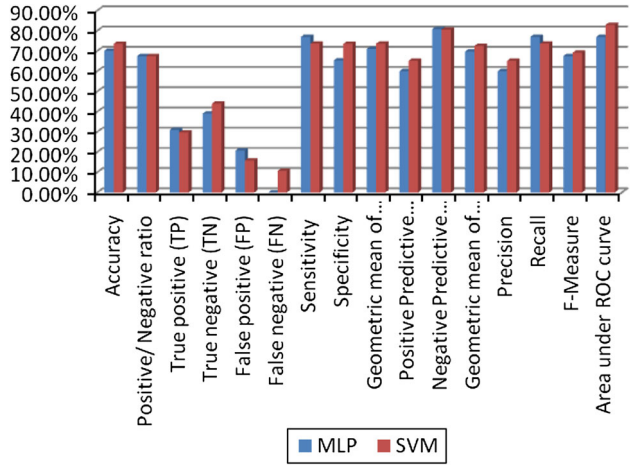

Fig. 9 Comparison of MLP and SVM on various criteria for training data

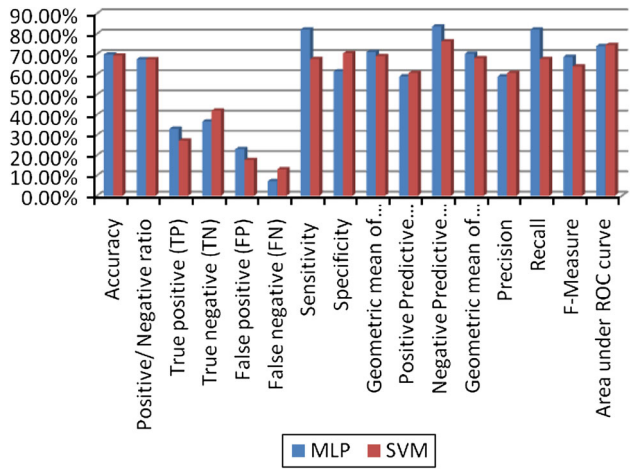

Fig. 10 Comparison of MLP and SVM on various criteria for validation data

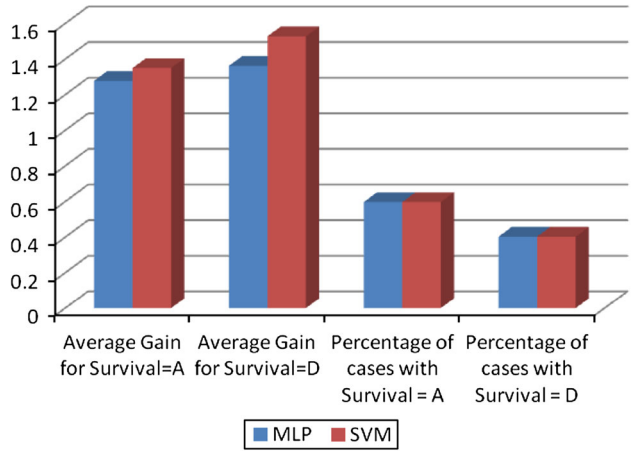

Fig. 11 Lift and gain of MLP and SVM for training data

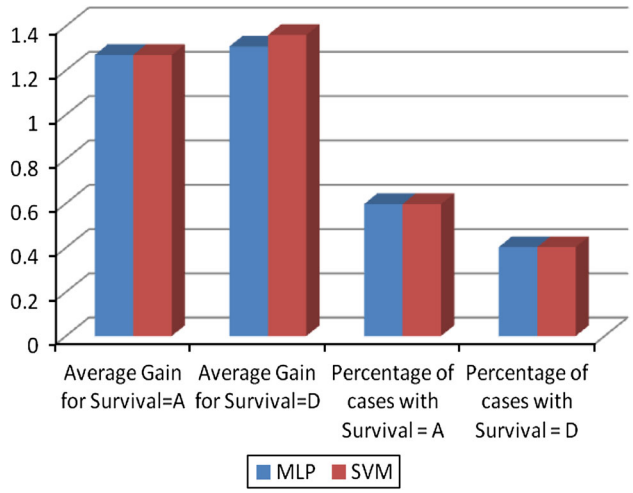

Fig. 12 Lift and gain of MLP and SVM for validation data

predictive value, geometric mean of PPV and NPV, precision, F-measure and area under ROC curve in comparison to MLP. Comparison of MLP and SVM models on various criteria is shown in Figs. 9 and 10 for training data and validation data.

Similarly, lift and gain chart of MLP and SVM shows that SVM model is better that MLP. Comparison is shown in Figs. 11 and 12 for training data and validation data of MLP and SVM.

8 Conclusion and future work

Support vector machine has certainly shown better results in terms of accuracy, true negative, false negative, specificity, geometric mean of sensitivity and specificity, positive predictive value, geometric mean of PPV and NPV, precision, F-measure and area under ROC curve in comparison to MLP. However, MLP displays better results in terms of true positive, false positive, negative predictive value, sensitivity and recall. Lift and gain chart also suggest SVM as a better predictive model. Our future work shall include finding out association between various fields.

Acknowledgments The authors devote their sincere thanks to the management and staff of Indian School of Mines, for their constant support and motivation.

References

Agrawal M, Pandey S, Jain S, Maitin S (2012) Oral cancer awareness of the general public in Gorakhpur City, India. Asian Pac J Cancer Prev 13:5195–5199

Christopher C (2010) Encyclopaedia Britannica: definition of data mining

Chuang LY, Wu KC, Chang HW, Yang CH (2011). Support vector machine-based prediction for oral cancer using four snps in DNA repair genes. In: Proceedings of the International MultiConference of Engineers and Computer scientists. March 16–18

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines. Cambridge University Press, Cambridge

Data Mining Curriculum (2006) ACM SIGKDD

Elango JK, Gangadharan P, Sumithra S, Kuriakose MA (2006) Trends of head and neck cancers in urban and rural India. Asian Pac J Cancer Prev 07(01):108–112

Exarchos KP, Rigas G, Goletsis Y, Fotiadis DI (2012) Modelling of oral cancer progression using dynamic bayesian networks, data mining for biomarker discovery, Springer optimization and its applications, pp 199–221

Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery: an overview. Advances in knowledge discovery and data mining (AAAI Press/MIT Press), p 1–36

Gadewal NS, Zingde SM (2011) Database and interaction network of genes involved in oral cancer: version II. Bioinformation 06(04):169–170

Ha SH, Joo SH (2010) A hybrid data mining method for medical classification of chest pain. World Acad Sci Eng Technol 37:608–613

Han J, Kamber M, Pei J (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers, San Francisco. ISBN 978-0123814791

HariKumar R, Vasanthi NS, Balasubramani M (2012) Performance analysis of artificial neural networks and statistical methods in classification of oral and breast cancer stages. Int J Soft Comput Eng (IJSCE) 2(3):263–269

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. 2nd edn. Springer, Vienna

Jemal A, Thimas A, Murray T, Thun M (2002) Cancer statistics. CA Cancer J Clin 52:181–182

Kaladhar DSVGK, Chandana B, Bharath Kumar P (2011) Predicting cancer survivability using Classification algorithms. Int J Res Rev Comput Sci (IJRRCS) 2(2):340–343

Kent S (1996) Diagnosis of oral cancer using genetic programming—a technical report. CSTR-96-14

Khandekar PS, Bagdey PS, Tiwari RR (2006) Oral cancer and some epidemiological factors: a hospital based study. Indian J Community Med 31(03):157–159

LeCun Y (1987) Modeles connexionnistes de l'apprentissage (connectionist learning models). Doctoral dissertation, Université P. et M. Curie (Paris 6)

Lippmann RP (1987) An introduction to computing with neural nets. IEEE ASSP Mag 4:4–28

Manoharan N, Tyagi BB, Raina V (2010) Cancer incidences in rural Delhi—2004–05. Asian Pac J Cancer Prev 11(01):73–78

Milovic B, Milovic M (2012) Prediction and decision making in health care using data mining. Int J Public Health Sci 01(02):69–78

Nahar J, Kevin ST, Ali ABMS, Chen YP (2009) Significant cancer prevention factor extraction: an association rule discovery approach. J Med Syst. doi: 10.1007/s10916-009-9372-8

Plamondon R, Shirari SN (2000) On-line and off-line handwriting recognition: a comprehensive survey. IEEE Trans Pattern Anal Mach Intell 22:63–84

Rosenblatt F (1957) The perceptron—a perceiving and recognizing automaton (Technical Report 85-460-1). Cornell Aeronautical Laboratory

Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by backpropagating errors. In Rumelhart D and Mc-Clelland J (eds), Parallel distributed processing: explorations in the microstructure of cognition, vol 1. MIT Press, Cambridge, pp 318–362

RuthRamya K, Anusha K, Chanti K, Vidya VS, Kumar PP (2012) A class based approach for medical classification of chest pain. Int J Eng Trends Technol 3(2):89–93

Sankaranarayanan R, Ramadas K, Thomas K (2005) Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. Lancet 365(9475):1927–1933

Scully C, Bagan JV, Hopper C, Epstein JB (2008) Oral cancer: current and future diagnostics techniques—a review article. Am J Dent 21(04):199–209

Shah S, Kusiak A (2007) Cancer gene search with data-mining and genetic algorithms. Comput Biol Med 37:251–261 http://www.intl.elsevierhealth.com/journals/cobm

Sharma N, Om H (2012) Framework for early detection and prevention of oral cancer using data mining. Int J Adv Eng Technol 4(2):302–310

Swami S, Thakur RS, Chandel RS (2011) Multi-dimensional association rules extraction in smoking habits database. Int J Adv Netw Appl 03(03):1176–1179

Vapnik V (1995) The nature of statistical learning theory. Second edition. Springer, Vienna

Werning JW (2007) Oral cancer: diagnosis, management, and rehabilitation. p 1. ISBN 978-1588903099

Woolgar JA, Scott J, Vaughan ED, Brown JS, West CR, Rogers S (1995) Survival, metastasis and recurrence of oral cancer in relation to pathological features. Ann R Coll Surg Engl 1995(77):325–331