Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 6 (2011) 279–284

Procedia Computer Science

Complex Adaptive Systems, Volume 1
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2011- Chicago, IL

A complex adaptive system using statistical learning theory as an inline preprocess for clinical survival analysis

Dan Margolis [a], Walker H. Land Jr. [a], Ron Gottlieb [b], Xingye Qiao [c]

[a]*Department of Bioengineering, Binghamton University, Binghamton, NY 13902-6000, USA.*
[b]*Department of Radiology, University of Arizona, Tuscan, AZ 85724, USA.*
[c]*Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA.*

Abstract

New advances in medicine have led to a disparity between the existing information about patients and the ability of clinicians to utilize it. Lack of training and incompatibility with clinical techniques has made the use of the complex adaptive systems approach difficult. To avoid this, we used statistical learning theory as an inline preprocess between existing data collection methods and clinical analysis of data. Clinicians would be able to use this system without any changes to their techniques, while improving accuracy. We used data from CT scans of patients with metastatic carcinoma to predict prognosis. Specifically, we used the standard for evaluating response to treatment, RECIST, and new qualitative and quantitative features. An Evolutionary Programming trained Support Vector Machine (EP-SVM), was used to preprocess the data for two traditional survival analysis techniques: Cox Proportional Hazard Models and Kaplan Meier curves. This was compared to Logistic Regression (LR) and using cutoff points. Analyses were also done to compare different inputs and different radiologists. The EP-SVM outperformed both LR and the cutoff method significantly and allowed us to both intelligently combine data from multiple sources and identify the most predictive features without necessitating changes in clinical methods.

"Keywords: SVM; Survival Analysis; Statistical Learning Theory; Radiology; RECIST; Cox Proportional Hazard Model;"

1. Introduction

The increasing pace of new advances in medical technologies and clinical data collection has led to a large disparity between the potential information about patients that exists and the ability of clinicians to understand and utilize this information (Marsland and Buchan, 2004). International standards have been even slower to change and often lack serious scientific evaluation of their efficacy (Michaelis and Ratain, 2006). Unfortunately, major barriers to introducing Complex Adaptive Systems (CAS) for clinical use, such as a lack of clinician training and an incompatibility with current clinical techniques, have led to great divide between the advances made by researchers and those advances being used to help real patients (Ludwig, 2005). When those difficulties have been overcome, as with usage of computer aided diagnostics for the early diagnosis of breast cancer, a significant improvement has occurred (Freer and Ulissey, 2001). Other advances in CAS research have led to the development of techniques that could be used at almost every stage of health care, such as diagnosis, prognosis, treatment selection, and surgery assistance, but have yet to be implemented in the clinic

1877–0509 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.procs.2011.08.052

(Sim *et al.*, 2001). Furthermore, CAS could be used to help overcome issues such as observer variability and medical errors which have plagued medicine and had few solutions (Brennan and Silman, 1992). Thus, our goal was not to develop a system to replace currently used technology in the clinic, but to insert new technology into the existing system in such a way as to avoid many of the aforementioned barriers to using these systems clinically. We tested using Statistical Learning Theory (SLT) as an inline preprocess between standard clinical data collection methods and traditional clinical analysis of this data. Thus, our CAS takes in the same inputs and gives the same outputs as clinicians are used to and trained for, seamlessly. Clinicians would be able to use this CAS without any current changes to their training or techniques, but would also have the benefit of additional knowledge gained from SLT and more accurate performance that comes from powerful, non-linear complex adaptive systems analysis.

2. Methods

When evaluating a patient, standard features such as Response Evaluation Criteria in Solid Tumors (RECIST), the internationally accepted standard for radiologists for prognosticating cancer, are used to categorize patients in good and bad prognosis groups (Eisenhauer *et al.*, 2009). This preprocess therefore creates a feature that is then used as an input to the two accepted standards for survival analysis, Cox proportional hazard models and Kaplan-Meier curves. These determine the risk of being in the poor prognosis group compared to the good prognosis group and provide a survival curve for each group. Our hypothesis was that replacing this feature with one generated from an SLT process will provide better categorization of patients. To test this hypothesis, we used data collected from 65 patients with metastatic melanoma to perform a prognostic analysis of survival. The features used to generate RECIST were used as one set of inputs for the SLT process. Other sets of inputs included the serum Lactic Dehydrogenase Level (LDH; baseline and follow-up), which is being looked at strongly in clinics as a good predictor of prognosis, and Visual Based Scoring (VBS), a simple yet non-traditional set of features attempting to harness the "gut instinct" of radiologists evaluations of visual net size change between baseline and follow-up CT scans. The dataset was collected by a board certified radiologist and included overall survival and progression free survival as outcomes. To create a gold standard to train the SLT technique, the data was split and placed into good prognosis (survival above median) and poor prognosis (survival below median) categories. The median for overall survival was 7.0 months and the median for progression free survival was 3.5 months. Three methods for preprocessing the data into a single categorization feature were chosen: (1) a traditional cutoff and/or rule method (such as RECIST), (2) Logistic Regression, and (3) an SLT method, EP-SVM.

2.1 Cox Proportional Hazard Models and Kaplan-Meier Curves

The two gold standards in survival analysis which clinicians understand, use, and are trained for are Cox Proportional Hazard (Cox PH) models and Kaplan-Meier (K-M) curves. The Cox PH model is a semi-parametric *linear* regression model which looks at the "hazard" or risk of an event. It assumes that the hazard of an observation is proportional to an unknown "baseline" hazard common to all observations, where this proportionality is modeled as an exponential of a linear function of the covariates. From this model, a single value is created called a Cox Hazard Ratio (CHR), which supposedly represents the hazard or risk of the event occurring over time between groups. The larger the CHR, the greater the risk of belonging to one group compared to the other (Cox, 1972). K-M curves measure the fraction of patients surviving at certain time intervals and give a plot of those points. A chi-squared test can then be performed on the K-M curves of two groups to determine if a significant difference in survival exists between them (Kaplan and Meier, 1958). Cox PH and K-M curves were used to evaluate the performance of the three preprocess techniques.

2.2 Traditional Method

Traditional method refers to a method that would be used in normal clinical practice for creating a feature categorizing patients. The most common method used by radiologists is RECIST, which compares the summed lengths of (up to) the five largest lesions taken from a baseline CT scan to the sum taken from a follow-up CT scan a couple months later. That percentage, specifically >20% increase, or the existence of any new lesions is considered progressive disease, while the percentage being <20% is considered stable disease or partial

recovery (Eisenhauer *et al.*, 2009). Thus, the method is basically a "cutoff" method, which uses a certain point (20%) in an original feature to create a categorization feature used as an input for Cox PH models and K-M curves. RECIST combines this cutoff method with a "rule" method, where the presence of a new lesion automatically causes the resultant categorization feature to become progressive disease. For LDH, a point of a baseline LDH of 250 IU/L was used as a cutoff, while VBS just used a score of 1 as the cutoff along with the same rule regarding new lesions as RECIST. All cutoff points rounded upwards.

2.3 Logistic Regression

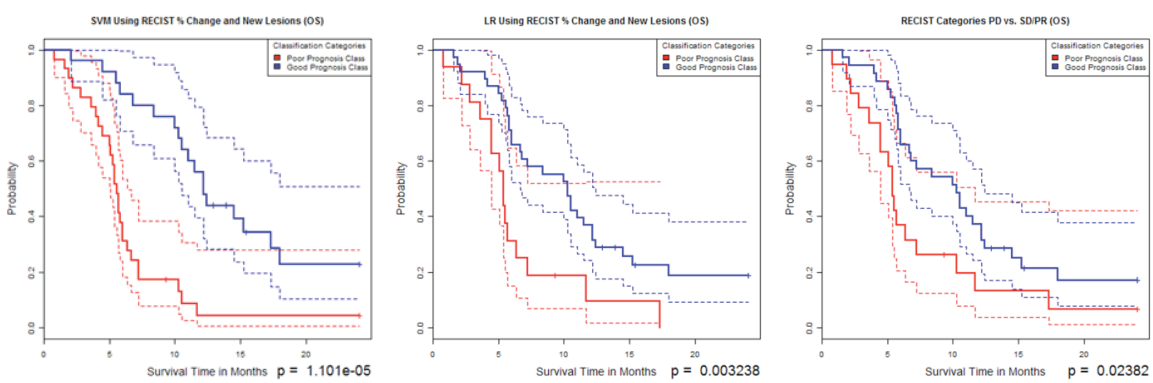
Logistic Regression (LR; Breslow and Day, 1980), a more statistically advanced linear technique that is currently used and accepted in clinics, was used in this paper. LR is a generalized linear model used for binary response variable, and is considered a staple of traditional statistical analysis. In particular, it models the probability of occurrence of certain event by a linear form of the covariates, connected through the logistic function. Maximum likelihood is usually employed to estimate the regression coefficients. The accuracy of LR was measured using the area under a Receiver Operating Characteristic curve (AUC) generated by one-hold-out cross validation.

2.4 Statistical Learning Theory

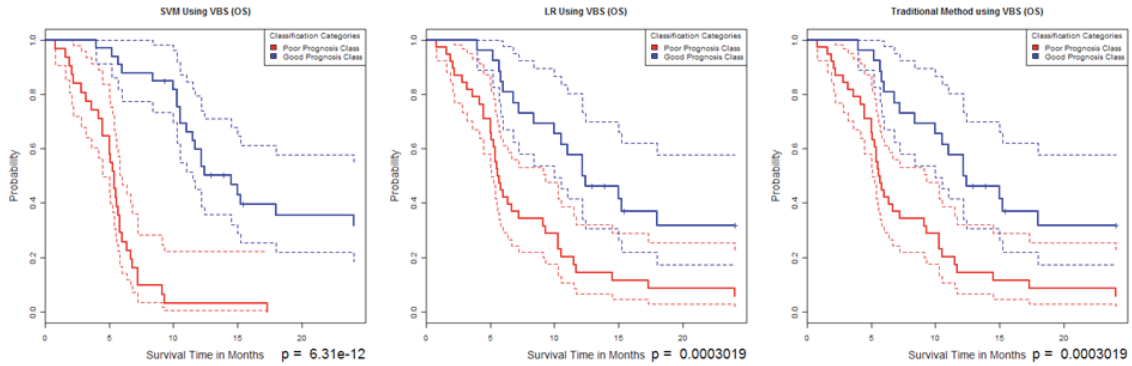
The SLT method chosen was an Evolutionary Programming / Evolutionary Strategies trained Support Vector Machine (EP-SVM) hybrid. The EP-SVM is a classification technique that uses a stochastic Evolutionary Programming/Evolutionary Strategies (EP) process to determine the best kernel and kernel parameters for use with a well-established non-linear binary classifier, a Support Vector Machine (SVM) (Vapnik, 1995). This technique was developed for and has been successfully tested for biomedical applications such as breast cancer diagnosis (Land *et al.*, 2003) and colorectal cancer prognosis (Land *et al.*, 2010). Similar to LR, the accuracy of the EP-SVM can be measured with an AUC generated by one-hold-out cross validation. The LibSVM package was used for the Support Vector Machine backend (Chang and Lin, 2001).

3. Results

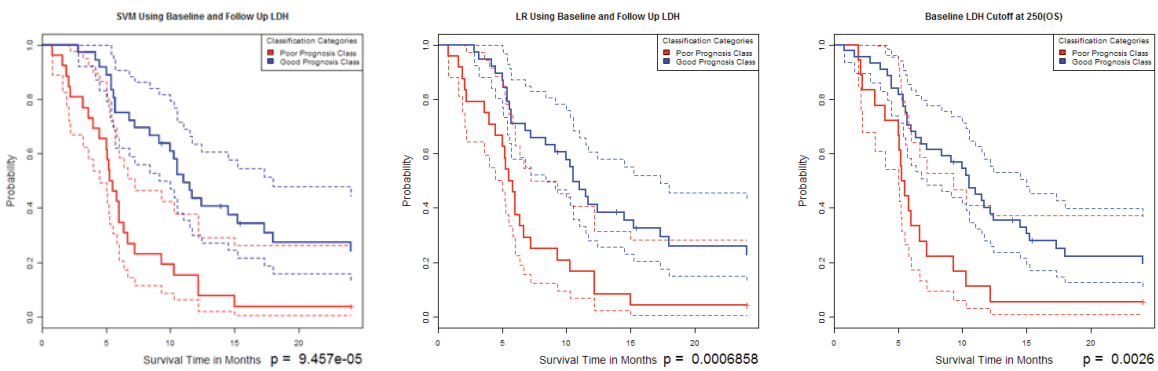
The EP-SVM outperformed LR, which in turn outperformed or equaled the traditional method for all experiments. Figures 1-3 show the Kaplan-Meier curves and associated chi-squared derived *p*-values for each of the three methods when using RECIST inputs, which consisted of the percent change in summed lesion lengths and the existence of new lesions. Figures 4-6 show the same curves but for VBS inputs, which consisted of a Likert-scale value for visual net size change and new lesions. Finally, Figures 7-9 show the curves based on LDH inputs, which consists of both baseline and follow-up LDH values as inputs, though only the baseline level for the traditional cutoff method. All of these figures utilize overall survival as the outcome, since the short median time for progression free survival made it difficult to visualize those curves. All censored survival times below the median were removed, and survival times above 24 months were truncated to make the graphs easier to view. The 95% confidence intervals are shown as dashed lines around each curve.



Figures 1-3: Kaplan-Meier curves for SVM, LR, and Traditional Method (RECIST) using Overall Survival as the outcome. The survival curve for the good prognosis group is blue, for poor prognosis group is red, and the chi-squared test based *p*-value showing the significance of the difference in the curves is in the bottom right corner. SVM creates a feature that can categorize patients into groups far better than LR; both of which are better than traditional RECIST categorization.



Figures 4-6: Kaplan-Meier curves for SVM, LR, and Traditional Method (VBS) using Overall Survival. SVM shows a vast improvement over LR and traditional cutoff methods for VBS, which were equal to each other in performance.



Figures 7-9: Kaplan-Meier curves for SVM, LR, and Traditional Method (LDH) using Overall Survival. SVM shows a significant improvement over LR and traditional cutoff methods for LDH.

Cox Hazard Ratios were produced for all experiments as a measure of the "risk" of death (or in the case of Progression Free Survival, PFS, progression) associated with being placed in the poor prognosis group over time. This could be considered a measurement of the usefulness or importance of the feature that the SVM, LR, and traditional method generate. Table 1 has the hazard ratio (Cox PH) and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for both Progression Free Survival (PFS) and Overall Survival (OS). Only EP-SVM and LR have associated ROC curves, the traditional methods do not.

	EP-SVM VBS	LR VBS	Trad VBS	EP-SVM RECIST	LR RECIST	Trad RECIST 1.1	EP-SVM LDH	LR LDH	Trad LDH (≥ 250 IU/L)
Cox PH PFS	*6.259*	6.259	6.259	**4.192**	3.617	3.655	**2.912**	2.865	1.801
Cox PH OS	***8.104***	2.841	2.841	**3.996**	2.541	1.997	**2.965**	2.595	2.454
AUC PFS	**0.7671**	0.753		***0.7971***	0.7357		**0.6355**	0.5785	
AUC OS	***0.8906***	0.5938		**0.7942**	0.7057		**0.7258**	0.718	

Table 1: Cox Hazard Ratios and Area under ROC Curves for RECIST, VBS, and LDH. Each column represents a technique (SVM, LR, and Traditional) for the three input sets (VBS, RECIST, and LDH). Each input/technique has a Cox Hazard Ratio for

each outcome; Progression Free Survival (PFS) and Overall Survival (OS). SVM and LR techniques also had an area under an ROC curve (AUC) value for each input and outcome. The higher CHR and AUC values, the better.

4. Discussion

The results demonstrate, both with Cox Hazard Ratios and Kaplan-Meier curves, a clear improvement when using EP-SVM to create a feature compared to Logistic Regression or traditional techniques in the clinic. The difference was far more pronounced with OS than PFS, which may have been due to the low median value (3.5 months) that was similarly mentioned as the reason for difficulty in showing the Kaplan-Meier curves for PFS experiments. While RECIST is considered the international standard for evaluating response to treatment, its prognostic value was lower than VBS in all experiments of comparable methods. This means that VBS outperformed RECIST, even when it was only comparing traditional methods. LDH generally did worse than RECIST or VBS, even though it did provide some prognostic power. The best result was obtained using the EP-SVM with VBS to create a feature for Overall Survival prognostication.

5. Conclusion

We can see that our CAS is a consistent improvement over existing statistical techniques used in the clinic and prognostication performed by using simple cutoff points and linear combinations of features. The creation of features based on statistical learning theory will allow more accurate results than current methods determined subjectively and provide a mechanism for the inclusion of new data easily. SLT is capable of combining data of multiple types and from multiple sources and observers without any additional work, and can capture non-linear relationships in the data which current methods cannot. This means that the long and heated process of generating and updating new international standards would be eliminated as the SLT method does the work of determining the best way to combine the data objectively. Future work should be done to test this CAS with more and diverse features, multiple observers, and non-binary categorization such as separating into treatment groups.

6. References

Brennan, P. and Silman, A. (1992). "Statistical methods for assessing observer variability in clinical measures". *Brit Med J.* 304:1491–4.

Breslow, N.E. and Day, N.E. (1980). Statistical Methods in Cancer Research: Volume 1–The Analysis of Case-Control Studies, The International Agency for Research on Cancer, Lyon, France.

Chang, C. and Lin, C. (2011). "LIBSVM: a library for support vector machines". ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Cox, D. (1972). "Regression Models and Life-Tables". *Journal of the Royal Statistical Society. Series B* 34 (2):187–220.

Eisenhauer, E., Therasse, P., Bogaerts, J., *et al.* (2009). "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)". *Eur J Cancer* 45:228–247.

Fraser, S. and Greenhalgh, P. (2001). "Coping with complexity; educating for capability". *BMJ.* 323:799-803 6th October.

Freer, T. and Ulissey, M. (2001). "Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center". *Radiology* 220:781-786.

Kaplan, E. and Meier, P. (1958). "Nonparametric estimation from incomplete observations". *J. Amer. Statist. Assoc.* 53:457–481.

Land, Jr., W., McKee, D., Velázquez, R., *et al.* (2003). "Application of Support Vector Machines to breast cancer screening using mammogram and clinical history data". Proceedings of SPIE (The International Society for Optical Engineering), Medical Imaging, Vol. 5032, San Diego, CA, February 17-20.

Land, Jr., W., Margolis, D., Gottlieb, R., *et al.* (2010). "Improving CT prediction of treatment response in patients with metastatic colorectal carcinoma using statistical learning". *I. J. Computational Biology and Drug Design.* 3(1): 15-18.

Ludwig, J. and Weinstein, J. (2005). "Biomarkers in cancer staging, prognosis and treatment selection". *Nature.* Reviews(5), 845-856.

Marsland, S. and Buchan, I. (2004). "Clinical quality needs complex adaptive systems and machine learning". In *Proceedings of the International Conference on Medical Informatics*, 644-647.

Michaelis, L. and Ratain, M. (2006). "Measuring response in a post-RECIST world: from black and white to shades of grey". *Nat Rev Cancer.* 6:409–414.

Sim, I., Gorman, P., Greenes, R., Haynes, R., Kaplan, B., Lehmann, H., *et al.* (2001). "Clinical Decision Support Systems for the Practice of Evidence-Based Medicine". *J Am Med Inform Assoc* 8(6):527-34

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York