ADAPTIVE COMPUTATION AND MACHINE LEARNING TERM PROJECT PROPOSAL

TITLE:
PATIENTS SURVIVABILITY PREDICTION AFTER BREAST CANCER SURGERY USING SUPERVISED MACHINE LEARNING TECHNIQUES

By: UZOR G. GOD'SGIFT Student Number: 1654209

And PATRICK T. TSHIABA Student Number: 1744742

ABSTRACT

The goal of this paper is to predict the survival of breast cancer patients, who had undergone surgery through machine learning techniques that will determine the chances of living or not living above five years post-surgery.

BRIEF DESCRIPTION OF PROJECT

The application of machine learning techniques has been proven productive in the analysis of diverse complex health related problems for quite some time now. In the past there often were situations where after surgery, it could not be predicted if a patient would live longer than five (5) years or lesser as a result of very little implementation of machine learning techniques. This paper is more particular about the prediction of the likelihood of patients who has gone through breast cancer surgery surviving above or below 5 years.

RELATED RESEARCH IN BREAST CANCER SURGICAL PREDICTIONS

In a research paper published in June 1995 by W. N. Street, O. L. Mangasarian, and W.H. Wolberg titled "An inductive learning approach to prognostic prediction", it was noted that machine learning was applied to a new breast cancer prognosis prediction method based on linear programming using backward propagation neural networks predicting how long after surgery we can expect the disease to recur [1].

In July 2005, from the department of management science and information systems of Oklahoma State University, Dursun Delen, Glenn Walker and Amit Kadam in an article title "Predicting breast cancer survivability: a comparison of three data mining methods", noted that breast cancer was a major cause of concern in the United States. They compared three data mining algorithms Artificial Neural Networks, Decision Trees and Logistic regression with 91.2%, 93.6% and 89.2% accuracy respectively, highlighting the Decision Tree method as the most suitable for their work [2]. Using machine learning techniques provided valuable assistance in making the diagnosis possible but not devoid of the physician's intuition and interpretive skills.

In an investigative research done in June 2016 by Rohit J. Kate and Ramya Nadig, from the University of Wisconsin's Department of Health Informatics and Administration, titled "Stage-specific predictive models for breast cancer survivability", it was correctly noted that breast cancer surgical survival rate has increased form 75.2% in 1980 to 90.6% in 2013, highlighting the importance of accurate prediction of breast cancer survivability aiding physicians and health-care providers to make informed decisions about a patients treatment and applying more aggressive therapies for patients with more tragic or critical prognosis[3]. They used a dataset publicly available from SEER. To achieve the goal of their research they used naive Bayes, logistic regression and decision tree machine learning classification methods to predict the survivability of breast cancer.

In June 2017, Lotsch J, Ultsch A, Kalso E wrote a paper on the Prediction of post-surgery pain by preoperative cold pain sensitivity using supervised machine-learning techniques to construct a classifier that could predict patients at risks of persistent pain [4]. They analyzed 763 patients who had undergone breast cancer surgery, of which 61 of the patients developed signs of persistent pain during three (3) years of follow up [4].

PROPOSED BREAST CANCER SURVIVABILITY PREDICTION SOLUTIONS

Several popular machine learning methods like naive Bayes, support vector machines (SVM), artificial neural networks (ANN), logistic regression and decision trees have been utilized in the past for predicting breast cancer survivability related courses.

In proffering a solution, this project will be required to go through a learning curve where the methods will be learning the Haberman survival dataset containing cases of patients from a study conducted from 1958 to 1970 [5]. Upon completion of the learning phase, different machine-learning methods will be applied to the Haberman dataset to enable a realization of the task of prognosis prediction. It is also important to note that the risk of death was higher at the time of the collection of these data than it is in our current time, hence one of the reasons we opted to evaluate how machine-learning methods could have been used in years past to correctly predict survivability after surgery.

The task in this project is to suggest techniques and methods that should be applied to better predict patient survival and increase the chances of them living longer.

We will be taking into consideration a couple of machine-learning methods to actualize our goal.

REFERENCES

1 Street, W. Nick, Olvi L. Mangasarian, and William H. Wolberg. "An inductive learning approach to prognostic prediction." In *Machine Learning Proceedings 1995*, pp. 522-530. 1995.

2 Delen, Dursun, Glenn Walker, and Amit Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods." *Artificial intelligence in medicine* 34, no. 2 (2005): 113-127.

3 Kate, Rohit J., and Ramya Nadig. "Stage-specific predictive models for breast cancer survivability." *International journal of medical informatics* 97 (2017): 304-311.

4 Lötsch, Jörn, Alfred Ultsch, and Eija Kalso. "Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis." *British Journal of Anaesthesia* 119, no. 4 (2017): 821-829.

5 Han, Hui, Wen-Yuan Wang, and Bing-Huan Mao. "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning." In *International Conference on Intelligent Computing*, pp. 878-887. Springer, Berlin, Heidelberg, 2005.