

University of Wollongong

## Research Online

---

Faculty of Engineering and Information  
Sciences - Papers: Part B

Faculty of Engineering and Information  
Sciences

---

2017

# A Comparison Study for Supervised Machine Learning Models in Cancer Classification

Huaming Chen

*University of Wollongong, hc007@uowmail.edu.au*

Hong Zhao

*Xiamen University*

Lei Wang

*University of Wollongong, leiw@uow.edu.au*

Jiangning Song

*Monash University, jiangning.song@monash.edu*

Jun Shen

*University of Wollongong, jshen@uow.edu.au*

Follow this and additional works at: <https://ro.uow.edu.au/eispapers1>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

---

### Recommended Citation

Chen, Huaming; Zhao, Hong; Wang, Lei; Song, Jiangning; and Shen, Jun, "A Comparison Study for Supervised Machine Learning Models in Cancer Classification" (2017). *Faculty of Engineering and Information Sciences - Papers: Part B*. 778.

<https://ro.uow.edu.au/eispapers1/778>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# A Comparison Study for Supervised Machine Learning Models in Cancer Classification

## Keywords

classification, cancer, learning, machine, supervised, comparison, models, study

## Disciplines

Engineering | Science and Technology Studies

## Publication Details

Chen, H., Zhao, H., Wang, L., Song, J. & Shen, J. (2017). A Comparison Study for Supervised Machine Learning Models in Cancer Classification. 16th International Conference on Bioinformatics (InCoB 2017) (pp. 1-2). China:

# A Comparison Study for Supervised Machine Learning Models in Cancer Classification

Huaming Chen<sup>1</sup>, Hong Zhao<sup>2</sup>, Lei Wang<sup>1</sup>, Jiangning Song<sup>3</sup>, Jun Shen<sup>1</sup>

<sup>1</sup>School of Computing and Information Technology, University of Wollongong, NSW, Australia

<sup>2</sup>Department of Physics, Xiamen University, Fujian

<sup>3</sup>Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia

Correspondence: hc007@uowmail.edu.au

## Background

The effective mining of the pattern detection problems underlying the gene expression data is an emerging research challenge, which has significant implications for disease diagnosis, such as fast and accurate cancer subtype classification. Meanwhile, different machine learning methods have been deployed to tackle the classification problem. However, deploying these methods on gene expression data, specifically for those cancer related data, to build a robust model among all of them is a cumbersome work for each individual researcher. Thus, in this work, we present a comprehensive study on the cancer related data, which considers five different scenarios and compares four different supervised machine learning models. Moreover, a newly introduced model named general vector machine is also taken into account and it achieves better results on two of these data sets when compared with other models.

## Results

In this study, we focus on presenting different cancer data sets to different supervised learning models whilst we also introduced GVM model to classify these tasks. The GVM model is further extended with *ReLU* transfer function and a revised cost function. Specifically, we have conducted ten times randomized 10-fold cross validation and reported the average accuracy results of each experiments. The joint GVM model achieves the best results on AML and colon cancer data sets, with the accuracy of  $60.00 \pm 7.45$  and  $87.5 \pm 2.01$ , respectively. For breast cancer and medulloblastoma data sets, the stacked auto encoder deep learning model demonstrates better performance than the others while linear SVM is most suitable for prostate cancer data set.

## Conclusions

It remains a significant research challenge to build a robust computational model for learning various types of gene expression data. In this work, we contribute a comprehensive study by experimenting with several different models to investigate their performance on gene expression data.

## **A Comparison Study for Supervised Machine Learning Models in Cancer Classification**

Huaming Chen *et al.*, University of Wollongong

General analyses of several supervised learning models, including a model named general vector machine, on various cancer data sets are conducted and discussed.