

BOOSTED FEATURE SELECTION FOR CLASS DEDICATED SVM  
AND ITS APPLICATION IN FETAL HEALTH PREDICTION

BY

JINPYO LEE

B.S.E. Seoul National University, 2003  
M.S.E. University of Michigan, 2011

DISSERTATION

Submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in Systems Science & Industrial Engineering  
in the Graduate School of  
Binghamton University  
State University of New York  
2019

## **Abstract**

This research developed improved classification methodology for Cardiotocography (CTG) data. CTG data has been widely used for diagnosing fetal health until delivery. However, the high dimension and multiclass characteristic of CTG data are barriers in improving the performance further in correct classification rate (CCR) and computational efficiency. This research implemented various experiments on feature selection and extraction methods, kernel selection in SVM and ensemble methods.

Firstly, this research experimented new methodology of classification model by using both feature selection and feature extraction, and SVM with 4 kernels by a wrapper method to obtain the highest CCR with reduced computational complexity. According to experimental results, the algorithm of 4 combination ensemble resulted in the highest CCR with 75% reduced time complexity. The alternative combinations reduced it further with slightly degraded CCR. More efficient ensemble depending on feature types reduced computation time further by 39% for 2-class and 70% for multiclass data.

Secondly, Linear Discriminant Analysis (LDA) and distance between classes were used as alternative feature ranking criteria for feature selection. According to comparison of performance, the CCRs of feature ranking by SVM, LDA, distance between classes and PCA are at almost same level in case of 2-class data. In case of multiclass data, the distance between classes is the most effective ranking criteria. The distance between classes is effective on multiclass and large-scale data with large instances or high dimension.

Finally, improved classification methodology for CTG data is proposed. As the first step, boosted feature selection was developed. One vs. all multiclass classification architecture with a wrapper composed of feature ranking by the same classifier SVM and distance between classes among misclassified instances resulted in increase of CCR by 3.5% compared to the highest in literature and 49.2% feature reduction compared to the case without feature selection. By applying boosted feature selection and feature extraction by K-means clustering, Class-dedicated SVM increased CCR and sensitivity by 6.9% and 0.131 compared to the highest in literature. This proposed methodology is expected to increase the efficiency of diagnosis based on CTG data by predicting the pathologic state more accurately.