

# CORRELATION FEATURE SELECTION WEIGHTING ALGORITHMS FOR BETTER SUPPORT VECTOR CLASSIFICATION: AN EMPIRICAL STUDY

*Doreen Ying Ying Sim*, Chee Siong Teh

<sup>1</sup>Department of Cognitive Science, Faculty of Cognitive Science and Human Development, Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300 Kota Samarahan, Sarawak, Malaysia.

Ahmad Izuanuddin Ismail

<sup>2</sup>Respiratory Medicine Unit, Department of Respiratory Medicine, UiTM Medical Specialist Centre, Universiti Teknologi MARA, Selangor, Malaysia.

dsdoreenyy@gmail.com, csteh@unimas.my, izuanuddin@uitm.edu.my

---

## ABSTRACT

Characteristics of Support Vector Machine (SVM) and its classifications are elaborated to show why incorporation of newly proposed and formulated regularization on feature selections based on correlation studies are necessary to achieve a better prediction or classification. Feature selections based on correlation studies are incorporated into the proposed formulations for the weighting portions of the objective functions for SVM. Proposed cfsw-SVM algorithms are then developed. Proposed formulations on SVM regularization parameter provides synergistic adjustments between prediction or classification accuracy and the level of correlations among features in the SVM implemented. Prediction and/or classification accuracies of cfsw-SVM algorithms are significantly improved.

**Keywords:** Support Vector Machine, Newly Proposed and Formulated Regularization, Feature Selections, Correlation Studies.

---

## 1. INTRODUCTION

### *1.1 Characteristics of Support Vector Classification*

Support Vector Machine (SVM) is a supervised and robust machine learning method [1]–[3], and hence its weaknesses, compared with other classifiers and the like, are not much and/or very prominent [6]–[8], [11]. SVM is for maximizing (i.e. supporting) the separating margin through vectors [10] so as to reduce the training (i.e. re-substitution) and testing (i.e. generalization) errors. The hinge loss function of SVM is used mainly for solving dual problems and this is done by significant maximum margin classification [3], [17]–[21]. Although SVM is a strong and stable classifier in terms of its classification and regression strength [1]–[4], [6], [10], [18]–[21], it may have certain weaknesses if it is being implemented