

# **KNN-Based Overlapping Samples Filter Approach for Classification of Imbalanced Data**

**Mar Mar Nwe**

**marmarnwe@ucsm.edu.mm**

**University of Computer Studies, Mandalay, UCSM**

**Khin Thidar Lynn**

**khinthidalynn@ucsm.edu.mm**

**University of Computer Studies, Mandalay, UCSM**

Imbalanced data classification is one of the most interesting problems in various real-world data sets. The class distribution of imbalanced data set strongly affects the classification rate of learning classifiers. If the class distribution problems can't be solved before implementing the learning algorithms, the predictions of learning classifiers tend to support a large number of samples (majority class) and ignore the other samples (minority class). In addition, the class overlapping problem can increase the difficulty to classify the minority class samples correctly. In this paper, we propose an effective under-sampling method for the classification of imbalanced and overlapping data by using KNN-based overlapping samples filter approach. Besides, this paper summarizes the performance analysis of three ensemble-based learning classifiers for the proposed method. Experimental results on fifteen imbalanced data sets indicate that the proposed under-sampling method can effectively improve the five representative algorithms in terms of three popular metrics; area under the curve (AUC), G-means and F-measure.