**Methods for feature selection**

The methods for feature selection considered in the study are the following: Conditional Mutual Information Maximization (CMIM) (Fleuret, 2004); Correlation-based Feature Selection (Hall, 1999), which considers correlation criteria in which the features should be highly correlated with the class label and uncorrelated with each other; Joint Mutual Information (JMI) (Yang & Moody, 1999); Maximum Relevance Minimum Redundancy (mRMR) (Peng, Long, & Ding, 2005) that attempts to select features that maximize the relevance to the class value and minimize the redundancy among them; RELIEF (Kira & Rendell, 1992), an algorithm that utilizes instance–based learning to give importance and assigns weights to each feature, where a weight represents the ability of a feature to distinguish between classes; T-test (Guyon & Elisseeff, 2003), that measures the statistical significance of difference in class relevance for a particular feature between the two classes; and Receiver Operating Characteristic (ROC) (Guyon & Elisseeff, 2003) curve, used to measure class discrimination by a feature using the area between the curve and the random classifier.

**REFERENCES**

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. *The Journal of Machine Learning Research, 5*, 1531-1555.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research, 3*, 1157-1182.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning.* The University of Waikato.

Kira, K., & Rendell, L. A. (1992). *The feature selection problem: Traditional methods and a new algorithm.* Paper presented at the AAAI.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27*(8), 1226-1238.

Yang, H., & Moody, J. (1999). *Feature selection based on joint mutual information.* Paper presented at the Proceedings of international ICSC symposium on advances in intelligent data analysis.