

# Abstract

**Sherine . Nagy**

## **A novel dynamic feature Selection and prediction algorithm for clinical decisions involving high-dimensional and varied patient data**

Predicting suicide risk for mental health patients is a challenging task performed by practitioners on a daily basis. Failure to perform proper evaluation of this risk could have a direct effect on the patient's quality of life and possibly even lead to fatal outcomes. Risk predictions are based on data that are difficult to analyse because they involve a heterogeneous set of patients' records from a high-dimensional set of potential variables. Patient heterogeneity forces the need for various types and numbers of questions to be asked regarding the individual profile and perceived level of risk. It also results in records having different combinations of present variables and a large percentage of missing ones. Another problem is that the data collected consist of risk judgements given by several thousand assessors for a large number of patients. The problem is how to use the associations between patient profiles and clinical judgements to generate a model that reflects the agreement across all practitioners. In this thesis, a novel dynamic feature Selection algorithm is proposed which can predict the risk level based only on the most influential answers provided by the patient. The feature Selection optimises the vector for predictions by Selecting variables that maximise correlation with the assessors' risk judgement and minimise mutual information within the ones already Selected. The final vector is then classified using a linear regression equation learned for all patients with a matching set of variables. The overall approach has been named the Dynamic Feature Selection and Prediction algorithm, DFSP. The results show that the DFSP is at least as accurate more accurate than alternative gold-standard approaches such as random forest classification trees. The comparison was based on accuracy and error measures applied to each risk level separately ensuring no preference to one risk over the other.