

# University of Durham

Prof. F.P.A. Coolen  
Professor of Statistics

## Department of Mathematical Sciences

Science Laboratories, South Road,  
Durham, DH1 3LE, England

Telephone:

Direct: Durham (0191) 334 3048

Department: Durham (0191) 334 3050

University: Durham (0191) 334 2000

E-mail: frank.coolen@durham.ac.uk

January 2, 2024

To whom it may concern

Dear Sir, Madam,

**RE:** Opponent's Review of Habilitation Thesis by Dr Jan Rabcan

**Thesis title:** Development of common methods from reliability analysis and machine learning.

**Author:** Dr Jan Rabcan, Faculty of Management Science and Informatics, University of Zilina, Slovakia.

**Opponent:** Professor Frank Coolen, Department of Mathematical Sciences, Durham University, United Kingdom.

### Summary of Review:

The work of Dr Rabcan is impressive and, I believe, amply satisfies all requirements for the status of Associate Professor to be awarded. I recommend the thesis for defence and am overall impressed by the thesis: it contains novel ideas and careful examinations which make suitable contributions to the literature. My detailed views on the thesis are described below, and I state a few questions resulting from reading the thesis.

### Aims:

This thesis explores opportunities to link up methods from Reliability Theory and Machine Learning, in particular Classification. It is interesting and, as far as I am aware, not done before in the literature in the way proposed. A major question for practical reliability theory, these days, is how to use large amounts of data, e.g. available from continuous condition monitoring with sensors which may however give faulty signals. As first aim of this thesis, learning a system's structure function on the basis of imperfect (incomplete, uncertain) data is considered. A second aim of this thesis is to improve statistical classification methods by first applying common reliability theory methods for quantifying importance of attributes.

### Relevance:

Overall, Reliability Theory and Classification are wide areas of methodology and application which are crucial for society. Hence, all genuine advances in these fields are of great importance. Theoretical research, when based on new ideas as presented in this thesis, is important and may lead to new methodology and, ultimately, to better reliability quantification of systems and to better classification methods. The combination of the two fields is of great interest and I can see further ways in which uncertainty quantification in both fields can be linked up, building on the foundations laid in this thesis.

### New Results and Methodology:

A proposal for the first aim of the thesis is based on the use of fuzzy methods, in particular fuzzy

decision trees, ultimately leading to multi-valued decision diagrams representing the system structure function. This is an interesting novel approach which, however, raises questions (see below) which I believe must be addressed before the method can be judged on its value in comparison to alternative methods. There are some claims on size of systems that could be addressed using this methodology but it is hard to see what the constraints are. The second aim is, in my view, somewhat more interesting: the idea to restrict the number of input attributes in classification problems by the use of importance measures, as typically used in reliability theory, is interesting, and on the basis of the initial investigation in this thesis, it may perform well in some large scale classification problems compared to methods that use all attributes without filtering of the attributes as first stage. So, the results presented are of interest and warrant further research on the presented approaches. The methodologies used are sound, where however the fuzzy methods raise some questions.

*Comments and Questions:*

With regard to the first aim, learning the system structure function using fuzzy decision trees, there are some issues that would require further clarification before the method can be strongly advocated for practical use. First, it is not clear what is meant by the word *data* in this thesis: it seems to encompass all information available (ideally) for specifying a system's structure function, which is often a combination of clear engineering knowledge and process data. If a system is a black box, one has to rely mostly on the latter, but most practical systems which are of interest to study with regard to their reliability, are not black-box. So, some more clarity on what kind of systems may benefit from fuzzy methods, and in case of what kind of data, would be useful. There is a statement that even precise information must first be fuzzified, which worries me to be fair. Maybe one has excellent knowledge about part of the structure function, less so on other parts, and possibly the proposed method could be developed to limit the use of the fuzzy methods to the latter?

I was further somewhat surprised that, following the use of fuzzy decision trees, the resulting structure function is precise: why is uncertainty not further reflected here, leading to a fuzzy structure function? The main issue here would be why the structure function is developed: this is never just for the sake of having a structure function, but typically in order to support decisions related to the practical use and maintenance of the system. Should uncertainty on the structure function not be fed forward into such decision support?

On the second aim, my main worry is that the filtering method may exclude variables which, on their own, have very minimal predictive power on the dependent variable of interest, but which in combination with one or more other attributes may have very strong predictive power. In reliability terms, this would e.g. be the case with two variables which each have very low importance measure values, but with very high joint importance. Many classification problems have such attributes, which typically shows in classification subtrees which often differ substantially after splitting on one or more variables. One could also consider similar approaches to joint importance measures, but doing so may reduce or eliminate any computational advantages the method has. It would be good to know the candidate's views on this.

Finally, I must comment on the quality of presentation of the thesis, which is unfortunately not good: there are many instances where the work gives the impression of being put together in a hurry, without proper proofreading, and this is not only with regard to use of the English language. Hence, if the thesis becomes public at some stage, I strongly recommend very careful proofreading.

Yours Sincerely,

Frank Coolen