

Modelling for understanding AND for prediction/classification - the power of neural networks in research

Eduardo Cascallar^{ab}, Mariel Musso^{acd}, Eva Kyndt^a and Filip Dochy^a

^aUniversity of Leuven, Belgium

^bAssessment Group International, USA / Belgium

^cNational Research Council (CONICET)/CIIPME, Argentina

^dUniversidad Argentina de La Empresa, Argentina

Article received 28 November 2014 / revised 18 January 2015 / accepted 18 January 2015 / available online 30 January 2015

Abstract

Two articles, Edelsbrunner and, Schneider (2013), and Nokelainen and Silander (2014) comment on Musso, Kyndt, Cascallar, and Dochy (2013). Several relevant issues are raised and some important clarifications are made in response to both commentaries. Predictive systems based on artificial neural networks continue to be the focus of current research and several advances have improved the model building and the interpretation of the resulting neural network models. What is needed is the courage and open-mindedness to actually explore new paths and rigorously apply new methodologies which can perhaps, sometimes unexpectedly, provide new conceptualisations and tools for theoretical advancement and practical applied research. This is particularly true in the fields of educational science and social sciences, where the complexity of the problems to be solved requires the exploration of proven methods and new methods, the latter usually not among the common arsenal of tools of neither practitioners nor researchers in these fields. This response will enrich the understanding of the predictive systems methodology proposed by the authors and clarify the application of the procedure, as well as give a perspective on its place among other predictive approaches.

Keywords: Artificial neural networks; Response to Commentaries; Methodology; Data Modelling

Corresponding authors: Eduardo Cascallar, KU Leuven, Leuven, Belgium, cascallar@msn.com and Mariel Musso, National Research Council (CONICET), Argentina and KU Leuven, Leuven, Belgium, mariel.musso@hotmail.com

Doi: <http://dx.doi.org/10.14786/flr.v2i5.135>



Research is the process of going up alleys to see if they are blind.
Marston Bates

Two articles, Edelsbrunner and, Schneider (2013), and Nokelainen and Silander (2014) comment on Musso, Kyndt, Cascallar, and Dochy (2013). Several relevant issues are raised and some important clarifications need to be made in response to both commentaries. This response will enrich the understanding of the predictive system methodology proposed by the authors and clarify the application of the procedure, as well as give a perspective on its place among other predictive approaches.

Edelsbrunner and Schneider (2013) in their commentary on Musso, Kyndt, Cascallar and Dochy (2013) argue that artificial neural networks (ANNs) should only be used as exploratory modelling techniques, in spite of being powerful statistical modelling tools with demonstrated ability to improve outcomes of classifications and predictions over traditional statistical methods (Marquez, Hill, Worthley, & Remus, 1991). Garson (1998, pp. 11-14) cites more than thirty-five articles which have shown the ability of ANNs to outperform traditional techniques in specific circumstances. In addition, Haykin (1994, pp. 4-5) summarizes some of the main favourable properties of ANNs which explain their advantages over traditional methods. The reasons Edelsbrunner and Schneider (2013) argue for their rather strong position are centred on two main arguments: (a) that the output from ANNs cannot be fully translated into a meaningful set of rules because of a lack of accessibility to the input-output relationships, and (b) that there is a lack of equivalent statistical parameters in ANNs when compared to more traditional statistical techniques. These are the two fundamental misconceptions that will be addressed.

One of the essential requirements for development and advancement in science is the willingness and vision to explore new conceptualizations and methods. In particular, as is the case in the study by Musso et al. (2013), the ability to bring together data from interdisciplinary domains (e.g., Decuyper, Dochy, & Van den Bossche, 2010), and to use new methodologies for analyses that are commonly applied in other disciplines such as business, finance, and the social sciences (Al-Deek, 2001; Detienne, Detienne, & Joshi, 2003; Laguna & Marti, 2002; Neal & Wurst, 2001; Nguyen & Cripps, 2001; White & Racine, 2001, and others as stated in Musso et al., 2003).

The literature still shows relatively few studies applying neural networks in education and in educational assessment in particular (Everson, Chance, & Lykins, 1994; Wilson & Hardgrave, 1995), although ANNs have been shown to improve the validity and the accuracy of the predictions and/or classifications, and also improve the predictive validity of test scores (Everson et al., 1994; Perkins, Gupta, & Tamanna, 1995; Weiss & Kulikowski, 1991). More recently, several studies have shown the applicability and use of this methodology in education (e.g., Cascallar, Boekaerts, & Costigan, 2006; Kyndt, Musso, Cascallar, & Dochy, 2011; Kyndt, Musso, Cascallar, & Dochy, 2015; Musso & Cascallar, 2009a; Musso & Cascallar, 2009b; Musso, Kyndt, Cascallar & Dochy, 2012; Musso et al., 2013; Pinninghoff Junemann, Salcedo Lagos, & Contreras Arriagada, 2007; Ramaswami & Bhaskaran, 2010; Zambrano Matamala, Rojas Díaz, Carvajal Cuello, & Acuña Leiva, 2011). These recent studies have used ANNs both for prediction/classification as well as for the understanding of the underlying variables involved in the educational outcomes studied. Now it



is important to show that recent advances in ANN analysis have addressed the main concerns expressed in Edelsbrunner & Schneider (2013).

First, the concerns regarding the presumed “opacity” of ANN in terms of their input-output relationships will be addressed. The authors undermine their own estimate of the value of ANNs as a “promising technique” by essentially arguing that it is contrary to good scientific practice for theory-building given the presumed “opaque” nature of their internal structure which makes interpretation difficult if not impossible. The often and now quite outdated argument of ANNs as “black boxes” (cf. Benitez, Castro & Requena, 1997) is therefore raised once again. However, these arguments are raised ignoring the vast amount of research that has been going on in this field to overcome this initial drawback of predictive systems analyses (e.g., Frey & Rusch, 2013; Intrator & Intrator, 2001; Lee, Rey, Mentele, & Garver, 2005; Tzeng & Ma, 2005; Yeh & Cheng 2010).

Considering the nature and centrality of modelling in science, as was clearly presented by Frigg and Hartmann (2006), models can perform two different representational functions, which are not mutually exclusive as scientific models. First, they can be a representation of an aspect or selected part of the world, what they call the “target system”. In this case, what can be modelled are either phenomena or data. The second notion of modelling is the representation of a theory in that it represents its rules, laws and axioms.

Clearly, ANNs contribute to the construction of better representational models consisting of “models of data” (Suppes, 1962). In particular, this contribution is based on ample research that has been crucial in making the link between ANNs representations and their relationship to the obtained outputs. As an anecdote, it is interesting and revealing that Edelsbrunner and Schneider (2013) cite the paper of Benitez, et al. (1997) which presents an addition to the usual ANN techniques which according to Benitez et al. (1997) provide “such an interpretation of neural networks so that they will no longer be seen as black boxes” (p. 1156), which clearly contradicts the use of the article of Benitez et al. (1997) as supporting the “black box” unique perception of ANNs. The proposed approach, in this case is based on the determination of the equality between multilayered perceptron ANNs, precisely the one used by Musso et al. (2013), and fuzzy rule-based systems. The operator derived from this equivalency concept results in the transformation of fuzzy rules into a format which can be easily understood. Thus, the knowledge generated by the ANN after the learning process is finished can be more easily and clearly explained, “so that they can no longer be considered as black boxes” (Benitez et al., 1997, p. 1156), while retaining all the advantages and power of the ANNs as very efficient computing representations as automated knowledge acquisition procedure models, and as universal approximators (Ripley, 1996). In fact, West, Brockett, and Golden (1997) state that neural networks “are a well-defined adaptive gradient search procedure for parameter fitting in a complex nonlinear model, and not a ‘black box’ at all” (p. 389).

In addition, the efforts to develop better and more comprehensive visualisation techniques for the complex interactions in an ANN, such as those suggested by Tzeng and Ma (2005) have contributed to open the “black box” and help the researcher in determining underlying dependencies between inputs and outputs of a neural network. As a consequence, they do not only facilitate the design of efficient ANNs, but also enable the use of ANNs for problem solving. It is true that



visualisation is not explanation, but they are powerful tools to guide the refinement of neural network structures for problem solving (e.g., classification tasks) using ANNs or other machine learning models. Another significant addition to the literature which “opens the box” in ANN analyses is the concept of structured neural network (SNN) techniques used for modelling (Lee, Rey, Mentele, & Garver, 2005). In this approach, the actual construction of the network is based on existing contextual and theoretical knowledge to assist in the design of the ANN structure of inputs. In fact, a similar approach was followed by Musso et al. (2013), by populating the inputs based solely on solid theoretical constructs derived from previous cognitive, motivational, and sociodemographic research and models, avoiding blind data mining techniques (Hand, Mannila & Smyth, 2001), and based on the factor analysis and structural equation modelling (SEM) of several variables to determine their potential weight in the problem.

Cause-and-effect relationships have been traditionally modelled, among others, by SEM and Partial Least Squares (PLS) approaches. But these procedures have their own shortcomings. In PLS, there is no theoretical rationale for all indicators to have the same weighting (Haenlein & Kaplan, 2004), and the PLS procedure does not take into account the fact that some indicators may be more reliable than others and should, therefore, receive higher weights (Chin, Marcolin, & Newsted (2003). In addition, there is the difficulty of interpreting the loadings of the independent latent variables in PLS (which are based on cross-product relations with the response variables). Regarding SEM several authors also point out some issues that require attention from the researcher or that are still awaiting further research (Lei & Qiong Wu, 2007; Schermelleh-Engel, Kerwer, & Klein, 2014; Weston & Gore, 2006). Among the issues noted with SEM are possible data problems, such as missing data, non-normality of observed variables, or multicollinearity; estimation problems that could be due to data problems or identification problems in model specification; or interpretation problems due to unreasonable estimates. These potential problems have led to suggestions involving the development of “mixture PLS” models (Hahn, Johnson, Herrmann, & Huber, 2002), hierarchical Bayesian methods in SEM models (Ansari, Jedidi, & Jagpal, 2000) and new ways of evaluating fit in non-linear multilevel structural equation models (Schermelleh-Engel et al., 2014). Even if nonlinear SEM and PLS models could handle asymmetric relationships, they still do not solve the problems associated with large data and complex interactions. The SNN approach takes into account these complexities and non-linearity in data sets, while maintaining the advantages of the ANN general model.

Another significant addition to the battery of approaches that researchers have explored to eliminate the “black box” risk of ANNs is the inclusion of sensitivity analysis for each of the variables in the model (Kim & Ahn, 2009) in order to extract the necessary information for model validation and process optimisation, from the relationships between inputs and outputs in the ANN. This method, based on the relative importance (RI) parameter estimate improves on Garson’s (1991) use of relative importance weights, and uses sensitivity analysis to determine the causal importance of the input variables on the outputs. The sensitivity is a measure of the increase in the error of the predicted value as each variable is excluded from the model, and demonstrates systematically the degree of influence on the network weights of each participating variable. The RI methods used in both classification and prediction models are another evidence of the fallacy of the



view of neural networks as black-boxes beyond human understanding. Incidentally, Kim and Ahn (2009) also compared the results from the ANN analysis with logistic regression and classification and regression trees (CART) analyses, with ANN models obtaining better results in both training and testing sets of data. Other authors (e.g., Blackard & Dean, 1999) have compared ANNs absolute accuracy and relative accuracy compared to predictions based on discriminant analysis (DA) models, with a consistent finding that ANN models outperformed the DA models.

A very interesting comparison of methods to accurately assess the contribution of variables in ANN architectures has been reported by Olden, Joy, and Death (2004). The authors compare nine different methods for quantifying variable importance in ANNs using simulated data with known properties. The use of simulated data, when the true importance of the variables is known, provides a solid base for future developments in this field, which are not possible with natural data as is the case with Gevrey, Dimopoulos, and Lek (2003). The nine methodologies studied by Olden et al. (2004) included: connection weights, Garson's algorithm, partial derivatives, input perturbation, sensitivity analysis, forward stepwise addition, backward stepwise elimination, improved stepwise selection 1, and improved stepwise selection 2 (see Olden et al., 2004 for details on these methods). The results indicated that the connection weights approach showed the best overall performance both in terms of accuracy (degree of similarity between true and estimated variable ranks) and precision (degree of variation in accuracy), when estimating the true importance of all the variables in the ANN. Partial derivatives, input perturbation, sensitivity analysis and both versions of the improved stepwise selection methods showed moderate performance in the simulations. When estimating the actual ranks, the connection weights approach once again was the method which exhibited the best performance. In addition, Olden and Jackson (2002) reviewed a randomisation approach to better evaluate and understand the contribution of predictors in ANN analysis. They conclude by stating: *"Thus, by coupling this new explanatory power of neural networks with its strong predictive abilities, ANNs promise to be a valuable quantitative tool to evaluate, understand, and predict ecological phenomena"* (Olden & Jackson, 2002, p. 135).

All of these examples demonstrate that using the appropriate techniques, the complexity of an ANN does not need to translate into "opacity", and researchers are not limited in their ability to gain insight into the explanatory factors of the prediction and classification processes performed efficiently by ANNs. Studies such as Olden et al. (2004), Gevrey et al. (2003), and Lek, Belaud, Baran, Dimopoulos, and Delacoste (1996), are but the beginnings of a vast number of applications that have "opened the box" in ANN analysis. In addition, regularisation approaches have been used to enhance the interpretation of ANN results (Intrator & Intrator, 2001), and the estimation of interaction effects in ANNs was used and demonstrated by Donaldson and Kamstra (1999). Therefore, contrary to what has been pointed out by Edelsbrunner and Schneider (2013) and quoted by Golino and Gomes (2014), the ANN approach offers the potential to examine the complex relationships amongst its components.

An additional important advantage of ANN analysis refers to the need to capture the complexity of the interaction of various factors in the understanding of also complex phenomena (Agrawal, 2001). It is difficult to find large-N studies with a large set of variables, particularly in the social and educational sciences. So, most studies attempt to develop causal models based on a



very limited set of variables, without the capacity to encompass a large number of predictors, and therefore not providing the possibility to observe their complex interactions (Boekaerts & Cascallar, 2006; Cascallar et al., 2006). A resulting problem is that meta-analyses trying to find general statistical correlations face very serious problems as interactions between the factors analysed are not known, which in turn leads to wrong estimations of relevance. Related to this problem is the fact that in all studies that knowingly or unknowingly exclude a relevant factor, the importance of all other variables shifts dramatically. This effect has been noted in very diverse fields ranging from natural resource estimation to self-regulated learning (Agrawal & Chhatre, 2006; Boekaerts & Cascallar, 2006). Studies which only take into account a few variables, in rather simple designs, and do not consider very important but complex interactions with a larger number of participating factors can and do often show contradictory results. This should not be considered a trivial problem for the conceptualisation of various effects and phenomena in every scientific field (Boekaerts & Cascallar, 2006). Frey and Rusch (2013) present an interesting study in the area of social-ecological systems which uses ANNs with an analytic approach that produces an open architecture in which it is possible to establish the input-output relationships which Edelsbrunner and Schneider (2013) seem to perceive are unachievable for ANNs. These analyses suggested by various authors (Thrush, Coco & Hewitt, 2008; Yeh & Cheng 2010) make the relationships among the various input-output variables explicit.

The second main argument regarding problems associated with the ANN methodology, as claimed by Edelsbrunner and Schneider (2013), has to do with the lack of some statistical parameters in ANNs. This ignores the evidence that there has also been an abundance of research to provide the ANN model with equivalent information. There have been increasing efforts for some time, to embed ANNs in general statistical frameworks (Cheng & Titterington, 1994), with Bridle (1992) comparing and blending ANNs with Markov-chain models, and applying Bayesian approaches and methods in the modelling of neural networks (MacKay, 1992). More recently, He and Li (2011) provide an interesting example of such work. They used the standard backpropagation algorithm derived in vector form, and they were successful in determining the confidence interval and prediction intervals for the ANN, while also exploring which neural network structural characteristics had more of an impact on such parameters. In particular, when the Levenberg-Marquardt backpropagation algorithm is used to train a neural network, since the Jacobian matrix has been calculated to update the weights and biases of the neural network, the confidence interval with the corresponding confidence level can be computed to evaluate the predictive capability of the ANN. In addition, on similar topics, Zapranis and Livanis (2005) state that given that ANNs are a good example of consistent non-parametric estimators with powerful universal approximation properties, they require that the development and implementation of neural network applications has to be based on established procedures for estimating confidence and especially prediction intervals. They go on to review the main state-of-the-art approaches for the construction of confidence and prediction intervals, and evaluate their strengths and weaknesses. After comparing them in a controlled simulation, the authors suggest that a combination of bootstrap and maximum likelihood approaches are superior to analytic approaches when constructing the prediction intervals (Zapranis & Livanis, 2005). On the other hand, other authors propose the construction of confidence intervals for neural networks based on least squares



estimations and using the linear Taylor expansion of the nonlinear model output, which also detects ill-conditioning of ANN candidates and can estimate their performance (Rivals and Personnaz, 2000).

In terms of the comparison between ANNs and logistic regression, in neural network analysis the purpose of the hidden layer is to map a set of patterns, which are linearly non-separable in the input space, into the so-called image-space in the hidden layer, where these patterns may become linearly separable. As in logistic regression, decision surfaces in the neural networks are hyperplanes in the input space. The key difference, though, between neural networks and logistic regression is that each hidden neuron (other than the bias neuron) produces an output that corresponds to a distinct, discriminating hyperplane in the input space. When these are weighted, summed, and transformed at an output neuron, the resulting output corresponds very closely to a multidimensional step function. It is found that the boundaries of regions of similar probability are defined by the discriminating hyperplanes, which crisscross the input space (Dreiseitl & Ohno-Machado, 2002).

Given the vast number of practical applications already mentioned in the original article by Musso et al. (2013), it is unfortunate that Edelsbrunner and Schneider (2013) choose to exemplify an unrealistic example of application of ANNs in a contrived situation in which a student is eliminated from a programme based on a neural network classification. ANNs, like any other methodology provides the researcher or applied scientist with information. As we have already shown from the literature cited, in the case of ANNs there are a number of methods to establish the necessary input-output relationships and to determine the confidence and prediction intervals provided by an ANN. Therefore, the contrived diagnostic example provided by Edelsbrunner and Schneider (2013, pp. 100) shows an underestimation/misinterpretation of the potential of ANNs. Furthermore, poor advice is always a problem, as would be the case in this example, with the unfortunately frequent decision-making of students' career paths determined by a single-point examination. On the other hand, a trusted result from a properly constructed and tested ANN could provide valuable diagnostic, educational, and public policy information. In fact, the research carried out by some of these authors (Cascallar et al., 2006; Kyndt et al., 2011, 2015; Luft, Gomes, Priori & Takase, 2013; Musso & Cascallar, 2009a; Musso et al., 2012, 2013) provides examples of useful diagnostic models in the educational field. It is a false dichotomy to present modelling for understanding versus modelling for prediction. In reality, both are achievable and in fact they should be integrated for the advancement of the field and the success of each application. Much insight has been gained by integrating understanding with predictive and classification models. As is good practice in various fields, especially in applied statistics and mathematical modelling, the various approaches constitute a toolbox that the professional has available in order to apply the best method for the problem at hand. The fact that our article (Musso et al., 2013) demonstrated the use of ANNs in a given academic application is not meant to be exclusionary. On the contrary, the field requires the integration of mathematical modelling and statistical techniques.

Regarding the comments in Nokelainen and Silander (2014) on the article by Musso et al. (2013), they can be summarized in two main points. The first point questions whether the methodology used was rigorous in its procedures, and the second suggests comparing the neural



network results with those obtained from another discriminative classifier in addition to the comparison to a generative classifier such as discriminant analysis.

It is very important to clarify that the data reported in Musso et al. (2013) rigorously followed the standards established by the Message Understanding Conferences (MUC) (Grishman & Sundheim (1996). As is clearly stated in the Musso et al. (2013) article, “the training and testing samples were selected at random from the existing data and the proportions were adjusted in order to maximize the training sample while preserving the appearance of all detected patterns in the testing sample, so as to be able to appropriately test the model” (p. 60). The two samples were chosen at random, precisely to avoid what Nokelainen and Silander (2014) put forward. These authors seem to have misinterpreted the sections on analyses procedures and architecture of the neural network (Musso et al., 2013, pp. 52-54) in which the process is described in detail, and they completely misjudge when they state that “The paper by Musso and her colleagues (2013) practically acknowledges that such a discipline was not rigorously followed.” (Nokelainen & Silander, 2014, p. 79). It is clearly stated in the above mentioned sections the way in which the sample was divided, the complete independence of the randomly selected training and testing subsets, and the criteria followed to determine the proportions of cases in each of the two subsets. Ironically, the procedures followed coincide with those suggested by (Nokelainen & Silander, 2014, p. 79). Let us state unequivocally that both subsets of cases in the training and testing samples were analyzed separately. In addition, all training of the neural network model was carried out on the training sample, as well as all parameter adjustments, until the desired level of precision was attained. Then, the model was independently tested on the testing sample, capturing the generalization of the network structure and the learning parameters. None of the model building took place on the testing sample as Nokelainen and Silander (2014) incorrectly assume. Thus, the performance of the model with the testing subset actually provides an indication of the generalization of the model, not just “fit” as Nokelainen and Silander (2014, pp. 79) also incorrectly state.

A related comment regarding the “ethical standards” of the Musso et al. (2013) paper is truly surprising. Do Nokelainen and Silander (2014) truly believe or imply that the authors could not “refrain from cheating (using the test data)” (Nokelainen & Silander (2014, p. 79) in developing the model? If so, it is alarming, because they are making a serious assumption regarding the authors or at best an implication of ignorance of basic rules of science and of this methodology in particular. Their fear of “cheating” and their implication that the testing sample analysis should be carried out by different researchers because of this assumed temptation to cheat could be extended to all research in all areas and all statistical methods. It is precisely part of the scientific method to follow any scientific finding with careful replications, not simply to avoid cheating, but to truly evaluate the generalizability of scientific results. It does not mean that we cannot trust researchers, at least a priori, with carrying out an ethically sound analysis. If not, all findings, including theirs, would be in question. Certainly, the Musso et al. (2013) article followed careful and rigorous methodological procedures. If their question has to do with the perfect classification obtained, it is the product both of the appropriate modelling process carried out, and of the granularity of the expected results given the available data; it should be noted that the correlation between the individual GPA scores of the



students in the whole testing sample and their predicted score (with data from one year in advance), was .86 (Musso et al., 2013, p. 64).

Regarding the suggestion to use other discriminative classifiers, such as logistic regression, to compare with the results obtained with the neural network model, it is a good suggestion which has already been carried out in the literature (Kim & Ahn, 2009), and it has been found that neural networks obtained better classification results. In fact, some of the authors in Musso et al. (2013) already have carried out such analyses in research currently underway, with the same results favourable to neural networks (Musso, Boekaerts, Segers, & Cascallar, in preparation).

The field of machine learning research and the related predictive systems is in constant development and new advances are introduced at a rapid pace (Monteith, Carroll, Seppi, & Martinez, 2011). Several methods have been suggested to improve the performance of machine learning algorithms and of neural network methods in particular, some of them using Bayesian approaches which have shown excellent potential (Aires, Prigent, & Rossow, 2004; Orre, Lansner, Bate, & Lindquist, 2000). We share the view expressed by Nokelainen and Silander (2014) that continued research in this field should be pursued, and ensemble methods (Rokach, 2010), such as those involving bootstrap aggregating (Sahu, Runger, & Apley, 2011), and Bayesian model combination (Monteith et al., 2011), together with multiple classifier systems (Roli, Giacinto, & Vernazza, 2001) are among those that should continue to be considered in certain applications.

In conclusion, we can state that as was very accurately stated by Anders and Korn (1996) in their work on model selection in neural networks, the process of model selection in ANN can be informed by statistical procedures and methods. Statistical methods can improve the model building and the interpretation of ANNs. What is needed is the courage and open-mindedness to actually explore new paths and new methodologies which can perhaps sometimes unexpectedly provide new conceptualisations and tools for theoretical advancement and practical applied research. This is particularly true in the fields of educational science and social sciences, where the complexity of the problems to be solved requires the exploration of proven methods and new methods, the latter usually not among the common arsenal of tools of neither practitioners nor researchers in these fields.

Keypoints

- Artificial Neural Networks are powerful mathematical modelling tools for classification and prediction.
- Advances in Artificial Neural Network methodologies have made them more transparent and useful, avoiding the original “black box” characteristics in their early development.
- There is a long history with significant recent advances which has achieved strong ties between traditional statistical constructs with their equivalent in Artificial Neural Networks.
- Artificial Neural Networks are a useful methodology that can advance our understanding of phenomena when modelling for understanding and modelling for classification/predictions are combined.



- Artificial Neural Networks are an additional important tool in the researcher's toolbox which can be particularly useful to tackle highly complex and large data sets with interactions among the variables which are not fully understood.

References

- Agrawal, A. (2001). Common property institutions and sustainable governance of resources. *World Development*, 29, 1649-1672. doi: 10.1016/S0305-750X(01)00063-8
- Agrawal, A., & Chhatre, A. (2006). Explaining success on the commons: Community forest governance in the Indian Himalaya. *World Development*, 34, 149-166. doi: 10.1016/j.worlddev.2005.07.013
- Aires, F., Prigent, C., & Rossow, W. B. (2004). Neural network uncertainty assessment using Bayesian statistics: A remote sensing application. *Neural Computing*, 16, 2415-2458. doi: 10.1162/0899766041941925
- Al-Deek, H. M. (2001). Which method is better for developing freight planning models at seaports – Neural networks or multiple regression? *Transportation Research Record*, 1763, 90- 97. doi: 10.3141/1763-14
- Anders, U., & Korn, O. (1996). Model selection in neural networks. *ZEW Discussion Papers*, 96-21. Retrieved from <http://hdl.handle.net/10419/29449>
- Ansari, A., Jedidi, K., & Jagpal, H. S. (2000). A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science*, 19, 328-347. doi: 10.1287/mksc.19.4.328.11789
- Benitez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks*, 8, 1156-1164. doi: 10.1109/72.623216
- Blackard, J. A. & Dean, D. J. (1999). Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24, 131–151. doi: 10.1016/S0168-1699(99)00046-0
- Boekaerts, M., & Cascallar, E. C. (2006). How far have we moved toward the integration of theory and practice in Self-regulation? *Educational Psychology Review*, 18, 199-210. doi: 10.1007/s10648-006-9013-4
- Bridle, J. S. (1992). Neural networks or hidden Markov models for automatic speech recognition: is there a choice? In P. LaFace (Ed.), *Speech Recognition and Understanding: Recent Advances, Trends and Application* (pp. 225-236). New York: Springer.
- Cascallar, E. C., Boekaerts, M., & Costigan, T. E. (2006) Assessment in the evaluation of self- regulation as a process. *Educational Psychology Review*, 18, 297-306. doi: 10.1007/s10648-006-9023-2



- Cheng, B., & Titterton, D. M. (1994). Neural networks: A Review from a statistical perspective. *Statistical Science*, 9, 1, 2-54. doi: 10.1214/ss/1177010638
- Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modelling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research*, 14, 189–217. doi: 10.1287/isre.14.2.189.16018
- Decuyper, S., Dochy, F., & Van den Bossche, P. (2010). Grasping the dynamic complexity of team learning: An integrative model for effective team learning in organisations. *Educational Research Review*, 5, 111-133. doi: 10.1016/j.edurev.2010.02.002
- Detienne, K. B., Detienne D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, 6, 236-265. doi: 10.1177/1094428103251907
- Donaldson, R. G., & Kamstra, M. (1999). Neural network forecast combining with interaction effects. *Journal of the Franklin Institute*, 336B, 227-236. doi: 10.1016/S0016-0032(98)00018-0
- Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, 35, 352–359. doi: 10.1016/S1532-0464(03)00034-0
- Edelsbrunner, P., & Schneider, M. (2013). Modelling for Prediction vs. Modelling for Understanding: Commentary on Musso et al. (2013). *Frontline Learning Research*, 2, 99-101.
- Everson, H. T., Chance, D., & Lykins, S. (1994, April). Exploring the use of artificial neural networks in educational research. *Paper presented at the Annual meeting of the American Educational Research Association*, New Orleans, Louisiana.
- Frey, U. J., & Rusch, H. (2013). Using artificial neural networks for the analysis of social-ecological systems. *Ecology and Society*, 18, 40. doi:10.5751/ES-05202-180240.
- Frigg, R. & Hartmann, S. (2006). Models in science. In E. N. Zalta (Ed.), *The Stanford Encyclopaedia of Philosophy*. Summer 2006 Edition. Stanford, CA: Stanford University Press.
- Garson, G. D. (1991). Interpreting neural-network connection weights. *AI Expert*, 6, 47-51.
- Garson, G. D. (1998). *Neural networks. An introductory guide for social scientists*. London: Sage Publications Ltd.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249-264. doi: 10.1016/S0304-3800(02)00257-0
- Golino, H. F., & Gomes, C. M. (2014). Four Machine Learning methods to predict academic achievement of college students: a comparison study. *Manuscript submitted for publication*.



- Grishman, R., & Sundheim, B. (1996). Message Understanding Conference - 6: A Brief History. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING), I, Copenhagen*, 466–471.
- Haenlein, M., & Kaplan, A. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, 3, 283–297. doi: 10.1207/s15328031us0304_4
- Hahn, C., Johnson, M. D., Herrmann, A., & Huber, F. (2002). Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review*, 54, 243- 269.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, MA: MIT Press.
- Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan.
- He, S., & Li, J. (2011). Confidence intervals for neural networks and applications to modeling engineering materials. In C. L. P. Hui (Ed.), *Artificial Neural Networks – Application*. Shanghai, China: InTech. doi: 10.5772/16097
- Intrator, O., & Intrator, N. (2001). Interpreting neural-network results: A simulation study. *Computational Statistics and Data Analysis*, 37, 373–393. doi: 10.1016/S0167-9473(01)00016-0
- Kim, J., & Ahn, H. (2009). A new perspective for neural networks: Application to a marketing management problem. *Journal of Information Science and Engineering*, 25, 1605-1616.
- Kyndt, E., Musso, M., Cascallar, E., & Dochy, F. (2011, August). Predicting academic performance in higher education: Role of cognitive, learning and motivation. *Symposium conducted at the 14th EARLI Conference*, Exeter, UK.
- Kyndt, E., Musso, M., Cascallar, E., & Dochy, F. (2015, in press). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. In V. Donche & S. De Maeyer (Eds.), *Methodological challenges in research on student learning*. Antwerp, Belgium: Garant.
- Laguna, M., & Marti, R. (2002). Neural network prediction in a system for optimizing simulations. *IIE Transactions*, 34, 273-282. doi: 10.1080/07408170208928869
- Lee, C., Rey, T., Mentele, J., & Garver, M. (2005). Structured neural network techniques for modeling loyalty and profitability. *Proceedings of the Thirtieth Annual SAS® Users Group International Conference*. Cary, NC: SAS Institute Inc.
- Lei, P. W., & Qiong Wu, Q. (2007). Introduction to structural equation modelling: Issues and practical considerations. Items – Instructional Topics in Educational Measurement - Fall 2007, *NCME Instructional Module*, 33-43.
- Lek, S., Belaud, A., Baran, P., Dimopoulos, I., & Delacoste, M. (1996). Role of some environmental variables in trout abundance models using neural networks. *Aquat. Living Resour*, 9, 23-29. doi: 10.1051/alr:1996004



- Luft, C. D. B., Gomes, J. S., Priori, D., & Takase, E. (2013). Using online cognitive tasks to predict mathematics low school achievement. *Computers & Education*, *67*, 219-228. doi: 10.1016/j.compedu.2013.04.001
- MacKay, D. J. C. (1992). A practical Bayesian framework for backpropagation networks. *Neural computation*, *4*, 448-472. doi: 10.1162/neco.1992.4.3.448
- Marquez, L., Hill, T., Worthley, R., & Remus, W. (1991). Neural network models as an alternative to regression. *Proceedings of the IEEE 24th Annual Hawaii International Conference on Systems Sciences*, *4*, 129-135. doi: 10.1109/HICSS.1991.184052
- Monteith, K., Carroll, J., Seppi, K., & Martinez, T. (2011). Turning Bayesian Model Averaging into Bayesian Model Combination. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN) 2011*, 2657-2663.
- Musso, M. F., & Cascallar, E. C. (2009a). New approaches for improved quality in educational assessments: Using automated predictive systems in reading and mathematics. *Journal of Problems of Education in the 21st Century*, *17*, 134-151.
- Musso, M. F. & Cascallar, E. C. (2009b). Predictive systems using artificial neural networks: An introduction to concepts and applications in education and social sciences. In M. C. Richaud & J. E. Moreno (Eds.). *Research in behavioural sciences (Volume I)*, (pp. 433-459). Buenos Aires, Argentina: CIIPME/CONICET.
- Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2012). Predicting mathematical performance: The effect of cognitive processes and self-regulation factors. *Education Research International*. Vol 2012, Article ID 250719, 13 pages. doi: 10.1155/2012/250719
- Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1*, 42-71. doi: 10.14786/flr.v1i1.13
- Musso, M. F., Boekaerts, M., Segers, M., & Cascallar, E. C. (in preparation). *A comparative analysis of the prediction of student academic performance*.
- Neal, W., & Wurst, J. (2001). Advances in market segmentation. *Marketing Research*, *13*, 14-18.
- Nguyen, N., & Cripps, A. (2001). Predicting housing value: A comparison of multiple regression and artificial neural networks. *Journal of Real Estate Research*, *22*, 313-336.
- Nokelainen, P. & Silander, T. (2014). Using New Models to Analyse True Complex Regularities of the World: Commentary on Musso et al. (2013). *Frontiers in Psychology*, *3*, 78-82. doi: .org/10.14786/flr.v2i1.107.
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*, 135-150. doi: 10.1016/S0304-3800(02)00064-9



- Olden, J. D., Joy, M. K. & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178, 389-397. doi: 10.1016/j.ecolmodel.2004.03.013
- Orre, R., Lansner, A., Bate, A., & Lindquist, M. (2000). Bayesian neural networks with confidence estimations applied to data mining. *Computational Statistics & Data Analysis*, 34, 473-493. doi: 10.1016/S0167-9473(99)00114-0
- Perkins, K., Gupta, L., & Tamanna (1995). Predict item difficulty in a reading comprehension test with an artificial neural network. *Language Testing*, 12, 34-53. doi: 10.1177/026553229501200103
- Pinninghoff Junemann, M. A., Salcedo Lagos, P. A., & Contreras Arriagada, R. (2007). Neural networks to predict schooling failure/success. In J. Mira & J. R. Alvarez (Eds.), *Nature Inspired Problem-Solving Methods in Knowledge Engineering, (Part II)*, (pp. 571–579). Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/978-3-540-73055-2_59
- Ramaswami, M. M., & Bhaskaran, R. R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science Issues*, 7, 10-18.
- Roli, F., Giacinto, G., & Vernazza, G. (2001). Methods for designing multiple classifier systems. In J. Kittler & F. Roli (Eds.), *Multiple Classifier Systems*, (pp. 78-87). Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/3-540-48219-9_8
- Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511812651
- Rivals, I., & Personnaz, L. (2000). Construction of confidence intervals for neural networks based on least squares estimations. *Neural Networks*, 13, 463-484. doi: 10.1016/S0893-6080(99)00080-5
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33, 1-39. doi: 10.1007/s10462-009-9124-7
- Sahu, A., Runger, G., Apley, D. (2011). Image denoising with a multi-phase kernel principal component approach and an ensemble version. *IEEE Applied Imagery Pattern Recognition Workshop*, 1-7.
- Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modelling. *Frontiers in Psychology*, 5, Article 181, 1-11. doi: 10.3389/fpsyg.2014.00181.
- Suppes, P. (1962). Models of Data. In E. Nagel, P. Suppes & A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 International Congress*. Stanford: Stanford University Press, 252-261.
- Thrush, S. F., Coco, G., & Hewitt, J. E. (2008). Complex positive connections between functional groups are revealed by neural network analysis of ecological time series. *American Naturalist* 171, 669-677. doi: 10.1086/587069



- Tzeng, F. Y., & Ma, K. L. (2005). Intelligent feature extraction and tracking for visualizing large-scale 4D flow simulations. In *DVD Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '05)*. November, 2005.
- Weiss, S. M., & Kulikowski, C. A. (1991). *Computer systems that learn*. San Mateo, CA: Morgan Kaufmann Publishers.
- West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science, 16*, 370-391. doi: 10.1287/mksc.16.4.370
- Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*, 719-751. doi: 10.1177/0011000006286345
- White, H., & Racine, J. (2001). Statistical inference, the bootstrap, and neural network modelling with application to foreign exchange rates. *IEEE Transactions on Neural Networks, 12*, 657-673. doi: 10.1109/72.935080
- Wilson, R. L., & Hardgrave, B. C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational and Psychological Measurement, 55*, 186-195. doi: 10.1177/0013164495055002003
- Yeh, I. C., & Cheng, W. L. (2010). First and second order sensitivity analysis of MLP. *Neurocomputing, 73*, 2225-2233. doi: 10.1016/j.neucom.2010.01.011
- Zambrano Matamala, C., Rojas Díaz, D., Carvajal Cuello, K., & Acu-a Leiva, G. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales. [Analysis of students' academic performance using data warehouse and neural networks] *Ingeniare. Revista Chilena de Ingeniería, 19*, 369-381. doi: 10.4067/S0718-33052011000300007
- Zapranis, A., & Livanis, E. (2005). Prediction intervals for neural network models. *Proceedings of the 9th WSEAS International Conference on Computers (ICCOMP'05)*. World Scientific and Engineering Academy and Society (WSEAS). Stevens Point, Wisconsin, USA.