

Journal Article /// Accepted manuscript (Postprint)

This version is available at <https://nbn-resolving.org/urn:nbn:de:kobv:526-opus4-18377>.

Artificial neural network in soft HR performance management: new insights from a large organizational dataset

Marc Roedenbeck, Petra Poljsak-Rosinski

This is an accepted manuscript of a journal article published by Emerald in *Evidence-Based HRM* (2023), available online at: <https://doi.org/10.1108/EBHRM-07-2022-0171>.



This author accepted manuscript is deposited under a Creative Commons Attribution Non-commercial 4.0 International (CC BY-NC) licence. This means that anyone may distribute, adapt, and build upon the work for non-commercial purposes, subject to full attribution. If you wish to use this manuscript for commercial purposes, please contact permissions@emerald.com.

Title

Artificial Neural Network in Soft HR Performance Management: New Insights from a Large Organizational Dataset

Abstract

Purpose

This study investigates whether the artificial neural network approach, when used on a large organizational soft HR performance dataset, results in a better (R^2 /RMSE) model compared to the linear regression. With the use of predictive modelling, a more informed base for managerial decision making within soft HR performance management is offered.

Design/methodology/approach

The study builds on a dataset ($n > 43k$) stemming from an annual employee MNC survey. It covers several soft HR performance drivers and outcomes (such as engagement, satisfaction and others) that either have evidence of a dual-role nature or non-linear relationships. This study applies the framework for artificial neural network analysis in organization research (Scarborough & Somers, 2006).

Findings

The analysis reveals a substantial artificial neural network model performance ($R^2 > 0.75$) with an excellent fit statistic ($nRMSE < 0.10$) and all drivers have the same relative importance (RMI [0.102; 0.125]). This predictive analysis revealed that the organization has to increase six of the drivers, keep two on the same level and decrease one.

Originality

Up to date, this study uses the largest dataset in soft HR performance management.

Additionally, the predictive results reveal that specific target values lay below the current levels to achieve optimal performance.

Keywords

Soft HRM, Performance, Drivers, Artificial Neural Network, Non-Linearity, Prediction

Introduction

Most management areas, including human resource management (HRM), have long focused on how managerial actions influence firm performance relative to its competitors, as well as its development over time (Richard et al., 2009). Modern organizations are interested in the contribution that particularly soft (collaborative) HRM can make to the overall performance (Caputo et al., 2019). Employees are considered as an added value to the organization (Guest, 1987). Thus, there is great interest in investigating employee performance drivers that evolve around their skills, motivation and work conditions (Becker & Huselid, 1998).

To study performance drivers and outcomes, organizations, consulting agencies and researchers predominantly develop and test causal relationships. They apply simple causality-based data analysis, such as (multiple/multivariate) linear regression (MMR), path analysis, structural equation modeling, or time series analysis (Grznar et al., 2007). Even when dealing with large-scale data, the analysis often relies on a linear or polynomial relationship between dependent and independent variables (George et al., 2016). The main aim is to identify drivers for achieving better performance outcomes. However, such relationships are often questioned by researchers because of poor to medium-range correlations and low variance explanations (Scarborough & Somers, 2006). Such questionable results are then used by HR managers to derive performance enhancing initiatives (Jiang et al., 2012). However, if the analysis and the results themselves are questionable, then the contribution of the subsequent initiatives may be only a matter of doing-something is better than nothing (Briner 2000).

It is increasingly evident that HRM practitioners and researchers should use more sophisticated methods of analysis. Artificial neural networks (ANNs) have been suggested to outperform current statistical methods (Stavrou et al., 2007). These can be used to explain complex phenomena more accurately. Unlike regressions, ANNs are a Data Mining technique without the need for causal theory and a priori model identification assumptions.

Their main strength is the ability to learn and detect non-linear patterns within datasets that can then be used for prediction (Kaastra & Boyd, 1996). Thus, such “data-driven” approaches can uncover more precise insights into an organization's soft HR performance management and produce more accurate decisions.

Research in the field of soft HRM and related organizational behaviour (OB) topics has used the ANN approach since the early 90s (Collins & Clark, 1993) and is currently the second most studied methodological approach in the general field of HR (Garg et al., 2022). These methods are however rarely employed in practitioner-oriented and evidence-based HRM research. Thus, the main contribution of this paper is to show how applying the ANN methodology may help practitioners in arriving to a better base for decisions regarding subsequent performance enhancing initiatives.

This study builds on a survey-based dataset with more than 49,000 employees and managers provided by a multinational pharmaceutical corporation (MNpC) with international offices in more than 100 countries. First, we create an ANN that simultaneously analyses all identified soft HR performance drivers and outcomes with evidence of nonlinear relationships (e.g. performance – work-life balance, performance – satisfaction). Second, we feed our ANN with a hypothetical dataset of all possible solutions, so we can predict the best possible soft performance management outcome. We believe ANNs have a unique value for scholars and practitioners when used as a predictive tool for context-specific soft performance management analyses. We thus pose the following research questions:

- 1) Applied to a large organizational soft HR performance dataset, does the ANN produce a better (R^2 /RMSE) model than linear regression?
- 2) Compared with the linear regression, what HR management fields of action are revealed by the ANN prediction approach?

Theoretical Background

In order to better position the ANN research within the soft HR performance field, we first outline studies that discuss non-linear relationships between soft HR performance drivers and outcomes. This is a fundamental argument for the use of ANNs instead of the MMR.

Thereafter, we summarize the existing limitations of the ANN research within the soft HR performance field.

Soft HR Performance Drivers and Outcomes: Non-linear and Dual-Role Arguments for the ANN Approach

According to this study's soft HRM perspective, employees perform best when they are fully engaged with the organization (Truss et al., 1997). Unlike hard objective performance measures (e.g. return on investment), soft performance measures tend to be subjective and rely on collective employee-based assessments of performance outcomes. These are comparable across the organization and often take into account specific organizational contexts, such as for example R&D (Kerssens-van Drongelen et al., 2000). Previous empirical research has however asserted a close correlation between subjective and objective performance measures (Wall et al., 2004).

In order to investigate individual and organizational performance outcomes, understanding the organizational context plays an important role (Den Hartog et al., 2004). Organizations strive to create a context and manage drivers that empower employees to have a positive experience at work and thus perform better. Linear approaches are more common when investigating soft HR performance drivers and outcomes. Table I outlines examples of studies using a non-linear perspective. Non-linearity includes curvilinear (continuous) as well as non-monotonous relationships (Losada & Heaphy, 2004). A sole focus on linear relationships may lead to small effect sizes, errors and non-supportive results (de Langhe et al., 2017).

With the ANN framework (detailed in the next section), it is possible to identify multiple non-linear relationships within datasets, which can then be used for prediction (Kaastra & Boyd, 1996). ANNs have been advocated to be a very fitting choice when dealing with large datasets that have multiple independent (inputs) and dependent (outputs) variables that interact in a linear as well as non-linear manner (Grznar et al., 2007).

In performance research, a number of constructs are recognized as drivers in some studies and outcomes in others (see Table I). Considering these dual-role constructs within one model is likely to cause multicollinearity. This can have a detrimental effect on traditional linear regression analysis, which makes the results at the very least questionable (Grznar et al., 2007). ANNs however have the power to overcome multicollinearity data concerns as they use a specific activation function that transforms the data so that the final inputs are uncorrelated (Morlini, 2006).

Using the Document Search function in the Web of Science, Table I presents studies based on a co-occurrence of performance and constructs. As well as aiming for a wide range of co-occurrence frequencies (Min: 655; Max: 46.885; Average: 10.794,6), we also reflected whether the constructs have non-linearity and dual roles (see below).

--- INSERT TABLE I HERE ---

As evidenced from Table I, employee engagement has been used as an outcome in performance studies (Zyman & Brott, 2002) but it is also commonly used as a driver for performance (Harter et al., 2002). Work-life balance has also been widely researched as a driver for performance (Perry-Smith & Blum, 2000). Other studies evidence that there are contextual factors that influence work-life balance (Keeton et al., 2007). Thus, both constructs can be considered to have a dual-role nature.

Furthermore, Somers and Casal (2009) have shown a (curvilinear) interdependency between performance (outcome) and satisfaction as well as commitment and engagement (drivers) resulting in a curved 3D surface. There is also evidence that engagement (driver) and performance (outcome) support curvilinear logic (Shimazu et al., 2018), as does the relationship between customer orientation (driver) and sales performance (outcome) (Homburg et al., 2011). Furthermore, there is support for the non-linear relationship (non-monotonous depending on task complexity) between occupational safety (driver) and performance-related outcomes (Wallace et al., 2008). Non-linearity (curvilinear) effects between work-life balance (driver) and performance (outcome) have also been discussed, particularly in relation to working time and life satisfaction (Drobnic et al., 2010). The theory has also indicated an unspecified non-linear perspective on leadership (driver) and innovation performance (outcome) (Goldstein et al., 2010).

In light of the discussed aspects and the benefits of the ANN approach, we now summarize the existing ANN studies within the field and current limitations.

ANN studies in soft HR performance management

ANNs are “non-linear, complex, and parallel” computing systems or machines that model any system according to the neurobiological brain (Haykin, 2005). They have been advocated to be a very fitting choice when dealing with large datasets that have multiple independent (inputs) and dependent (outputs) variables and interact in a linear as well as non-linear manner (Grznar et al., 2007). The general structure of ANNs (see Figure 1) consists of a number of given output and input neurons. These are equivalent to the number of independent and dependent variables. There is at least one hidden layer between them, which has a set number of hidden neurons. Every neuron is interlinked with other neurons on the subsequent layer with a defined activation function.

--- INSERT FIGURE 1 HERE ---

Our analysis of the studies outlined in Table II identified three types of limitations that affect ANN results and its future application in soft HR performance management: 1) standardized ANN framework limitations and updates, 2) sample limitations and 3) prediction limitations.

--- INSERT TABLE II HERE ---

First, the ANN approach, which comes from computer science and biology, was rather unstructured in its early days in OB / HRM research (e.g. Collins & Clark, 1993). Thus, Scarborough and Somers (2006) introduced a standard framework for the application and evaluation of ANNs for prediction in organization research. It entails 6 steps.

--- INSERT FIGURE 2 HERE ---

It is rather straightforward to select a multilayer perceptron network for OB/HRM rather than a convolutional network for image analysis (step 1), define the input/output structure of the network based on the number of variables (step 2) and pre-process data with normalization and outlier deletion (step 3). As for the quantitative assessment of an ANN, it is compared to an ordinary least squares regression model (OLS). This involves examining the root mean squared error (RMSE) and explained variance (R^2) for the training and test data (step 5). It is also straightforward to evaluate qualitatively using a graphical sensitivity analysis as a 3D response surface mapping two criteria and one predictor (step 6).

It is however more challenging to estimate hidden neurons and their activation function in the network (see Figure 1) as well as determine the number of input/output neurons (step 2).

Model performance is affected by the number of hidden nodes on a hidden layer and the number of hidden layers in the ANN (step 2). Following the selection of the activation function and random initial parameters (step 4), steps 5 and 6 consist of the initial evaluation.

In case of a poor evaluation, the researcher can adjust the model design (step 2) and the model parameters (step 4). A good model will be identified via this recursive process. It is transparent to report parameters changed during the process or at least the final parameters. Unfortunately, the framework does not explain how to use the trained model for prediction.

In Table II, only five of eight articles (62.5%) published since 2006 cited and related their method to the framework. Even though the framework gets mentioned, its application is often imprecise. Many studies failed to indicate the final ANN used (2 of 8), failed to include outlier analysis (7 of 8), failed to display how changes in model parameters affected model performance (7 of 8), omitted the RMSE/MSE (4 of 8) or overlooked the response surface analysis (2 of 8).

It has been suggested that step 2 can be enhanced by starting with the Kolmogorov rule [1] for hidden neurons instead of a rule-of-thumb [2] and model improvements can be achieved by experimentation (stepwise increases/decreases) or genetic algorithms. Step 4 can be improved by showing how the hidden nodes and learning rates affect the evaluation (step 5 & 6). Step 5 can be enhanced by adding regression techniques (e.g. Logistic Regression or Robust Regression), adding further criteria (e.g. SSE, F-value, MAPE, CVR, Accuracy), as well as performing statistical tests between the different models (via Wilcoxon matched pairs signed ranks test, Fisher Z-test, or Student t-test).

The second limitation deals with sample size. In general, regression methods can cope with the sample sizes reported in Table II (average 391.1, median 277). ANNs however, are particularly “data hungry”, where a sample size of 740 is just enough for demonstration purposes (Shah et al., 2020). There is only one dataset in the table that exceeds the 740 by 263 ($n = 1,003$). The main problem of small sample sizes is that ANNs need training and test-data for validation. In order to test cross-validation, ANNs require at least 10% (Collins &

Clark, 1993) of the input data as test data. Based on the studies reviewed, the average test sample size was 89.6, leaving 76.6% for the training sample. Unfortunately, the smaller the sample, the smaller the test data, and the higher the risk of overfitting. This means that although the ANN will learn to replicate the data exactly, it would not be suitable for predictive tasks.

Last limitation pertains to the use of trained ANNs in prediction (Kaastra & Boyd, 1996), which is lacking in the reviewed studies. Only Montagno, Sexton and Smith (2002) used the trained and tested ANN to predict the best management system: they tested an artificial combinatorial input dataset and found the best management system via the trained ANN. Unfortunately, the prediction uses dichotomies (Yes/No) instead of a qualitative evaluation (e.g. 5-point Likert scale). In all other studies, analysis stopped at comparing training and test performance without even attempting prediction.

We now apply the discussed ANN framework to our dataset.

Methodological Framework

We first describe the large dataset at hand, then check for validity and reliability, and conduct a congruence analysis (Drost, 2011). Finally, we explain important decisions that were taken whilst applying the ANN research framework of Scarborough and Somers (2006). This should in turn aid in further standardization of the framework's application in research and practice.

Dataset

The data was contributed by the MNpC for research purposes, which had international offices in more than 100 countries and approximately 51,500 employees at the time of analysis. The HR department of the MNpC collected the data via an annual employee survey that entails 49,680 managerial and employee responses. This equals to a response rate of 96.4% which is

well above the reported average (48%) of HR related studies (Baruch & Holtom, 2008). The survey consisted of 82 items pertaining to 9 constructs measured, which is more than four million data points. The quota with a minimum of one missing value within all of the 82 items was 0.92% (i.e., 461 missing values). Thus, the final size of the dataset for further analysis amounted to 49,219 as all cases with missing values were excluded.

Measures

The annual employee survey itself was designed as a joint project by members of the MNpC, a consultancy firm as well as a team of researchers. The scale and item discussion relied on numerous contributions in the field of soft HR performance management research. The selection of scales, the sub-selection and redesign of items as well as answer styles were conducted in a participative HR process. Thus, adjustments have been made based on the feedback received from general employees, members of the working council, and from managers across all levels and functions.

At the scale level, all constructs reported above to have a non-linear or dual-role nature in soft HR performance research have been included. Two major changes have been made: First, the construct of general Job Satisfaction had been split up into specific aspects of the job (Spector, 1985), namely Immediate Manager, Pay & Reward System, Individual Development and Information Sharing. Second, the MNpC insisted to include Organizational Effectiveness as a construct, although we could not identify any evidence about non-linearity or a dual-role nature. In line with Richard et al. (2009) the MNpC regarded the Effectiveness construct crucial for performance and, thus, wanted it to be part of the study.

At the item level, Contextual Performance (CP) was covered by ten items linked to the three dimensions of interpersonal (e.g. 'In the teams in which I work, we are held accountable for the results we deliver'), organizational (e.g. 'I fell my ideas are actively considered') and job-

task ('My last performance review discussion helped me improve my job performance') performance (Johnson, 2001). Customer Orientation (CO) was addressed by eight positive items related to the SOCO scale (Saxe & Weitz, 1982), but the items were formulated in relation to the team (e.g. 'Teams in which I work, constantly look for better ways to serve their customers'). Employee Engagement (EN) was measured via nine items covering emotional ('I am proud to work for the organization'), behavioural ('I am willing to put a great deal of effort beyond what is normally expected') and cognitive ('I believe strongly in the organizations future direction') engagement (Shuck et al., 2017). Measuring Leadership (LE) was done by an evaluation of core behavioural leadership competencies (Oyinlade, 2006), linked to the companies' leadership model (in total nine, e.g. 'Senior Leaders make effective decisions that drive forward' or 'Senior leaders provide a clear direction'). Organizational Justice (OJ) focused on procedural ('I have not encountered any bias or discrimination towards myself and/or others') and relational ('Managers in my function support equal opportunity for all colleagues') justice with seven items in total (Sweeney & McFarlin, 1993). While the safety construct entails several different aspects (Díaz-Cabrera et al., 2007), the MNpC measured Organizational Safety (OS) using five items centred around the existence of rules and practices (e.g. 'There are adequate security measures' and 'My function encourages colleagues to put safety first'). As mentioned above, the general Job Satisfaction was split into four constructs; Immediate Manager (IM, in total ten, e.g. 'My immediate manager coaches me to develop my strengths'), Pay & Reward System (PR, in total four, e.g. 'How satisfied are you with your salary'), Individual Development (ID, in total four, e.g. 'I have the opportunity for personal development and growth') and Information Sharing (IS, in total four, e.g. 'The organization does an excellent job of keeping colleagues informed'). Work-Life Balance (WL) was measured with a scale linked to its three major indicators (Pichler, 2008), by three corresponding items (e.g. 'My work schedule allows

sufficient flexibility to meet my personal/family needs’). Finally, Organizational Effectiveness (EF) was measured on the basis of the employee evaluation of a mix of structural (e.g. ‘The organization has reduced the amount of bureaucracy in the last 12 months’) as well as operational (e.g. ‘Actions that have been agreed are then implemented effectively’) aspects, but it did not include any additional financial or attitudinal items (Sharma & Reetesh K., 2019). Items were measured using a five-point Likert scale (i.e., Agree/.../Neutral/.../Disagree).

Validity, reliability and congruence analysis

All participants involved in the development phase agreed that the final survey offered good content and face validity. Since Confirmatory Factor Analysis (CFA) does not require meeting the distributional data assumption (Kaplan, 2009), we performed it for testing the construct validity using statistical software R with lavaan. Ideally, standardized loading estimates should be 0.7 or higher with 0.5 as the recommended minimum cut-off value (Hair et al., 2006). The findings in Table III reveal that 10 items fall short of the minimum loading threshold. These were consequently removed as they did not fit the expected model.

(INSERT TABLE III ABOUT HERE)

The reliability analysis was conducted using R with psych package, and Cronbach’s alpha statistics are reported in Table IV. The minimum threshold of acceptability lies between 0.95 and 0.65, values are poor and questionable if between 0.5 and 0.65 (DeVellis, 2012). Most of our scales were reliable (alpha statistics were between 0.95 and 0.65). Two scales were questionable (PR, WL) as their alpha statistic did not meet the 0.65 threshold. Nevertheless, both scales were above the absolute minimum of 0.5. The MNpC considered these two scales particularly important for soft performance, and since the discussed measures (i.e., WL) used

in other scholarly work (Sturges & Guest, 2004) also rely on one or two-item scales, we cautiously decided to keep the two scales for the subsequent analysis.

(INSERT TABLE IV ABOUT HERE)

For the final congruence analysis, we mean-centred items to corresponding scales, whereby the scales were recoded so that the highest values show agreement. Table V summarizes the descriptive statistics with all mean scores (M) close to 4.0, all standard deviations (SD) close to 0.8, skewness close to -1.0 and a kurtosis in the range between 0.09 and 1.17. The Kolmogorov-Smirnov test (KST) values (Table V) show that all scales are non-normally distributed.

(INSERT TABLE V ABOUT HERE)

Prior to correlation analysis, the data had been Z-transformed. The correlation matrix in Table VI shows significant inter-correlations among the different constructs with one very strong correlation coefficient (0.9) and thus implies a high risk of multicollinearity (Dohoo et al., 1997). This was confirmed with the $|X'X|$ determinant, the Farrar Chi-Square, and the Red indicator tests. However, as mentioned above, ANNs have the power to overcome multicollinearity (Morlini, 2006).

(INSERT TABLE VI ABOUT HERE)

Preparations for the ANN Approach

As introduced above and summarized in Figure 2, within the domain of performance-related HR research, there is a standardized framework for the ANN approach given by Scarborough and Somers (2006). Their framework is anchored in data mining, which means that no model assumptions are required and tested. It calls for several straightforward decisions. Thus, we first used the suggested multilayer perceptron network (with a feed-forward, back-propagation algorithm). Second, we defined the network architecture based on the discussed

drivers and outcomes with nine input and five output nodes. To determine the number of hidden layers as well as nodes per layer we followed the suggested rule-of-thumb rather than the above outlined alternatives, because of a smaller interval for the number of used neurons [3]. Third, we normalized the data and did an outlier analysis that resulted in a final sample size of 43,567 cases. As introduced in our review above, a cross-validation is done with at least 10% of the dataset. Due to the large sample size at hand, we decided to use 20% of the data as the training dataset (i.e., 8,707 cases), exceeding the current total sample size with the factor $k=8$, and the rest was used for the test dataset (i.e., 34,860 cases) to evaluate model performance. Applying the rule-of-thumb resulted in a maximum of 124 hidden nodes $((20\%*43,567/5)/(9+5))$ in one hidden layer, or spread across two. In the fourth step, we went with the network parameters that were randomly defined by the software (learning rate 0.7 and momentum 0.7) and made use of the standard sigmoid activation function.

The fifth step normally entails the report of results and model parameter adjustments, but as covered in the theoretical part some important changes were introduced that have to be discussed. First, in order to evaluate the model performance, we report the RMSE and variance explained (R^2). R^2 ranges between substantial (>0.75), moderate (>0.5) or weak (>0.25) model performance (Hair et al., 2006). Since our data had been normalized in the pre-processing, we were able to interpret the RMSE outputs as a normalized RMSE [nRMSE] as a percentage measure (Loague & Green, 1991). An excellent fit is achieved with nRMSE value below 10%, good fit with nRMSE below 20%, and a fair fit with nRMSE below 30% (Jamieson et al., 1991). A further check of the model fit entails comparing the similarity of the nRMSE between test and training datasets. If the values are similar, then the problem of overfitting has been avoided (Rosin & Fierens, 1995). Other criteria, that were previously introduced, were not used, because they are either linked to the RMSE (SSE) or they are used for classification tasks (F, Accuracy). Second, Scarborough and Somers (2006) suggested an

OLS regression for comparing NN results to a base-line model, while the above mentioned studies proposed logistic regression (classification) or robust regression (if the non-linear function is known). Since our aim is not about classification and the model function is unknown, we decided to use the OLS regression as suggested in the framework. These results are only useful as a baseline because regression assumptions such as linear relationships, multivariate normality and homoscedasticity are violated. Third, the framework was enhanced with statistical tests that compare the regression and ANN results. In line with Scarborough & Somers (2006), we focus on the measures of R^2 and RMSE instead of the results comparison. In order to obtain more information out of the ANN, we suggest the use of the Garsons' RMI score. It enables the evaluation of the individual factor's relevance within the ANN model (Zhang et al., 2018) and can be interpreted in a similar manner to regression coefficients. Starting at 124 hidden nodes and after several runs, our final model consisted of 84 nodes on 1 hidden layer and had been run for 137 epochs.

In step 6, a qualitative 3D surface analysis was conducted without changes to the suggested procedure. This analysis step entails the examination of the graphical relationships (Edwards & Parry, 1993), with the aim of checking whether the factors analysed follow a flat (non-curved) or a curved 3D surface. The existence of the latter serves as evidence of a non-linear relationship in the data.

After the outlined six steps, the final step involves the use of the trained network for prediction. Following (Montagno et al., 2002), we generated an artificial combinatorial dataset for our input variables. This means that, our dataset had to cover 9 input scales, each with the same possible spread [1 to 5]. One combination of inputs with positive integers would look like [1/5/5/4/2/5/1/1/1] and this procedure results in $5^9=1.9$ million combinatorial cases.

Results

Table VII summarizes the OLS regression and ANN results. Step 5 results indicate a substantial model performance for ANN in terms of the explained variance ($R^2 > 0.75$) and an excellent fit statistic ($nRMSE < 0.10$) for the training and test datasets. The training dataset model performance is better than that of the test dataset model, which indicates that overfitting has been avoided. The regression results in a poor model fit ($nRMSE > 0.30$) and a substantial variance explained ($R^2 > 0.75$). The R^2 similarity likely stems from the multicollinearity of the variables used. The difference in the $nRMSE$ can be potentially rooted in the non-linear relationships among the drivers and outcomes.

(INSERT TABLE VII ABOUT HERE)

The qualitative analysis in step 6 (i.e. graphical non-linearity check) would need to cover 360 [4] combinatorial 3D analyses to show whether curved 3D surfaces are present instead of flat (non-curved) ones. To exemplify our results, two of the 360 solutions are presented in Figure 3. There are clear non-linear relationships in both graphs. On the left side, the relationships between Immediate Manager and Contextual Performance are either represented by a solid line (i.e. when Work-Life Balance is low) or the dotted line (i.e. when Work-Life Balance is high). This provides evidence for different (upward or downward) curving effects. The overall 3D surface (grid area) results in the highest R^2 (0.82) when a polynomial of 3rd degree is used. On the right side, the relationships between Engagement and Contextual Performance are either represented by a solid line (i.e. when Work-Life Balance is low) or the dotted line (i.e. when Work-Life Balance is high). Again, this points to different (downward or polynomial) curving effects, with the highest R^2 (0.62) achieved when a polynomial of 4th degree is used. More than 90% of these 3D surfaces have a higher R^2 in the case of a 2nd degree polynomial or higher, which ascertains non-linearity.

(INSERT FIGURE 3 ABOUT HERE)

Within the predictive analysis, we computed the sum of all predicted output variables based upon our combinatorial dataset. Then we picked the combination of input values with the highest sum for their corresponding outputs. This combination reflects the optimal values of performance drivers and outcomes that can then be aspired by the management. For example, if the current mean of Individual Development (4.09) is lower than the predicted optimal value (5), then management needs to invest organizational resources so that a higher ID score is achieved in the next annual survey. Six of the current means of the drivers are below the predicted ANN optimal values, 2 are nearly there and 1 is even higher than the predicted optimal value. Garsons' RMI scores attest to highly similar relative importance of all factors within the model [0.102; 0.125], which means that all factors need to be considered. These will be further addressed in the next section.

(INSERT TABLE VIII ABOUT HERE)

Discussion

In this study, we outline the current state of research on ANNs in the soft HR performance field, highlight their advantages when analyzing complex non-linear organizational models with multiple dual-role constructs, and showcase their underutilized predictive usage. Using the largest known sample size in the soft HR performance field (43,567 cases), we attempted to mediate a potentially serious small sample size issue for ANN application and results.

To address our first research question, we show that the ANN provides a better fit ($R^2 > 0.75$; $nRMSE < 0.10$), leading to more informed managerial decisions than OLS regressions ($nRMSE > 0.30$, $R^2 > 0.75$). We indeed found evidence for nonlinear relationships between drivers, such as Work-Life Balance and Satisfaction with Immediate Manager, and Contextual Performance outcome.

Predictive use of the ANN approach revealed interesting results for research question two. The close range of the drivers' relative importance [0.102; 0.125] showed that the organization should pay attention to all performance drivers when deriving soft HRM initiatives. According to the predictive analysis, there are several drivers that need be targeted in different ways. The following arguments only serve as an example of what these possible performance initiatives may look like for our MNpC. Engagement and Leadership should be increased for superior Contextual Performance. Increasing Engagement could entail rethinking job design and person-environment fit (Attridge, 2009). Despite the current values not reaching the highest possible levels, Information Sharing and Satisfaction with Immediate Manager have to be maintained. Scientific work is the biggest stimulus for scientists at the MNpC, so it may be that they pay less attention to the organizational social conditions (Keller et al., 1996). As such, the organization should monitor levels of these drivers. Interestingly, to improve performance, the MNpC has to decrease Work-Life Balance. One possible explanation may lie in the organizational context of MNpC. Given the complicated and uncertain nature of the tasks scientists carry out (Ryan & Hurley, 2007), their work already entails a substantial amount of autonomy and thus long working hours.

Conclusion

Applying the ANN offers the company's HR department an evidence-based recipe for success in terms of facilitating data-driven decision making in the field of soft HRM performance. It is beneficial for HR managers to make use of the predictive ANN approach as a tool for more informed decision making (i.e. by identifying the best input combination to achieve the highest outcome possible). This way the managers have a better idea which performance drivers need to be increased or decreased in line with the employee perception. ANN, however, requires a fairly advanced analytical skillset that is rarely available today in HR departments and imposes certain organizational costs. Investing in training (current HR

employees trained in data analysis methods), recruitment (hire new HR data scientists) or external consultants (as specialists in HR or people analytics) is thus required.

Although our method revolved around the use of the ANN approach and its stepwise application for prediction, we acknowledge we cannot rule out that our single-organization approach limits generalizability and external validity. These data mining results are valid for the analysed organisation. Thus, we recommend that future studies use the ANN approach across different organizational settings (e.g., geographies, sizes, or industries), to further investigate its predictive function and compare the results with the study at hand. The generalizability of such data mining approaches would increase, if different organizational datasets were used in independent studies using the same or similar survey instruments. In addition, a meta-analysis could provide insights from a comparative perspective.

Notes

[1] Number of hidden neurons is equal to $(input*2)$ vs the General approach $((input + output) / 2)$.

[2] At least five training examples for every connection in the neural network should be available (Scarborough & Somers, 2006).

[3] See note 2, while the number of connections equals the number of hidden nodes multiplied by the sum of inputs and outputs; that is $cases/5 = hidden\ nodes*(inputs+outputs)$. The number of hidden layers shall not exceed two, in case of using backpropagation.

[4] There are $9*8$ possibilities and if two out of nine input variables are selected, the result is then combined with one out of five output variables equals to $9*8*5=360$.

Bibliography

- Attridge, M. (2009). Measuring and Managing Employee Work Engagement: A Review of the Research and Business Literature. *Journal of Workplace Behavioral Health, 24*, 383–398.
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations, 61*(8), 1139–1160.
- Becker, B. E., & Huselid, M. A. (1998). High performance work systems and firm performance: A synthesis of research and managerial implications. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (pp. 53–101). JAI Press.
- Caputo, F., Cillo, V., Candelo, E., & Liu, Y. (2019). Innovating through digital revolution: The role of soft skills and Big Data in increasing firm performance. *Management Decision, 57*(8), 2032–2051.
- Collins, J., & Clark, M. (1993). An application of the theory of neural computation to the prediction of workplace behavior: An illustration and assessment of network analysis. *Personnel Psychology, 46*, 503–524.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*, 386–400.
- de Langhe, B., Puntoni, S., & Larrick, R. (2017). Linear thinking in a nonlinear world. *Harvard Business Review, 95*(3), 130–139.
- Den Hartog, D. N., Boselie, P., & Paauwe, J. (2004). Performance Management: A Model and Research Agenda. *Applied Psychology, 53*(4), 556–569.
- DeVellis, R. F. (2012). *Scale development: Theory and applications*. Sage.

- Díaz-Cabrera, D., Hernández-Fernaud, E., & Isla-Díaz, R. (2007). An evaluation of a new instrument to measure organisational safety culture values and practices. *Accident Analysis and Prevention, 39*, 1202–1211.
- Dohoo, I., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine, 29*(3), 221–223.
- Drobnic, S., Beham, B., & Präg, P. (2010). Good Job, Good Life? Working Conditions and Quality of Life in Europe. *Social Indicators Research, 99*, 205–225.
- Drost, E. (2011). Validity and Reliability in Social Science Research. *International Perspectives on Higher Education Research, 38*(1), 105–124.
- Edwards, J., & Parry, M. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577–1614.
- Feyer, A.-M., & Williamson, A. (2004). *Occupational Injury. Risk, Prevention and Intervention*. Taylor & Francis.
- Garg, S., Sinha, S., & Kar, A. K. (2022). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management, 71*(5), 1590–1610.
- George, G., Osinga, E. C., Dovev, L., & Scott, B. A. (2016). Big Data and Data Science Methods for Management Research. *Academy of Management Journal, 59*(5), 1493–1507.
- Goldstein, J., Hazy, J. K., & Lichtenstein, B. B. (2010). *Complexity and the Nexus of Leadership—Leveraging Nonlinear Science to Create Ecologies of Innovation*. Palgrave MacMillan.

- Grznar, J., Prasad, S., & Tata, J. (2007). Neural networks and organizational systems: Modeling non-linear relationships. *European Journal of Operational Research*, 181, 939–955.
- Guest, D. (1987). Human resources management and industrial relations. *Journal of Management Studies*, 24(5), 503–521.
- Hair, J., William, C., Barry, J., Rolph, E., & Ronald, L. (2006). *Multivariate data analysis*. Pearson Prentice Hall.
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology*, 87(2), 268–279.
- Haykin, S. (2005). *Neural Networks—A Comprehensive Foundation*. Pearson.
- Homburg, C., Mueller, M., & Klarmann, M. (2011). When should the customer really be the king? On the optimum level of salesperson customer orientation in sales encounters. *Journal of Marketing*, 75(2), 55–74.
- Jamieson, P. D., Porter, J. R., & Wilson, D. R. (1991). A test of computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. *Field Crops Research*, 27, 337–350.
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How Does Human Resource Management Influence Organizational Outcomes? A Meta-Analytic Investigation of Mediating Mechanisms. *Academy of Management Journal*, 55, 1264–1294.
- Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgements of overall performance. *Journal of Applied Psychology*, 86(5), 984–996.
- Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(215–236).

- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions*. SAGE.
- Karanika-Murray, M., & Cox, T. (2010). The use of artificial neural networks and multiple linear regression in modelling work–health relationships: Translating theory into analytical practice. *European Journal of Work and Organizational Psychology, 19*(4), 461–486.
- Keeton, K., Fenner, D. E., Johnson, T. R. B., & Hayward, R. A. (2007). *Predictors of Physician Career Satisfaction, Work–Life Balance, and Burnout. 109*(4), 949–955.
- Keller, R., Julian, S., & Kedia, B. (1996). A multinational study of work climate, job satisfaction, and the productivity of R&D teams. *IEEE Transactions on Engineering Management, 42*(1), 48–55.
- Kerssens-van Drongelen, I., Nixon, B., & Pearson, A. (2000). Performance measurement in industrial R&D. *International Journal of Management Reviews, 2*(2), 111–143.
- Ladstätter, F., Garrosa, E., & Dai, J. (2014). Neural Network Analysis of Nonlinear Effects of Hardiness on Burnout in Chinese Nurses. *Open Journal of Social Sciences, 2*, 96–99.
- Ladstätter, F., Garrosa, E., Moreno-Jiménez, B., Ponsoda, V., Aviles, M., & Dai, J. (2016). Expanding the occupational health methodology: A concatenated artificial neural network approach to model the burnout process in Chinese nurses. *Ergonomics, 59*(2), 207–221.
- Loague, K., & Green, R. E. (1991). Statistical and graphical methods for evaluating solute transport models; overview and application. *Journal of Contaminant Hydrology, 7*(1–2), 51–73.
- Losada, M., & Heaphy, E. (2004). The Role of Positivity and Connectivity in the Performance of Business Teams: A Nonlinear Dynamics. *American Behavioral Scientist, 47*(6), 740–765.

- Minbashian, A., Bright, J., & Bird, K. (2010). A Comparison of Artificial Neural Networks and Multiple Regression in the Context of Research on Personality and Work Performance. *Organizational Research Methods, 13*(3), 540–561.
- Montagno, R., Sexton, R. S., & Smith, B. N. (2002). Using neural networks for identifying organizational improvement strategies. *European Journal of Operational Research, 142*, 382–395.
- Morlini, I. (2006). On multicollinearity and concavity in some nonlinear multivariate models. *Statistical Methods and Applications, 15*, 3–26.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence That Task Performance Should Be Distinguished from Contextual Performance. *Journal of Applied Psychology, 79*(4), 475–480.
- Ogbonna, E., & Harris, L. C. (2000). Leadership style, organizational culture and performance: Empirical evidence from UK companies. *International Journal of Human Resource Management, 11*(4), 766–778.
- Oyinlade, A. O. (2006). A Method of Assessing Leadership Effectiveness. *Performance Improvement Quarterly, 19*(1), 25–40.
- Perry-Smith, J. E., & Blum, T. C. (2000). Work-Family Human Resource Bundles and Perceived Organizational Performance. *Academy of Management Journal, 43*(6), 1107–1117.
- Pichler, F. (2008). Determinants of Work-Life Balance: Shortcomings in the Contemporary Measurement of WLB in Large-scale Survey. *Social Indicators Research, 92*, Art. 449.
- Richard, P. J., De Vinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring Organizational Performance: Towards Methodological Best Practice. *Journal of Management, 35*(3), 718–804.

- Rosin, P. L., & Fierens, F. (1995). Improving neural network generalisation. *International Geoscience and Remote Sensing Symposium*, 95(2), 1255–1257.
- Ryan, J. C., & Hurley, J. (2007). An empirical examination of the relationship between scientists' work environment and research performance. *R&D Management*, 37(4), 345–354.
- Saxe, R., & Weitz, B. A. (1982). The SOCO Scale: A Measure of Customer Orientation of Salespeople. *Journal of Marketing Research*, 19(3), 343–351.
- Scarborough, D., & Somers, M. J. (2006). *Neural Networks in Organizational Research. Applying Pattern Recognition to the Analysis of Organizational Behavior*. American Psychological Association.
- Shah, S. A. A., Uddin, I., Aziz, F., Ahmad, S., Al-Khasawneh, M. A., & Sharaf, M. (2020). An Enhanced Deep Neural Network for Predicting Workplace Absenteeism. *Hindawi Complexity*, 2020, 1-12 (ID5843932).
- Sharma, N., & Reetesh K., S. (2019). A unified model of organizational effectiveness. *Journal of Organizational Effectiveness: People and Performance*, 6(2), 114–128.
- Shimazu, A., Schaufeli, W., Kubota, K., Watanabe, K., & Kawakami, N. (2018). Is too much work engagement detrimental? Linear or curvilinear effects on mental health and job performance. *PLoS ONE*, 13(12), e0208684.
- Shuck, B., Adelson, J. L., & Reio, T. G. (2017). The employee engagement scale: Initial evidence for construct validity and implications for theory and practice. *Human Resource Management*, 56(6), 953–977.
- Somers, M. J. (1999). Application of Two Neural Network Paradigms to the Study of Voluntary Employee Turnover. *Journal of Applied Psychology*, 84(2), 177–185.

- Somers, M. J. (2001). Thinking differently: Assessing nonlinearities in the relationship between work attitudes and job performance using a Bayesian neural network. *Journal of Occupational and Organisational Psychology*, 74, 47–61.
- Somers, M. J., Birnbaum, D., & Casal, J. (2021). Supervisor support, control over work methods and employee well-being: New insights into nonlinearity from artificial neural networks. *The International Journal of Human Resource Management*, 32(7), 1620–1642.
- Somers, M. J., & Casal, J. C. (2009). Using Artificial Neural Networks to Model Nonlinearity—The Case of the Job Satisfaction–Job Performance Relationship. *Organizational Research Methods*, 12(3), 403–417.
- Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the Job Satisfaction Survey. *American Journal of Community Psychology*, 13, 693–713.
- Sturges, J., & Guest, D. (2004). Working to live or living to work? Work/life balance early in the career. *Human Resource Management Journal*, 14(4), 5–20.
- Sweeney, P. D., & McFarlin, D. B. (1993). Workers' Evaluations of the “Ends” and the “Means”: An Examination of Four Models of Distributive and Procedural Justice. *Organizational Behavior and Human Decision Processes*, 55, 23–40.
- Truss, C., Gratton, L., Hope-Hailey, V., McGovern, P., & Stiles, P. (1997). Soft and Hard Models of Human Resource Management: A Reappraisal. *Journal of Management Studies*, 34(1), 53–73.
- Wall, T. D., Michie, J., Patterson, M., Wood, S. J., Sheehan, M., Clegg, C. W., & West, M. (2004). On the Validity of Subjective Measures of Company Performance. *Personnel Psychology*, 57, 95–118.

- Wallace, J. C., Little, L. M., & Shull, A. (2008). The moderating effects of task complexity on the relationship between regulatory foci and safety and production performance. *Journal of Occupational Health Psychology, 13*(2), 95–104.
- Yavas, U., & Babakus, E. (2010). Relationships between organizational support, customer orientation, and work outcomes: A study of frontline bank employees. *International Journal of Bank Marketing, 28*(3), 222–238.
- Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: Methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine, 6*(11), 216.
- Zyman, S., & Brott, A. (2002). *The End of Advertising as We Know It*. John Wiley & Sons.

Tables

Table I – Scholarly evidence for non-linear and dual-role perspective on soft HR performance drivers and outcomes

Construct	Frequency of Co-occurrence (Performance/construct)	Construct role (Performance driver/outcome)	Dual role	Non-linear relationship (Performance/driver)
Contextual Performance (CP) (Motowidlo & Van Scotter, 1994)	-	Outcome	No	-
Customer Orientation (CO) (Yavas & Babakus, 2010)	1.130	Outcome	No	curvilinear (Homburg et al., 2011)
Employee Engagement (EN) (Harter et al., 2002)	1.144	Driver (Harter et al., 2002). Outcome (Zyman & Brott, 2002)	Yes	curvilinear (Shimazu et al., 2018)
Leadership (LE) (Ogbonna & Harris, 2000)	31.951	Driver	No	non-linear (Goldstein et al., 2010)
Organizational Justice (OJ) (Colquitt, 2001)	1.290	Driver	No	
Occupational Safety (OS) (Feyer & Williamson, 2004)	2.462	Driver	No	non-linear relationship (non-monotonous) (Wallace et al., 2008)
Satisfaction (Spector, 1985)	46.885	Driver	No	curvilinear (Somers & Casal, 2009)
Work-life balance (WL) (Keeton et al., 2007)	655	Driver (Perry-Smith & Blum, 2000). Outcome (Keeton et al., 2007).	Yes	Curvilinear (Drobnic et al., 2010)

Table II – Literature review in the field of soft HR performance management using ANNs for data analysis

Article	Performance Construct	Limitation 1		Limitation 2		Train vs. Test; Training %	Limitation 3 Prediction (step 7)
		Framework cited	Framework Additions	Sample Size	Sample Source		
(Collins & Clark, 1993)	Job Performance	impossible	2) Kolmogorov-Rule (hidden = inputs*2) vs. (hidden = ((inputs+outputs)/2))	81	Mid-Level Manager	54 vs. 27; 67%	-
				649	Upper-Level Managers	435 vs. 214; 67%	-
(Somers, 1999)	Job Withdrawal	impossible	2) Hidden Kohonen-layer and learning vector quantization 5) LogReg & Accuracy	577	Nurses	462 vs. 115; 80%	-
(Somers, 2001)	Job Performance	impossible	-	232	Nurses	185 vs. 47; 80%	-
(Montagno et al., 2002)	Organizational Performance	impossible	2/4) Genetic Algorithm for model improvement used 5) Wilcoxon Matched Pairs Signed Ranks Test between estimates of ANN & OLS	137	Organizations	127 vs. 10; 93%	Predict combinations of Improvement Categories
(Grznar et al., 2007)	Team Effectiveness	no	2/4) Experimentation (hidden neurons, learning rate), h = 30 3) outliers proportion 0-30% 5) OLS & LMS Regression, MSE/SSE/F value	102	Teams	2/3 vs. 1/3; 66%	-
(Somers & Casal, 2009)	Job Performance	yes	1) MLP & RBF NN 2/4) Experimentation (hidden neurons), Comparison of Performance 5) OLS & Tobit Regression	176	Nurses & Technicians	134 vs. 39; 80%	-

(Karanika-Murray & Cox, 2010)	Worn-Out	yes	2) 2 models, 2 hidden layer, hidden = each 14 & 9 5) MAPE, Fisher z test	1,003	Employees of 5 different Organization	Merged: 802 vs. 201; 80%	-
(Minbashian et al., 2010)	Performance	yes	2) 3 models, hidden = 6, 16, 26 4) Initial parameters described 5) CVR	120	University Graduates	2/3 vs. 1/3; 66%	-
				286	NZ Police Recruits		
				486	US Bus Drivers		
(Ladstätter et al., 2014)	Consequences of Burnout	no	5) t-test	268	Nurses	227 vs. 49; 82%	-
(Ladstätter et al., 2016)	Consequences of Burnout	yes	2) hidden = [70;10] stepwise 5) t-test	465	Nurses	395 vs. 70; 85%	-
(Somers et al., 2021)	Stress, Well-Being	yes	-	235	Nurses	178 vs. 57 75%	-
(Shah et al., 2020)	Absenteeism	no	1) DNN 2) NN (hidden = 100) + DNN (hidden = 200, 150, 100, 50, 10, 5) 5) Accuracy, Precision, Recall, F1, Recursive Operating Characteristic, Confusion Matrix	740	Courier Company	553 vs. 187 75%	-
Average				391.1	76.6%, 89.6		
Median				277.0	80%, 57		

Notes: * LMS (Least Median Squares), MSE (Mean Squared Error), SSE (Sum of Squared Errors), RBF (Radial Basis Functions), MAPE (Mean

Absolute Percentage Error), CVR (Cross Validation Coefficient), DNN (Deep ANN~multiple hidden layers).

Table III – Results of confirmatory factor analysis

	CP	EN	EF	LE	IM	PR	ID	IS	CO	OJ	OS	WL
Item1	0.456	0.580	0.712	0.428	0.611	0.725	0.601	0.794	0.077	0.532	0.611	0.757
Item2	0.503	0.815	0.628	0.644	0.726	0.617	0.783	0.809	0.745	0.484	0.592	0.573
Item3	0.550	0.773	0.748	0.711	0.562	-0.09	0.586	0.623	0.606	0.691	0.732	0.450
Item4	0.742	0.678	0.676	0.733	0.712	0.325	0.621	0.644	0.788	0.761	0.743	
Item5	0.509	0.692	0.746	0.615	0.606				0.641	0.770	0.729	
Item6	0.728	0.587	0.637	0.361	0.751				0.504	0.705		
Item7	0.688	0.773	0.720	0.734	0.788				0.564	0.607		
Item8	0.659	0.341	0.592	0.673	0.533				0.606	0.644		
Item9	0.724	0.638		0.687	0.403							
Item10	0.706				0.787							

Notes: Contextual Performance (CP), Employee Engagement (EN), Organizational Effectiveness (EF), Leadership (LE), Immediate Manager (IM), Pay & Reward System (PR), Individual Development (ID), Information Sharing (IS), Customer Orientation (CO), Organizational Justice (OJ), Occupational Safety (OS), Work-life balance (WL). Values in bold are below the 0.5 suggested limit of acceptable factor loading. Model fit indices: N=49,680, degrees of freedom (dF): 3,173, $p < 0.0001$, comparative fit index (CFI): 0.773, Tucker–Lewis index (TLI): 0.76, Standardized Root Mean Square Residual (SRMR): 0.062, root mean square error of approximation (RMSEA): 0.062.

Table IV – Reliability analysis

	CP	EN	EF	LE	IM	PR	ID	IS	CO	OJ	OS	WL
Total	0.86	0.88	0.88	0.86	0.88	0.62	0.74	0.81	0.83	0.85	0.81	0.58
Item1	-	0.87	0.86	-	0.87	0.45	0.70	0.75	/	0.85	0.79	0.41
Item2	0.86	0.85	0.86	0.85	0.86	0.45	0.64	0.74	0.80	/	0.79	0.41
Item3	0.86	0.86	0.86	0.84	0.88	/	0.72	0.78	0.82	0.82	0.76	/
Item4	0.85	0.87	0.86	0.83	0.87	/	0.67	0.78	0.79	0.81	0.76	
Item5	0.86	0.86	0.86	0.85	0.87				0.80	0.83	0.76	
Item6	0.84	0.87	0.87	/	0.86				0.82	0.83		
Item7	0.84	0.85	0.86	0.83	0.86				0.81	0.84		
Item8	0.85	/	0.87	0.83	0.88				0.81	0.84		
Item9	0.84	0.87		0.83	/							
Item10	0.84				0.86							

Notes: Contextual Performance (CP), Employee Engagement (EN), Organizational Effectiveness (EF), Leadership (LE), Immediate Manager (IM), Pay & Reward System (PR), Individual Development (ID), Information Sharing (IS), Customer Orientation (CO), Organizational Justice (OJ), Occupational Safety (OS), Work-life balance (WL). Standardized alphas are reported per scale and in terms of changes in std. alpha if item is dropped. Fields marked with / have not been calculated due to CFA exclusion. Alpha-values in bold are questionable > 0.95 or <0.65.

Table V – Descriptive statistics

	n	M	SD	MIN	MAX	Skewness	Kurtosis	KST
CP	49013	4.05	0.75	1	5	-0.81	0.29	0.10*
EN	48865	3.89	0.81	1	5	-0.76	0.25	0.08*
EF	48614	4.11	0.74	1	5	-0.94	0.75	0.11*
LE	48475	4.08	0.8	1	5	-0.70	0.09	0.09*
IM	49004	4.1	0.79	1	5	-0.97	0.61	0.12*
PR	49259	4.09	0.87	1	5	-1.14	1.17	0.20*
ID	49158	4.05	0.84	1	5	-0.95	0.56	0.12*
IS	48714	3.95	0.85	1	5	-0.79	0.3	0.11*
CO	48717	3.97	0.79	1	5	-0.79	0.3	0.09*
OJ	49134	4.04	0.8	1	5	-0.95	0.58	0.11*
OS	49149	4.1	0.82	1	5	-1.04	0.77	0.13*
WL	49519	4.24	0.82	1	5	-1.21	1.27	0.20*

Notes: Contextual Performance (CP), Employee Engagement (EN), Organizational Effectiveness (EF), Leadership (LE), Immediate Manager (IM), Pay & Reward System (PR), Individual Development (ID), Information Sharing (IS), Customer Orientation (CO), Organizational Justice (OJ), Occupational Safety (OS), Work-life balance (WL), sample size without missing values (n), mean (M), standard deviation (SD). Descriptive statistics had been calculated before z-transformation; Kolmogorov-Smirnov Test (KST) for normal

distribution shows test values that are all significant with $p < 0.001$ to reject the null hypothesis.

Table VI – Correlation matrix

	CP	EN	EF	LE	IM	PR	ID	IS	CO	OJ	OR
EN	0.74										
EF	0.73	0.87									
LE	0.76	0.86	0.85								
IM	0.90	0.70	0.70	0.72							
PR	0.68	0.77	0.76	0.77	0.64						
ID	0.85	0.69	0.68	0.70	0.86	0.63					
IS	0.74	0.84	0.83	0.83	0.71	0.73	0.69				
CO	0.78	0.85	0.83	0.83	0.75	0.78	0.72	0.82			
OJ	0.88	0.73	0.71	0.74	0.88	0.65	0.82	0.73	0.78		
OS	0.87	0.67	0.66	0.69	0.87	0.62	0.82	0.69	0.72	0.85	
WL	0.70	0.70	0.74	0.72	0.69	0.63	0.63	0.65	0.68	0.69	0.63

Notes: Contextual Performance (CP), Employee Engagement (EN), Organizational Effectiveness (EF), Leadership (LE), Immediate Manager (IM), Pay & Reward System (PR), Individual Development (ID), Information Sharing (IS), Customer Orientation (CO), Organizational Justice (OJ), Occupational Safety (OS), Work-life balance (WL). All correlations are significant at $p < 0.001$

Table VII – ANN analysis

Method	nRMSE *	R ²
MMR (OLS, Train)	0.351812	0.804
MMR (OLS, Test)	0.353506	0.8052
ANN (Train)	0.043698	0.87388
ANN (Test)	0.04375	0.87219

Notes: Variance Explained (R²). * Only normalized Root Mean Squared Error (nRMSE) is reported because it is independent of the unit of the data analysed and it is naturally the result of ANN analysis, because normalized data needs to be used.

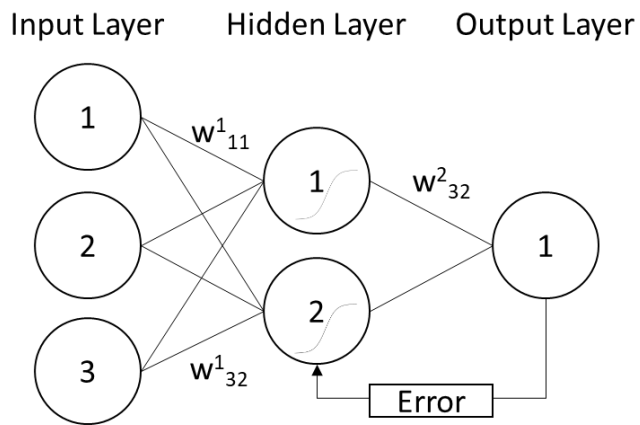
Table VIII – Predictions based on ANN (NN)

	Current means	MMR optimal combination	NN optimal combination	NN based managerial decision	Garson's RMI
EN	4.05	5	5	Increase	0.125
LE	4.11	5	5	Increase	0.103
IM	4.08	5	4	Keep steady	0.102
PR	4.1	5	5	Increase	0.106
ID	4.09	5	5	Increase	0.126
IS	4.05	5	4	Keep steady	0.106
OJ	3.97	5	5	Increase	0.119
OS	4.04	5	5	Increase	0.109
WL	4.1	5	3	Decrease	0.104

* Note: Employee Engagement (EN), Leadership (LE), Immediate Manager (IM), Pay & Reward System (PR), Individual Development (ID), Information Sharing (IS), Organizational Justice (OJ), Occupational Safety (OS), Work-life balance (WL), Relative magnitude of factor importance for the ANN outcome (RMI)

Figures

Figure 1 – Simple ANN Model



Notes: Simple ANN Structure adapted and modified from Ladstätter et al (2016) and Somers et al (2018). w^x_y are edge weights for level x [e.g. input to hidden layer =1] and connection y [e.g. node 3 on input layer to node 2 on hidden layer], each edge has a corresponding activation function (e.g. sigmoid).

Figure 2 – ANN framework for organizational research (according to Scarborough and Somers, 2006)

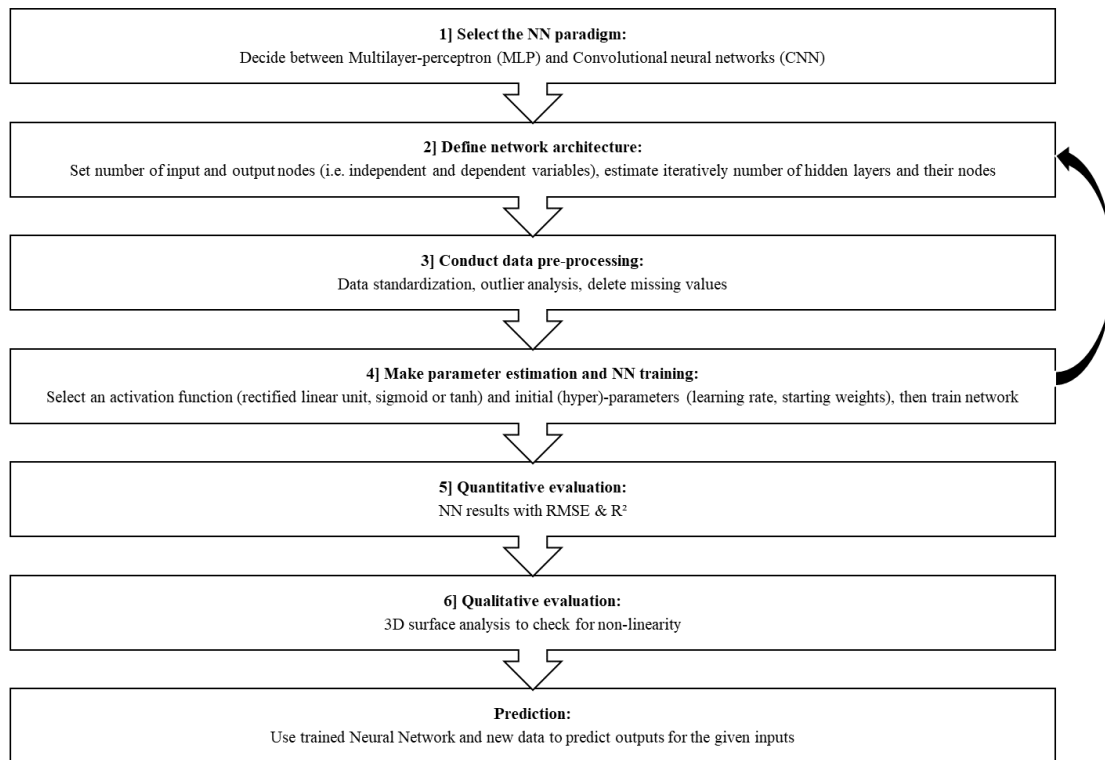
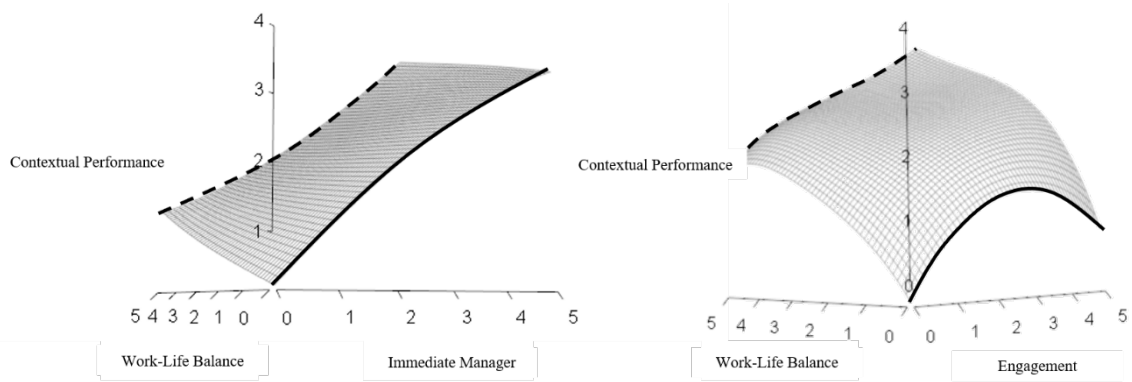


Figure 3 – 3D Surface Analysis Sample



Notes: Immediate Manager (IM) served as an input, Work-Life Balance (WL) and Engagement (EN) as dual-role constructs, and Contextual Performance (CP) as an output.