Performance Modelling on Banking System: A Data Envelopment Analysis-Artificial Neural Network Approach

Bangarwa Preeti, Supriyo Roy

Performance modelling on Banking System: A Data Envelopment Analysis-Artificial Neural Network Approach

Preeti[1[0000-0001-6702-2827]] and Supriyo Roy[2[0000-0003-0600-0696]]

[1,2] Department of Management, BIT Mesra, Ranchi, India
supriyo.online@gmail.com

Abstract. With changing banking environment, the efficiency of the operational function of bank is of critical importance and needs timely watch. Apart from measuring the operational performance of banks using DEA approaches, the banking sector today is more inclined to predictive analytics to identify their future performance and improve their competitiveness well in advance. In this sequel, the present paper proposes hybridisation of Data Envelopment Analysis and Artificial Neural Network Approaches for operational performance measurement and prediction for Indian banks using the five-year (2015 to 2019) dataset. Non-oriented non-radial DEA model is adopted in the present study, attempting to provide decision-makers the discretion to identify slacks in performance by maximising outputs and minimising inputs. This can identify causes of inefficiency and suggest necessary steps for improvement. In addition to DEA findings, the paper performs prediction task for obtained efficiency scores. Finding of will be advantageous for policymakers, managers of banking industry for predicting future operational performance of banks until they are able to make required changes for its improvement.

Keywords: Data Envelopment Analysis, Artificial Neural Network, Operational Performance, Efficiency, Banking.

1 Introduction

Banking system holds immense significance in any countries economy upliftment (Bhattacharya et al., 1997). With advancement of digital banking, the competition between banks have only risen further. Banks are constantly under pressure to improve performance to withstand the change (Reserve Bank of India, 2018). As a result, the performance analysis of banks has become integral issue for management concerned. Top managers are constantly putting effort to identify inefficiencies to eliminate them to achieve competitive advantage and face challenges. Moreover, operation risk management is pertinent to any commercial banks to ensure operation effectiveness and competitiveness. How to improve the operational performance of commercial banks has become an emerging question that commercial banks now faces prominently. The basic research question boils down to: *the need for performance evaluation to measure, improve and predict the operating efficiency levels of banking system*.

Traditionally, the bank managers made use of multiple ratios to analyse different aspects of banking operations. However, ratios are subjected to limited information while making economies of scale assumptions, framing benchmarking policies or estimating overall performance of banks (Yeh, 1996). Alternatively, frontier approaches as opposed to ratio approaches, allows objective measurement of performance within complex operational environments. Frontier efficiency approaches, namely parametric and non-parametric, are two categories of approach differing on the basis of assumption of functional form of variables and presence of computational error (Berger and Humphrey, 1997). The problem of performance measurement is well studied by using non-parametric approaches (Cooper *et al.,* 2007). As per literature, Data Envelopment Analysis (DEA) is a non-parametric benchmarking technique to identify efficient and inefficient banks based on multiple input and output data derived from banking dataset (Charnes *et al.,* 1978).

Apart from measuring the operational performance of banks using DEA approaches, the banking sector today is more inclined to predictive analytics to identify their future performance and improve their competitiveness well in advance. The basic capability of DEA modelling is of performance measurement, but not predictive capacity of performance. Also, DEA results do not provide potential solutions for allocation of resources to inefficient units leading to inappropriate support to managerial decision making. Therefore, there is need for flexible modelling to measure and predict performance is of significant interest for practical concerns (Wu et al., 2006). The progressive use of various machine learning approaches is proved to be indispensable. Artificial neural network, a machine learning technique is increasingly used to assist DEA findings to estimate efficiency (Wang, 2003). Considering similarities between DEA and ANN, as both belong to non-parametric category as well as both approaches do not makes assumption related to functional form of inputs and output variables. At first, Athanassopoulos and Curram (1996) applied the combination of DEA and ANN to classification and prediction problems. The study revealed that DEA proves better than ANN for measurement purpose and latter can be utilised for prediction analysis. Since then, many researchers applied DEA-ANN approach towards various domains like banking, education, industries, hospitals (Sreekumar and Mahapatra, 2011; Tosun, 2012). To the best of our knowledge, there is limited study developing hybrid model which utilises DEA and neural network in context of Indian banks.

To summarise, the plenty of DEA-related studies is limited to measuring performance without prediction. Therefore, it is imperative to integrate DEA and data mining techniques such as ANN for Indian banking system. Present paper adopts hybrid modelling approach to estimate, improve and predict operational performance of Indian banks with a period of five years from 2015 to 2019. The two main contribution of the study can be listed as follows:

- *Development of DEA model* - Operational performance is measured and improvement identified for Indian banks
- *Development of ANN model* – ANN model is trained using five year banking dataset (2015 to 2019) with efficiency score as output and input-output variables as utilized in DEA model as inputs.

Remaining study is structured as follows. Most relevant literature pertaining to use to DEA and DEA-ANN approach in various sectors, especially banking is discussed in section 2. Next, section 3 describes DEA and ANN approaches explicitly. Section 4 explains data, proposed framework and empirical findings of the paper. Section 5 concludes with summary, limitations and future direction of the study.

2 Related Studies

In recent past, plenty of studies has resorted to quantitatively measure performance of banks by developing several DEA models. Although, many researchers applied DEA to estimate bank's operational performance and only few studies attempted to using Artificial Intelligence techniques like ANN to predict operational performance of banking system.

2.1 DEA applications to banking sector

Initially, the stress was on development of mainly traditional (radial) DEA models suggesting proportional changes in input and output variables (Wild, 2016). Sahoo and Tone (2009a) and Sahoo and Tone (2009b) introduced radial and non-radial DEA models for Indian banking system to study the effect of financial sector reforms on the efficiency growth of banks. The studies concluded that public sector banks do not reflect the learning experience compared to private sector banks. In 2016, Stewart *et al.* used radial DEA models to study the bank performance in Vietnam for the period of 1999 to 2009. The study concluded that efficiency rose over the observed period. Defung *et al.* (2016) studied the impact of regulatory reforms on the performance of Indonesian banking industry, confirming statistical significant relationship between regulatory reforms and technical efficiency score. Profit efficiency is compared across different Indian banking ownership groups by developing DEA model to assess the impact of Global financial crisis (Gulati and Kumar, 2016). As per the findings, the Global financial crisis did not much effect the profit efficiency of Indian banks.

Kumar *et al.* (2016) evaluated the efficiency, productivity and return to scale of Indian banks for the period of post reform and global financial crisis. The study answered question regarding the impact of global financial crisis on the performance of banks. Azad *et al.* (2017) used sample of 43 Malaysian commercial banks to compare the performance across bank ownership and nature. Sathye and Sathye (2017) confirmed this relationship by testing the impact of ATM intensity, Bank size, soundness, ownership and risk on the performance of Indian banks using bootstrap DEA model. Covering the period from 2012 to 2016, the efficiency of Brazilian banks was estimated using radial DEA models (Henriques *et al.,* 2018). The study identified the efficient and inefficient banks by estimating pure technical and scale efficiency. Most recent study (Davidovic *et al.,* 2019) on the implementation of radial DEA model to estimate the efficiency trends of the Croatian banking industry for the period from 2006 to 2015. Variables like relative ownership structure, market size, and origin of capital are studied to test the relationship with efficiency score. Mohapatra *et al.* (2019) estimated the operating efficiency of Indian banks for the year 2011 to 2015 using radial DEA model.

Study investigated the relationship between intellectual capital and performance of banks.

2.2 DEA-ANN

Despite of many data mining techniques that have been used in literature for predicting certain output, the ANN technique has been sporadically used in literature. Since the pioneering work of combining DEA and ANN by Athanassopoulos and Curram (1996), plenty of studies have adopted DEA-ANN approach in various field such as banking, healthcare, supply chain and manufacturing industries. After extensive literature review of studies, particularly those studies that utilised DEA in conjunction with ANN is discussed below. Wu et al. (2006) developed DEA-ANN model for branch efficiency of big Canadian banks. Based on its findings, the study offers guidelines to improve the performance of inefficient branches. Moreover, short-term efficiency prediction is performed using developed model. Emrouznejad and Shale (2009) generated back propagation neural network in conjunction with DEA to estimate efficiency scores of DMUs with large datasets with many input/output variables. This paper used five large dataset to propose that developed model is better alternative than using conventional DEA which uses large amount of computer resources like CPU time and memory.

Mostafa (2009) quantified the performance of top Arab banks using DEA-ANN approach. The study develops probabilistic neural network to perform classification function of banks depending on best accuracy score of the model. The study highlighted flexibility and robustness advantages of utilising developed model for classifying banks performance. Sreekumar and Mahapatra (2011) developed integrated DEA and neural network model for assessing performance of Indian B-school. The study identified input-output variables to measure and suggest improvement in technical efficiency of B-schools. Performance prediction using neural network is made for effective decision making. Tosun (2012) combined DEA-NN model to measure the efficiency of hospitals and overcome the limitations of DEA approach. The developed model is trained using DEA results and prediction is made on test data to categorise banks as efficient and inefficient banks. Moreover, the obtained results are compared with results using Discriminant analysis (DA). The study confirms that ANN is best approach to perform classification function compared to DA as it requires lesser computer resources and CPU time.

Barros and Wanke (2014) analysed performance of Insurance companies by developing two-stage model using DEA and neural network. The study identified ceded reinsurance as potential output for increasing efficiency score of companies and also predicted performance using the developed neural network model. Kwon and Lee (2015) enhanced the two-stage DEA model by adding predictive capacity using BPNN approach. The study applied the developed model to the datasets from large US banks to empirically demonstrate constructive performance modelling. To overcome the predictive capability of DEA model, Shokrollahpour (2016) developed combined DEA and ANN model to forecast future benchmark for Iranian commercial banks. The study performed five year efficiency forecast to suggest strategies to improve efficiency and its causes. Kwon (2017) developed performance measurement and prediction model for railways using DEA and NN approach. Efficiency trend of railroad is estimated using

CCR model of DEA and prediction is made for efficiency score and projected output using NN for each railroad. The proposed framework is beneficial for decision making and benchmarking prac2tises for railroads. Tavana et al. (2018) adopted two data mining models such as ANN and Bayesian networks to assess bank liquidity risk. Dataset from large US banks is utilised for implementing the proposed model for liquidity risk assessment. Petropoulos (2020) adopted multiple machine learning techniques to forecast bank insolvencies of sampled US-based financial institution. The result confirmed that Random Forest as well as Neural Network are superior methods compared to others. For prediction of bank failures, the study suggested CAMELS evaluation framework offers higher marginal contribution. Le and Viviani (2020) proved that ANN is more accurate than traditional statistical technique to predict bank failure. A sample of 3000 banks was investigated for period of 5 years before bank becomes inactive. Five significant ratios that study identified as significant includes capital quality, liquidity, loan quality, operations efficiency and profitability.

Only few studies demonstrated the integration of DEA-ANN approach within banking system, particularly in Indian banking system. Most DEA related studies measured the performance based on different DEA models, only few of them included prediction of performance. Prediction of bank performance is of utmost importance to restrict banks from being insolvent or weaker. The hybrid model is developed to enhance the predictability dimension to black-box like DEA model. The empirical finding provided by DEA for performance improvement may not always be action-oriented due to lack of prediction capability. Therefore, the hybrid DEA-NN model provide managers plausible capabilities to predict optimal operational performance for setting advance improvement goals and progress. In this sequel, the objective framed for the study is as follows:

1. To *develop* DEA model to measure operational efficiency scores and identify areas of improvement of Indian banks for the period of five years.
2. To *train* ANN model using estimated operational efficiency scores and input-output variables.
3. To *predict* the operational performance of banks using developed ANN model.

3 Research Methodology

3.1 Data Envelopment Analysis

DEA, a linear programming managerial tool used for measuring productivity and efficiency for any Decision Making Units (DMU). It is excessively applied for both public and private sector including airlines, banks, hospitals, manufacturers, transportations and universities. Consecutively, variations in model development, new applications, variable differences are ongoing research fields in area of DEA. As name suggests, DEA "envelops" input-output production function as closely as possible by developing efficient frontier that identifies best and worst performing DMUs. Original DEA models like, Charnes-Cooper-Rhodes model and Banker-Charnes-Cooper model measures the radial technical and pure technical efficiency of DMUs, respectively. More sophisticated model like Slack Based Measure (SBM) model, as proposed by Tone (2001), supersedes these traditional DEA models by measuring efficiency of an inefficient DMU by referring to the furthermost point on the benchmark frontier of an inefficient DMU by referring to the furthermost point on the frontier. SBM model are non-radial models measuring non-proportional input excesses and output shortfalls, unlike CCR and BCC measuring only radial (proportional) efficiencies (weaker efficiency). Certain properties that SBM model satisfies includes acceptance of semi-positive data and unit invariance of variables, scalar value for reported efficiency score. Also, non-oriented modelling of SBM model allows to report inefficiencies on both side of production function - input excesses and output shortfalls.

Mathematical equations

Notations for development of non-oriented SBM model is as under:

n = Total DMUs each having inputs(m) and outputs(v); j=1,2,…,n)

$$X = \left(x_{ij}\right) \text{ set of inputs; } i = 1,2,...,m$$

$$Y = \left(y_{uj}\right) \text{ set of outputs; } u = 1,2,...,v$$

ρ = non-radial slack indicator (efficiency value), s^{-}=input slack, s^{+}=output slack, λ =intensity vector; X>0 and Y>0

With notations as clarified, the production possibility set A for DMU $\left(x_{i0}, y_{u0}\right)$ is defined as:

$$A = \left\{ \left(x, y\right) | x \geq X\lambda, y \leq Y\lambda, \lambda = 1 \right\} \tag{1}$$

Non-oriented SBM model:

$$\rho = \min \frac{1 - \frac{1}{m} \sum_{i=1}^{m} \frac{s_i}{x_{i0}}}{1 - \frac{1}{v} \sum_{u=1}^{v} \frac{s_u^{+}}{y_{u0}}} \tag{2}$$

Subject to:

$$x_0 = X\lambda + s^{-}$$

$$y_0 = Y\lambda - s^{+}$$

$$\lambda = 1; s^{-}, s^{+} \geq 0 \tag{3}$$

The objective of the study is to optimise the objective equation (2) subject to constraints (3) such that value of ρ is equal to 1 and value of input and output slack is 0.

3.2 Artificial Neural Network

Influenced by biological neural network, ANN is popularly known machine learning method that captures non-linear patterns in data, utilised mainly for credit rating classification problems. Such problems require big dataset, explanatory variables. Literatures arguments numerous neural networks with structural variations depending on information flow, hidden layers and algorithm differences used to train them. These layers of neural network are connected by connection weights. A typical neural network is series of interconnected neuron layers. The information transfer in neural network occurs in two ways: feedforward and back-propagation. Feed forwarding functions to processes the information from input layer to output layer resulting into error, whereas back propagation tries to optimise output by fixing errors by sending information back in the network. On designing multilayer neural network, it is pertinent to decide upon number of hidden layer. Also, the number of hidden layers depends on complexity of problem to be predicted.

Back Propagation Neural Network is considered to most popular neural network used for both classification and prediction purpose (Rumelhart et al., 1986). All ANN uses multilayer feed-forward neural network to learn the parameters to form non-linear function between input-output variables. The typical structure of ANN comprises of input layer, hidden layer and output layer. Inputs are fed into input layer simultaneously in units. Weighted output of units from input layer are fed into second layer called hidden layer. Weighted output of this hidden layer is input to second hidden layer, and so on depending on the network. The output of the hidden layer is fed as input to output layer which provides prediction for DMUs. Basically, input data related to back prorogation is fed to neural network, the output is compared to the desired output to estimate the error for each iteration. This estimated error is back propagated to adjust weights in order to decrease error for each iteration. Learning process continues till acceptable range of output error is reached after considerable weight adjustment to train the model to produce the desired output. Figure 1. Illustrates the basic structure of ANN

4 Empirical Analysis and results

4.1 Data and variables

Data on 39 Indian banks for the period of 5 years was obtained from annual publications of Reserve Bank of India and Indian Banks Association websites. Total of 39 banks from both public and private sector is involved in present analysis covering the study period of five years (2015 to 2019). As per DEA rule of thumb, minimum number of DMUs to be included in the study must be three times the sum of input-output variables (Charnes et al. 1978). Input-output dataset for 39 banks (DMUs) is used for period for 5 years (2015 to 2019) totalling to 195 observations is used to develop ANN model. The dataset is split into training and testing data in 7:3 ratio. The implementation of model includes training the model and using the developed to test the outputs. There exists no formal rule of thumb regarding optimal dataset for neural network. However, usually the training dataset should be 10 times the sum of independent variables used as inputs in the model (Kwon and Lee, 2015).

For variable selection, the two approaches are mainly followed- production and intermediation. Production approach, developed by Benston (1965), consider banks as service providers for customers using physical inputs like labor, capital and assets to make services available like deposits and advances, whereas intermediation approach, as developed by Sealey and Lindley (1977) portrays the intermediary functioning of banks by collecting funds from customers and converting into loans. Berger and Humphrey (1997) and Fethi and Pasiouras (2010) advocated that intermediation approach is more appropriate for measuring performance of banking system. Considering the literature relevance, the present study follows intermediation approach in selecting variables for developing non-radial SBM DEA model.

The study uses investments, performing loans and advances, non-interest income as desirable outputs, whereas operating expenses and deposits are included as input variables. Operating expenses is one of the prominent factors that effects profitability and also improves efficiency. In intermediation function of banks, deposit is an important input factor. Investments includes sum of long-term investments as well as marketable securities. Adjusting total loan amount with non-performing loans creates creates new output variable referred as performing loans. Performing loans and advances represents the source of interest income that helps to maintain security and liquidity for normal banking operations by conveying the message of banking stability. If unadjusted loan amount is considered into the model, the efficiency score might be overestimated. Non-interest income- represents additional source of income for banks that helps in improving efficiency. Summary statistics of dataset for entire study period is presented in table1.

4.2 Proposed Framework

Idea behind utilising DEA in conjunction with ANN is that it can complement DEA approach by capturing the non-linear relationship between selected variables and performs optimisation. Continuous operational changes in banking environment calls for evaluation for operational performance to ensure its competitiveness, to predict its operation in delivering objective and to work upon their future development. In this sequel, the proposed hybrid model, two-stage Hybrid model including non-oriented DEA model and BPNN is used to measure the operational efficiency of banks, incorporating two inputs and three outputs. Fig 2. Illustrates the framework for hybrid model. Present hybrid model is built on five datasets of 39 banks for year 2015 to 2019. The model tries to predict efficiency scores as estimated using DEA model.

Steps for hybrid modelling:

1. *Non-oriented SBM model* - The model aims to minimise input levels as well as maximise output levels to measure the non-radial operational efficiency of banks.
2. *DEA result analysis* - Identifying benchmarks banks and performance improvement requirement for inefficient banks
3. *Data Preparation for ANN* - The sample dataset is divided into train set and test set. There exists no formal rule of thumb regarding optimal dataset for neural network. However, usually the training dataset should be 10 times the sum of independent

variables used as inputs in the model (Kwon and Lee, 2015). At first, 70 % of data from 5-year period is used to train the model and rest 30% is used to test the prediction accuracy of the model.

4. *Data Normalisation* - As variables included in the sampled dataset differs in ranges, data must be normalised. In this study, logarithmic transformation is used to normalise the variables to be considered for ANN training.
5. *Network Structure* - Before initiating to train the sample, the network typology is pre-decided in terms of number of variables in input-layers, hidden layers and output layers. In present model, input and output variables in DEA model is fed as variables in input layer, and efficiency score is fed as only output in the output layer. The optimum number of hidden layers depends on trial and error process to minimise error and increase accuracy of trained sample.
6. *Model Training* - BPNN learns by repeated iterations on training sample and comparing the predicted efficiency scores with actual efficiency scores. Weights for each training sample is so adjusted (back propagated) to minimise mean square error between predicted efficiency scores derived from BPNN and actual efficiency score derived from DEA model.
7. *Model Validation* - BPNN efficiency score (predicted) is compared to DEA efficiency scores (actual) Validating the ANN model using random samples of banks from dataset. BPNN efficiency score (predicted) is compared to DEA efficiency scores (actual) and shown via plot in fig. Correlation proves that predicted scores are good estimate of actual DEA efficiency scores.

The DEA efficiency score is obtained using SBM model in DEA SOLVER LEARNING VERSION 13.

4.3 Empirical findings

Mean and standard deviation of efficiency from non-oriented SBM model under variable return to scale from 2015-’19 is 0.82334 and 0.13912, respectively (*table 2*) On average, out of the sample banks of 39, only 6 banks is observed to be fully efficient. The empirical findings confirm continuous rise and fall in efficiency of banks. Over the period of last five years, Indian banking system is 82.33% efficient, with public sector banks as 77.51% efficient and private sector banks as 87.41% efficient. Therefore, inefficiency present in the banks calls for in-depth analysis into inefficient input or output variables

Such deviation from efficiency can be improved by working on improvable spaces as highlighted by slack values for each input-output variable. Difference between projected value and actual value is termed as slack value, particularly as input excesses and output shortfalls. For each variable, the banking system must target to reduce input values and expand output values as mentioned in table 4. Conclusion drawn upon slack analysis gives future direction for strategic decision making.

The functional relationship that represents the developed ANN model can be interpreted as:

$$Efficiencyscore = f(I_{Operatingexpenses}, I_{Deposits}, O_{TotalAdvances}, O_{Investment}, O_{NonInterestIncome})$$

Where, I and O stands for inputs and outputs used for non-oriented SBM model. The above production function is used to train the ANN model using input and output variables to predict the scalar efficiency scores. The package named "NeuralNet" in R software is used for developing the proposed ANN model. The most arduous task is to select the best network structure for the performance of NN model by trail-and-test method until minimum error is obtained for model training. Table 5 shows different network structures compared based on number of nodes in hidden layer, error scores and number of epochs (iterations) before finding the best trained model to predict the efficiency scores of banks. Since error is minimum with network structure (15,10), it is considered to be the best network with 98.5 % accuracy after 9,259 iterations.

After training of the network, it is tested with 30% of the dataset to prove generalisation capability of the developed model (*table 6*). It is observed that Mean Square Error (MSE) and Root Mean Square Error (RMSE) on actual and predicted datasets with value equal to 0.6197 and 0.7872, respectively. The correlation between actual performance (DEA efficiency scores) and predicted performance (ANN predicted scores) is exhibited in fig 3, with pattern indicating sufficient correlation. Predictive potential of model is shown in fig 4, exhibiting stable mean absolute error (MAE). In conclusion, implementing the proposed ANN and using the given definition of efficiency score (Eq. (1)), we were able to predict operational efficiency with a 98.5 % accuracy.

5 Concluding remarks

Proposed hybrid model is basically combination of traditional approach (DEA) and non-traditional approach (Neural Network) to develop prediction model for operational performance of Indian banks. Empirical results show that proposed model possess good predictive power by having low error and high accuracy. Paper focuses on measurement and prediction of operational performance of Indian banks using DEA and ANN approaches. Firstly, the efficiency score is measured for each bank along with slack analysis. Secondly, the estimated efficiency scores are used to predict performance of banks using developed ANN model. In this paper, a Hybrid model is proposed, incorporating DEA and ANN, to estimate and predict operational performance of banks. The study addressed the problem of measuring and predicting the operational efficiency performance by using two-stage hybrid approach.

The result of the study posits important policy implications. Developed model will help banks to predict performance with different data according to their capacity to test performance. Nevertheless, the empirical findings of this study offer constructive insights to regulator and policymakers to investigate the health of banking system based individual banks and augment policy responses. The proposed model can assist in managerial decision-making process proving its practical and innovative. The study provides useful contribution in terms of utilising the proposed model for small dataset problems.

Identification of benchmark banks will help to investigate best business decision, effective and innovative strategies and operating procedures. Conclusion drawn upon slack analysis gives future direction for strategic decision making. Managers can use the propose model to infer about future operational performance of banks and buy time to frame relevant strategies.

Also, the study does not include macroeconomic variable that could enhance performance prediction. Inclusion of more contextual variables into BPNN model would enhance the predicting accuracy of the developed model. Present paper postulated hybrid model only on Indian banks, thereby excluding foreign banks. Future scope of the study points to developing model on enriched dataset including foreign banks and thereby larger dataset. ANN model developed is trained on 195 observation (5-year dataset). However, use of larger datasets is recommended to avoid the problem of under training. Additionally, the future directions to present modelling framework can be explored with dynamic setting DEA models to capture more realistic performance prediction. Also, the proposed model can be expanded to other relevant sectors like manufacturing, research & development, supply chain and across different industries. To check the robustness of the developed model by testing its predictive applicability with other data mining models. Future study can utilise the existing model for benchmarking process by exploiting classification function of BPNN model. Also, to investigate the number of hidden layers and number of units in the hidden layer that further reduces prediction error of the proposed model is one of the areas for future investigation.

References

1. Azad, M. A. K., Munisamy, S., Masum, A. K. M., Saona, P., Wanke, P.: Bank efficiency in Malaysia: a use of malmquist meta-frontier analysis. Eurasian Business Review 7(2), 287-311 (2017).
2. Benston, G. J.: Branch banking and economies of scale. The Journal of Finance 20(2), 312-331 (1965).
3. Berger, A.N., Humphrey, D.B.: Efficiency of financial institutions: International survey and directions for future research. European journal of operational research 98(2), 175-212 (1997).
4. Bhattacharyya, A., Lovell, C.K. Sahay, P.: The impact of liberalization on the productive efficiency of Indian commercial banks. European Journal of Operational Research 98(2), 332–345 (1997).
5. Barros, C. P., Wanke, P.: Insurance companies in Mozambique: A two-stage DEA and neural networks on efficiency and capacity slacks. Applied Economics 46(29), 3591-3600 (2014).
6. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision-making units. European Journal of Operational Research 2(6) 429-444 (1978).
7. Cooper, W. W., Seiford, L. M. Tone, K.: Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Springer, New York (2007).
8. Cooper, W. W., Huang, Z., Li, S., Zhu, J.: A response to the critiques of DEA by Dmitruk and Koshevoy, and Bol. Journal of Productivity Analysis 29(1), 15-21 (2008).
9. Davidovic, M., Uzelac, O., Zelenovic, V. Efficiency dynamics of the Croatian banking industry: DEA investigation. Economic research-Ekonomska istraživanja 32(1), 33-49 (2019).
10. Doyle, J., Green, R. Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. Journal of the operational research society 45(5), 567-578 (1994).

11. Emrouznejad, A., Shale, E.: A combined neural network and DEA for measuring efficiency of large scale datasets. Computers & Industrial Engineering 56(1), 249-254 (2009).
12. Fausett, L.: Fundamentals of neural networks: Architectures, algorithms, and applications. Englewood Cliffs, NJ: Prentice Hall (1994).
13. Fethi, M. D., Pasiouras, F.: Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. European journal of operational research 204(2), 189-198 (2010).
14. Kwon, H. B., Lee, J. Two-stage production modeling of large US banks: A DEA-neural network approach. Expert Systems with Applications 42(19), 6758-6766 (2015).
15. Kwon, H. B., Lee, J., Roh, J. J. Best performance modeling using complementary DEA-ANN approach. Benchmarking: An International Journal 23(3), 704-721 (2016).
16. Kwon, H. B.: Exploring the predictive potential of artificial neural networks in conjunction with DEA in railroad performance modeling. International Journal of Production Economics 183, 159-170 (2017).
17. Han, Y. M., Geng, Z. Q., Zhu, Q. X.: Energy optimization and prediction of complex petrochemical industries using an improved artificial neural network approach integrating data envelopment analysis. Energy Conversion and Management 124, 73-83 (2016).
18. Henriques, I. C., Sobreiro, V. A., Kimura, H., Mariano, E. B.: Efficiency in the Brazilian banking system using data envelopment analysis. Future Business Journal 4(2), 157-178 (2018).
19. Lam, M.: Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. Decision Support System, 37, 567–581 (2004).
20. Le, H. H., Viviani, J. L.: Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Research in International Business and Finance 44, 16-25 (2018).
21. Mostafa, M. M.: Modeling the efficiency of top Arab banks: A DEA–neural network approach. Expert systems with applications 36(1), 309-320 (2009).
22. Petropoulos, A., Siakoulis, V., Stavroulakis, E., Vlachogiannakis, N. E.: Predicting bank insolvencies using machine learning techniques. International Journal of Forecasting, 36(3) 1092-1113 (2020).
23. Pendharkar, P. C., Rodger, J. A.: Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumption. Decision Support Systems 36(1), 117–136 (2003).
24. Reserve Bank of India, Report on Trend and Progress of Banking in India 2017-18 Retrieved from https://www.rbi.org.in. last accessed 20/07/18
25. Samoilenko, S., Osei-Bryson, K.-M.: Determining sources of relative inefficiency in heterogeneous samples: Methodology using cluster analysis, DEA and neural networks. European Journal of Operational Research 206(2), 479–487 (2010).
26. Sathye, S., Sathye, M.: Do ATMs increase technical efficiency of banks in a developing country? Evidence from Indian banks. Australian Accounting Review 27(1), 101-111 (2017).
27. Sealey Jr, C. W., Lindley, J. T.: Inputs, outputs, and a theory of production and cost at depository financial institutions. The Journal of Finance 32(4), 1251-1266 (1977).
28. Sexton, Thomas R.: Data envelopment analysis: Critique and extension. New Directions for Program Evaluation 32, 73-105 (1986).
29. Sreekumar, S., Mahapatra, S. S.: Performance modeling of Indian business schools: a DEA-neural network approach. Benchmarking: An International Journal 18(2), 221-239 (2011).
30. Sherman, H.D. Gold, F.: Bank branch operating efficiency: Evaluation with data envelopment analysis, Journal of Banking and Finance 9(2), 297-315 (1985).

31. Shokrollahpour, E., Lotfi, F. H., Zandieh, M.: An integrated data envelopment analysis–artificial neural network approach for benchmarking of bank branches. Journal of Industrial Engineering International 12(2), 137-143 (2016).
32. Tavana, M., Abtahi, A. R., Di Caprio, D., Poortarigh, M.: An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking. Neurocomputing 275, 2525-2554 (2018).
33. Tone, K.: A slacks-based measure of efficiency in data envelopment analysis. European journal of operational research 130(3), 498-509 (2001).
34. Tosun, Ö.: Using data envelopment analysis–neural network model to evaluate hospital efficiency. International Journal of Productivity and Quality Management 9(2), 245-257 (2012).
35. Tsolas, I. E., Charles, V., & Gherman, T.: Supporting Better Practice Benchmarking: A DEA-ANN Approach to Bank Branch Performance Assessment. Expert Systems with Applications, 113599 (2020).
36. Wu, D. D., Yang, Z., Liang, L.: Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. Expert systems with applications 31(1), 108-115 (2006).
37. Wu, D.: Supplier selection: A hybrid model using DEA, decision tree and neural network. Expert systems with Applications 36(5), 9105-9112 (2009).
38. Yeh, Q. J.: The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation. Journal of the Operational Research Society 47(8), 980-988 (1996).

Table 1. Summary statistics of selected variables

Variables	Mean	Median	Minimum	Maximum
Operating Expenses	5333.83	536.89	4619.32	6024.36
Deposits	230605.02	30603.16	183213.68	264097.57
Performing Loans	175430.44	9260.53	166770.86	190928.55
Investments	74503.06	6814.34	67207.92	83831.35
Non-interest income	15711.55	27406.69	2912.99	64731.16

Table 2. Efficiency score measurement for five years

Particulars	2015	2016	2017	2018	2019	Overall
Average efficiency score	0.8217	0.7930	0.8652	0.7354	0.8955	0.8233
Maximum Efficiency Score	1	1	1	1	1	1
Minimum Efficiency score	0.5362	0.4610	0.4708	0.08802	0.5542	0.5061
Standard Deviation	0.1667	0.1737	0.1422	0.3458	0.1416	0.1391

Table 3. DEA parameters with slack analysis

Variables	Mean
Efficient DMUs	6
Average Efficiency score	0.8233
Slack value analysis (Input excesses and output shortfall):	
Operating Expenses	2,300.5385

Deposits	3,60,109.7535
Performing Loans	1,13,282.9135
Investments	83,320.7701
Non-interest income	1,00,305.4249

Table 4. Comparison of different network structures

Model Decision	Model Network	Error	Epochs	Accuracy
Excluded	(5,5,3)	0.019	15,656	96.15%
Excluded	(5,5,5)	0.026	19,188	98.01%
Included	(15,10)	0.007	9,259	98.5%

Table 5. Actual and predicted operational efficiency scores

Banks	Actual	Predicted	MAE	SE	Banks	Actual	Predicted	MAE	SE
B39	0.990000	0.986001	0.000077	0.000016	**B50**	0.611896	0.568766	0.000829	0.001860
B40	0.675409	0.654931	0.000394	0.000419	**B51**	0.711288	0.699880	0.000219	0.000130
B41	0.771183	0.716970	0.001043	0.002939	**B52**	0.663660	0.600759	0.001210	0.003956
B42	0.990000	0.768494	0.004260	0.049065	**B54**	0.592869	0.615368	0.000433	0.000506
B44	0.625537	0.584095	0.000797	0.001717	**B56**	0.703195	0.708060	0.000094	0.000024
B45	0.990000	0.860430	0.002492	0.016788	**B57**	0.597082	0.584757	0.000237	0.000152
B46	0.445490	0.353427	0.001770	0.008476	**B58**	0.665522	0.623967	0.000799	0.001727
B47	0.820521	0.814866	0.000109	0.000032	**B59**	0.990000	1.053788	0.001227	0.004069
B48	0.482282	0.463190	0.000367	0.000365	**B60**	0.990000	0.948562	0.000797	0.001717
B49	0.699680	0.673277	0.000508	0.000697	**B61**	0.836759	0.794670	0.000809	0.001771
								MSE	**0.619768**
								RMSE	**0.787254**

Table 6. Detail of average efficiency scores with bank codes

Bank Name (Public)	Bank Code	Efficiency score	Bank Name (Private)	Bank Code	Efficiency score
Allahabad Bank	B1	0.6300	City Union Bank Ltd.	B21	1.0000
Andhra Bank	B2	0.8620	Tamilnad Mercantile Bank Ltd.	B22	0.8599
Bank of Baroda	B3	0.9530	The Catholic Syrian Bank Ltd.	B23	0.8876
Bank of India	B4	0.6838	Dhanlaxmi Bank Ltd	B24	1.0000
Bank of Maharashtra	B5	0.5659	The Federal Bank Ltd.	B25	0.7456
Canara Bank	B6	0.9188	The Jammu & Kashmir Bank Ltd.	B26	0.6431
Central Bank of India	B7	0.5299	The Karnataka Bank Ltd.	B27	0.8358
Corporation Bank	B8	0.9286	The Karur Vysya Bank Ltd.	B28	0.7406

Dena Bank	B9	0.5061	The Lakshmi Vilas Bank Ltd.	B29	0.9197
Indian Bank	B10	0.8736	RBL Bank Ltd.	B30	0.9003
Indian Overseas Bank	B11	0.6446	The South Indian Bank Ltd.	B31	0.8436
Oriental Bank of Commerce	B12	0.7711	Axis Bank Ltd.	B32	0.9204
Punjab & Sind Bank	B13	0.7817	DCB Bank Ltd.	B33	0.9352
Punjab National Bank	B14	0.7637	HDFC Bank Ltd.	B34	0.8461
Syndicate Bank	B15	0.7968	ICICI Bank Ltd.	B35	1.0000
UCO Bank	B16	1.0000	Indusind Bank Ltd.	B36	0.7950
Union Bank of India	B17	0.8775	Kotak Mahindra Bank Ltd.	B37	0.7497
United Bank of India	B18	0.7616	YES Bank Ltd.	B38	1.0000
Vijaya Bank	B19	0.6540	IDBI Ltd.	B39	0.9848
State Bank of India (SBI)	B20	1.0000			
Average efficiency score		0.7751	Average efficiency score		0.8741
Standard Deviation		0.1527			0.1047

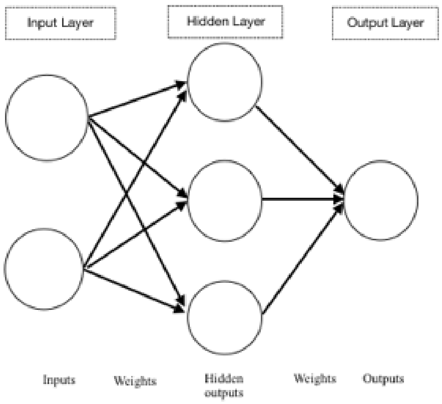

Fig.1. Typical structure of Artificial Neural Network

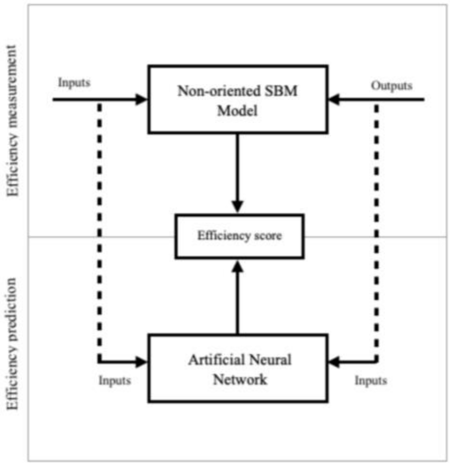

Fig. 2. Framework of Hybrid Modelling

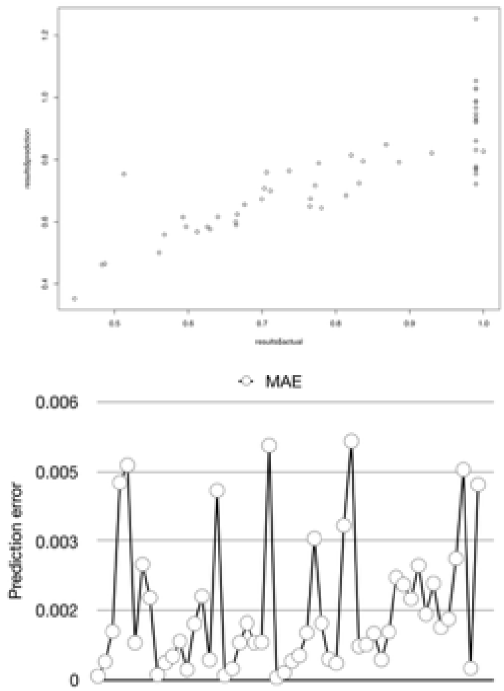

Fig. 3. Correlation and error for actual and predicted efficiency score

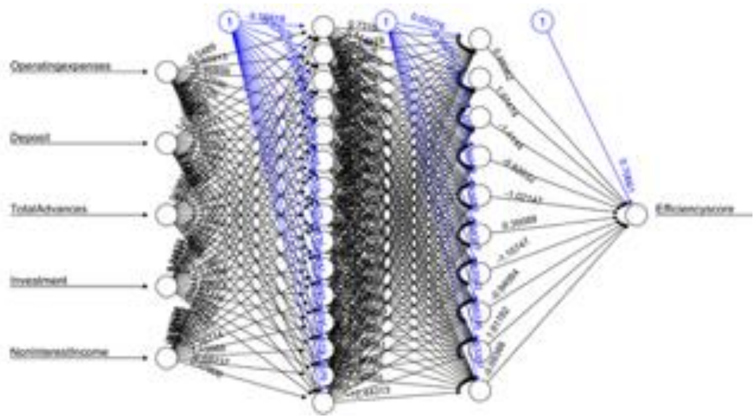

Fig.4 Neural Network Structure