

León, Carlos; Barucca, Paolo; Acero, Oscar; Gage, Gerardo; Ortega, Fabio

## Article

# Pattern recognition of financial institutions' payment behavior

Latin American Journal of Central Banking (LAJCB)

## Provided in Cooperation with:

Center for Latin American Monetary Studies, Mexico City

*Suggested Citation:* León, Carlos; Barucca, Paolo; Acero, Oscar; Gage, Gerardo; Ortega, Fabio (2020) : Pattern recognition of financial institutions' payment behavior, Latin American Journal of Central Banking (LAJCB), ISSN 2666-1438, Elsevier, Amsterdam, Vol. 1, Iss. 1/4, pp. 1-14, <https://doi.org/10.1016/j.latcb.2020.100011>

This Version is available at:

<https://hdl.handle.net/10419/336300>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Latin American Journal of Central Banking

journal homepage: [www.elsevier.com/locate/latchb](http://www.elsevier.com/locate/latchb)Pattern recognition of financial institutions' payment behavior <sup>☆</sup>Carlos León <sup>a,e,\*</sup>, Paolo Barucca <sup>c</sup>, Oscar Acero <sup>b</sup>, Gerardo Gage <sup>d</sup>, Fabio Ortega <sup>a</sup><sup>a</sup> Banco de la República, Colombia<sup>b</sup> Stratio, Colombia<sup>c</sup> University College London, United Kingdom<sup>d</sup> Centro de Estudios Monetarios Latinoamericanos (CEMLA), México<sup>e</sup> Tilburg University, the Netherlands

## ARTICLE INFO

## JEL Classification:

C45  
E42  
G21

## Keywords:

Payments  
Neural networks  
Feature selection  
Machine learning  
Pattern recognition

## ABSTRACT

We present a general supervised machine-learning methodology to represent the payment behavior of financial institutions starting from a database of transactions in the Colombian large-value payment system. The methodology learns a feedforward artificial neural network parameterization to represent the payment patterns through 113 features corresponding to financial institutions' contribution to payments, funding habits, payment timing, payment concentration, centrality in the payment network, and systemic effects due to failure to pay. We then use the representation to compare the coherence of out-of-sample payment patterns of the same institution to its characteristic patterns. The performance is remarkable, with an out-of-sample classification error around three percent. The performance is robust to reductions in the number of features by unsupervised feature selection. In addition, we confirm that network centrality and systemic effect features definitively contribute to enhancing the performance of the methodology. For financial authorities, this is a major step towards the automated detection of individual financial institutions' anomalous behaviors in payment systems.

## 1. Introduction

The amount of electronic transactions occurring in financial systems is constantly growing. Identifying usual and unusual patterns of behavior is crucial to monitoring the safe and efficient functioning of payment systems. Machine-learning methodologies have demonstrably learned complex patterns and can provide accurate data representations. In the context of payment systems, machine learning is leading the development of automated monitoring tools that enhance the ability of financial authorities to supervise and oversee payment systems and financial market infrastructures.

Starting from a large-value payment systems dataset, [Triepels et al. \(2017\)](#) and [Sabeti and Heijmans \(2020\)](#) have shown that accurate representations can be learned to identify payment vectors that deviate from their norm. This is accomplished by an entirely data-driven, unsupervised neural network methodology (i.e., an autoencoder architecture) that can consistently recognize a typical—yet potentially complex—pattern of financial transactions. [León \(2020\)](#) accomplishes a similar representation based on principal

<sup>☆</sup> Opinions and statements in this article are the sole responsibility of the authors and do not represent those of the institutions to which the authors are affiliated. We thank Serafín Martínez, Raúl Morales, and Deisy Zambrano for their contributions to the methodology's development. We thank Clara Machado, Serafín Martínez, and Raúl Morales for their comments on and suggestions for this article. We thank the two anonymous reviewers for their contributive comments and suggestions. Any remaining errors are our own.

\* Corresponding author at: Banco de la República, Colombia. Postal address: Carrera 7 # 14-78, Bogotá-110321, Colombia.  
E-mail address: [cleonrin@banrep.gov.co](mailto:cleonrin@banrep.gov.co) (C. León).

<https://doi.org/10.1016/j.latchb.2020.100011>

Received 9 September 2020; Accepted 30 October 2020

Available online 21 November 2020

2666-1438/© 2020 The Authors. Published by Elsevier B.V. on behalf of Center for Latin American Monetary Studies. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

component analysis (PCA) and clustering methods. However, to the best of our knowledge, no study has reported the representation of individual financial institutions' behaviors in payment systems.

Financial institutions do not act in isolation but in a complex network of interactions (see Farmer et al., 2012; Caldarelli, 2020). Consequently, in such representations of financial institutions' behaviors, integrating systemic features of the entire financial network is now considered crucial. This became evident as the macroprudential approach to financial stability (see Crockett, 2000; Borio, 2003) was called for in the aftermath of the Global Financial Crisis—when evidence showed the importance of risk propagation channels that amplify losses in unforeseen manners (see Haldane and May, 2011). This approach emphasizes how financial relationships create networks that link institutions that may not have considered their mutual interdependence and risk exposures. It is opposed to a purely institution-centric approach, where institutions are evaluated individually, neglecting most network effects.

In this vein, the recent data-driven machine-learning approach applied to payment systems grants a new perspective on the supervision and oversight of financial systems. Machine-learning methodologies can be trained to represent not only financial institutions' payment behaviors but also to take into account financial institutions' interrelations in payment networks as determinants of that behavior.

We use a feedforward artificial neural network to represent the payment behavior of individual financial institutions starting from a dataset of transactions in the Colombian large-value payment system. We use features corresponding to financial institutions' payment patterns, namely their contributions to payments, funding habits, payment timing, payment concentration, centrality in the payment network, and systemic effects due to failure to pay. We use an individual representation of institutions to compare the coherence of out-of-sample payment patterns of the same institution to its characteristic patterns. That is, we address the financial institutions' payment behavior problem as a supervised pattern-recognition problem. We test whether classification performance of financial institutions by their payment patterns is robust to reductions in the number of features by unsupervised feature selection. In addition, we measure the extent to which a macroprudential approach to the monitoring of payment systems, in the form of network centrality and systemic effect features, contributes to enhancing the performance of the classification methodology.

We achieve high-performance out-of-sample pattern recognition. Based on 113 selected features, we classify 26 banks with a three percent out-of-sample mean classification error. After reducing the number of features by unsupervised feature selection, the misclassification doubles—yet is still noticeably low. Discarding features related to a macroprudential approach to financial institutions' behaviors (i.e., network and simulation methods) increases the mean classification error around 29 percent.

For financial authorities, this is a major step towards the automated detection of individual financial institutions' anomalous behaviors in payment systems. A sizeable change in model's performance can flag a change in financial institution's behavior within the payment system. That is, financial authorities may use changes in the model's individual or joint classification performance as warning signals of behavioral alterations that should be further studied. With financial stability in view, this adds to and enhances existing tools for the supervision and oversight of financial markets and payment systems.

## 2. Literature review

Our choice of an artificial neural network for this pattern recognition problem is based on three main strengths highlighted in related literature. First, given enough hidden layers and enough training samples, artificial neural networks can closely approximate any function, thus they can deal with nonlinear relationships between factors in the data (see Bishop, 1995; Han and Kamber, 2006; Fioramanti, 2008; Demyanyk and Hasan, 2009; Eletter et al., 2010; Sarlin, 2014; Hagan et al., 2014). Second, artificial neural networks make no assumptions about the statistical distribution or properties of the data (see Zhang et al., 1999; McNelis, 2005; Demyanyk and Hasan, 2009; Nazari and Alidadi, 2013; Sarlin, 2014). Third, artificial neural networks have proven to be very effective classifiers, even better than the state-of-the-art models based on classical statistical methods (see Wu, 1997; Zhang et al., 1999; McNelis, 2005; Han and Kamber, 2006).

Artificial neural networks have been applied to classification and anomaly detection problems in the financial domain. They have been used in credit card fraud detection (see Aleskerov et al., 1997; Ghosh and Reilly, 1994; Dorransoro et al., 1997), which presents the challenge of providing a prompt classification to be able to act and stop an on-going fraudulent activity. They have been used to counter money laundering (see Brause et al., 1999), where the absence of labeled data about account holders and transactions is a significant challenge. In auditing, financial ratios and artificial neural networks have been used together to identify potential tax evasion cases (see Wu, 1997). In credit risk, artificial neural networks have been used to improve loan decisions by classifying potential borrowers' credit quality (see Angelini et al., 2008; Eletter et al., 2010; Nazari and Alidadi, 2013; Bekhet and Eletter, 2014). They have also been used to classify firms according to their likelihood of bankruptcy or failure (see Tam and Kiang, 1990; Tam, 1991; Salchenberger et al., 1992; Wilson and Sharda, 1994; Olmeda and Fernández, 1997; Zhang et al., 1999; Atiya, 2001; Brédart, 2014). Fioramanti (2008), Sarlin (2014), and Holopainen and Sarlin (2016) use artificial neural networks in macro early-warning systems. In addition, artificial neural networks have been used to classify banks as domestic or foreign (see Turkan et al., 2011) and Islamic or conventional (see Khediri et al., 2015), as well as to classify banks' balance sheets into their corresponding bank (see León et al., 2017).

In recent work related to the oversight of payment systems, Triepels et al. (2017) and Sabetti and Heijmans (2020) developed methodologies based on a particular artificial neural network (i.e., an autoencoder) to detect anomalous payments networks. Triepels et al. (2017) worked on large-value payment networks corresponding to the Dutch partition of the Eurosystem payment network (TARGET2). Sabetti and Heijmans (2020) worked on data from the Automated Clearing Settlement System, a Canadian retail payment system. León (2020) developed a methodology to detect anomalous payment networks in the Colombian large-value payments system (CUD) based on a mixture of PCA and clustering methods. These three works share a common aim; namely, to detect

anomalous payments networks. They do not address the individual behaviors of financial institutions participating in the payment systems but rather the overall structure of the payment networks. Further, they share an unsupervised approach and the construction of synthetic anomalous payments to test out-of-sample performance.

We aim to understand the individual behaviors of financial institutions in payment networks. Our approach is a supervised methodology to represent the payment behavior of financial institutions in the Colombian large-value payment system. In this vein, our work is related to León et al. (2017), who used a supervised methodology to represent banks' characteristic financial structures through features corresponding to their balance sheets. Because it is a pattern-recognition problem based on banks' balance sheet data, León et al. (2017) showed an artificial neural network could accurately classify out-of-sample banks by learning the main features of their balance sheets. Similarly, our work is a pattern-recognition problem based on financial institutions' individual and systemic behaviors in the large-value payment system.

### 3. Methods

The base case model uses an artificial neural network for pattern recognition on a set of 113 features that capture the behavior of 26 banking institutions participating in the Colombian large-value payment system during 2019. That is, we train an artificial neural network to classify banking institutions based on their individual payment behavior as captured by the selected 113 features. Although the model is generalized to work with all financial institutions, focusing on banking institutions not only allows for an easier and more tractable presentation of results but also acknowledges that banks are the most contributive type of institution from several viewpoints (e.g., contribution to payments and importance in the payment network).<sup>1</sup>

In this section, we discuss the methodologies we use for pattern recognition and dimensionality reduction. First, we introduce the feedforward artificial neural network architecture and explain how it can be applied to our dataset. Second, starting from transactional data, we describe the feature selection procedure that defines the set of features for each institution in our system. Third, we present PCA and describe how we apply it to our data for the dimensionality and noise reduction of the set of features.

#### 3.1. The feedforward artificial neural network architecture

We employ a feedforward artificial neural network to classify financial institutions as in a multiclass classification problem.<sup>2</sup> We consider a set of distinctive features that characterize each financial institution. In our case,  $p_{t,v}$  is the  $v$  feature of the institution corresponding to example  $t$ , defining a matrix  $P$  of size  $T \times V$ , where  $V$  is the number of features per example and  $T$  is the number of examples; an example corresponds to a financial institution and its features on a certain date. The feedforward artificial neural network is trained to recognize the pattern of features of each of the financial institutions within the training set.

As usual with multiclass classification problems, where classification targets are discrete variables, we use the *cross-entropy* ( $CE$ ) error function to train the artificial neural network (see Bishop, 1995). Where  $A^{(t)}$  and  $Q^{(t)}$  represent two binary variables corresponding to the predicted and target (i.e., observed) classifications for all examples  $t$ , respectively, the cross-entropy error is calculated as shown in (1) (see Hagan et al., 2014).<sup>3</sup>

$$CE(A^{(t)}, Q^{(t)}) = - \sum_{t=1}^T Q^{(t)} \log \frac{A^{(t)}}{Q^{(t)}} \quad (1)$$

Nevertheless, for interpretability and comparability purposes, the performance of the model is measured by its classification error. In our case, the classification error or misclassification is calculated as the percentage of financial institutions incorrectly classified by the model. The lower the misclassification, the better the performance of the model.

In our case, feature matrix  $P$  is a two-dimensional  $6369 \times 113$  matrix that contains 113 leading indicators or features ( $V = 113$ ) for the financial institutions that participated in the large-value payment system during 2019. For training purposes, we chose 2019 because it was the most recent tranquil 1-year period before the COVID-19 pandemic. Further, the 1-year period was chosen to balance performance and the relevance of data: Shorter periods (e.g., 3 and 6 months) attained lower classification performance, whereas longer periods (e.g., 18, 24, and 36 months) increased performance slightly but could be questionable because financial institutions' early behavior could be outdated (i.e., unrepresentative) for pattern-recognition purposes.

For each of the 6369 examples ( $T = 6369$ ), we have the corresponding identity of the financial institution (i.e., it is a supervised machine learning model). Therefore,  $P$  is the input to our artificial neural network, with the element  $p_{t,v}$  corresponding to the  $v$  feature of the  $t$  example. As usual, to avoid issues related to the scale of features across different financial institutions and days,

<sup>1</sup> For illustrative purposes, the main results for all financial institutions (i.e., banking and non-banking) are reported in Appendix 2.

<sup>2</sup> In a feedforward artificial neural network, the connections between nodes do not form cycles or feedback loops; those cycles, typical of recurrent neural networks, are customary for time-series or dynamic problems, which are absent in our case. Thus, the outputs can be expressed as deterministic functions of the inputs, and so the whole network represents a multivariate nonlinear functional mapping (see Bishop, 1995).

<sup>3</sup> Where the targets are in the form of continuous variables (e.g., prediction of prices), the sum (or the mean) of squared errors is used—instead of the cross-entropy.

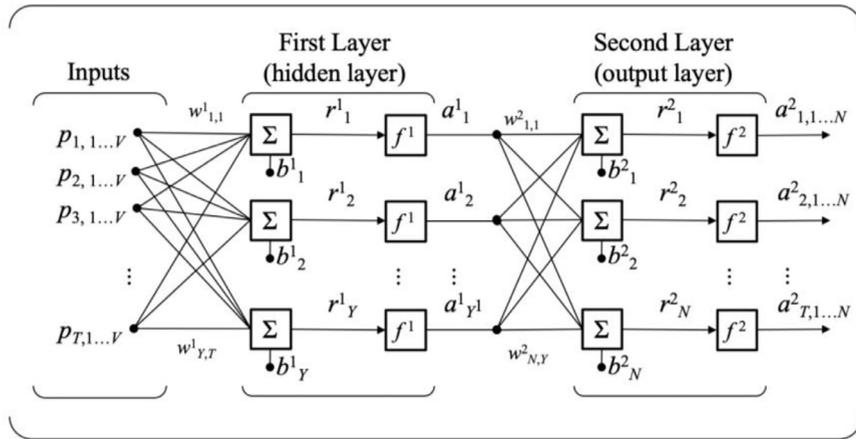


Fig. 1. Neural network architecture.

matrix  $P$  is row normalized.

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,V} \\ p_{2,1} & p_{2,2} & & \\ \vdots & & \ddots & \vdots \\ p_{T,1} & & \dots & p_{T,V} \end{bmatrix} \tag{2}$$

The target matrix  $Q$  contains the actual target class. It is a binary array with  $T = 6369$  rows (i.e., the same as  $P$ ) and  $N$  columns that identify the class (i.e., financial institution identity) to which each row belongs; in each row of  $Q$ , there is a single column with a value of 1 matching the financial institution’s identity, and the rest are zeros. For instance,  $q_{t,n} = 1$  corresponds to features in example  $t$  belonging to financial institution  $n$ .

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \dots & q_{1,N} \\ q_{2,1} & q_{2,2} & & \\ \vdots & & \ddots & \vdots \\ q_{T,1} & & \dots & q_{T,N} \end{bmatrix} \tag{3}$$

Training the feedforward artificial neural network will attain a prediction matrix  $A$ , which has the same dimensions as  $Q$ . However, unless complete certainty is achieved, it is a non-binary array. In our case, the neural network architecture achieves for each row in  $A$  the probability associated with each class; this is achieved by using a *softmax* transfer function—as explained next. Therefore, elements in each row of  $A$  are in the  $[0,1]$  range, and the sum of each row equals 1.

A standard two-layer network, with one hidden layer and one output layer, is the simplest and most commonly used in economic and financial applications (Zhang et al., 1999; McNelis, 2005; Witten et al., 2011). Therefore, we selected this architecture for parsimony. Fig. 1 depicts the architecture, in which the superscripts identify the layer (i.e., first or second);  $p_{v,t}$  are the elements of the input matrix  $P$ ;  $Y$  is the number of neurons in the hidden layer; and  $r$  represents a net input vector that results from the weighted sum ( $\Sigma$ ) involving the weights and bias terms that change while training  $w$  and  $b$ , respectively. Regarding activation or transfer functions in the hidden and output layers,  $f^1$  and  $f^2$ , respectively, the first accommodates a customary log-sigmoid function, whereas the second is a *softmax* function.<sup>4</sup> The number of neurons in the output layer is the number of financial institutions,  $N$ . The number of neurons in the hidden layer,  $Y$ , may be adjusted. As before, because this is a classification problem, the cross-entropy fitness measure is preferred.

<sup>4</sup> The log-sigmoid function in the hidden layer takes the form  $f(x) = 1/(1 + e^{-x})$ . This is a standard choice because its output range is between 0 and 1, it is able to map a very large input domain into a small range of outputs, and it is easily differentiable. Regarding the output layer, the softmax function is interesting for our classification purposes because its outcome can be interpreted as the probabilities associated with each class, which is particularly convenient because we use cross-entropy as our error measure (see Bishop, 1995; Hagan et al., 2014). After dropping the superscript corresponding to the second layer, the softmax function in the output layer of Fig. 1 is of the form  $f(r_n) = e^{r_n} / \sum_{j=1}^N e^{r_j}$ .

Our work focuses on out-of-sample classification. Therefore, we should avoid overfitting when training the neural network.<sup>5</sup> The overfitting problem may be described as the model's ability to successfully fit in-sample but fail at fitting out-of-sample (see Shmueli, 2010; Varian, 2014). Overfitting occurs when the parameters in the artificial neural network, namely  $w$  and  $b$ , depend too strongly on the details of the particular examples used to produce them (Witten et al., 2011). In this sense, artificial neural networks provide general non-linear mappings between a set of input variables and a set of output variables. The goal is not to memorize the training data but to model the underlying generator of the data (Bishop, 1995).

In our case, we avoid overfitting through early stopping. Instead of allowing the algorithm to attain the minimal in-sample error (at the expense of out-of-sample error), we stop the minimization process before the solution's complexity inhibits its generalization capability. We stop training before the minimum in-sample is reached, so the network will be less likely to overfit (Hagan et al., 2014).

An intuitive and customary early-stopping criterion is cross-validation.<sup>6</sup> For this purpose, the data are divided into three independent datasets: a training dataset, a validation dataset, and a test dataset, which are randomly selected with approximately 70 percent, 15 percent, and 15 percent allocations, respectively (see Hagan et al., 2014). With this partition, the training, validation, and test datasets comprise 4459, 955, and 955 examples, respectively. The training dataset is used to train the artificial neural network, i.e., to minimize the difference between the predicted and the actual target value, measured with cross-entropy. As the neural network is trained, the validation dataset is used simultaneously to check how the estimated parameters fit out-of-sample data. Because we expect the validation error to start increasing as overfitting arises (Hastie et al., 2013), the validation dataset is used to determine when to stop training. After training, the error obtained from the test dataset is used to evaluate the out-of-sample performance of the artificial neural network, i.e., its generalization capability.

Regarding the number of neurons in the hidden layer,  $Y$ , we tried several options in the 20–110 range at 10-neuron increments. As shown in the results (Section 4), the classification error in the test data set becomes stable when using 60 neurons or more.

### 3.2. Feature selection

In payment systems, financial institutions send and receive funds to and from each other over time using different payment concepts (e.g., (un)secured lending, securities purchases, foreign exchange transactions, third-party transfers, and central bank repos). For each payment concept, we can create a time series of payment networks. Let  $N$  be the number of financial institutions,  $D$  be the number of days, and  $M$  be the number of payment concepts, and the  $N \times N \times D \times M$  hypermatrix  $X$  accommodates all transactions between financial institutions. Each element  $x_{p,q,d,m}$  corresponds to the value of the funds sent by financial institution  $p$  to financial institution  $q$  on day  $d$  under the type of payment  $m$ .

Hypermatrix  $X$  contains each financial institution's behavior. However, this behavior comprises each financial institution as the sender and receiver of payments for each type of payment. The behavior of financial institution  $p$  as a sender of payments is contained in elements  $x_{p,1\dots N,1\dots D,1\dots M}$ , whereas its behavior as a receiver is contained in elements  $x_{1\dots N,p,1\dots D,1\dots M}$ . Therefore, a traditional manner of detecting an anomaly in a financial institution's behavior involves determining a normal pattern from its interactions with all other financial institutions for all payment types during a sample period. Nevertheless, this may become computationally burdensome: each financial institution's behavior at day  $d$  depends on  $2 \times M \times (N - 1)$  relations, and the behavior of all financial institutions is a  $N \times M \times (2 \times (N - 1))$  problem. There is a dimensionality problem.

In the Colombian case, the large-value payment system is non-tiered, with all kinds of financial institutions participating directly. The large-value payment system included 111 financial institutions in 2019, comprising commercial banks (26), other banking institutions (22), securities broker-dealer firms (19), insurance firms (13), private pension funds (4), and trust companies (26).<sup>7</sup> There are 20 payments concepts worth studying because of their contribution to total payments and to the study of payment behavior ( $M = 20$ ).<sup>8</sup> Therefore, a single financial institution's behavior on day  $d$  consists of  $4400 = 2 \times M \times (N - 1)$  elements, whereas the behavior of all financial institutions on day  $d$  is contained in about half a million elements. Because we are focusing on banks, a single bank's behavior on day  $d$  consists of  $1000 = 2 \times M \times (N - 1)$  elements, whereas the behavior of all banks on day  $d$  is contained in 26,000 elements.

Payment systems have been studied extensively (McAndrews and Rajan, 2000; Becher et al., 2008; Bernal et al., 2012; Diehl, 2013; Denbee et al., 2014; Martínez and Cepeda, 2018); thus, some measures to monitor financial institutions have already been well-established as leading indicators of their behavior. Accordingly, we have selected several measures, which may be conveniently depicted under the following typology.<sup>9</sup>

<sup>5</sup> This feedforward artificial neural network is trained through back-propagation. That is, the training involves an iterative procedure for minimizing the classification error, with adjustments made in the parameters in a backwards sequence (see Bishop, 1995).

<sup>6</sup> As will be presented next, implementing cross-validation is enough to avoid overfitting in our case, i.e., to attain high out-of-sample classification performance. Therefore, we regard additional or more sophisticated resampling techniques to avoid overfitting as unnecessary.

<sup>7</sup> Other institutions participate directly in the large-value payment system, such as the Central Bank, the Ministry of Finance, information processors, and financial infrastructures. They have been excluded because our aim is to represent financial institutions' payment behavior.

<sup>8</sup> About 200 payments types have been identified as uninformative or as having limited informational content about financial institutions' behavior, such as fees and fines paid to financial infrastructures, accounting adjustments, and taxes. They are not excluded but are considered in an aggregated manner.

<sup>9</sup> The exact features, including their description and calculation, are not revealed for confidentiality reasons. However, most of the features correspond to standard measures from related literature that are based on financial institutions' activity in large-value payment systems.

The first type of measures assesses financial institutions' contribution to payments. For the 20 most important payment concepts (e.g., repos with the central bank, secured and unsecured interbank lending, and sovereign securities purchases) and their aggregate, we measured the amount of net and gross payments—sent and received. The second type assesses how financial institutions fund their payments. We use the initial balance in central bank's accounts, the contribution of liquidity recycling (i.e., using received funds to make payments), the contribution of liquidity savings mechanisms (e.g., multilateral offsetting), and borrowing sources (e.g., central bank intraday and term repos; interbank lending). The third type assesses the timing of payments during the day, in the form of the distribution of intraday hourly payments from 7:00 to 20:00. The fourth type assesses the concentrations of sent and received payments, in the form of Herfindahl–Hirschman indexes by payment type and counterparty. Fifth are those aimed at evaluating the importance in the payment network, measured by different network-centrality metrics (degree, strength, authority, hub, SinkRank, and PageRank).<sup>10,11</sup> Sixth are those intended to gauge the systemic footprint in case of failure, in the form of the simulated decrease in overall payments due to an inability to make discretionary payments—payments that financial institutions can deliberately delay or refuse to make.<sup>12</sup> We refer to the last two types as the network features and the simulation-based features, respectively, which pertain to a systemic approach to monitoring financial and payment systems.

Based on these six classes of measures, a set of 113 leading indicators of financial institutions' payment behavior was created from hypermatrix  $X$ . Using this set of leading indicators reduces the dimensionality from 4400 elements per financial institution on day  $d$  to 113. If all financial institutions are considered, the input is contained in a two-dimensional matrix  $P$  with 24,234 rows and 113 columns; if financial institutions are limited to banking institutions (i.e., the most contributive type of financial institution), then  $P$  is a 6369-row and 113-column input matrix.<sup>13</sup>

Although working on 113 features is suitable in our case, such a non-small number of features may contain potentially redundant or noisy data. Further reducing the number of elements may contribute to testing the robustness of the chosen features and the classification model. Instead of subjectively discarding leading indicators, we implemented PCA dimensionality reduction on the 113 selected features.

### 3.3. Principal component analysis dimensionality reduction

PCA is an unsupervised feature-selection method in which the data are orthogonally transformed to find high-variance directions, while low-variance ones are discarded (Mehta et al., 2019). PCA finds an agnostic mapping from original features into a new lower-dimensional set of features, with minimum information loss.<sup>14</sup>

In our case, PCA feature selection worked on the  $113 \times 113$  covariance matrix from input matrix  $P$ —namely,  $C = P'P$ . As in (4), PCA decomposes  $C$  into eigenvector and eigenvalue matrices,  $\Gamma$  and  $\Lambda$ , respectively.

$$C = \Gamma \Lambda \Gamma' \quad (4)$$

Here,  $\Lambda$  is a diagonal matrix, with diagonal elements corresponding to the eigenvalues of  $C$ , such that the eigenvalues are in decreasing order of magnitude. The matrix  $\Gamma$  contains the eigenvectors as columns paired to the eigenvalues, such that the  $i$ th column contains the  $i$ th eigenvector. The first principal component, corresponding to the first column in  $\Gamma$ , lies in the direction of the samples' maximum variance, whereas the second column corresponds to the direction of maximum variance in the remaining data, and so on (Ding and Tian, 2016).

By choosing a subset of eigenvectors,  $\Gamma^*$ , it is possible to map the original set of features into a lower-dimensional set,  $X^*$ . As in (5),  $X^*$  is attained by calculating a projection of  $X$ . If  $\Gamma^*$  is a subset that contains the first  $i$  eigenvectors of  $\Gamma$ , then  $X^*$  is a mapping of the 113 features into  $i$  features. The retained variance in this mapping is calculated as the contribution of the corresponding eigenvalues in  $\Lambda^*$  to the sum of eigenvalues in  $\Lambda$ .

$$X^* = X' \Gamma^* \quad (5)$$

We selected the number of eigenvectors to attain  $X^*$  by choosing a minimum retained variance target. With a 0.90 minimum retained variance target, the number of features in  $X^*$  is 26. Therefore, the number of features decreased from  $V = 113$  to  $V^* = 26$ . When the PCA-based feature-selection method is implemented, matrix  $P$  changes—but  $Q$  remains unaltered. Based on [5], with a

<sup>10</sup> See Kleinberg (1998), Brin and Page (1998), Newman (2010), Soramaki and Cook (2013), and León et al. (2018) for a comprehensive review of these centrality measures; we use the Financial Network Analytics (FNA) platform to calculate these centrality measures.

<sup>11</sup> Regarding payment networks, centrality measures tend to be correlated (see Martínez-Jaramillo et al., 2014). In our case, the PCA dimensionality reduction performed as a feature-selection procedure was designed to avoid redundancies across all features—including centrality measures.

<sup>12</sup> In this case, we use the Financial Network Analytics (FNA) platform to simulate payments under the assumption of a financial institution being unable to make discretionary payments during the day. The amount of payments left unsettled with respect to the observed payments is a measure of that financial institution's systemic footprint in case of failure. This simulation was performed for each financial institution, for each day in the sample. For more details, refer to the FNA documentation at <https://fna.fi/technologies/#Simulation>.

<sup>13</sup> Not all financial institutions participated every day in the large-value payment system during the period under analysis. Thus, when all financial institutions are considered, the number of rows is 24,234, instead of  $26,973 = N \times T$ ; when only banks are considered, the number of rows is 6369 instead of  $6370 = N \times T$ .

<sup>14</sup> References on using PCA for dimensionality reduction are Vishwanathan et al. (2010), Sree and Venkata (2014), Alpaydin (2014), Ding and Tian (2016), and Mehta et al. (2019).

**Table 1**

Mean classification error for different choices of the number of neurons in the hidden layer. Calculated on 100 independent training processes; the standard deviations are reported in parentheses. The lowest mean classification errors in the test set are in bold.

Set	Number of neurons in the hidden layer									
	20	30	40	50	60	70	80	90	100	110
Training	1.87 (4.60)	0.99 (0.52)	0.84 (0.41)	0.84 (0.38)	0.80 (0.37)	0.80 (0.82)	0.77 (0.36)	<b>0.74 (0.34)</b>	0.81 (0.46)	0.86 (0.85)
Validation	4.94 (4.47)	3.46 (0.69)	3.17 (0.63)	3.02 (0.60)	2.90 (0.62)	2.94 (0.95)	2.80 (0.55)	<b>2.64 (0.58)</b>	2.70 (0.65)	2.87 (1.07)
Test	5.20 (4.47)	3.65 (0.72)	3.37 (0.67)	3.25 (0.69)	3.08 (0.58)	2.96 (0.88)	2.88 (0.50)	<b>2.80 (0.50)</b>	2.89 (0.62)	3.07 (1.18)

**Table 2**

Lowest classification error for different choices of the number of neurons in the hidden layer. The lowest overall classification error is in bold.

Set	Number of neurons in the hidden layer									
	20	30	40	50	60	70	80	90	100	110
Test	2.83	2.09	1.68	1.88	1.68	<b>1.47</b>	1.68	1.78	1.78	1.57

0.90 minimum retained variance target,  $P^*$  is the 26-column projection of 113-column matrix  $P$ .

$$P^* = \begin{bmatrix} P_{1,1}^* & P_{1,2}^* & \dots & P_{1,V^*}^* \\ P_{2,1}^* & P_{2,2}^* & & \vdots \\ \vdots & & \ddots & \\ P_{T,1}^* & \dots & & P_{T,V^*}^* \end{bmatrix} \tag{6}$$

#### 4. Results

We present the three main pattern-recognition experiments we performed. First, in the base case experiment, we use the full set of 113 selected features to train and validate the artificial neural network while varying the number of neurons in the hidden layer. Second, we removed the network- and simulation-based features and repeated the training and validation to compare the resulting performance with the full model. Third, we explored the possibility of reducing dimensionality and noise by redefining our full set of features with PCA, i.e., identifying a smaller set of linear combinations of our initial features that maximize the fraction of explained variance in the feature space.

The full model’s out-of-sample classification performance was remarkable. In the base-case scenario, which corresponds to the classification of 26 banks based on 113 features, the out-of-sample mean classification error was about 3 percent when using 60 or more neurons, and errors did not cluster in any particular bank. When we discarded features related to a systemic approach to financial institutions’ behavior (i.e., network and simulation methods), the mean classification error increased to around 29 percent, which verifies these features’ contribution to the model’s performance. When PCA-feature selection was implemented, we reduced the number of features from 113 to 26, but the mean classification error doubled with respect to the base-case scenario. We present each of these three cases in detail next.

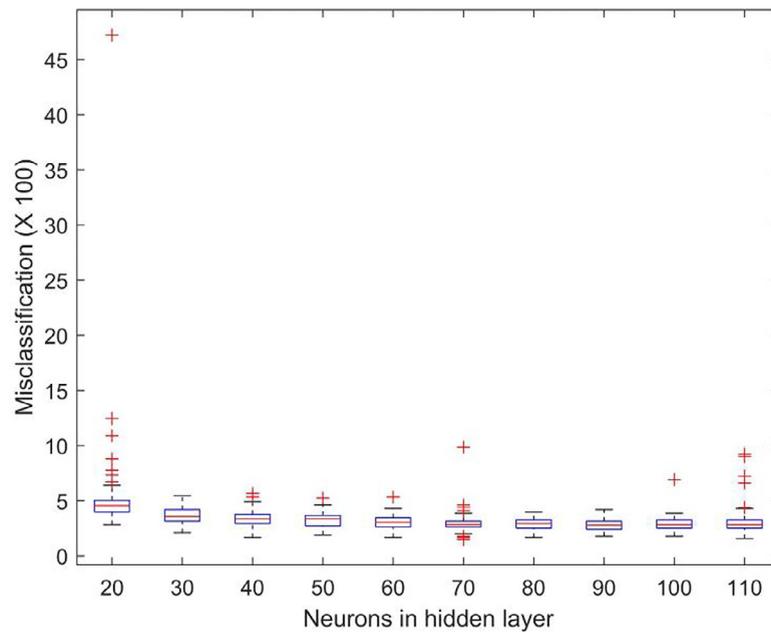
##### 4.1. Base-case scenario

As artificial neural networks’ results depend on their initialization parameters (see Bishop, 1995; Han and Kamber, 2006; Sarlin, 2014), we ran 100 independent training processes for several choices of the number of neurons in the hidden layer.<sup>15</sup> We report the mean classification error attained for each choice, for the training, validation, and test sets. The overall performance attained with the selected architecture is displayed in Table 1. The lowest mean classification error in the test set, 2.80 percent, was attained with 90 neurons in the hidden layer.

For the test set, Fig. 2 shows the boxplot for our choices of the number of neurons. The top and bottom of each box is the 25th and 75th percentile of the samples, respectively; the line inside the box is the median, and the whiskers mark  $a \pm 2.7$  standard deviation limit that reveals outliers (shown with +). As expected, for low numbers of neurons, the classification error is higher and more dispersed. Based on Fig. 2 and Table 1, the lowest mean classification error and dispersion occurred in the 80–90 neurons range.

Table 2 shows the lowest classification error for each choice of neurons for the test set—that is, the best of the 100 independent runs for each number of neurons. The lowest classification error, 1.47 percent, was attained in a run with 70 neurons.

<sup>15</sup> In our case, not only were the results dependent on the initial values of the weights ( $w$  and  $b$ ) but also on the random partition of the dataset required by the cross-validation procedure to avoid overfitting.



**Fig. 2.** Boxplot of classification errors for the test set, for different choices of the number of neurons in the hidden layer. Calculated on 100 independent training processes.

**Table 3**

Mean classification error for different choices of the number of neurons in the hidden layer, excluding network- and simulation-based features. Calculated on 100 independent training processes; standard deviations are reported in parentheses. The lowest mean classification errors in the test set are in bold.

Set	Number of neurons in the hidden layer										
	20	30	40	50	60	70	80	90	100	110	
Training	2.11 (0.93)	1.49 (0.65)	1.21 (0.51)	1.30 (0.98)	1.10 (0.42)	1.03 (0.43)	<b>1.09 (0.41)</b>	1.07 (0.40)	1.08 (0.46)	1.13 (0.73)	
Validation	5.76 (1.16)	4.70 (0.79)	4.10 (0.60)	3.94 (1.15)	3.74 (0.66)	3.60 (0.68)	<b>3.67 (0.64)</b>	3.57 (0.72)	3.53 (0.83)	3.60 (0.89)	
Test	6.00 (1.25)	4.65 (0.72)	4.19 (0.77)	4.15 (1.18)	3.81 (0.61)	3.84 (0.67)	<b>3.61 (0.57)</b>	3.71 (0.64)	3.66 (0.72)	3.79 (0.92)	

In addition to the classification error, we report the confusion matrix.<sup>16</sup> The confusion or misclassification matrix is a square table that relates the target class (in rows) to the output class achieved by the model (in columns). For a classifier to have good accuracy, most of its predictions must be represented along the diagonal of the confusion matrix (i.e., the predicted class must match the observed class, also referred to as true positives), with the rest of the entries (i.e., below or above the diagonal) being close to zero (Han and Kamber, 2006).

Fig. 3 exhibits the confusion matrix corresponding to the lowest classification error attained reported in Table 2. As expected from a good classifier, most of the predictions matched the target classification in the confusion matrix's diagonal, whereas a few (14 out of 955) were outside of the diagonal. Furthermore, because predictions outside the diagonal were not only scarce but also not clustered in any bank, the model attained a noticeably good degree of generalization for all banks in the sample.

#### 4.1.1. The contributions of network- and simulation-based features

One of the most important traits of our dataset is the usage of network- and simulation-based features corresponding to a systemic approach to financial institutions' payment behavior. There are 10 (out of 113) such features. Table 3 exhibits the mean classification error attained for each choice of the number of neurons, for the training, validation, and test sets, after removing these 10 features.

According to Table 3, the lowest mean misclassification in the test set (3.61 percent) was achieved when using 80 neurons. Because the lowest mean classification in the test set in Table 1 is 2.80 (i.e., the base-case scenario), the gain in classification performance from including 10 network- and simulation-based features was about 22.44 percent. That is, by increasing the number of features by

<sup>16</sup> We also report the receiver operating characteristic (ROC) curve (see Appendix 1). Performance shown by the ROC curve concurs with reported classification error and the confusion matrix.

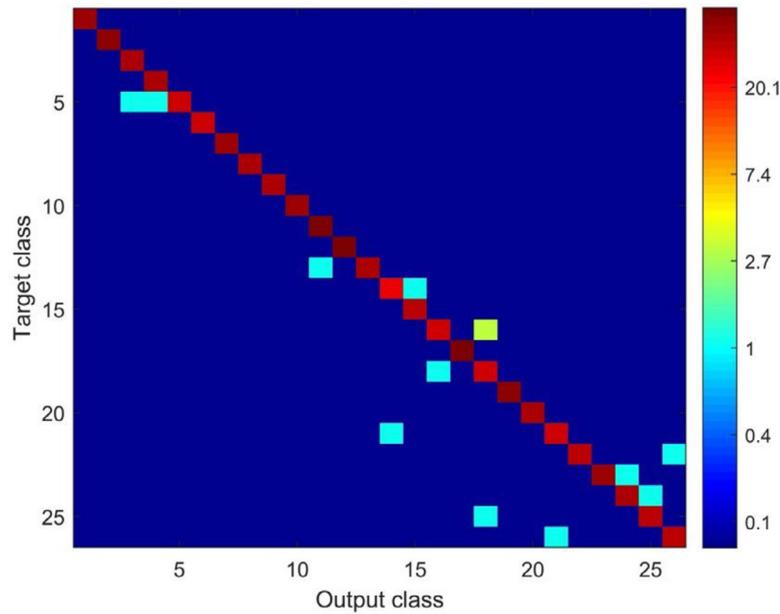


Fig. 3. Confusion matrix of lowest classification error. The lowest classification error was achieved in a run with 70 neurons. For readability, the colors are in logarithmic scale.

Table 4

Mean classification error for different choices of the number of neurons in the hidden layer, after feature selection. Calculated on 100 independent training processes; standard deviations are reported in parentheses. The lowest mean classification errors in the test set are in bold.

Set	Number of neurons in the hidden layer									
	20	30	40	50	60	70	80	90	100	110
Training	4.30 (0.88)	3.73 (0.83)	3.72 (0.67)	3.62 (0.69)	3.39 (0.71)	3.36 (0.68)	3.34 (0.61)	3.38 (0.61)	<b>3.32 (0.70)</b>	3.31 (0.68)
Validation	7.05 (0.99)	6.61 (0.88)	6.32 (0.96)	6.16 (0.85)	6.03 (0.75)	5.99 (0.75)	5.80 (0.78)	5.94 (0.74)	<b>5.70 (0.79)</b>	5.79 (0.78)
Test	7.52 (0.88)	6.85 (0.76)	6.45 (0.81)	6.25 (0.76)	6.22 (0.84)	6.13 (0.73)	6.05 (0.75)	6.19 (0.80)	<b>5.99 (0.85)</b>	6.03 (0.83)

9.71 percent, the mean classification error decreased by 22.44 percent.<sup>17</sup> Thus, it is fair to say that network- and simulation-based features definitively enhanced the classification performance.<sup>18</sup>

Interestingly, the converse is not true. If we remove all features that are not related to network or simulation-based methods, the classification performance is poor. The lowest mean classification error was about 43 percent when using 110 neurons. Therefore, network- and simulation-based features enhanced the model’s performance but were not sufficient by themselves.

#### 4.1.2. Principal component analysis feature selection

When PCA feature selection was implemented on the dataset, we worked with 26 features instead of 113. Table 4 shows the results attained when feature selection was implemented. As expected, misclassification error increased but remained low. The lowest mean misclassification error in the test set (5.99 percent) was achieved when using 100 neurons. This is about 2.1 times the lowest mean misclassification in the base-case scenario.

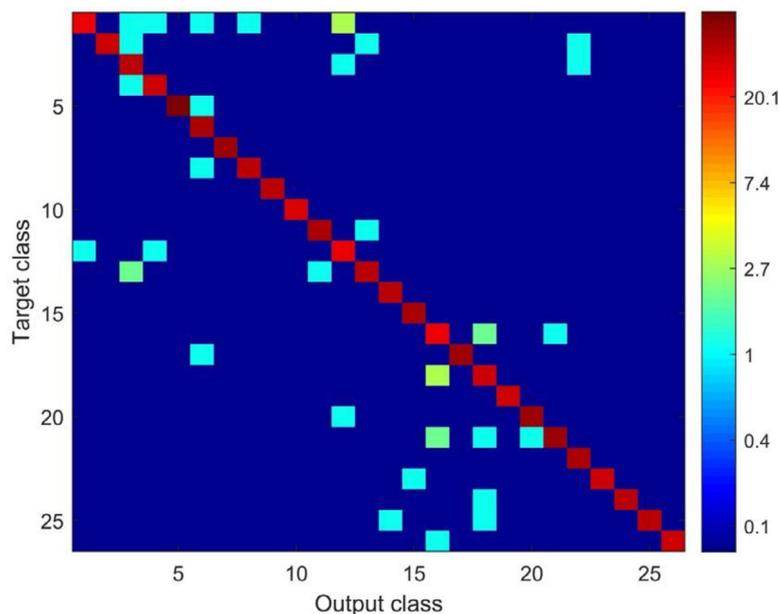
The lowest classification error after feature selection (3.98 percent) was attained in a run with 110 neurons, whereas the lowest misclassification before feature selection was 1.47 percent. Correspondingly, the confusion matrix for the lowest classification error after feature selection (in Fig. 4) showed higher misclassification than when using all 113 features (i.e., before feature selection). However, the classification performance is good.

The decrease in performance, corresponding to a mean classification error about 2.1 times that obtained in the base-case scenario, comes with some gains in computational time. Running the base-case scenario (with 113 features) lasts about 1.5 h, whereas running the lower-dimensional feature matrix attained with PCA feature selection (26 features) lasts about 0.4 h.<sup>19</sup>

<sup>17</sup> Comparing the lowest classification error achieved in each case also shows an enhancement in classification performance from including network- and simulation-based features: the lowest classification error decreased from 1.99 to 1.47 (about 26.13 percent).

<sup>18</sup> Individually, network-based features contribute more to enhancing the classification performance than simulation-based features do. Classification error when excluding simulation-based features is about 17 percent lower than when excluding network-based features.

<sup>19</sup> Measured as the number of hours to achieve 1000 independent runs, i.e., 100 runs for each of the 10 choices of numbers of neurons. The model was run on a desktop computer with 128 GB RAM and a 3.60 GHz CPU (no parallel computing was used).



**Fig. 4.** Confusion matrix of the lowest classification errors, after feature selection. The lowest classification error was achieved in a run with 110 neurons. For readability, the colors are in logarithmic scale.

The chosen model and its architecture enable us to classify banks based on a set of features that characterize their participation in the large-value payment system. In the base-case scenario, with 113 features, the out-of-sample mean classification error attained was about 3 percent, and errors did not cluster in any particular bank. That is, the out-of-sample classification performance was remarkable, and the methodological approach suggested for characterizing banks by their payment patterns is suitable.<sup>20</sup>

In addition, removing features related to network and simulation methods causes the mean classification error to increase by around 29 percent. That is, their contribution to our classification model is noteworthy. Finally, when we implement a PCA feature selection that retains 90 percent of the variance in the feature matrix, we can reduce the number of features from 113 to 26. As a result, the mean classification error doubles with respect to the base case model. Nevertheless, an error twice as large is around 6 percent, which is consistent with a good-performance classification model. Furthermore, with PCA feature selection, the time spent running the model decreased by 73 percent with respect to the base case model—without compromising the features' informational content.

## 5. Conclusions

We built an artificial neural network parameterization that represents the characteristic payment patterns of financial institutions in the Colombian large-value payment system. The types of features that serve as building blocks of payment patterns are the contributions to payments, funding habits, payment timing, payment concentration, centrality in the payment network, and systemic impact due to failure to pay. We achieved high out-of-sample classification performance, with about 3 percent error. The model's robustness comes in the form of stable performance after about 60 neurons and good (yet lower) performance when implementing a PCA feature-selection procedure. Additionally, we showed definitively that network centrality and systemic impact features contribute to enhancing the method's performance.

From a payment-system oversight and supervision perspective, the main objective of this model is to serve as a monitoring tool for detecting anomalies. In our case, a sizable change in the model's ability to classify a financial institution is a signal of a change in its behavior within the payment system.<sup>21</sup> In this vein, variations in individual or joint classification performance may be used as warning signals for behavioral changes that should be studied further. Therefore, this model is a novel and promising tool for supervising and overseeing financial markets and payment systems.

There are some challenges related to implementing this model as a monitoring tool. The first is deciding on the neural network's training frequency. Frequent (e.g., daily) training may quickly convert anomalous into normal behavior, whereas sporadic training

<sup>20</sup> The main results for all financial institutions are reported in [Appendix 2](#) (Table A1, Figures A3 and A4). As expected, due to the non-banking institutions' less consistent and contributive participation in the large-value payment system, the performance with all financial institutions was lower yet adequate, with about an 11 percent out-of-sample classification error. However, evidence in [Appendix 2](#) shows that the classification error was particularly high for some financial institutions; most of them are insurance firms, other banking institutions, or trust companies.

<sup>21</sup> Changes in a financial institution's or system's payment behavior are neither negative nor positive per se.

(e.g., yearly) may overlook the system's dynamics. This trade-off should be addressed. Second is deciding on a threshold with which to determine what a sizable change in individual classification performance would be. Based on the attained results, we foresee addressing these two challenges with an empirical approach involving an iterative process to fine-tune the outcomes of the monitoring process. Third, the relationships between behavioral changes in payment systems and individual or systemic events must be studied.

Finally, as in most artificial neural network models, the importance of the features is concealed. Other machine-learning methods could shed some light on the features' importance and interactions. This pending challenge will be addressed in a forthcoming research project based on random forest models,<sup>22</sup> which will enable us to understand further how features drive the classification process.

### Declaration of Competing Interest

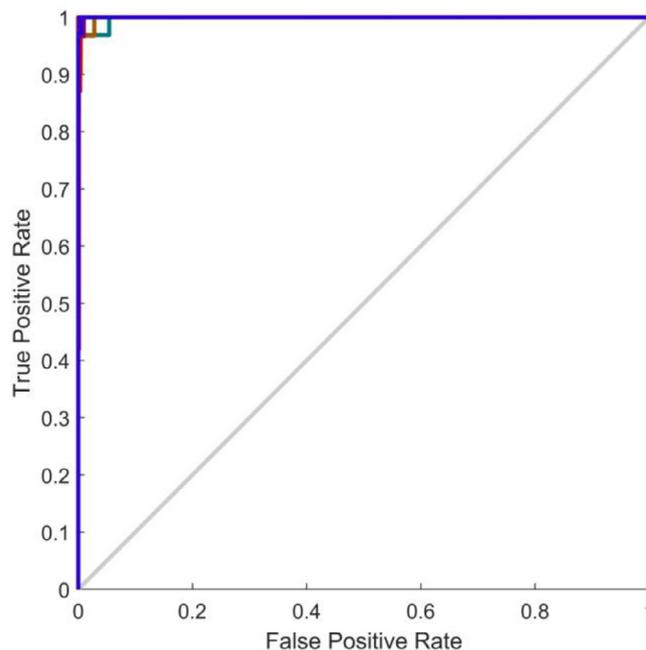
None.

### Acknowledgments

The opinions and statements in this article are the sole responsibility of the authors. O. Acero contributed to this article while working at Banco de la República. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: none.

### Appendix

#### Appendix 1. ROC curves<sup>23</sup>



**Fig. A1.** ROC curve of the lowest classification error. The lowest classification error was achieved in a run with 70 neurons.

<sup>22</sup> A random forest classification model uses a large number of uncorrelated decision trees (i.e., classification with the use of hierarchical decisions on the features, arranged in a tree-like structure) to obtain an overall classification. Unlike artificial neural networks, the outcome of random forest models allows for easy inspection of the roles of features in the classification process.

<sup>23</sup> The ROC shows the trade-off between the true positive rate (y-axis) and the false positive rate (x-axis) for a given model (Han and Kamber, 2006). When accurate, a model is more likely to encounter true positives than it is false positives. Hence, a good model is expected to show a steep ROC curve (i.e., close to the y-axis). On the other hand, a poor classifier is expected to show a curve close to the x-axis, whereas a random guess corresponds to the 45° line in the plot.

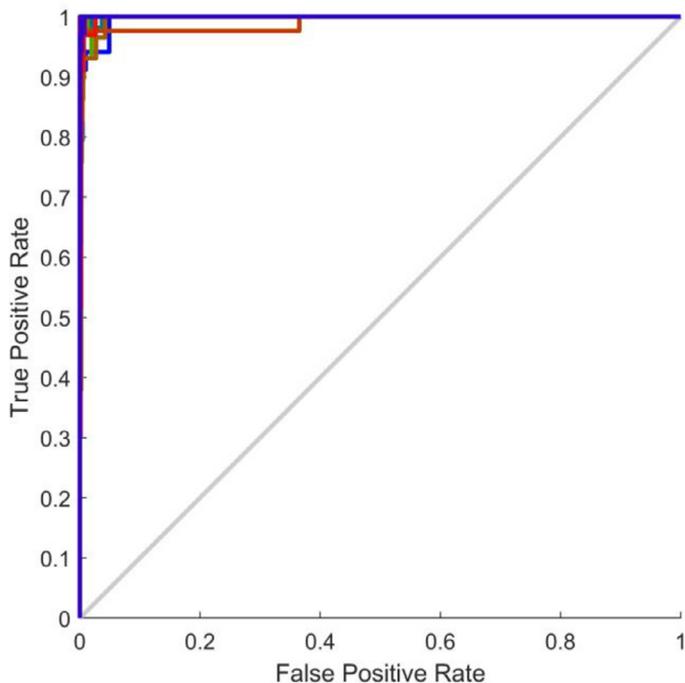


Fig. A2. ROC curve of the lowest classification error, after feature selection. The lowest classification error was achieved in a run with 110 neurons.

Appendix 2. Including all financial institutions (banking and non-banking)

Table A1

Mean classification error for different choices of the number of neurons in the hidden layer, including all financial institutions. Calculated on 100 independent training processes; the standard deviations are reported in parentheses. The lowest mean classification errors in the test set are in bold.

Set	Number of neurons in the hidden layer									
	20	30	40	50	60	70	80	90	100	110
Training	10.51 (2.55)	9.15 (0.62)	8.73 (0.55)	8.56 (0.54)	8.49 (0.48)	8.45 (0.56)	8.39 (0.44)	8.35 (0.46)	8.35 (0.48)	<b>8.31 (0.42)</b>
Validation	13.76 (2.39)	12.25 (0.66)	11.79 (0.64)	11.52 (0.53)	11.37 (0.64)	11.21 (0.57)	11.23 (0.50)	11.06 (0.54)	11.01 (0.54)	<b>11.00 (0.51)</b>
Test	13.80 (2.29)	12.38 (0.72)	11.83 (0.73)	11.55 (0.55)	11.46 (0.64)	11.43 (0.60)	11.18 (0.61)	11.21 (0.56)	11.22 (0.44)	<b>11.09 (0.57)</b>

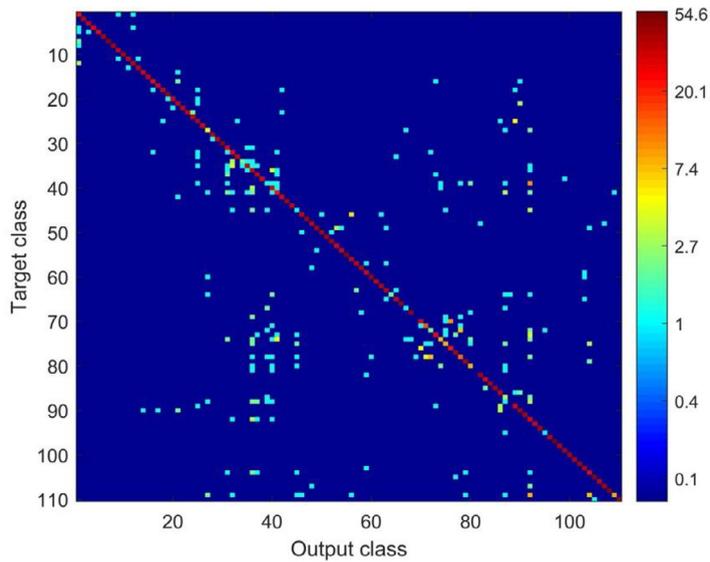


Fig. A3. Confusion matrix of the lowest classification errors, including all financial institutions. The lowest classification error was achieved in a run with 80 neurons. For readability, the colors are in logarithmic scale.

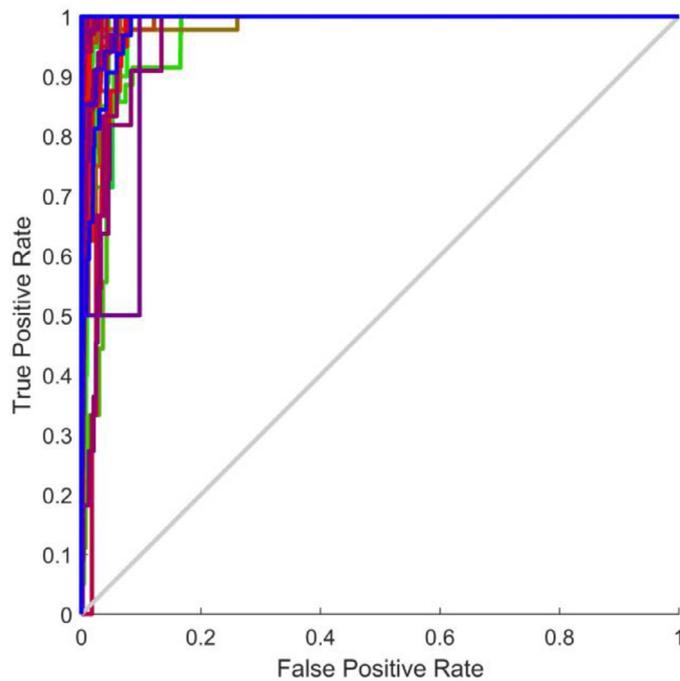


Fig. A4. ROC curve of the lowest classification error, including all financial institutions. The lowest classification error was achieved with 80 neurons.

**References**

Aleskerov, E., Freisleben, B., Rao, B., 1997. Cardwatch: a neural network-based database mining system for credit card fraud detection. In: Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering (CIFER). IEEE, pp. 220–226.

Alpaydin, E., 2014. Introduction to Machine Learning. The MIT Press, Cambridge.

Angelini, E., di Tollo, G., Roli, A., 2008. A neural network approach for credit risk evaluation. Q. Rev. Econ. Financ. 48, 733–755.

Atiya, A.F., 2001. Bankruptcy prediction for credit risk using neural networks: a survey and new results. IEEE Trans. Neural Netw. 12 (4), 929–935.

Becher, C., Galbiati, M., Tudela, M., 2008. The timing and funding of CHAPS Sterling payments. In: Economic Policy Review. Federal Reserve Bank of New York, pp. 113–133.

Bekhet, A.H., Eletter, S.F., 2014. Credit risk assessment model for Jordanian commercial banks: neural scoring approach. Rev. Dev. Financ. 4, 20–28.

- Bernal, J., Cepeda, F., Ortega, F., 2012. Estimating the contribution of liquidity sources in the Colombian large-value real-time gross settlement payment: a preliminary approach. *J. Paym. Strateg. Syst.* 6 (2), 159–182.
- Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
- Borio, C., 2003. Towards a Macroprudential Framework for Financial Supervision and Regulation?. Bank for International Settlements *BIS Working Papers*, 128.
- Brause, R., Langsdorf, T., Hepp, M., 1999, November. Neural data mining for credit card fraud detection. In: *Proceedings of the 11th International Conference on Tools with Artificial Intelligence*. IEEE, pp. 103–106.
- Brédart, X., 2014. Bankruptcy prediction model using neural networks. *Account. Financ. Res.* 3 (2), 124–128.
- Brin, S., Page, L., 1998. Anatomy of a large-scale hypertextual web search engine. In: *Proceedings of the 7th International World Wide Web Conference*.
- Caldarelli, G., 2020. A perspective on complexity and network science. *J. Phys. Complex.* 1, 021001.
- Crockett, A., 2000. Marrying the Micro- and Macro-Prudential Dimensions of Financial Stability. Bank of International Settlements *BIS Speeches*.
- Demyanyk, Y., Hasan, I., 2009. Financial Crises and Bank Failures: A Review of Prediction Methods. Federal Reserve Bank of Cleveland *Working Papers Series*, 09-04R.
- Denbee, E., Garratt, R.J., Zimmerman, P., 2014. Variations in Liquidity Provision in Real-Time Payment Systems, 513. Bank of England *Working Paper Series*.
- Diehl, M., 2013. Measuring free-riding in large-value payment systems: the case of TARGET2. *J. Financ. Market Infrastruct.* 1 (3), 31–53.
- Ding, M., Tian, H., 2016. PCA-based network traffic anomaly detection. *Tsinghua Sci. Technol.* 21 (5), 500–509.
- Dorransoro, J.R., Ginel, F., Sanchez, C., Cruz, C.S., 1997. Neural fraud detection in credit card operations. *IEEE Trans. Neural Netw.* 8 (4), 827–834.
- Elter, S.F., Yaseen, S.G., Elrefae, G.A., 2010. Neuro-based artificial intelligence model for loan decisions. *Am. J. Econ. Bus. Adm.* 2 (1), 27–34.
- Farmer, J.D., Gallegati, M., Hommes, C., Kirman, A., Ormerod, P., Cincotti, S., Sanchez, A., Helbing, D., 2012. A complex systems approach to constructing better models for managing financial markets and the economy. *Eur. Phys. J.* 214, 295–324.
- Fioramanti, M., 2008. Predicting sovereign debt crises using artificial neural networks: a comparative approach. *J. Financ. Stab.* 4, 149–164.
- Ghosh, S., Reilly, D.L., 1994. Credit card fraud detection with a neural-network. In: *Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*, 3. IEEE, pp. 621–630.
- Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O., 2014. *Neural Network Design*. Martin Gagan, Oklahoma.
- Haldane, A.G., May, R.M., 2011. Systemic risk in banking ecosystems. *Nature* 469 (7330), 351–355.
- Han, J., Kamber, M., 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco.
- Hastie, T., Tibshirani, R., Friedman, J., 2013. *The Elements of Statistical Learning*. Springer, New York.
- Holopainen, M., Sarlin, P., 2016. Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty. European Central Bank *Working Paper*, 1900.
- Khediri, K.B., Charfeddine, L., Youssef, S.B., 2015. Islamic versus conventional banks in the GCC countries: a comparative study using classification techniques. *Res. Int. Bus. Financ.* 33, 75–98.
- Kleinberg, J.M., 1998. Authoritative sources in a hyperlinked environment. In: *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*.
- León, C., 2020. Detecting anomalous payment networks: a dimensionality reduction approach. *Lat. Am. J. Cent. Bank.* 1. doi:10.1016/j.latcb.2020.100001.
- León, C., Machado, C., Sarmiento, M., 2018. Identifying central bank liquidity super-spreaders in interbank funds networks. *J. Financ. Stab.* 35, 75–92.
- León, C., Moreno, J.F., Cely, J., 2017. Whose balance sheet is this? Neural networks for banks' pattern recognition. *Wilmott* 91, 34–47.
- Martínez-Jaramillo, S., Alexandrova-Kabadjova, B., Bravo-Benitez, B., Solórzano-Margain, J.P., 2014. An empirical study of the Mexican banking system's network and its implications for systemic risk. *J. Econ. Dyn. Control* 40, 242–265.
- Martínez, C., Cepeda, F., 2018. Freeriding on liquidity in the Colombian large-value payment system. *J. Financ. Market Infrastruct.* 6 (4), 19–40.
- McAndrews, J., Rajan, S., 2000. The timing and funding of fedwire funds transfers. In: *Economic Policy Review*. Federal Reserve Bank of New York, pp. 17–32 July.
- McNelis, P.D., 2005. *Neural Networks in Finance*. Elsevier, Burlington.
- Mehta, P., Bukov, M., Wang, C.-H., Day, A.G.R., Richardson, C., Fisher, C.K., Schwab, D.J., 2019. A high-bias, low-variance introduction to machine learning for physicists. *Phys. Rep.* 810, 1–124.
- Nazari, M., Alidadi, M., 2013. Measuring credit risk of bank customers using artificial neural network. *J. Manag. Res.* 5 (2), 17–27.
- Newman, M.E.J., 2010. *Networks: An Introduction*. Oxford University Press, New York.
- Olmeda, I., Fernández, E., 1997. Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction. *Comput. Econ.* 10, 317–355.
- Sabetti, L. & Heijmans, R. (2020). Shallow or deep? Detecting anomalous flows in the Canadian Automated Clearing and Settlement System using an autoencoder. *DNB Working Paper*, 681.
- Salchenberger, L., Mine, C., Lash, N., 1992. Neural networks: a tool for predicting thrift failures. *Decis. Sci.* 23, 899–916.
- Sarlin, P., 2014. On biologically inspired predictions of the global financial crisis. *Neural Comput. Appl.* 24 (3), 663–673.
- Shmueli, G., 2010. To explain or to predict? *Stat. Sci.* 25 (3), 289–310.
- Soramaki, K., Cook, S., 2013. SinkRank: an algorithm for identifying systemically important banks in payment systems. *Econ. Open Access Open Assess. E J.* 7 (2013–28), 1–27.
- Sree, A., Venkata, K., 2014. Anomaly detection using principal component analysis. *Int. J. Comput. Sci. Technol.* 5 (4), 124–126.
- Tam, K.Y., 1991. Neural network models and the prediction of bank bankruptcy. *Omega* Westport 19, 429–445.
- Tam, K.Y., Kiang, M., 1990. Predicting bank failures: a neural network approach. *Appl. Artif. Intell.* 4, 265–282.
- Triepels, R., Daniels, H., Heijmans, R., 2017. Anomaly detection in real-time gross settlement systems. In: *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017)*, 1, pp. 433–441.
- Turkan, S., Polat, E., Gunay, S., 2011. Classification of domestic and foreign commercial banks in Turkey based on financial performances using linear discriminant analysis, logistic regression and artificial neural network models. In: *Proceedings of the International Conference on Applied Economics – ICOAE 2011*, pp. 717–723.
- Varian, H.R., 2014. Big data: new tricks for econometrics. *J. Econ. Perspect.* 28 (2), 3–28. doi:10.1257/jep.28.2.3.
- Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M., 2010. Graph kernels. *J. Mach. Learn. Res.* 11, 1201–1242.
- Wilson, R.L., Sharda, R., 1994. Bankruptcy prediction using neural networks. *Decis. Support Syst.* 11 (5), 545–557.
- Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman Publishers, Burlington.
- Wu, R.C., 1997. Neural network models: foundations and applications to an audit decision problem. *Ann. Oper. Res.* 75, 291–301.
- Zhang, G., Hu, M.Y., Patuwo, B.E., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. *Eur. J. Oper. Res.* 116, 16–32.